



KI im Superwahljahr 2024

Generative KI im Umfeld demokratischer Prozesse

GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung

 **acatech**

DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN

WHITEPAPER

Bieber, C., Heesen, J.,
Grunwald, A. & Rostalski, F.
AG IT-Sicherheit, Privacy,
Recht und Ethik

Inhalt

Zusammenfassung	3
1 Einleitung: Einflussnahme auf demokratische Wahlen	4
2 Auswirkungen generativer KI auf Meinungsbildung und Wahlen	6
2.1 Beispiele für KI-unterstützte Einflüsse auf Wahlen	6
2.2 Relevanz von KI-basierten Einflussnahmen	15
3 Maßnahmen zur Sicherung der Integrität demokratischer Wahlen	17
4 Fazit und Ausblick	21
Literatur	22
Über dieses Whitepaper	24

Zusammenfassung

Die zunehmende Verbreitung generativer Künstlicher Intelligenz (KI) wie ChatGPT stellt Demokratien vor neue Herausforderungen und Chancen. Denn für diese gilt: Demokratien benötigen einen öffentlichen Diskurs, der wiederum Vertrauen und Vertrauenswürdigkeit voraussetzt. Falsche oder ungenaue, zugleich aber überzeugende Informationen in den sozialen Medien nehmen mit zunehmendem Aufkommen generativer KI zu und weichen so diese Basis gesellschaftlicher Diskurse auf. Damit sind die potenziellen Auswirkungen von generativer KI auf Demokratien zu einem wichtigen Thema – gerade im Superwahljahr 2024 – geworden. Welche Rolle generative KI als nicht mehr wegzudenkende Technologie in unserem Alltag in politischen Prozessen zur Sicherung der Integrität von Wahlen und Stärkung des Vertrauens in die Demokratie einnehmen wird, ist letztlich abhängig vom Umgang und der Art und Weise, wie wir als demokratische Gesellschaft sie nutzen und einfordern.

Die Autorinnen und Autoren der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik beleuchten in diesem Papier die wachsende Bedeutung generativer KI für Wahlen und Demokratien sowie deren potenzielle Auswirkungen auf die politische Meinungsbildung. Anhand einer Bandbreite von konkreten wie möglichen Beispielen werden Herausforderungen und Chancen aufgezeigt: Von täuschend echt gefälschten Bildern, Videos oder Audioaufnahmen als sogenannte Deepfakes im Vorfeld von Wahlkämpfen über Personalisierung von (falschen) Wahlinformationen, Unterstützung der Parteien beim Wahlkampf per Chatbot bis hin zu neuen Teilhabechancen bei der Zugänglichkeit von Wahlen. Welchen (direkten) Einfluss generative KI dabei auf Wahlen und die politische Meinungsbildung tatsächlich hat, lässt sich zum jetzigen Zeitpunkt aufgrund der unzureichenden Datenlage noch nicht belegen. Jedoch unabhängig davon sind demokratische Gesellschaften gefordert, sich mit dieser in unserem Alltag bereits fest verankerten Technologie mit all ihren Vor- wie Nachteilen auseinanderzusetzen und geeignete Maßnahmen wie Herkunftspflicht oder Stärkung der gesellschaftlichen KI-Kompetenzen zur Eindämmung von Manipulationen und Einflussnahme – auch mithilfe generativer KI – zu ergreifen, so die Einschätzung der Autorinnen und Autoren.

1 Einleitung: Einflussnahme auf demokratische Wahlen

Demokratische Wahlen bilden den Kern einer freien Gesellschaft: Durch sie haben die Bürgerinnen und Bürger eines Landes die Möglichkeit, politisches Geschehen entsprechend der jeweiligen Mehrheitsverhältnisse zu gestalten. Eingebettet in Prinzipien der Rechtsstaatlichkeit manifestieren sich so die Menschen- und Bürgerrechte jeder und jedes Einzelnen in aktiver Politik. Die Integrität von Wahlen zu wahren, ist deshalb immer wieder eine aktuelle Herausforderung demokratischer Gesellschaften – insbesondere bei anstehenden Wahlen.

Diese Herausforderung ist 2024 so relevant wie kaum je zuvor. Denn in diesem Jahr werden für mehr als vier Milliarden Menschen in rund 40 Ländern Wahlen stattfinden, darunter einige der bedeutendsten Wahlen für die demokratische Welt: Die Wahlen zum Europäischen Parlament im Juni 2024 und die US-Präsidentschaftswahlen im November 2024 betreffen allein schon knapp 780 Millionen Menschen. Auch für Deutschland ist 2024 ein Superwahljahr, das mit mehreren Landtags- und Kommunalwahlen für politische Weichenstellungen in den nächsten Jahren steht. Einflussnahme auf Wahlen ist kein neues Phänomen. Seit jeher haben politische Akteure versucht, (die Durchführung von) Wahlen zu manipulieren oder die Meinungsbildung von Wählerinnen und Wählern durch falsche oder irreführende Informationen zu lenken. Auch der Einfluss von Technologie auf Wahlen ist dabei nichts Neues: Fake News verbreiten sich durch soziale Medien rasend schnell, Wahlempfehlungs-Software kann dagegen den Zugang zu politischer Willensbildung erleichtern.

Im Vorfeld der nun anstehenden Wahlen wird diskutiert, ob KI-Systeme und insbesondere die seit 2023 rasant an Bedeutung gewinnende generative KI die Einflussnahme auf Wahlen in ganz neuen Dimensionen ermöglichen könnten. Vorstellbar ist unter anderem: Sprachmodelle werden als Redenschreiber genutzt oder treten als Chatbots in den Dialog mit potenziellen Wählerinnen und Wählern – Politikerinnen und Politiker können so klassische Wahlkampfwerkzeuge erweitern. Mit Deepfakes wiederum können Akteure des In- und Auslands Politikerinnen und Politikern Worte gleichsam in den Mund legen, um sie zu diskreditieren oder ihnen nicht geteilte politische Meinungen unterzuschieben. Und auch die Menge an täuschend echt wirkenden Falschinformationen in der Berichterstattung zu politischen Geschehnissen und Wahlen lässt sich nahezu unbegrenzt skalieren. Es ist kein Zufall, dass immer häufiger prägnante Beispiele dafür in die öffentliche Wahrnehmung gelangen. Bisher weniger im öffentlichen Fokus, aber nicht weniger problematisch können sich Auswirkungen auf die tatsächliche Organisation und Durchführung von Wahlen erweisen, denn KI-Werkzeuge werden auch im „Back-Office“ von Verwaltungsprozessen eingesetzt.

KI-Tools können somit an unterschiedlichen Stellen im Wahlprozess eingesetzt werden und sowohl Bedrohungsszenarien auslösen als auch unterstützend oder absichernd wirken. Die zahlreichen Wahlen im Jahr 2024 zeigen deutlich, wie Wahlprozesse in verschiedenen Phasen immer stärker digital unterstützt, aber auch beeinflusst werden können. Von besonderer Bedeutung sind hybride Bedrohungen, die sich aus dem Zusammenspiel von traditionellen und technischen Ansatzpunkten ergeben und häufig die Infrastrukturen sozialer Netzwerke nutzen. Immer öfter werden dabei auch KI-Werkzeuge in bereits vorhandene Angriffstaktiken eingebunden, die im Vorfeld von Wahlen auf irreführende, täuschende Aktivitäten setzen, um destabilisierende Wirkungen zu entfalten (Pawelec, 2022).

Mit dem vorliegenden Whitepaper wollen die Autorinnen und Autoren eine Bestandsaufnahme der möglichen Einflussnahme auf Wahlen durch (generative) KI machen und anhand realer Fallbeispiele erläutern. Von der Darstellung theoretischer Möglichkeiten und einzelner Fallbeispiele muss die Betrachtung unterschieden werden, ob und wie sich der Einfluss von KI-basierter Wahlmanipulation nachweisen lässt und in welchem Umfang er ein tatsächliches Problem darstellt. Gleichzeitig soll beleuchtet werden, dass (generative) KI als anwendungsoffene Technologie auch Chancen für die Unterstützung der Integrität und Zugänglichkeit von Wahlen haben kann, etwa durch den Abbau von Sprachbarrieren, und Demokratien in die Lage versetzen kann, sich gegen Einflussnahmen von manipulativen Akteuren zu schützen.

Die im Folgenden dargestellten Überlegungen müssen in einen größeren Kontext von legitimen wie illegitimen Einflussnahmen auf Wahlen durch Techniken, Kommunikationsstrategien und Kampagnen eingeordnet werden. Dies beginnt bei herkömmlichen „alt-digitalen“ Formen der Einflussnahme (wie Social Bots). Cybersicherheit und der Kampf gegen „herkömmliche“ Manipulationen von demokratischen Wahlkämpfen und Wahlen haben auch jenseits der KI-Diskussion große Bedeutung. Das Papier erweitert und aktualisiert die Betrachtungen der Arbeitsgruppe Recht und Ethik zum Einfluss von KI-Systemen auf die individuelle Wahlentscheidung aus dem Jahr 2021 (Heesen et al., 2021).

2 Auswirkungen generativer KI auf Meinungsbildung und Wahlen

Im folgenden Kapitel geben die Autorinnen und Autoren anhand realer Beispiele einen Überblick über Auswirkungen von generativer KI auf Wahlen. Diese Beispiele zeigen eine Bandbreite von Szenarien auf, wie generative KI Einfluss auf Wahlen haben kann und das Potenzial mitbringt, Wahlkämpfe, wie sie aus der Vergangenheit bekannt waren, zu beeinflussen. Dabei muss nicht jeder KI-Einsatz auf eine illegale Beeinflussung oder sogar Manipulation hinweisen. Doch auch wenn Politikerinnen und Politiker, Parteien oder auch Institutionen der öffentlichen Meinungsbildung wie Verbände oder Kampagnennetzwerke KI einsetzen, um Interessen zu stärken, stellen sich an derartige Einsätze besondere Anforderungen etwa hinsichtlich der Transparenz der KI-Beteiligung (Chennupati, 2024). Viele dieser Anforderungen ergeben sich unter anderem durch die KI-Verordnung der Europäischen Union, die für KI-Anwendungen mit hohem Risiko, das etwa aus dem Einsatz im Umfeld demokratischer Wahlen entstehen könnte, besondere Verpflichtungen an Transparenz oder Datenqualität einfordert (siehe dazu auch: [KI-Kompakt: AI Act der Europäischen Union](#), Plattform Lernende Systeme, 2024).

2.1 Beispiele für KI-unterstützte Einflüsse auf Wahlen

Unterstützung von Wahlkämpfen durch Anrufe von KI-Chatbots

Wahlkämpfe waren schon seit jeher auch eine Frage des finanziellen Budgets. Wahlplakate, Werbungen in Radio, Fernsehen oder Kino wie auch Give-aways an Wahlkampfständen: Kampagnen kosten Geld. Auch ohne die zusätzlichen Möglichkeiten durch KI-Systeme ist das Idealbild eines fairen und gleichberechtigten Zugangs aller zu politischen Entscheidungsprozessen deshalb nur bedingt realisiert. Dieser Faktor variiert stark hinsichtlich des Bezugslands: Während beispielsweise in Deutschland die Finanzierung von Wahlkämpfen mit kostenfreien Sendeplätzen im Rundfunk für Wahlwerbung und der öffentlichen Teilkostenerstattung von Wahlkämpfen zumindest anteilig durch die öffentliche Hand getragen wird, sind in anderen demokratischen Staaten wie den USA Kandidatinnen und Kandidaten auf eigene Mittel sowie Spenden und die Unterstützung durch Kampagnen-Organisationen, die sogenannten Political Action Committees (PACs), angewiesen.

Ein wesentliches Element von Wahlkämpfen ist und war schon immer die persönliche Begegnung mit Wählerinnen und Wählern etwa am Infostand und im Haustürwahlkampf oder auch via Telefon: In den USA ist es üblich, dass Freiwillige Wählerinnen und Wähler anrufen, um für bestimmte Kandidatinnen und Kandidaten zu werben. In den letzten Jahren wurden diese Anrufe bereits von Telefon-Bots unterstützt. Nun steigen KI-Systeme in diese Praxis ein. Ein Beispiel: Die KI-Stimme „Ashley“ rief im US-Wahlkampf tausende von Menschen an, um für die Kandidatin Shamaine Daniels zu werben (Bomke et al., 2024).

Ashley kann dabei viel mehr als die bisherigen Bots. Am Anfang des Gesprächs weist sie selbst darauf hin, dass sie eine KI sei – ihre Stimme ist dazu erkennbar nicht menschlich. Sie kann in 20 Sprachen auf die jeweilige Person zugeschnittene Gespräche führen. Diese digitale Vervielfältigung des Outputs und die damit verbundene Personalisierung stellt aus der Perspektive der Kampagnentreibenden einen deutlichen Vorteil dar – doch auch jenseits des transparenten Umgangs mit der KI-Unterstützung müssen Kandidatinnen und Kandidaten darauf achten, dass sich keine ungewollten Trainingseffekte einstellen. Ashley sollte keinen „Bias“

entwickeln und keine Wählergruppen bevorzugen oder benachteiligen. Ob dies in der Kampagnenpraxis funktioniert, wird sich erst noch erweisen müssen.

Durch das rasante Aufkommen generativer KI-Technologien braucht es oft noch Zeit, um relevante Gesetzgebungen anzupassen: So gab es in Europa bislang keine verbindliche und allgemeine Verpflichtung, KI-Chatbots so zu gestalten, dass sie sich ihren menschlichen Gesprächspartnern gegenüber zu erkennen geben. Erst der AI Act der Europäischen Union, der die Entwicklung und den Einsatz von KI in Europa regelt, wird für generative KI-Systeme (wie Sprachbots), die mit Menschen direkt interagieren, Transparenzpflichten einführen (Art. 52, Abs. 1 AI Act). Für das bereits angelaufene Wahljahr 2024 werden diese Regelungen jedoch nicht mehr greifen.

Die Vorteile für Kandidatinnen und Kandidaten liegen auf der Hand: Durch die Verwendung von KI-Bots können sie mit „Anrufen“ eine bislang nicht mögliche Reichweite erzielen: Das KI-System kann problemlos mit zahlreichen Bürgerinnen und Bürgern gleichzeitig kommunizieren und dabei unter Umständen auch die Aussagen personalisieren. Wenn diesen zudem nicht klar ist, dass sie nicht mit einer echten Person sprechen, kann dies Vorteile für die Wirkung einer Kampagne bringen. Aber auch wenn transparent gemacht wurde, dass das Gespräch mit einer KI geführt wird, kann das Telefongespräch ein niedrighschwelliges und interaktives Informationsangebot sein. In Deutschland werden „Robo-Calls“, sogenannte Roboteranrufe, allerdings auch in Zukunft keine Rolle in Wahlkämpfen spielen, da Werbeanrufe ohne vorherige Zustimmung der Angerufenen grundsätzlich nicht zulässig sind.

Sprachmodelle und Chatbots als Basis für (falsche) Wahlinformationen

Mit großen Sprachmodellen wie ChatGPT ist KI in der Mitte der Gesellschaft angekommen. Der spielerisch leichte Zugang zu diesen Arten von generativer KI und ihr Alltagsnutzen haben dazu beigetragen, die Nutzung von KI-Systemen zu einem Massenphänomen jenseits von Forschung und Industrie zu machen. Schnell wurde dabei klar, dass generative KI nicht nur ein harmloser Zeitvertreib ist, sondern ganz konkrete Diskurse über die verantwortliche Entwicklung und Anwendung solcher Systeme aufwerfen muss.

Ein qualitatives Problem von Sprachmodellen sind sogenannte „Halluzinationen“ – also beispielsweise die irreführende Vermischung von Inhalten oder die Erfindung von Quellen. Solche Fehler entstehen – aufgrund der generativen Anwendung – aus der statistisch wahrscheinlichsten Verknüpfung einzelner Elemente. Die Sprachmodelle können „Halluzinationen“ nicht zuverlässig vermeiden, weil ihre Aussagen nicht standardmäßig überprüft werden und sie nicht immer auf aktuelle Internetinhalte zugreifen können. Dadurch ist das mit „Halluzinationen“ verbundene Schadenspotenzial bislang kaum zu begrenzen.

Auf Wahlen und Wahlkämpfe bezogen kann dies zu einer unabsichtlichen Falschinformationsbasis führen. Eine beispielhafte Untersuchung von AlgorithmWatch und AI Forensics zu Antworten von Sprachmodellen auf Anfragen zu den Landtagswahlen 2023 in Bayern und Hessen hat gezeigt, dass KI-Sprachmodelle zum Teil irreführende Inhalte hervorbringen (AlgorithmWatch, 2023). So wurde ein kurz vor der Landtagswahl in Bayern aufgekommener politischer Skandal mit falschen inhaltlichen Hintergründen wiedergegeben. Auch ordnete der Chatbot die Spitzenkandidatinnen und -kandidaten der Landtagswahl in Hessen konsequent der falschen Partei zu oder gab Politikerinnen und Politiker an, die schon lange nicht mehr aktiv sind. Auch erfundene Umfragewerte waren ein Problem.

Im Vorfeld der Europawahl fand das Recherchemagazin Correctiv heraus, dass Google Gemini, Microsoft Copilot und ChatGPT falsche oder ungenaue politische Informationen bereitstellten (Simon et al., 2024). So produzierte ChatGPT mitunter falsche Informationen darüber, wer überhaupt wahlberechtigt ist. Das Sprachmodell orientiert sich an den Vorgaben für vergangene Wahlen und geht zum Beispiel für Deutschland vom Mindestwahlalter von 18 Jahren aus – für die Europawahl 2024 können jedoch alle deutschen Staatsbürgerinnen und -bürger wählen, wenn sie mindestens 16 Jahre alt sind. Stichproben von Forscherinnen und Forschern brachten auch weitere Fehlinformation zutage, etwa eine falsche Anzahl der zu wählenden Abgeordneten in das EU-Parlament oder falsche Angaben zum Wahlsystem (in Italien) (Simon et al., 2024).

ANWENDUNGSBEISPIEL**Irreführende ChatGPT-Antwort zur Europawahl 2024****You**

Ich lebe in Deutschland und habe am 8. Mai 2008 Geburtstag. Darf ich bei der Europawahl am 9. Juni 2024 wählen?

**ChatGPT**

Da du in Deutschland lebst und am 8. Mai 2008 geboren wurdest, bist du im Jahr 2024 voraussichtlich mindestens 16 Jahre alt. In Deutschland liegt das Mindestwahlalter für die Europawahl bei 18 Jahren. Daher wirst du leider nicht berechtigt sein, bei der Europawahl am 9. Juni 2024 zu wählen. Du müsstest 18 Jahre alt sein, um an dieser Wahl teilnehmen zu können.

Quelle: Exemplarischer ChatGPT-Chat (GPT 3.5) vom 12. April 2024

Falsche oder irreführende Informationen können als Grundlage für die Wahlentscheidungen einzelner Personen gravierende Auswirkungen auf die Meinungsbildung für demokratische Wahlen haben. Denn Wählerinnen und Wähler verlassen sich unter Umständen auf diese falschen Aussagen und verzichten auf eine zusätzliche Überprüfung an anderer Stelle, etwa durch die Nutzung einer Suchmaschine oder spezialisierte Informationsangebote (im Beispielfall etwa im Angebot der Bundeswahlleiterin) (Simon et al., 2024). Zudem zeigt sich die Tendenz, dass große Sprachmodelle als Echokammern wirken, indem sie dazu neigen, den Meinungen ihrer Nutzerinnen und Nutzer zuzustimmen, und somit menschliche Vorurteile verstärken können (Nehring et al., 2024). Fehlerhafte Informationen werden auch durch Menschen und andere Technologien verbreitet. Das Besondere an generativer KI ist dabei jedoch eine gesteigerte Anfälligkeit der Wählerinnen und Wähler, vor allem bei mangelnder Aufklärung über das große Fehlerpotenzial von generativer KI. Verstärkt wird diese Anfälligkeit durch das Phänomen des „automation bias“ im Umfeld von KI-Entscheidungen. Damit wird der psychologische Effekt bezeichnet, Algorithmen und deren Empfehlungen und Entscheidungen ein Übermaß an Vertrauen entgegenzubringen und diesen gegenüber als unsicher betrachteten menschlichen Einschätzungen den Vorrang zu geben. Übertriebenes Vertrauen in die täuschend echt wirkenden und überzeugend dargelegten Informationen von Sprachmodellen kann so Wahlen und Wahlergebnisse beeinflussen, besonders, wenn generative KI immer stärker in die Alltagsnutzung der Menschen Einzug hält und so eine immense Breitenwirkung erzielen kann.

Auf der anderen Seite könnten zuverlässig regulierte und qualitativ hochwertige Chatbots Wählerinnen und Wähler mit auf ihre individuellen Bedürfnisse hin zugeschnittenen Information versorgen. Dies kann den Zugang zu politischen Inhalten verbessern und Politikverdrossenheit begegnen. Denn die politische Landschaft wird zunehmend komplexer und unübersichtlicher. So wird beispielsweise die Parteienlandschaft in Deutschland immer kleinteiliger; Wählerinnen und Wählern steht mittlerweile eine ganze Reihe an Parteien zur Auswahl, die in Umfrageergebnissen über oder an der Fünf-Prozent-Hürde gesehen werden. Aufgearbeitete Informationen über die Programme dieser Parteienangebote mit passgenauen Antworten auf spezifische Fragen zu erhalten, wäre ein echter Mehrwert für Wählerinnen und Wähler, die eine informierte Wahlentscheidung suchen, und könnte beispielsweise das bereits vorhandene Angebot wie den „Wahl-O-Mat“ der Bundeszentrale für politische Bildung in der Breite verbessern und eine individualisierte Nutzung ermöglichen. Allerdings wäre hierfür eine zuverlässige Verzahnung der Chatbots mit tagesaktuellen Informationsangeboten notwendig.

Nutzung von generativer KI durch Parteien: Erstellung zielgruppenorientierter Inhalte

Ein spezieller Anwendungsfall von generativer KI, insbesondere von großen Sprachmodellen, ergibt sich mit Blick auf deren Potenziale für politische Akteure. Politikerinnen und Politiker sind seit jeher daran interessiert, politische Inhalte, die aus den weltanschaulichen und politischen Überzeugungen von Personen und Parteien entspringen, mit den je aktuellen Stimmungen der Wählerinnen und Wähler abzugleichen. Anpassungen in der Programmatik von Parteien beispielsweise nach Wahlniederlagen, die Wünsche und Bedürfnisse aus der Wählerschaft aufgreifen, sind deshalb ein bekanntes und notwendiges Mittel.

Populistische Akteure versuchen dagegen, unter Ausnutzung und Verstärkung von Stimmungsbildern in der Bevölkerung zu polarisieren und eine Spaltung zwischen verschiedenen Bevölkerungsgruppen (in der Regel zwischen „oben“ und „unten“) aufzumachen.

Sowohl für die Anpassung an Trends als auch für die Verstärkung populistischer Agitation können KI-Sprachmodelle genutzt werden: Mit diesen könnten Parteien Inhalte von Wahlprogrammen in Ausdruck und Sprache bestimmter Zielgruppen übersetzen oder sich basierend auf aktuellen Trendanalysen neue Inhalte schreiben lassen. Während auf der einen Seite so eine zielgruppengerechte Informationsbasis geschaffen werden kann – man denke beispielsweise an Versionierungen von für bestimmte Zielgruppen relevanten Inhalten aus Parteiprogrammen oder deren Übersetzung in Leichte oder Einfache Sprache –, könnten populistische Parteien Sprachmodelle gezielt dafür nutzen, um mit auf unterschiedliche Zielgruppen angepassten Inhalten und sprachlichen Stilen politische Stimmungen auszunutzen.

Microtargeting und der Digital Services Act

In diesem Zusammenhang muss ein Blick auf die Möglichkeiten geworfen werden, mit denen politische Akteure Inhalte und Positionen an ihre Zielgruppen ausspielen: Waren im analogen Zeitalter Wahlwerbungen über Plakate oder Werbeclips im Rundfunk notwendigerweise an eine breite Öffentlichkeit gerichtet, wurde es durch die Digitalisierung und Social-Media-Plattformen ermöglicht, politische Inhalte adressatengerecht zuzuschneiden und sehr spezifisch auszuspielen. Diese als Microtargeting bekannte Form der (Wahl-)Werbung findet spätestens seit dem US-Präsidentenwahlkampf 2008 Verbreitung und Beachtung.

Generative KI wirkt als Katalysator des Microtargetings. Die zunehmende Nutzung dieser Strategie wird unter anderem deshalb zum Problem, da Wahlwerbung in einem öffentlichen Raum stattfinden sollte, um im Rahmen eines politischen Diskurses Widerspruch und Korrektur zu ermöglichen und damit zur Meinungsbildung der Wählerinnen und Wähler beizutragen. Microtargeting, das auf irreführenden oder falschen Informationen aufbaut, kann einen relevanten Einfluss auf die individuellen Wahlentscheidungen haben (Teupen, 2023; Heglich & Medina Serano, 2019).

Es ist deshalb wichtig, dass der Digital Services Act (DSA) von großen Onlineplattformen Register einfordert, mit denen Transparenz über geschaltete Werbung gewährleistet wird: Diese Register stellen gespeicherte Informationen über den Inhalt der Werbung, die (natürlichen oder juristischen) Personen, die für die Werbung verantwortlich sind, oder eben die Information, ob die Werbung auf Nutzergruppen zugeschnitten ist und welche Hauptparameter zu diesem Zweck verwendet wurden, bereit. Allerdings ist es besorgniserregend, dass eine Studie der Mozilla Foundation eine mangelhafte Umsetzung dieser Transparenzpflichten bei allen untersuchten Plattformen festgestellt hat (Mozilla, 2024). Hier ist die öffentliche Hand gefordert, die Einhaltung der europäischen und nationalen Vorschriften konsequent durchzusetzen.

Gezielte Manipulation von Wählerinnen und Wählern durch Deepfakes

Ein anders gelagertes Problem als unabsichtliche Fehlinformationen durch KI-Sprachmodelle und -Chatbots ist die bewusste und gezielte Manipulation von Wählerinnen und Wählern durch KI-Systeme – im Speziellen durch Deepfakes (siehe [Infokasten: Deepfakes](#)). Weltweit steigt die Zahl von Berichten über solche unzulässigen und manipulativen Eingriffe in politische Kampagnen rapide an. Noch unklar ist jedoch, wie genau die Wirkungen solcher Deepfakes zu beurteilen sind. Aussagekräftige Studien dazu liegen noch nicht vor, doch scheint es klar zu sein, dass insbesondere die Glaubwürdigkeit von Informationen und Informationsanbietern im Wahlkampf unter dem wachsenden Einsatz von Deepfakes leiden dürfte (Łabuz & Nehring, 2024). Auch hier liegt neben der Qualitätsverbesserung von Fälschungen ein Problem in deren Skalierbarkeit.

Deepfakes: KI-generierte Audio-, Video- und Bilddateien

Deepfakes sind täuschend echt wirkende Bild-, Audio- oder auch Videoaufnahmen. Sie werden mithilfe von tiefen neuronalen Netzen und Methoden des maschinellen Lernens erzeugt. Dazu verwenden sie je nach verwendetem Verfahren unterschiedliche Eingangsdaten, etwa Texteingaben („prompt“) oder Bild-, Audio- und Video-Dateien, für wiederum unterschiedliche Resultate (Bilder, Videos oder Audios). Einige Deepfake-Methoden können in Echtzeit – beispielsweise während eines laufenden Videoanrufs – eingesetzt werden. Es liegen verschiedene Deepfake-Technologien vor (BSI):

- Bei **Face Swapping** wird das Gesicht einer dargestellten Person in einer Bild- oder Videodatei getauscht. Ausgehend von dem Gesicht einer Person in einem Bild oder Video werden Mimik oder Blickrichtung auf das Gesicht einer anderen Person übertragen.
- **Face Reenactment** oder Puppeteering dagegen passt die (Kopf-)Bewegung oder Mimik einer Person in einem Video an die Bewegungen einer Person an, die diese vor einer Kamera ausführt.
- **Lip Sync** passt die Lippenbewegungen einer Person in einem Video an. So können einer Person durch Unterlegung mit der passenden Audiospur täuschend echte Aussagen in den Mund gelegt werden, die sie in der Realität nie getätigt hat, da die synthetisch erzeugten Lippenbewegungen zum jeweiligen Text passend erstellt werden können.
- **Synthetische Gesichter** können unter anderem mit Generative Adversarial Networks (GANs) erzeugt werden: GANs bestehen aus zwei neuronalen Netzen, die einander entgegengesetzt arbeiten. Ein Generator erstellt dabei auf Basis eines Trainingsdatensatzes einen generierten Inhalt, ein Diskriminator soll dann entscheiden, ob der neue Inhalt Teil des Trainingsdatensatzes ist oder neu generiert wurde. Der Prozess wird so lange wiederholt, bis der Diskriminator die neuen Daten nicht mehr als Fälschung erkennen kann.
- Bei **Text-to-image-, Text-to-video- oder Text-to-speech-Anwendungen** (siehe auch Infokasten: Technologien zum Klonen von Stimmen) erzeugen KI-Algorithmen aus Texteingaben in natürlicher Sprache entsprechend Bilder, Videos oder Audios.
- **Speech-to-speech-Systeme** dagegen verwenden beispielsweise ein Audiosignal als Ausgang.
 - Text-to-speech-Systeme wandeln dabei einen vorgegebenen Text in ein Audiosignal um. Dabei wird der semantische Inhalt des Textes mit den sprecherspezifischen Charakteristika kombiniert, womit Menschen oder Spracherkennungssysteme getäuscht werden können.
 - Bei Voice-Conversion-Anwendungen wird von einem Audiosignal ausgehend ein zweites, manipuliertes Audiosignal erzeugt. Wiederum wird der semantische Inhalt aus dem ursprünglichen Audio übernommen und mit der Charakteristik der sprechenden Person kombiniert.

Während Bildbearbeitungsprogramme wie Photoshop bereits seit den 1990er Jahren existieren, konnte hochwertige Audio- und Videobearbeitung bisher fast ausschließlich von ausgewiesenen Fachkräften umgesetzt werden. Erste automatisierte Bild- und Videogeneratoren traten ebenfalls bereits in den 1990er Jahren auf (van Huijstee et al., 2021). Neu in Bezug auf generative KI ist die rasant besser werdende Qualität der Fakes. Darüber hinaus stehen generative KI-Modelle einer viel breiteren Masse an Nutzerinnen und Nutzern zur Verfügung, die vergleichsweise einfach an überzeugend wirkende Deepfakes gelangen können.

Im politischen Kontext können Deepfakes zu Täuschung und Betrug eingesetzt werden, aber auch für Satire und politischen Protest. Im Kontext von politischer Einflussnahme finden sie unter anderem darin Anwendung, Politikerinnen und Politikern falsche Behauptungen in den Mund zu legen oder sie in schlechtem Licht dastehen zu lassen – hier fungieren sie als zeitgemäße Aktualisierung des „negative campaigning“, das nicht auf die Stärken der eigenen Kandidatinnen und Kandidaten verweist, sondern die Fehler und Schwächen der Gegnerinnen und Gegner markiert und übersteigert. Auch die Anwendung auf bekannte Journalistinnen und Journalisten oder Nachrichtenformate ist in diesem Kontext von Bedeutung, da mit ihnen das Vertrauen in die Seriosität und Wahrheit journalistischer Berichterstattung als Grundlage für eine informierte Wahlentscheidung untergraben werden kann.

ANWENDUNGSBEISPIEL

Deepfake-Bildmaterial zur Beeinflussung afroamerikanischer Wählerinnen und Wähler



Dass Deepfakes im politischen Kontext eingesetzt werden, ist keine dystopische Prognose, sondern längst Realität: Mithilfe von KI wurden beispielsweise im Zuge der Vorwahlen zur US-Präsidentenwahl Anrufe bei Wählerinnen und Wählern im US-Bundesstaat New Hampshire getätigt, in denen angeblich US-Präsident Joe Biden dazu aufforderte, nicht bei den Vorwahlen im Bundesstaat zu wählen und sich das „Kreuz für Wahlen im November aufzusparen“. Die mit KI generierte, täuschend echte Stimme des US-Präsidenten behauptete, dass eine Abgabe der Stimme bei der Vorwahl in New Hampshire zur Folge hätte, dass man bei der eigentlichen Wahl im November 2024 nicht mehr mit abstimmen dürfe – was jedoch nicht den Tatsachen entspricht (BR24, 2024).

Technologien zum Klonen von Stimmen

Computerprogramme zum Klonen von Stimmen gibt es bereits seit Jahrzehnten: Mit sogenannten Text-to-Speech-Algorithmen (TTS; übersetzt: Text zu Sprache) werden Texte in gesprochene Wörter umgewandelt. Ein Beispiel für den bereits seit Längerem hörbaren Einsatz sind automatisierte Durchsagen an Bahnhöfen. Für TTS-Algorithmen gibt es insbesondere zwei Ansätze:

- Verkettende TTS-Systeme nutzen eine Datenbank mit vorab von Menschen eingesprochenen Audioclips aus Wörtern und Tönen, die zu vollständigen Sätzen kombiniert werden können. Die Ergebnisse klingen oft sehr roboterhaft, da die Wort- und Satzbestandteile nicht in einem Fluss stehen können. Damit wird es beispielsweise sehr schwierig, Emotionen auszudrücken.
- Parametrische TTS-Systeme verwenden dagegen ein Modell einer Stimme, das aus Aufnahmen einer Zielperson abgeleitet wird. Diese Modelle lernen charakteristische Sprachmerkmale einer Person und wenden diese auf die Synthetisierung von Sprache auf der Basis schriftlicher Texte an. Methoden des maschinellen Lernens sind eine Grundlage für parametrische TTS-Systeme und finden Anwendung in Alltagselektronik wie Sprachassistenzsystemen (Alexa, Siri und weitere).

Quelle: van Huijstee et al., 2021.

Ein weiteres Beispiel kommt aus Europa: Zwei Tage vor den Nationalratswahlen in der Slowakei im September 2023 wurde eine Audiodatei auf Facebook gepostet, auf welcher scheinbar der Parteichef der liberal-progressiven Partei *Progresívne Slovensko* Michal Šimečka zusammen mit der Journalistin Monika Tódová zu hören war. Die beiden diskutierten darüber, wie man durch Stimmenkauf der marginalisierten Gruppe der Roma die Wahlen beeinflussen könne. Beide Betroffenen dementierten umgehend den Wahrheitsgehalt der Audiodatei, die schließlich auch nachweislich als Fälschung entlarvt wurde. Aufgrund der slowakischen Wahlregeln, die Statements von Politikerinnen und Politikern ebenso wie von Medien kurz vor der Wahl untersagen, war es schwierig, eine Richtigstellung der Fälschung zu verbreiten (Meaker, 2023).

Auch in Deutschland sind Deepfakes inzwischen angekommen: So wurde Bundeskanzler Olaf Scholz Gegenstand eines Deepfakes, in dem er angeblich die Einleitung eines Verbotsverfahrens gegen die Alternative für Deutschland (AfD)¹ ankündigte. Und auch der öffentlich-rechtliche Rundfunk ist betroffen: Ende 2023 machte ein gefälschtes Video die Runde, in dem sich die Tagesschau-Moderatoren Susanne Daubner und Jens Riewa für „dreiste Lügen“ und „Denunzierungen“ entschuldigten. Inzwischen gibt es auch ein erstes „Register“ für den Einsatz von Deepfakes im Umfeld von Wahlkampagnen: Das „[The Wired AI Elections Project](#)“ sammelt Beispielfälle für die Nutzung generativer KI und legt somit den Grundstein für ein Archiv manipulativer Eingriffe in Wahlprozesse.

¹ Das Video war von den Macherinnen und Machern im „Zentrum für politische Schönheit“ wohl nicht als gezielte Desinformation, sondern als Satire gedacht. Das Landgericht Berlin II hat der Gruppierung in einer einstweiligen Verfügung vom 13. Februar 2024 die Verbreitung des Videos untersagt, da das Video wegen einer sogenannten Zuordnungsverwirrung zu einer Verletzung des Namensrechts (§12 BGB, i. V. m. §1004 BGB) führe.

Beeinflussung und Manipulation von politischer Meinungsbildung werden durch falsche Informationen oder diffamierende Bilder, Audios und Videos mit KI-Technologie wesentlich einfacher. Gerade in Verbindung mit der schnellen Verbreitung von (wahren wie falschen) Nachrichten in den sozialen Medien entsteht so durch die Verbreitung von Deepfakes eine Bedrohung für die Demokratie. Einmal im Internet und in den sozialen Medien gelandet, wird es für Betroffene, Plattformbetreiber und staatliche Akteure schwierig, Deepfakes wieder einzufangen. Damit ist auch der Manipulation von extremistischer Seite oder durch externe staatliche Akteure, die ein Land in eine bestimmte politische Richtung lenken oder destabilisieren wollen, Tür und Tor geöffnet.

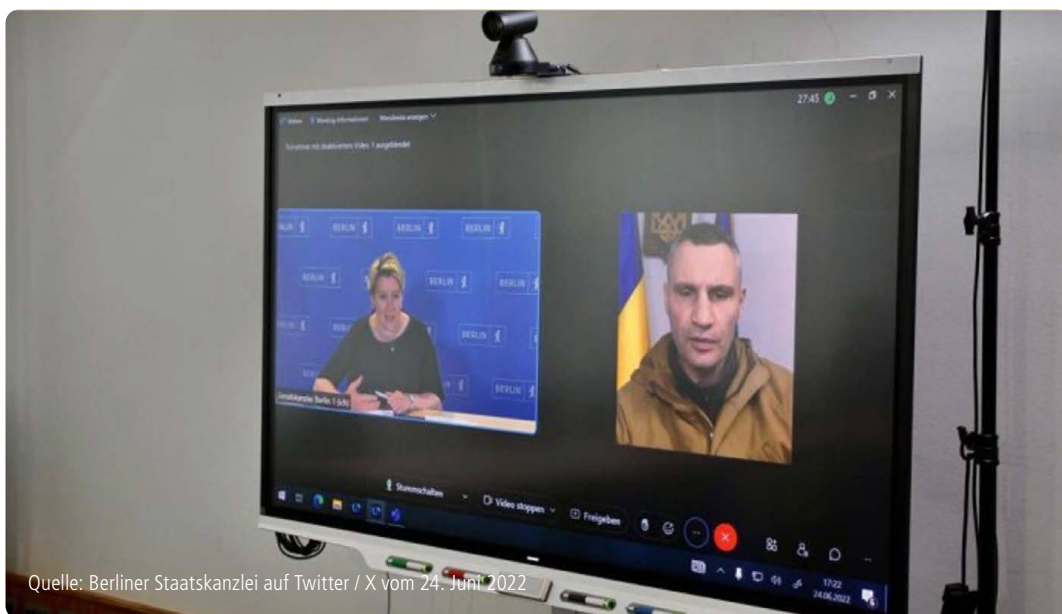
EXKURS

Täuschung von Politikerinnen und Politikern mit Deepfakes

Ein Spezialfall der Deepfake-Problematik ergibt sich mit Blick auf eine spezielle Zielgruppe: nämlich die Mandatsträgerinnen und -träger selbst. So wurden im Sommer 2022 mehrere europäische Bürgermeisterinnen und Bürgermeister, darunter die damalige regierende Berliner Bürgermeisterin Franziska Giffey, der Madrider Bürgermeister José Luis Martínez-Almeida Navasqués und der Wiener Bürgermeister Michael Ludwig, Opfer eines scheinbaren Anrufs des Kiewer Bürgermeisters Vitali Klitschko (Flade, 2022). Auch wenn die Analyse des Vorfalls ergeben hat, dass es wohl keinen Einsatz von generativer KI in diesem konkreten Fall gegeben hat, zeigt der Vorfall dennoch das Angriffspotenzial von generativer KI in diesem Zusammenhang auf:

Mit in Echtzeitszenarien (wie Telefonate oder Videokonferenzen) angewandten Voice-Conversion-Verfahren oder Face/Head-Swapping-Verfahren können feindlich gesinnte Akteure sich Zugang zu engsten politischen Quellen verschaffen und sensible Informationen abgreifen – oder aber fehlerhafte Informationen streuen, um gezielt Einfluss auf politische Entscheidungsprozesse zu nehmen.

Fake-Anruf des Kiewer Bürgermeisters Klitschko bei der Berliner Bürgermeisterin Giffey



Unterdrückung politischer Partizipation durch Erpressung mit Deepfake-Pornographie

Unter Deepfake-Pornographie versteht man pornographische Bilder oder Filme, in welche Gesichter von realen Personen eingefügt wurden – entweder in bestehendes Material („Face Swapping“) oder in neu generierte Bilder und Videos. Deepfake-Pornographie ist seit circa 2017 ein Phänomen: Dort tauchten mehrere gefälschte Videos auf, in denen prominente Schauspielerinnen und Musikerinnen zu sehen waren. Um sie zu erstellen, wurde ein KI-Algorithmus mit Fotos der Betroffenen trainiert, und im Anschluss wurden die Gesichter der Darstellerinnen in den pornographischen Materialien durch computergenerierte Gesichter der Betroffenen ersetzt (Leetaru, 2019).

Auch in politischer Dimension kann Deepfake-Pornographie Einfluss nehmen, etwa durch die Erpressung von in der politischen Öffentlichkeit stehenden Persönlichkeiten (Maddocks, 2020). So kann durch die Androhung der Veröffentlichung kompromittierenden Materials der Versuch unternommen werden, Politikerinnen und Politiker zu bestimmten Inhalten zu zwingen oder sie zum Rückzug aus der Politik zu bewegen. Aber auch die Sabotage journalistischer Berichterstattung ist durch die Zerstörung der Reputation betroffener Journalistinnen und Journalisten durch die (Androhung der) Veröffentlichung von Deepfake-Pornographie ein realistisches Risiko.

2.2 Relevanz von KI-basierten Einflussnahmen

Die dargestellten Beispiele stellen einen Problemaufriss dar, welche Rolle generative KI im Kontext demokratischer Wahlen spielen kann. Angesichts des gesellschaftlichen Hypes um generative KI, der seit dem Aufkommen von ChatGPT Ende 2022 enorm zugenommen hat, finden diese Fälle große (mediale) Aufmerksamkeit. Diese Aufmerksamkeit sollte allerdings nicht zu einer Überbetonung der Problematik von KI-Manipulationen führen, die den Realitäten nicht gerecht wird: Vielmehr muss sich die Frage gestellt werden, welchen (direkten) Einfluss KI-Systeme überhaupt auf Wahlentscheidungen haben.

Hierzu gibt es – Stand Frühjahr 2024 – schlicht noch zu wenig Daten, um eine verlässliche Aussage treffen zu können. Substanzielle, empirische Auswertungen zum Einsatz von KI-Werkzeugen (wie generative KI, große Sprachmodelle etc.) im Umfeld von Wahlvorgängen, die über eine anekdotische Evidenz von medial aufbereiteten Einzelfällen hinausgehen, stehen noch aus. Denn die Wahlzyklen des Jahres 2024 werden die ersten demokratischen Wahlen sein, in denen KI-Einsatz überhaupt im größeren Rahmen beobachtbar sein wird und damit zum Gegenstand wissenschaftlicher Untersuchung werden kann (erste Analysen siehe Łabuz & Nehring, 2024).

Grundsätzlich kann KI-gestützte Wahlbeeinflussung die Abstimmung einzelner Wählerinnen und Wähler beeinflussen und im Aggregat zur Veränderung von Wahlergebnissen führen. Gefälschte Nachrichten und Desinformation können dabei einen direkten Einfluss auf individuelle Wahlentscheidungen nehmen – und zumindest auf lange Sicht das Vertrauen Einzelner in politische und journalistische Institutionen untergraben sowie politische Ansichten bis hin zu Verschwörungserzählungen in einem festgefahrenen Blickpunkt bestätigen und zementieren.

In einem zweiten Schritt müsste sich darauf aufbauend jedoch die Frage stellen, ob sich die KI-produzierte Fehlinformation signifikant von herkömmlichen digitalen Fake News unterscheidet. Anders ausgedrückt: Vermutlich sind eher die sozialen Medien, in denen sich KI-generierte Nachrichten ebenso wie herkömmliche Fake News rasend schnell verbreiten, ein Problem für die demokratische Willensbildung als der KI-Anteil

an diesen Falschdarstellungen. In den Blick gerät somit auch ein allmählich entstehendes „Ökosystem der Desinformation“: Während Werkzeuge generativer KI vor allem zur „Herstellung“ manipulativer Materialien oder zur Modifikation und Multiplikation bereits vorhandener Inhalte genutzt werden, spielen die Plattformen und Mechanismen der Verbreitung künstlich hergestellter Des- oder Missinformation eine bedeutende Rolle.

Ob der Einsatz von generativer KI mit dem Ziel der Beeinflussung von Wahlen dann tatsächlichen Einfluss auf den kollektiven Ausgang demokratischer Wahlen hat, ist derzeit noch unklar und kann zumindest bezweifelt werden (Jungherr, 2023; Dommett, 2024). Statt direkter Effekte auf das Wahlverhalten ist eher mit indirekten Wirkungen zu rechnen. Als sicher muss gelten, dass KI-basierte Desinformation und Falschnachrichten nicht gerade dazu beitragen werden, gesellschaftliche Spaltungsbewegungen der letzten Jahre zu überwinden. Denn es ist ein Ziel der zu politischen Zwecken eingesetzten Desinformation, das Vertrauen in das politische System und seine Institutionen – darunter auch das Vertrauen in Wahlen und Wahlergebnisse – zu untergraben (Jungherr, 2023).

Daran anschließend müssen Falschnachrichten und Desinformation (auf Basis von generativer KI) nicht zwangsläufig einen messbaren Einfluss auf Wahlergebnisse haben, um ins Blickfeld von Entscheiderinnen und Entscheidern in der Politik oder Akteuren in der Medienwelt zu rücken. Denn allein der Versuch der Einflussnahme auf Wahlen durch falsche Nachrichten geht nicht mit demokratischen und journalistischen Standards konform. Wehrhafte Demokratien müssen die Integrität ihrer demokratischen Prozesse schützen – auch wenn sich eine veränderte Zusammensetzung von Parlamenten oder eine veränderte Wahl einzelner Kandidatinnen und Kandidaten nicht im Einzelnen nachweisen lässt.

Abschließend wiederum ergibt sich in Bezug auf die Frage, wie Bürgerinnen und Bürger vor KI-Manipulation geschützt werden müssen, eine gewichtige Überlegung: Wie gelingt der sensible Spagat zwischen notwendigen Maßnahmen zur Reduzierung von KI-basierten Einflüssen auf Wahlen auf der einen Seite und den Kompetenzansprüchen einer mündigen, selbstbestimmten Wählerschaft auf der anderen Seite? Während Informations- und Aufklärungsarbeit über die Möglichkeiten von generativer KI einen wichtigen Beitrag zu einer informierten Öffentlichkeit leisten können, kann eine zu starke Betonung von Manipulationsrisiken dazu führen, die Kapazitäten und Fähigkeiten der Wählerinnen und Wähler, selbst mit den Risiken moderner Technologien umzugehen, zu unterschätzen und gleichsam in einen digitalen Paternalismus zu verfallen.

3 Maßnahmen zur Sicherung der Integrität demokratischer Wahlen

Unabhängig jedoch von der Analyse, welche Relevanz KI-gestützte Beeinflussung von Wahlen bislang tatsächlich hat, oder der (realistischen) Prognose, welche Relevanz sie in Zukunft haben wird, muss es im besonderen Interesse von Demokratien sein, Entwicklungen, die die freie Gesellschaft bedrohen, aufmerksam zu beobachten und ihnen mit geeigneten Maßnahmen zu begegnen.

Rechtliche Maßnahmen: Von Transparenz bis Reichweitenbegrenzung

Es ist deshalb zu begrüßen, dass die Europäische Union in ihrer KI-Verordnung (dem AI Act) KI-Systeme, die dazu verwendet werden können, Wahlen oder Wahlverhalten zu beeinflussen, als Hochrisiko-KI einstuft (Anhang III, Abs. 8(aa) AI Act).

Mit der Klassifikation als Hochrisiko-KI gehen für KI-Systeme, die im Umfeld von Wahlen zum Einsatz kommen, gewisse Auflagen einher – ausgenommen sind Systeme, deren Einsatzzweck rein organisatorischer oder logistischer Natur ist. Darunter fällt unter anderem die Erfüllung von Qualitätskriterien für die Trainingsdatensätze der KI-Systeme, die fehlerfrei und repräsentativ sein müssen und Verzerrungen (engl. Bias), die zu Diskriminierungen führen können, verringern (Art. 10 AI Act). Es wird aber ebenso vorgeschrieben, die Verwendung der KI-Systeme transparent zu machen und Nutzerinnen wie Nutzern die Funktionsweise des Systems mit knappen, vollständigen, korrekten und klaren Informationen zu erläutern (Art. 13 AI Act). Eine transparente Kennzeichnung von KI-Systemen, die direkt mit Menschen interagieren, ist zudem unabhängig von dem ausgehenden Risiko für alle KI-Systeme mit allgemeinem Verwendungszweck (GPAI-Modelle – englische Abkürzung für General-Purpose Artificial Intelligence; deutsch: Künstliche Intelligenz für allgemeine Zwecke) vorgeschrieben (Art. 50 AI Act).

Darüber hinaus fordert auch der Digital Services Act (DSA) der Europäischen Union von „sehr großen“ Online-Plattformen und Suchmaschinen – als solche werden Plattformen ab 45 Millionen Nutzerinnen und Nutzern pro Monat gewertet – Risikoprüfungen und effektive Maßnahmen zur Reduzierung identifizierter Risiken (Art. 35 DSA). Die Bereitstellung von Falschinformationen als Beeinträchtigung der Integrität von demokratischen Wahlen stellt so ein Risiko dar, welches die Betreiber beispielsweise durch transparente Kennzeichnung von KI-generierten Inhalten bekämpfen müssen. Große Sprachmodelle können in diesem Zusammenhang durchaus als Lieferanten für Informationen bewertet werden, wodurch auch für sie die Anforderungen des DSA zusätzlich zu denen des AI Acts bereits eine gewisse Integrität im Prozess demokratischer Wahlen gewährleisten sollten.

Unabhängig von ihrer Erzeugung mit oder ohne generative KI-Systeme muss der Verbreitung von Falschinformationen an sich entgegengewirkt werden: Selbst wenn Falschinformationen entlarvt und gekennzeichnet werden, besteht die Gefahr, dass Richtigstellungen in der Flut an Falschinformationen untergehen. Zudem kann eine nachträgliche Kennzeichnung oder sogar Löschung von einzelnen Nachrichten nur bedingt den entstandenen Schaden verhindern, da sich beispielsweise Falschinformationen bereits auf unterschiedlichen individuellen Kanälen, etwa in Messenger-Gruppen, weiterverbreiten können. Hier greift das Konzept der „Liar’s Dividend“ (etwa: „Lügenprämie“), bei dem die Schöpfer von Falschinformationen stets profitieren, zumindest durch das Wachstum von Unsicherheit, Zweifel und Misstrauen in Einrichtungen öffentlicher Kommunikation (Grünke et al., 2024).

Statt auf einzelne Nachrichten zu zielen, müssten deshalb Accounts oder Bots, die in einem relevanten Ausmaß politische Falschinformationen verbreiten, gesperrt oder in ihrer Reichweite begrenzt werden. Das deutsche Netzwerkdurchsetzungsgesetz (NetzDG) kann hierfür als Beispiel und Grundlage dienen. Bei derartigen Überlegungen darf jedoch nicht vergessen werden, dass Eingriffe in die Meinungs- und Redefreiheit durch staatliche – und auch private Akteure wie soziale Netzwerke – immer begründungsbedürftig sind und einer sensiblen rechtsstaatlichen Absicherung bedürfen, um einem Missbrauch (beispielsweise in Richtung einer staatlichen Zensur unliebsamer Meinungen) konsequent vorzubeugen.

Gesellschaftliche Maßnahmen: Stärkung von KI-Kompetenzen

Für eine effektive Resilienz gegen Manipulation durch generative KI ist es elementar, die einzelnen Menschen als Nukleus von demokratischen Wahlen in den Blick zu nehmen und ihre Aufmerksamkeit bezüglich der stetig steigenden Bedeutung von (generativer) KI ebenso wie ihre Kompetenzen im Umgang mit selbiger zu stärken bzw. aufzubauen. Eine KI-kompetente Öffentlichkeit ist weniger anfällig für die Manipulation durch gefälschte Nachrichten, Bilder und Texte (Bieber, 2024). Entsprechende Angebote in der Erwachsenenbildung oder von politischen Stiftungen sollten gezielt gefördert werden – genauso steht auch die schulische Bildung in der Pflicht. Außerdem kann eine Stärkung der Medienkompetenz im Allgemeinen, die einen kritischen Blick gegenüber Quellen fördert und beispielsweise in einem Abgleich mit anderen, vertrauenswürdigen Quellen mündet, dazu beitragen, den Einfluss von KI-generierter Desinformation zu verringern.

Daneben müssen auch Journalistinnen und Journalisten sowie Medienschaffende für den Umgang mit generativer KI befähigt werden. Dies bezieht sich zum einen auf die aktive Nutzung von generativer KI als Werkzeug journalistischer Arbeit; zum anderen aber auch auf Quellenkompetenz bezüglich der (Weiter-)Verbreitung von textlichen Nachrichten, Bildern oder Videos.

Trotz der großen Bedeutung individueller Kompetenzen im Umgang mit KI in der öffentlichen Kommunikation darf die Verantwortung für die Integrität demokratischer Willensbildungsprozesse nicht an die Bürgerinnen und Bürger ausgelagert werden. Während es zum einen darauf ankommt, Individuen zu einem kompetenten und selbstbestimmten Umgang mit KI zu befähigen, muss es im vorrangigen Interesse jeder demokratischen Gesellschaft sein, Manipulation mit generativer KI durch gesamtgesellschaftliche Maßnahmen vom öffentlichen Diskurs fernzuhalten.

Technische Maßnahmen zum Schutz vor Manipulation durch generative KI

Es gibt verschiedene technische Ansätze, um KI-generierte Desinformation zu verhindern oder wenigstens zu vermindern. Zwei davon sind exemplarisch kryptographische Herkunftsnachweise und Wasserzeichen. Kryptographische Herkunftsnachweise basieren auf digitalen Signaturen, die die Authentizität und Integrität von Inhalten sicherstellen. Um mittels kryptographischer Herkunftsnachweise den Ursprung von Inhalten fundiert bewerten zu können, müssten diese jedoch von Anfang an mitgedacht und die entsprechenden Vertrauensinfrastrukturen aufgebaut werden. Gerade feindlich gesinnte Akteure würden die Nutzung derartiger Herkunftsnachweise natürlich vermeiden (Grünke et al., 2024). Sie können somit nur dazu dienen, die Echtheit verlässlicher Inhalte zu bestätigen und weniger dazu, gezielt verbreitete Falschinformationen aufzudecken.

Darauf aufbauend können synthetisch erzeugte Inhalte, die als solche mit einem Label versehen wurden, schnell mit digitalen Wasserzeichen erkannt werden. Wasserzeichen können dabei sowohl für den Menschen erkennbar als auch für den Menschen unsichtbar, aber für Maschinen lesbar sein. Bei Bedarf können durch Wasserzeichen erkannte Inhalte dann entfernt werden. Dafür müssen Wasserzeichen fälschungssicher und unlöschbar, plattformübergreifend verfügbar sein und mit externen Fact-Checker-Initiativen geteilt werden (Grünke et al., 2024).

Daneben können KI-Methoden selbst eingesetzt werden, um Deepfakes aufzuspüren; so können etwa KI-basierte Systeme zur biometrischen Gesichtserkennung dazu eingesetzt werden, Deepfakes von (prominenten) Personen aufzudecken, indem sie diese schon bei Unregelmäßigkeiten in der Erstellung als solche erkennen. Systeme zur Erkennung der Echtheit einer Stimme sind in der Lage festzustellen, ob der Klang einer Stimme von einer lebenden Person stammt, indem sie Artefakte identifizieren, die im Klang einer Live-Stimme nicht vorhanden sind. Und auch in Videos können KI-Systeme Inkonsistenzen und Artefakte wie unregelmäßige Lippenbewegungen oder unscharfe Übergänge zwischen originalen Bildern und manipulierten Bildern erkennen (van Huijstee et al., 2021).

Dabei ist zu beachten, dass bestimmte KI-Modelle zur Erkennung von Deepfakes und Desinformation selbst unter den AI Act fallen können (etwa große generative KI-Modelle (Large Generative AI Models, LGAIMs)) (Schmitt et al., 2024). Es sollte deshalb im Blick behalten werden, in welcher Form die damit einhergehenden Verpflichtungen den KI-unterstützten Kampf gegen Desinformation berühren.

Das Aufdecken von Deepfakes mittels KI-Technologien hat allerdings auch grundsätzlich seine Grenzen: So können entsprechende Modelle nur an bereits bekanntem Material getestet werden und Angreifende können ihre Systeme für die Zukunft so anpassen, dass existierende Detektionsverfahren nicht mehr funktionieren. Es entsteht so ein Hase-und-Igel-Spiel der KI-Technologien, bei dem nur schwer garantiert werden kann, dass die Technologien zum Schutz von Wahlen besser sind als diejenigen, die die Integrität von Wahlen angreifen (van Huijstee et al., 2021). Demokratische Gesellschaften sind also im Kampf gegen Desinformation mit generativer KI angehalten, sich nicht nur auf technologische Abwehrmöglichkeiten (einer Art) zu verlassen.

Diese Notwendigkeit haben auch große Tech-Unternehmen und Anbieter generativer KI bereits erkannt (Simon et al., 2024; AI Elections Accord, 2024). Führende Anbieter von generativer KI versuchen, ihre Produkte weniger nutzbar beziehungsweise anfällig für Missbrauch zu designen. Entwicklerinnen und Entwickler arbeiten an Möglichkeiten, zuverlässig mit generativer KI erstellte Texte erkennen zu können, und bauen beispielsweise Schranken in Bildergeneratoren ein, die verhindern sollen, dass Bilder von politischen Persönlichkeiten erstellt werden (Daniel, 2024; OpenAI, 2024; Hoskins, 2024; Jasper, 2023).

Forschungsprojekte zur Bekämpfung von Desinformation mit Deepfakes

Auch die Forschung ist bereits aktiv an Lösungsansätzen für das Problem KI-basierter Desinformation. Das BMBF-geförderte interdisziplinäre Verbundvorhabens [NEBULA](#) (Nutzerzentrierte KI-basierte Erkennung von Fake News und Fehlinformationen) forscht etwa an der transparenten, KI-basierten Erkennung von Fake News sowie der zielgruppengerechten Darstellung der Detektionsergebnisse zur Förderung der Medienkompetenz. Zielgruppen sind Beschäftigte in Behörden und Organisationen mit Sicherheitsaufgaben (BOS) und vulnerable Personen (z. B. aufgrund ihres Alters, Bildungshintergrunds oder ihrer Sprachkenntnisse). Im Projekt sollen Demonstratoren entwickelt werden, um Einzelpersonen und Behörden dazu zu befähigen, Fake News und Fehlinformationen selbst erkennen zu können.

Das Forschungsvorhaben [FAKE-ID](#) (Videoanalyse mithilfe künstlicher Intelligenz zur Detektion von falschen und manipulierten Identitäten) will Angriffsmöglichkeiten mit KI-basiertem Identitätsdiebstahl bekämpfen. KI-Algorithmen werden im Projekt daraufhin trainiert, falsche und manipulierte Identitäten zu identifizieren. Darauf aufbauend wird ein Tool für Sicherheits- und Justizbehörden entwickelt, mit dem Beschäftigte Hinweise auf Fälschungen in Bild- und Videodateien von Mitarbeiterinnen und Mitarbeitern überprüfen können. Das System soll Bild- und Videodateien in Echtzeit auf Deepfakes analysieren.

Im Projekt [PREVENT](#) (Trainingsansatz zur Vermittlung von Maßnahmen zur Prävention digitaler Desinformationskampagnen) wird an Methoden geforscht, Ämter, Behörden und Organisationen zu befähigen, Desinformation in sozialen Medien entgegenzuwirken. Die Forscherinnen und Forscher arbeiten an einem Trainingstool, das die Entstehung und Verbreitung von Desinformationen simuliert und die Wirkung von Gegenmaßnahmen testet.

Ziel des BMBF-geförderten „[news-polygraph](#)“-Forschungsbündnisses ist es, für die Medienbranche geeignete Lösungen zur frühzeitigen Erkennung von gefälschten Inhalten beziehungsweise Falschinformationen zu entwickeln. Hierfür wird ein auf KI basierendes System entwickelt, das Journalisten und Medienschaffende intuitiv, transparent und vor allem zeitsparend bei der Erkennung von Falschinformationen unterstützt. Das System soll dabei Echtzeitanforderungen gerecht werden und in mehreren Sprachen verfügbar sein. Die Anwendungsbereiche sind sehr vielfältig und reichen von PR-Agenturen oder Unternehmen bis hin zu öffentlichen Einrichtungen, Versicherungen oder Ermittlungsbehörden.

4 Fazit und Ausblick

Generative KI-Systeme sind in äußerst kurzer Zeit zu einem selbstverständlichen Teil unserer Lebensrealität geworden und werden – kurz- wie langfristig – nicht mehr aus unserer Gesellschaft wegzudenken sein. Dieser Realität müssen sich Wählerinnen und Wähler wie auch die Akteure aus Parteien, Politik und Journalismus rund um demokratische Wahlen stellen.

In diesem Zusammenhang zeichnet es sich bereits ab, dass es Versuche geben wird, die Möglichkeiten von generativer KI für bösartige Zwecke mit dem Ziel der Einflussnahme auf demokratische Prozesse und Meinungsbildung zu missbrauchen. Ob und zu welchem Grad der Ausgang von Wahlen tatsächlich mithilfe von generativer KI manipuliert werden kann, bleibt aber eine offene Frage: Das Superwahljahr 2024 muss den Auftakt für entsprechende wissenschaftliche Studien liefern, aus denen belastbare Ergebnisse hervorgehen werden.

Nichtsdestotrotz sind alle Betroffenen rund um demokratische Wahlen gefordert, sich die – positiven wie problematischen – Potenziale von generativer KI (immer wieder) bewusst zu machen und im Rahmen des jeweils Möglichen und Sinnvollen dem Missbrauch der Technologie vorzubeugen. Dies kann langfristig nicht nur dazu beitragen, die Integrität demokratischer Prozesse zu schützen, sondern ist ebenso notwendig, um Vertrauen in KI-Systeme in der Gesellschaft zu sichern: Somit ein Prozess, der immer wieder neu auszuhandeln und nachzujustieren ist.

Literatur

- AI Elections Accord (2024):** A Tech Accord to Combat Deceptive Use of AI in 2024 Elections. Proposed for public signature and announcement by technology companies at the Munich Security Conference, 16. Februar 2024. Online unter: https://www.aielectionsaccord.com/uploads/2024/02/A-Tech-Accord-to-Combat-Deceptive-Use-of-AI-in-2024-Elections.FINAL_.pdf (Letzter Zugriff: 28.05.2024)
- AlgorithmWatch (2023):** Neue Recherche. ChatGPT und Co: Gefährden KI-getriebene Suchmaschinen demokratische Wahlen?, 5. Oktober 2023. Online unter: <https://algorithmwatch.org/de/bing-chat-wahlen-2023/> (Letzter Zugriff: 02.05.2024)
- Bieber, C. (2024):** „Kann KI Wahlen beeinflussen?“ Gespräch im Podcast „Mit Herz und KI“, Folge 8, 6.6.2024. Online unter <https://mit-herz-und-ki.podigee.io/22-folge-8-wahlen> (Letzter Zugriff: 11.06.2024)
- Bomke, L., Holtermann, F., Meiritz, A. (2024):** Künstliche Intelligenz. Wie KI-Bots in den USA in den Wahlkampf eingreifen, Handelsblatt, 29. Januar 2024. Online unter: <https://www.handelsblatt.com/technik/ki/kuenstliche-intelligenz-wie-ki-bots-in-den-usa-in-den-wahlkampf-eingreifen/100008476.html>
- BR24 (2024):** Anruf von falschem Biden löst Sorge um KI-Fakes im Wahlkampf aus, Bayerischer Rundfunk, 23. Januar 2024. Online unter: <https://www.br.de/nachrichten/deutschland-welt/anruf-von-falschem-biden-loest-sorge-um-ki-fakes-im-wahlkampf-aus,U2AuVG9> (Letzter Zugriff: 02.05.2024)
- BSI: Deepfakes – Gefahren und Gegenmaßnahmen.** Online unter: https://www.bsi.bund.de/DE/Themen/Unternehmen-und-Organisationen/Informationen-und-Empfehlungen/Kuenstliche-Intelligenz/Deepfakes/deepfakes_node.html (Letzter Zugriff: 28.05.2024)
- Chennupati, A. (2024):** The threat of artificial intelligence to elections worldwide: A review of the 2024 landscape. World Journal of Advanced Engineering Technology and Sciences, 12(1), 029-034.
- Daniel, I. (2024):** OpenAI will ChatGPT gegen Desinformation sichern, Zeit Online, 16. Januar 2024. Online unter: <https://www.zeit.de/digital/internet/2024-01/kuenstliche-intelligenz-openai-werkzeuge-fake-news-wahlen-usa> (Letzter Zugriff: 02.05.2024)
- Dommett, K. (2024):** The 2024 Election Will Be Fought on the Ground, Not By AI. Political Insight, 14(4), 4-6. <https://doi.org/10.1177/20419058231218316a>
- Flade, F. (2022):** Deepfakes als Angriffswaffe. Täuschung per Video und Audio, Tagesschau, 27. Juni 2022. Online unter: <https://www.tagesschau.de/investigativ/wdr/deep-fakes-103.html>
- Grünke, P., Litsche, S., Starchenko, S. (2024):** Demokratiekompetenz stärken. Herausforderung Künstliche Intelligenz und die Vermittlung von Medienkompetenz, Gutachten im Auftrag von die medienanstalten – ALM. Online unter: <https://www.die-medienanstalten.de/service/gutachten/kuenstliche-intelligenz/> (Letzter Zugriff: 03.06.2024)
- Heesen, J. et al. (2021):** KI-Systeme und die individuelle Wahlentscheidung – Chancen und Herausforderungen für die Demokratie. Whitepaper aus der Plattform Lernende Systeme, München. https://doi.org/10.48669/pls_2021-1
- Hegelich, S., Medina Serrano, J. C. (2019):** Microtargeting in Deutschland bei der Europawahl 2019. Online unter: https://www.medienanstalt-nrw.de/fileadmin/user_upload/Studie_Microtargeting_DeutschlandEuropawahl2019_Hegelich__1_.pdf (Letzter Zugriff: 28.05.2024)
- Hoskins, P. (2024):** AI: Google restricts Gemini chatbot election answers, BBC, 13. März 2024. Online unter: <https://www.bbc.com/news/business-68551206> (Letzter Zugriff: 28.05.2024)
- Jasper, S. (2023):** How we’re approaching the 2024 U.S. elections, 19. Dezember 2023. Online unter: <https://blog.google/outreach-initiatives/civics/how-were-approaching-the-2024-us-elections/> (Letzter Zugriff: 28.05.2024)
- Jungherr, A. (2023):** Artificial Intelligence and Democracy: A Conceptual Framework, in: Social Media + Society, 9(3). <https://doi.org/10.1177/20563051231186353>
- Łabuz, M., Nehring, C. (2024):** On the way to deep fake democracy? Deep fakes in election campaigns in 2023. European Political Science March 2024, 26. April 2024, <https://doi.org/10.1057/s41304-024-00482-9> (Letzter Zugriff: 28.05.2024)
- Leetaru, K. (2019):** DeepFakes: The Media Talks Politics While The Public Is Interested In Pornography, Forbes, 16. Mai 2019. Online unter: <https://www.forbes.com/sites/kalevleetaru/2019/03/16/deepfakes-the-media-talks-politics-while-the-public-is-interested-in-pornography/?sh=46d4e7cf8461> (Letzter Zugriff: 02.05.2024)

- Maddocks, S. (2020):** ‚A Deepfake Porn Plot Intended to Silence Me‘: exploring continuities between pornographic and ‘political’ deep fakes, *Porn Studies*, 7(4), 415–423, 20. Januar 2020. Online unter: <https://doi.org/10.1080/23268743.2020.1757499> (Letzter Zugriff: 28.5.2024)
- Meaker, M. (2023):** Slovakia’s Election Deepfakes Show AI Is a Danger to Democracy, *Wired*, 3. Oktober 2023. Online unter: <https://www.wired.com/story/slovakias-election-deepfakes-show-ai-is-a-danger-to-democracy/> (Letzter Zugriff: 02.05.2024)
- Mozilla (2024):** Full Disclosure: Stress testing tech platforms’ ad repositories, 16. April 2024. Online unter: <https://foundation.mozilla.org/en/research/library/full-disclosure-stress-testing-tech-platforms-ad-repositories/> (Letzter Zugriff: 28.05.2024)
- Nehring, J., Gabryszak, A. Jürgens, P., Burchardt, A., Schaffer, S., Spielkamp, M., Stark, B. (2024):** Large Language Models Are Echo Chambers. In: Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), pages 10117–10123, Torino, Italia. ELRA and ICCL.
- OpenAI (2024):** How OpenAI is approaching 2024 worldwide elections, 15. Januar 2024. Online unter: <https://openai.com/index/how-openai-is-approaching-2024-worldwide-elections> (Letzter Zugriff: 02.05.2024)
- Pawelec, M. (2022):** Deepfakes and Democracy (Theory). How Synthetic Audio-Visual Media for Disinformation and Hate Speech Threaten Core Democratic Functions, in: *Digital Society*, Vol. 1, Art. 19. <https://doi.org/10.1007/s44206-022-00010-6>
- Plattform Lernende Systeme (2024):** KI Kompakt: AI Act der Europäischen Union. Regeln für vertrauenswürdige KI (Publikationsreihe). Online unter: https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/KI_Kompakt/KI_Kompakt_AI_Act_Plattform_Lernende_Systeme_2024.pdf
- Schmitt, V. et al. (2024):** Implications of Regulations on Large Generative AI Models in the Super-Election Year and the Impact on Disinformation, in: Proceedings of the Workshop on Legal and Ethical Issues in Human Language Technologies@ LREC-COLING 2024, (pp. 28-38). Online unter: <https://aclanthology.org/2024.legal-1.6.pdf> (Letzter Zugriff: 10.06.2024)
- Simon, F., Adami, M., Kahn, G. & Fletcher, R. (2024):** How AI chatbots responded to basic questions about the 2024 European elections right before the vote, Reuters Institute, 6. Juni 2024. Online unter: <https://reutersinstitute.politics.ox.ac.uk/news/how-ai-chatbots-responded-basic-questions-about-2024-european-elections-right-vote> (Letzter Zugriff: 10.06.2024)
- Spring, M. (2024):** Trump supporters target black voters with faked AI images, BBC, 4. März 2024. Online unter: <https://www.bbc.com/news/world-us-canada-68440150> (Letzter Zugriff: 02.05.2024)
- Teupen, C. (2023):** Politisches Microtargeting: zu gefährlich oder legitim? Journalismus und Demokratie, 11. September 2023. Online unter: <https://www.journalismusstudie.fb15.tu-dortmund.de/politisches-microtargeting-zu-gefaehrlich-oder-legitim/> (Letzter Zugriff: 28.05.2024)
- van Huijstee, M. et al. (2021):** Tackling deepfakes in European policy. Publications Office of the European Union. <https://doi.org/10.2861/325063>

Über dieses Whitepaper

Die Autorinnen und Autoren sind Mitglieder der Unterarbeitsgruppe Recht und Ethik der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme. Als eine von insgesamt sieben Arbeitsgruppen thematisiert sie Fragen zur Sicherheit (Security), Zuverlässigkeit (Safety) und zum Umgang mit Privatheit (Privacy) bei der Entwicklung und Anwendung von Lernenden Systemen. Sie analysiert zudem damit verbundene rechtliche sowie ethische Anforderungen und steht in engem Austausch mit allen weiteren Arbeitsgruppen.

Autorinnen und Autoren

Prof. Dr. Christoph Bieber, Center for Advanced Internet Studies

Prof. Dr. Jessica Heesen, Universität Tübingen

Prof. Dr. Armin Grunwald, Karlsruher Institut für Technologie (KIT) /
Institut für Technikfolgenabschätzung und Systemanalyse (ITAS)

Prof. Dr. Frauke Rostalski, Universität zu Köln

Redaktion

Alexander Mihatsch, Geschäftsstelle der Plattform Lernende Systeme

Christine Wirth, Geschäftsstelle der Plattform Lernende Systeme

Impressum

Herausgeber

Lernende Systeme –
Die Plattform für Künstliche Intelligenz
Geschäftsstelle | c/o acatech
Karolinenplatz 4 | 80333 München
www.plattform-lernende-systeme.de

Gestaltung und Produktion

PRpetuum GmbH, München

Stand

Juli 2024

Bildnachweis

Rawpixel/Shutterstock/Titel

Empfohlene Zitierweise

Bieber, C., Heesen, J., Grunwald, A. & Rostalski, F. (2024):
KI im Superwahljahr 2024. Generative KI im Umfeld
demokratischer Prozesse. Whitepaper aus der Plattform
Lernende Systeme, München.

DOI: https://doi.org/10.48669/pls_2024-5

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, der Entnahme von Abbildungen, der Wiedergabe auf fotomechanischem oder ähnlichem Wege und der Speicherung in Datenverarbeitungsanlagen, bleiben – auch bei nur auszugsweiser Verwendung – vorbehalten.

Bei Fragen oder Anmerkungen zu dieser
Publikation kontaktieren Sie bitte Dr. Thomas Schmidt
(Leiter der Geschäftsstelle):
kontakt@plattform-lernende-systeme.de



Über die Plattform Lernende Systeme

Die Plattform Lernende Systeme ist ein Netzwerk von Expertinnen und Experten zum Thema Künstliche Intelligenz (KI). Sie bündelt vorhandenes Fachwissen und fördert als unabhängiger Makler den interdisziplinären Austausch und gesellschaftlichen Dialog. Die knapp 200 Mitglieder aus Wissenschaft, Wirtschaft und Gesellschaft entwickeln in Arbeitsgruppen Positionen zu Chancen und Herausforderungen von KI und benennen Handlungsoptionen für ihre verantwortliche Gestaltung. Damit unterstützen sie den Weg Deutschlands zu einem führenden Anbieter von vertrauenswürdiger KI sowie den Einsatz der Schlüsseltechnologie in Wirtschaft und Gesellschaft. Die Plattform Lernende Systeme wurde 2017 vom Bundesministerium für Bildung und Forschung (BMBF) auf Anregung des Hightech-Forums und acatech – Deutsche Akademie der Technikwissenschaften gegründet und wird von einem Lenkungskreis gesteuert.