Lernende Systeme

DIE PLATTFORM FÜR KÜNSTLICHE INTELLIGENZ

KI-Systeme und die individuelle Wahlentscheidung

Chancen und Herausforderungen für die Demokratie

GEFÖRDERT VOM

Bundesministerium für Bildung und Forschung

acatech
DEUTSCHE AKADEMIE DER TECHNIKWISSENSCHAFTEN

WHITEPAPER

Jessica Heesen et al.
AG IT-Sicherheit, Privacy, Recht und Ethik

Inhalt

Zusammenfassung

Fake News, Desinformationskampagnen, Uploadfilter oder auch Wahlempfehlungs-Apps sind keine neuen Phänomene. Durch den Einsatz von KI-Systemen steigt aber deren Effizienz. Hinzu kommt, dass oftmals nicht ersichtlich ist, dass hier gerade eine Maschine – und kein Mensch – handelt. Die Diskussion über Künstliche Intelligenz (KI) wird deshalb intensiv geführt und ruft in der Debatte zwei entgegengesetzte Reflexe hervor: einerseits weckt ihr Potenzial große Erwartungen, andererseits löst ihr Einsatz zahlreiche Befürchtungen aus wie Kontrollverlust, Machtmissbrauch und Manipulation – vor allem im Hinblick auf die individuelle Meinungsbildung.

Der Fokus des vorliegenden Whitepapers richtet sich daher darauf, was KI-Systeme unter welchen Bedingungen dazu beitragen können, um einerseits die Meinungsbildung im Zuge von demokratischen Wahlen zu unterstützen und andererseits in diesem Kontext eventuell auftretenden Problemen abzuhelfen. Mit einem bedarfs- und nachfrageorientierten Ansatz verfolgt das Autorenteam der Unterarbeitsgruppe Recht und Ethik der Arbeitsgruppe IT-Sicherheit, Privacy, Recht und Ethik der Plattform Lernende Systeme ganz gezielt eine nüchterne und ergebnisoffene Analyse, die sich aufgrund des stetig fortschreitenden gesellschaftlichen wie technologischen Wandels als Momentaufnahme versteht.

In einem ersten Schritt werden Potenziale von KI-Systemen im Zusammenhang mit Wahlen vorgestellt. So können diese Informations- und Mobilisierungsprozesse vereinfachen und damit effizienter gestalten. Drei Einsatzbereiche werden aufgezeigt, bei denen KI teilweise schon zum Tragen kommt und die zugleich das Potenzial aufweisen, mittels KI-Systemen weiterentwickelt zu werden: als Wahlempfehlungs-Apps wie der Wahl-O-Mat oder Wahlkampf-Apps der Parteien und bei Wahlprognosen (Kapitel 2.1). So können Parteimitglieder per Wahlkampf-App den Wahlkampf effizienter koordinieren, indem besonders erfolgversprechende Regionen und Zielgruppen durch automatisierte Auswertungsprozesse identifiziert werden. Oder Algorithmen-basierte KI-Systeme könnten aufgrund der zunehmenden Datenmengen Wahlprognosen bei der Auswertung verbessern. Auch wenn es noch die ein oder andere Schwachstelle wie beispielsweise Sicherheitslücken gibt – wie jüngst bei der CDU-Connect-App aufgezeigt wurde –, zeigen diese Anwendungsbereiche Potenzial für einen nutzbringenden Einsatz von KI im Kontext von Wahlen und der individuellen Meinungsbildung.

Diesen Potenzialen werden im nächsten Schritt mögliche Risiken gegenübergestellt.
Ein Risiko besteht in der Einwirkung auf die Wahlentscheidung bis hin zur Manipulation durch den Einsatz von KI-Systemen. So funktioniert die zunehmend auf KI-gestützte Informationsverbreitung nach anderen Kriterien, als dies bei der Presse oder dem Rundfunk der Fall ist. Über die Nutzung von KI-Systemen für eine personalisierte Ansprache von Wählerinnen und Wählern mittels Microtargeting – ähnlich wie in der Werbung – kann darüber hinaus die Wahlentscheidung beeinflusst werden, ebenso wie durch den Einsatz

von Social Bots. Hinzu kommt, dass mithilfe von KI-Systemen täuschend echt gefälschte Videos, Audios und Bilder, sogenannte Deepfakes, erstellt werden können (Kapitel 2.2).

Diese verzerrenden und manipulierenden Einwirkungen können teilweise durch den Einsatz von KI-Systemen im Rahmen einer Risikomanagement-Strategie mittels Electoral Content Management aufgedeckt sowie gegebenenfalls behoben werden. Möglichkeiten sind hier beispielsweise Uploadfiter sowie das Auffinden und Labeln von Beiträgen, die von Faktencheckern als Desinformation identifiziert wurden. Der Einsatz dieser Instrumente ist jedoch umstritten – unter anderem, weil er wiederum im Wechselspiel zu neuen Problemen führen kann (Kapitel 2.3).

Im Folgenden wird die gesellschaftliche Bedeutung der Einsatzmöglichkeiten von KI-Systemen im Zusammenhang mit Wahlen einer juristischen Bewertung unterzogen (Kapitel 3.1). Das europäische wie das nationale Recht versuchen den Einsatz von KI-Technologien in diesem Kontext bisher durch allgemeine Vorgaben zu beschreiben, die auf ein verträgliches Zusammenleben abzielen. Dennoch existieren beispielsweise für Plattformbetreiber Vorgaben wie das Netzwerkdurchsetzungsgesetz (NetzDG), die zur Löschung von rechtswidrigen Informationen verpflichten. Im Weiteren werden die skizzierten Veränderungsprozesse in einem größeren Rahmen eingeordnet. Denn die KI-gestützte Verbreitung von Desinformation wirkt bereits gegenwärtig auf die Wahlentscheidungen ein und bedarf daher gesteigerter Aufmerksamkeit – auch mit Blick auf zukünftige mögliche Gegenmaßnahmen. Aufgrund der zunehmenden Verfügbarkeit von Daten und deren Verarbeitung wird sich demzufolge zunehmend auch unser Verständnis von „Staat" und „Demokratie" infrage stellen (Kapitel 3.2).

Um die Chancen von KI-Systemen im Kontext von Wahlen und für eine individuelle Meinungsbildung zu stärken und Risiken abzuschwächen, formuliert das Autorenteam mögliche Gestaltungsoptionen (Kapitel 4.3). So sollten beispielsweise Plattformbetreiber transparente und niederschwellige Beschwerdemechanismen installieren. Politische Verantwortliche sollten Microtargeting gesetzlich weiter einschränken (per Kennzeichnungspflicht). Forschungs- und forschende Nichtregierungsorganisationen sollten KI-Anwendungen zur Detektion von Falschinformationen weiterentwickeln. KI-Entwickelnde sollten bei der Entwicklung der Produkte verantwortungsvoll vorgehen und die (kritische) Öffentlichkeit sollte ihre medialen Kompetenzen zur Bewertung von Informationen weiter auf- und ausbauen. Grundlegend sind darüber hinaus Transparenz, Vertrauenswürdigkeit und ein verantwortungsvoller Umgang.

1. Einleitung

„Hype-Charakter" der öffentlichen Debatte um KI als Schlüsseltechnologie

In der Digitalisierung hat sich in den letzten Jahrzehnten die Gesellschaft stark verändert und ist weiter in rascher Entwicklung. Fake News, Desinformationskampagnen, Uploadfilter oder auch Wahlempfehlungs-Apps sind keine neuen Phänomene in der digitalen Kommunikation. Durch den Einsatz von KI-Systemen steigt aber ihre Effizienz und Bedeutung. Hinzu kommt, dass der Einsatz von KI häufig nicht den Transparenzansprüchen einer demokratischen Öffentlichkeit genügt und maschinelles Handeln das menschliche ablöst. Die Diskussion über Künstliche Intelligenz (KI) wird deshalb intensiv geführt. Die öffentliche Debatte nimmt gelegentlich jedoch einen „Hype-Charakter" an und schwankt zwischen zwei Extremen: weitreichende Hoffnungen auf bessere Systeme auf der einen und vielfältige Befürchtungen wie Kontrollverlust, Überwachung, Abhängigkeit und Diskriminierung auf der anderen Seite. Dieser „Hype-Charakter" der Debatte lässt sich auch am Verhältnis von Demokratie und KI erkennen. Hier gelten KI-Systeme in vielen Fällen vorrangig als Bedrohung der Demokratie. Vor allem ist hier die befürchtete Manipulation des Wählerwillens zu nennen, etwa durch lernfähige Social Bots, durch individuelle Wählerbeeinflussung unter Datenmissbrauch (vgl. den Cambridge Analytica-Skandal 2018) oder mittels der Beeinflussung von Wahlen durch externe Geheimdienste (vgl. die letzten beiden Präsidentschaftswahlen in den USA; vgl. Süddeutsche Zeitung; Deutschlandfunk). Andererseits werden immer wieder Hoffnungen geäußert, dass KI die Demokratie „fördern" kann (vgl. Johnson 2020; Polonski 2017), bis hin aber auch zu spekulativen Überlegungen, dass KI die Demokratie in einiger Zeit schlichtweg ersetzen könne, weil sie dem politischen Geschäft durch optimierende Berechnungen überlegen sei (vgl. Bischof 2019; Jonsson & de Tena 2021).

Das Ziel: eine sachliche Perspektive auf die Debatte

Bei vielen „KI-Enthusiasten" ist die Haltung eines *technology push* weitverbreitet. Nach dieser Logik können KI-Systeme, allein weil sie verfügbar und flexibel sind, praktisch jede Aufgabe lösen und sollten sehr breit – also auch im Zusammenhang mit der Verbesserung demokratischer Wahlvorgänge – zum Einsatz kommen. Dieser Sichtweise stehen weitreichende Befürchtungen möglicher demokratiegefährdender Folgen auf einer sehr allgemeinen Ebene gegenüber, in denen – dies ist die Parallele zu den Enthusiasten – die Befürchtungen schon allein deswegen entstehen, nur weil KI dabei im Spiel ist. Der zentrale Beitrag dieses Papiers liegt darin, diese kontrovers geführte Debatte zu strukturieren und die Veränderungsprozesse sachlich zu analysieren: Entgegen pauschalen Erwartungen wie Befürchtungen fragt das Autorenteam ganz konkret, was KI-Systeme unter welchen Bedingungen beitragen können, um die Meinungsbildung im Zuge von demokratischen Wahlen einerseits in bestimmten Hinsichten zu unterstützen und andererseits eventuell auftretenden Problemen abzuhelfen. Ein bedarfs- oder nachfrageorientiertes Vorgehen

(*demand pull*) ist daher gefordert. Hierbei fokussiert das Autorenteam den Prozess der Debatte und Meinungsbildung vor der Wahl – der konkrete Wahlvorgang selbst ist kein Betrachtungsgegenstand der Analyse. Denn Pauschalurteile – im Rahmen des aktuellen Hypes – sind nicht hilfreich. Stattdessen analysiert das Autorenteam ergebnisoffen und auf wissenschaftlicher Basis KI-basierte technische Möglichkeiten zur Unterstützung der Meinungsbildung im Kontext von Wahlen.

2. KI-Systeme im Zusammenhang mit Wahlen

KI-Systeme könnten im Zusammenhang mit Wahlen theoretisch vielfältig eingesetzt werden: als Wahlempfehlungs-Apps, um Orientierung zu geben, oder als App, um Parteien bei der Organisation des Wahlkampfes zu unterstützen. Möglicherweise kann ihr Einsatz auch zu genaueren Wahlprognosen führen. KI-Systeme werden zum gegenwärtigen Zeitpunkt allerdings vor allem zur Einwirkung auf die Wahlentscheidung in der Plattformkommunikation (insbesondere in Sozialen Medien) eingesetzt. KI kann jedoch auch dazu beitragen, manipulativen oder missbräuchlichen Einwirkungen auf die Wahlentscheidung entgegenzuwirken. Kapitel 2 gibt einen Überblick über die vorgestellten Einsatzmöglichkeiten.

2.1 Potenziale: Vereinfachung von Informations- und Mobilisierungsprozessen

Im Zusammenhang mit Wahlen weisen KI-Systeme zumindest theoretisch einige Potenziale auf, um Informations- und Mobilisierungsprozesse zu vereinfachen und effizienter zu gestalten (siehe Abbildung 1). Aktuell werden diese Potenziale jedoch überwiegend kaum realisiert, da KI-Systeme bis dato hier meist nur sehr punktuell eingesetzt werden.

Abbildung 1: Potenziale von KI-Systemen bei Wahlen

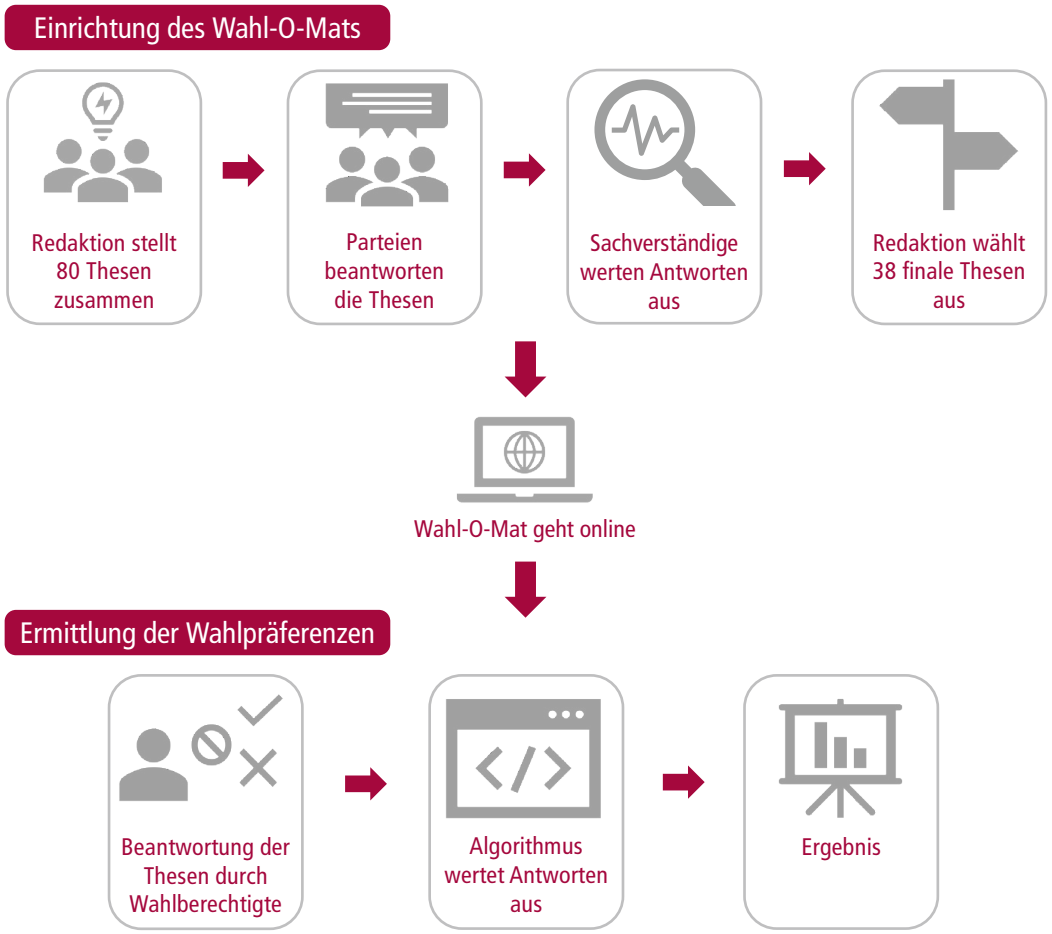
2.1.1 Wahlempfehlungs-Apps

Bei der Information von Wahlberechtigten finden Ansätze der KI in digitalen Medien bislang kaum Anwendung. Daher ist zu fragen, welche KI-Mechanismen und KI-Systeme künftig im Vorfeld von Wahlen eingesetzt werden könnten, um Informations- und Entscheidungsprozesse zu unterstützen. Ein Anwendungsbeispiel, im Rahmen dessen KI verstärkt eingesetzt werden könnte, sind Wahlempfehlungs-Apps – dies ist insofern naheliegend, da „recommender systems" (wie individualisierte Suchmaschinenergebnisse oder Vorschläge von Streaming-Diensten) ein typischer Anwendungsbereich von „machine learning"-Verfahren sind. Politische Empfehlungsmaschinen werden zumeist durch neutrale Dritte entwickelt und zur Verfügung gestellt. Sie verzichten noch auf automatisierte Prozesse. Die bekannteste App ist vermutlich der von der Bundeszentrale für politische Bildung betriebene Wahl-O-Mat. Daneben existieren noch einige weitere Wahlempfehlungs-Apps (vgl. Garzia & Marschall 2019).

Die Anwendungen ermitteln aus der Einstellung der Wahlberechtigten zu bestimmten Programminhalten Positionen, die Zustimmung oder Ablehnung entweder zu einer Partei oder auch zu einer oder einem Kandidierenden zum Ausdruck bringen. Hierfür werden zunächst programmatische Aussagen von Kandidierenden oder auch ganzer Parteien ausgewertet. Diese Auswertung erfolgt in der Regel händisch durch ein zuständiges Redaktionsteam, das häufig von Fachleuten unterstützt wird – also: nicht automatisiert. In einem zweiten Schritt kommen dann die Nutzenden und ein (einfacher) Algorithmus ins Spiel: Die Nutzenden geben an, ob sie bestimmten Aussagen eher zustimmen oder diese eher ablehnen. Der Algorithmus wertet dann den Grad der Zustimmung der Nutzenden zu bestimmten Aussagen aus, indem er die „Nähe" beziehungsweise „Ferne" zu den Aussagen der untersuchten Kandidierenden oder auch Parteien berechnet. Die Berechnungsverfahren (Algorithmen) unterscheiden sich von Anwendung zu Anwendung und werden nicht in allen Anwendungsfällen offengelegt.[1] Für die Funktionsweise des Wahl-O-Mat siehe Abbildung 2.

1 Ein Rechenbeispiel ist hier zu finden: https://team-tomorrow.org/2021/01/25/algorithmus/.

Abbildung 2: Funktionsweise des Wahl-O-Mat



Die aktuellen Wahlempfehlungsanwendungen enthalten nur wenige automatisierte Verfahren und noch kaum maschinenbasierte Lernprozesse. So lernen diese „Voting Advice Apps" aktuell weder aus der Gesamtheit der aggregierten Daten noch in Bezug auf einzelne Nutzende dazu. Deshalb beginnt jeder einzelne Nutzungsvorgang wieder „bei null". Auch während eines Empfehlungsprozesses ist der Einsatz zusätzlicher automatisierter Elemente denkbar: Die Führung durch das Programm könnte in Begleitung eines Chatbots geschehen, der die einzelnen Aussagen mit Beispielen erläutert oder auf Fragen der Nutzenden eingehen kann. Durch ein Monitoring solcher Dialoge wird die Qualität der Benutzerführung überprüft und es können Impulse für die weitere Entwicklung des Empfehlungssystems abgeleitet werden.

Zukünftig sind solche Weiterentwicklungen beziehungsweise der Einsatz fortgeschrittener KI-Systeme denkbar. So könnten die Systeme als Lernende Systeme (siehe Infobox) gestaltet werden und über einen intensiven Dialog mit den Nutzenden stärker auf die individuellen Bedürfnisse in der jeweiligen Nutzungssituation eingehen (etwa durch einen entsprechend programmierten Chatbot als „digitales Assistenzsystem", das eigenständig

die Wahlkampfkommunikation verfolgt und für die einzelnen Nutzenden auswertet). Eine Weiterentwicklung der Wahlempfehlungssysteme könnte auch in Richtung eines „digital twin" zielen. Dies würde bedeuten, dass ein persönliches (lernendes) Assistenzsystem die Online-Aktivitäten eines Nutzenden begleitet und aufzeichnet. Dabei ließe sich entlang der Nutzung digitaler Plattformen, dem Gebrauch von Online-Informationsangeboten (z. B. Nachrichtenmedien, Mediatheken), digitalen Kaufentscheidungen, Besuchen von Informationsangeboten politischer Akteure oder möglicherweise auch der Auswertung von Kontaktnetzwerken eine Datengrundlage erstellen, die zur Modellierung politischer Präferenzen genutzt wird. Auf Basis eines solchen Profils könnten Nutzende dann wahlrelevante Informationen erhalten, die vorab durch das personalisierte KI-System gefiltert und sortiert worden sind. Anschließend könnte ein solcher „politischer Zwilling" die Wahlberechtigten im Vorfeld der Stimmabgabe mit passenden Materialien versorgen und einen gut informierten Wahlvorgang unterstützen. Allerdings ist gegenwärtig unklar, ob für solch eine Anwendung eine ausreichende Transparenz hergestellt werden kann.

Lernende Systeme
Lernende Systeme sind Maschinen, Roboter und Softwaresysteme, die abstrakt beschriebene Aufgaben auf Basis von Daten, die ihnen als Lerngrundlage dienen, selbstständig erledigen, ohne dass jeder Schritt spezifisch vom Menschen programmiert wird. Um ihre Aufgabe zu lösen, setzen sie von Lernalgorithmen trainierte Modelle ein. Mithilfe des Lernalgorithmus können viele Systeme im laufenden Betrieb weiterlernen: Sie verbessern die vorab trainierten Modelle und erweitern ihre Wissensbasis. Lernende Systeme basieren auf Methoden der Künstlichen Intelligenz (KI), genauer: des maschinellen Lernens. Vor allem durch die Fortschritte im Deep Learning entwickelten sich Lernende Systeme in den letzten Jahren zum dynamischsten Bereich der KI-Forschung und -Anwendung (vgl. Plattform Lernende Systeme).

Ähnliche Mechanismen gehören bereits jetzt zur Realität der Online-Kommunikation, allerdings erreichen sie die Nutzenden eher unbemerkt – so bieten zum Beispiel Suchmaschinen personalisierte Suchergebnisse und eine Personalisierung der Ergebnisse (z. B. Anpassung der Überschrift) an, die sich u. a. an bisherigen Suchanfragen auf einem Endgerät orientieren. Auch auf digitalen Plattformen verrichten „Empfehlungsalgorithmen" ihre Dienste („Personen, die dieses Buch gelesen haben, interessierten sich auch für…") oder schlagen unmittelbar neue Inhalte zum Abruf vor (z. B. Video-Plattformen oder Streamingdienste). Manche digitalen Plattformen nutzen algorithmische Steuerungssysteme bereits als zentrale Mechanik zur Navigation durch die Inhalte (z. B. TikTok), wobei jeder Nutzungsvorgang den Einstiegspunkt für den jeweils nächsten darstellt. Die produktive Einbettung solcher Mechanismen in politische Informationsprozesse stellt ein mögliches Einsatzfeld für KI-Systeme in der politischen Bildung dar, wobei die Entwicklung transparenter, datenschutzkonformer, vertrauenswürdiger und fairer Assistenzsysteme hierbei eine ganz wesentliche Aufgabe ist.

Dass Wahlempfehlungs-Apps komplexe Sachverhalte auf Ja-/Nein-Fragen herunterbrechen, kann auch zu einem Kontextverlust in der Darstellung führen. Sie könnten in der Konsequenz zu einer Verfälschung der individuellen Wahlentscheidung führen. Werden KI-Systeme in Wahlempfehlungs-Apps eingesetzt, so muss Transparenz über deren Einsatz, die verwendeten Daten sowie die Funktionsweise des Algorithmus herrschen (siehe auch Kapitel 4.3). Gleichzeitig sollten auch potenzielle Risiken, wie beispielsweise Datenschutzprobleme sowie unbeabsichtigte Verzerrungen oder mutwillige Manipulationen des Algorithmus sowie Beeinflussungen durch den Algorithmus, nicht außer Acht gelassen werden (siehe auch Kapitel 2.2).

2.1.2 Organisation des Wahlkampfes

Der digitale Werkzeugkasten für die politische Kommunikation und Organisation verändert sich ständig, Jungherr et al. (vgl. 2020) fassen die verschiedenen Möglichkeiten unter dem Dachbegriff „Retooling Politics" zusammen.[2] KI-Systeme gehören zum gegenwärtigen Zeitpunkt in Deutschland aber noch kaum zur Ausstattung der Online-Wahlkämpfenden. Allerdings liefern sogenannte „Wahlkampf-Apps" einen Ausgangspunkt für die Entwicklung und den Einsatz von KI-Systemen und anderen automatisierten Verfahren in der Wahlkampforganisation. Bislang setzen vier im Bundestag vertretene Parteien eigene Apps im Wahlkampf ein (CDU: „Connect", Grüne: „Wahlkampf-App", SPD: „Tür-zu-Tür", Die Linke: „Partisanin"; Stand: April 2021). Ziel der Entwicklung und des Einsatzes solcher Apps ist die verbesserte Kommunikation mit Mitgliedern im Rahmen konkreter Wahlkämpfe.[3]

Wie aber funktionieren diese Apps? Zumeist enthalten die Apps Basisinformationen über die Kampagne und bieten konkrete Anleitung für das Sammeln weiterer Daten über potenzielle Unterstützende und Wahlberechtigte. Insbesondere sollen Informationen zum Verlauf der „Haustürgespräche" protokolliert werden, um Rückschlüsse über die Wählerunterstützung in bestimmten Gebieten zu erhalten. Personenbezogene Daten werden dabei in der Regel nicht gespeichert. Deutlich sichtbar wurde das Innenleben solcher Anwendungen durch den Hinweis auf ein Sicherheitsleck bei der CDU-Wahlkampf-App „Connect" im Mai 2021. Die Software-Entwicklerin Lilith Wittmann hatte darauf hingewiesen, dass zahlreiche bei der Nutzung eingegebenen Daten unzureichend geschützt seien und durch die Kombination von anonymisierten Informationen, Notizen zum Gesprächsverlauf und dem konkreten Einsatzort Rückschlüsse auf einzelne Personen möglich seien (vgl. Wittmann 2021).

Einige dieser Apps nutzen das Prinzip der „Gamification" (siehe Infobox), um die Anwendenden zum Einsatz der App zu motivieren. Die wenigsten dieser Prozesse sind automati-

2 Es existieren bereits Diskussionen zum Verhalten der Parteien im digitalen Raum und den damit einhergehenden Chancen und Risiken (vgl. z. B. D64 2019).

3 Die Initiative geht hierbei meist von Parteiuntergliederungen oder einzelnen Entwickelnden aus – nicht unbedingt durch die Parteizentrale.

siert und zum gegenwärtigen Zeitpunkt ist noch unklar, inwiefern die über die Apps erhobenen Daten organisationsintern weiterverwendet werden. Werden solche Datensätze systematisch erfasst – und davon ist nach einem mehrjährigen, wenn auch nicht flächendeckenden Einsatz solcher Apps auszugehen –, dann lassen sich auf dieser Basis automatisierte Auswertungsprozesse mithilfe von Verfahren des maschinellen Lernens entwickeln. Diese können dazu beitragen, die Kampagnentätigkeit von Parteien zu optimieren. Dies gelingt gegenwärtig bereits beispielsweise, indem die Algorithmen besonders erfolgversprechende Regionen und Zielgruppen identifizieren und das Wahlkampfteam so seine Ressourcen effizienter einsetzen und bei Bedarf auch spezialisierte Kampagneninhalte entwickeln und verteilen kann.[4]

Gamification
Unter Gamification (dt. „Spielifizierung") wird die Einbettung von spielerischen Elementen in spielunverwandte Kontexte verstanden. Ziele sind eine gesteigerte Interaktion mit den Nutzenden und damit eine höhere Motivation. Beispiele für solche Elemente sind die Vergabe von Punkten für die Erledigung unterschiedlicher Aufgaben oder Rankings, in denen sich die Nutzenden miteinander vergleichen können (z. B. in Hinsicht von Jogging-Leistungen). Gamification kann auch zur Förderung politischer Partizipation von Bürgerinnen und Bürgern eingesetzt werden. Für eine kritische Betrachtung der Gamification politischer Partizipation siehe Loh 2019.

Nicht zuletzt machen die Sicherheitslücken wie bei der CDU-Connect-App deutlich, dass solche Werkzeuge dazu beitragen, Datenbanken aufzubauen, die Informationen sowohl über die aktiven App-Anwendenden enthalten als auch mithilfe der App erhobene Daten zu potenziellen Wählerinnen und Wählern. Nach dem Erkennen der Sicherheitslücke wurde die CDU-App am 12. Mai 2021 vorläufig abgeschaltet. Zu diesem Zeitpunkt befanden sich persönliche Daten von 18.500 Wahlkampfhelferinnen und Wahlkampfhelfern sowie 1.350 Datensätze weiterer Personen, die im Haustürwahlkampf angeworben wurden, in der Datenbank (vgl. Wittmann 2021).

2.1.3 Wahlprognosen

KI-Systeme können auch zur Entwicklung und Verbesserung von Wahlprognosen eingesetzt werden. Adressaten von Wahlprognosen können zwei unterschiedliche Gruppen sein. Erstens: die Wahlkampforganisationen selbst, die ihre Anstrengungen aufgrund von Wahlprognosen beispielsweise auf Gebiete, in denen ein besonders knapper Ausgang vorhergesagt wird, verstärkt konzentrieren. Zweitens: die Wahlberechtigten, die ihre Stimmabgabe – je nach Ausgestaltung des Wahlsystems – möglicherweise ebenfalls von Vorhersagen abhängig machen. Insbesondere im Vorfeld der US-Präsidentschaftswahl 2020

4 Ein solches Angebot ist beispielsweise auf der Website wahlkreisprognose.de zu finden.

sind solche Verfahren vorgestellt worden, die in Konkurrenz zu den „herkömmlichen" Panel-Studien und den darauf basierenden Modellierungen treten sollen, wie das Tool von expert.ai oder das Tool von Makse und Zhou (vgl. 2019).

In der Literatur wird jedoch das Buzzword „Künstliche Intelligenz" vergleichsweise unpräzise im Zusammenhang mit unterschiedlichen Analysetechniken genutzt. Häufig ist hierbei die Nutzung unterschiedlicher Datenquellen für die Modellierung von Prognosen gemeint – in diesem Fall treten häufig Kommunikationsinhalte von Social-Media-Plattformen wie Twitter oder Facebook an die Stelle herkömmlicher Umfragedaten und Gewichtungsmodelle. Automatisierte („lernende") Auswertungsverfahren sollen durch eine „sentiment analysis" die „Stimmungslage" der Postings erkennen und daraus positive oder negative Bewertungen einer oder eines Kandidierenden durch die Nutzenden ableiten (vgl. Makse & Zhou 2019; Zhou et al. 2021). Die „Prognosequalität" solcher Verfahren kann auf Basis der bisher vorliegenden Untersuchungen nicht seriös beurteilt werden. Eine direkte Verbindung zwischen „social media traffic" und der Wahlentscheidung herzustellen, scheint mit Blick auf das Forschungsdesign als zumindest gewagt.

Es wird durchaus versucht, den Datensatz zu bereinigen und etwa automatisierte Kommunikationsakteure („bots") sowie fehlerhafte oder fehlgeleitete Informationen („fake news", „spam") zu identifizieren und auszuschließen. Das Einbeziehen von Zensusdaten in die Wahlvorhersage soll Fehler reduzieren, die durch die Selektionseffekte der digitalen Plattformen entstehen können. Allerdings bleibt offen, ob – und wenn ja, wie – gewährleistet werden kann, dass die analysierten Inhalte überhaupt von wahlberechtigten Personen oder registrierten Wählerinnen und Wählern stammen. Angesichts vermehrt kritisch hinterfragter traditioneller Meinungsumfragen bleibt diese Linie der Modernisierung datengestützter Prognoseverfahren dennoch beachtenswert und scheint vor allem mit Blick auf die schnellere Anpassungsfähigkeit Algorithmen-basierter Modelle ausbaufähig.

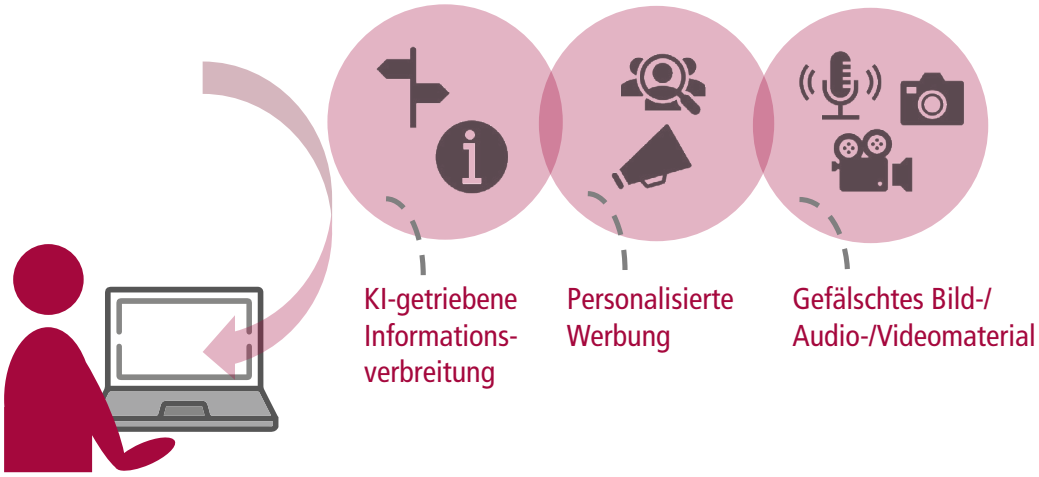
2.2 Herausforderungen: Einwirkungen auf die Wahlentscheidung mithilfe von KI-Systemen

Eine mögliche Gefahr besteht in der unbewussten Manipulation der individuellen Wahlentscheidung (der Meinungsbildung vor der Wahl, des Wahlkampfes oder der Motivation der Stimmabgabe) mithilfe von KI-Systemen. Der Prozess der Wahl direkt (also die Stimmabgabe und Auszählung an sich) wird nur sehr selten mit KI-Verfahren in Verbindung gebracht.[5] Stattdessen sind nicht zuletzt die Plattformen selbst als potenzielle Gefahr für eine unbeeinflusste Wahl in den Fokus geraten: Sie verfügen über die meisten Informationen und einen direkten Zugang zu allen Inhalten. Facebook hat in einer umstrittenen Studie selbst gezeigt, dass die Plattform mit relativ einfachen Mitteln – in diesem Fall ein

5 Der Prozess der Wahlen wurde verschiedentlich bezüglich technischer Risiken untersucht – vor allem im Hinblick auf Wahlcomputer, die in Deutschland aber nicht eingesetzt werden. Für Deutschland wurden lediglich potenzielle Gefahren in der Datenverarbeitung bei der Stimmauswertung gesehen (vgl. z. B. Chaos Computer Club e. V. 2017). All das betrifft aber klassische Fragen der IT-Sicherheit und steht nicht im direkten Zusammenhang mit neuen Manipulationsmöglichkeiten durch KI-Systeme.

freiwilliger „I voted"-Button, der Freundinnen und Freunden angezeigt wird – die Wahlbeteiligung erhöhen kann (vgl. Bond et al.). Auch wenn in diesem Fall die Nutzenden direkt – und keine KI-Systeme – involviert und die Effekte minimal waren, ist es ein gutes Beispiel für die Macht der Plattformen. Deshalb tauchen die Plattformen selbst auch regelmäßig in Analysen über die Gefahren für Wahlen auf (vgl. z. B. Helbing 2015). Bei der indirekten Beeinflussung einer Wahl müssen vor allem – aber nicht nur mit Blick auf Social-Media-Plattformen – drei Aspekte betrachtet werden: Erstens: KI zur Informationsverbreitung. Zweitens: KI zur Erstellung von Persönlichkeitsprofilen für personalisierte Werbung. Drittens: KI zur Erstellung von gefälschtem Bild-, Audio- oder Videomaterial (siehe Abbildung 3).

Abbildung 3: Einwirkungen auf die Wahlentscheidung mithilfe von KI-Systemen



2.2.1 KI-getriebene Informationsverbreitung

Suchmaschinen und soziale Netzwerke verbreiten Informationen nach anderen Kriterien, als dies bei der Presse und dem Rundfunk der Fall ist. Suchmaschinen wie soziale Netzwerke stellen hauptsächlich Inhalte dritter, diverser Urheber dar und bieten wenig eigene Inhalte an. So können auch nicht-professionell erstellte Informationen schnell eine große Reichweite erzielen. Zudem werden die Informationen durch Algorithmen gesichtet und gewichtet. Die Relevanz einzelner Inhalte wird hierbei in Bezug auf die Interessen der Nutzenden, aber auch auf die Interessen der Plattformen und ihrer Werbekundinnen und -kunden bestimmt (vgl. von Ungern-Sternberg et al. 2018). Die Nutzenden tragen durch Likes, Retweets, Sharing und ähnliche Funktionen zur Bewertung und Verbreitung der Inhalte bei.

Vor allem diese statischen Funktionsweisen der Plattformökonomien (vgl. Gerlitz & Helmond 2013) lassen sich leicht automatisiert beeinflussen. Automatisch ablaufende Skripte können Inhalte mit „Likes" versehen oder über die entsprechenden Funktionen der Plattformen verteilen (Share, Retweet etc.) Dafür ist noch keine KI nötig. Allerdings sind solche einfachen, automatisiert ablaufenden Prozesse durch die Plattformbetreiber inzwischen leicht zu erkennen. Es geht also darum, das manipulative Verhalten so aussehen zu lassen, als stamme es von einem Menschen. Hier kommt nun KI ins Spiel: Mit ihrer Hilfe kann auf bestimmte Muster in der Kommunikation anderer reagiert werden. Damit können Inhalte abgeändert oder sogar ganz neu generiert werden. Dies geschieht, um beispielsweise den Eindruck zu erwecken, eine bestimmte Position hätte sehr viel Unterstützung, oder auch um Falschinformationen überhaupt erst Geltung zu verschaffen. Auch wenn KI-Systeme inzwischen in der Lage sind, selbst relativ überzeugende Inhalte zu generieren (Stichworte: Textgeneratoren, Deepfakes, Chatbots), wird die Gefahr (vorerst) vor allem in dieser relativ kleinteiligen Unterstützung der Verbreitung menschlich erzeugter Inhalte gesehen (vgl. Kind et al. 2017; von Ungern-Sternberg et al. 2018).

2.2.2 Erstellung von Persönlichkeitsprofilen für personalisierte Werbung

Eine weitere Möglichkeit, mithilfe von KI-Systemen Wahlen zu beeinflussen oder sogar zu manipulieren, liegt in der Erstellung und Verwendung personalisierter Inhalte: der Personalisierung/dem Microtargeting (siehe Infobox). Diese Erstellung von Persönlichkeitsprofilen wird vor allem für die personalisierte Werbung eingesetzt und ist damit Teil des zentralen Geschäftsmodells vieler digitaler Plattformen. Aus diesen Möglichkeiten der Werbung wird immer wieder eine mögliche Gefahr für die Manipulation im Bereich des Politischen und eben auch der Wahlen abgeleitet. Dann geht es um die Befürchtung, dass individuelle Dispositionen durch personalisierte Inhalte so angesteuert werden, dass die Wahlentscheidung besonders effizient in die gewünschte Richtung manipuliert werden kann (vgl. Helbing 2021). Dieses Potenzial steht auch hinter dem Skandal um Cambridge Analytica.[6]

Personalisierung/Microtargeting
Durch Personalisierung/Microtargeting ist eine gezielte Zielgruppenansprache möglich. Mithilfe von KI-Systemen sollen aus vielen vorliegenden Daten umfassende Persönlichkeitsprofile gebildet werden, auf deren Basis kleine Gruppen mit fein abgestimmter Kommunikation angesteuert werden. Dieses Prinzip kommt aus der Werbung und lässt sich bis zu einem gewissen Grad auch auf die Dynamiken der Öffentlichkeit vor Wahlen übertragen. Weitere Informationen zu Microtargeting in der Politik: Borgesius et al. 2018; Chester & Montgomery 2017.

6 Bei dem Datenskandal wurden über eine Persönlichkeitstest-App Informationen von bis zu 87 Millionen Facebook-Nutzenden und deren Kontakten an die Analysefirma Cambridge Analytica weitergereicht, die unter anderem für das Wahlkampfteam von US-Präsident Donald Trump arbeitete.

Es stellt sich allerdings die Frage, wie weit diese Prinzipien von der Werbung auf die Dynamiken der Öffentlichkeiten vor Wahlen übertragen werden können. Für Werbung reichen schon relativ kleine Effekte: Es genügt, wenn nur einige von tausenden Personen auf die Werbung reagieren. Insofern könnten Wahlsysteme, in denen oft sehr kleine Stimmvorteile entscheiden, dann anfälliger für diese Form der Manipulation sein als andere. Aber auch die theoretischen Grundannahmen dieser Analyse stehen hinsichtlich ihrer Übertragbarkeit auf Wahlen zur Debatte. Andere Analysen betonen zum Beispiel die starken emotionalen oder affektiven Aspekte der Meinungsbildung und Diskussion in neuen Öffentlichkeiten (vgl. Papacharissi 2015). Wenn Affekte also ohnehin ein wichtiger Aspekt des öffentlichen Austauschs sind, würde, diese durch bestimmte Inhalte gezielt anzusprechen, dies nicht unbedingt einen Verdacht der Manipulation begründen: Menschen beziehen sich beispielsweise oft ganz bewusst und affirmativ auf diese Emotionen und machen sie selbstbewusst zur Grundlage ihres Handelns. Eine andere Perspektive betont wiederum, dass das Festhalten auch an sehr irrational erscheinenden Positionen (wieder stark affektiv aufgeladen) weniger aus psychologischen/kognitiven Dispositionen, sondern aus der sozialen Situation der Menschen her zu erklären sei. Aus ihrer Warte ist es „sinnvoller", an etwas festzuhalten, das gegen ihre Interessen spricht, als Grundhaltungen oder Werte des Lebens infrage zu stellen (vgl. Berlant 2012; Hochschild 2016). Vor diesem Hintergrund wäre die Manipulationsgefahr durch KI-Systeme dann zusammen mit detaillierten soziokulturellen Analysen zu betrachten.

2.2.3 Erstellung von gefälschtem Bild-, Audio- oder Videomaterial

Mithilfe von KI-Systemen können auch Deepfakes (siehe Infobox) erstellt werden, welche die Grenzen zwischen Realität und Fiktion verschieben. Diese manipulierten Bilder, Videos oder Audiodateien werden von unterschiedlichen Akteuren mit unterschiedlichen Zielsetzungen verwendet. Die Auswirkungen und Implikationen sind oft vom Kontext abhängig. Viele Deepfakes zeigen bekannte politisch aktive Personen bei Handlungen und Äußerungen, die sie nie getätigt haben. Deepfakes werden zu 96 Prozent im pornografischen Bereich verwendet und betreffen fast ausschließlich Frauen (vgl. Ajder et al. 2019). Diese können dann mit dem angeblichen kompromittierenden Bildmaterial diffamiert oder auch erpresst werden. Ziel kann es sein, politische Inhalte zu beeinflussen oder den Rückzug der Betroffenen aus der Politik zu erwirken (vgl. Aider et al. 2019).

Gleichzeitig bietet diese Technologie aber auch eine Hintertür für Personen, die aufgrund tatsächlich getätigter Handlungen oder Aussagen in der Kritik stehen: Sie behaupten zunehmend, dass die sie belastenden Bild-, Video- oder Audiodateien ein Deepfake seien (vgl. Chesney & Citron 2019). So beispielsweise auch der ehemalige US-amerikanische Präsident Donald Trump über das ihn belastende „Access Hollywood"-Video, in welchem er mit der Belästigung von Frauen angab.

Deepfake
Ein Deepfake (auch: machine-manipulated media – synthetic media – digital content forgeries) ist eine mithilfe von KI erstellte Bild-, Audio- oder Videofälschung. Diese Fälschungen lassen sich meist nur sehr schwer vom Original unterscheiden. Für ihre Erstellung, also z. B. den Austausch eines Gesichts oder Gegenstandes auf einem Bild, sind Algorithmen notwendig, die anhand des zur Verfügung stehenden Datenmaterials einen Deepfake erstellen. Je mehr Datenmaterial vorhanden ist, umso täuschend echter gelingt ein Deepfake. Der Begriff setzt sich aus den Begriffen „Deep Learning" und „Fake" zusammen.

2.3 Risikomanagement-Strategie mithilfe von KI-Systemen

In Kapitel 2.2 wurden einige Risiken dargelegt, die beim Einsatz von KI-Systemen im Kontext von Wahlen für die individuelle Wahlentscheidung entstehen können. Einigen dieser Risiken – aber auch anderen Risiken, die nicht auf KI basieren – kann wiederum mithilfe von KI-Instrumenten begegnet werden. Dies kann jedoch zu einem Wechselspiel führen, da der Einsatz von KI-Systemen in diesem Zusammenhang wieder zu neuen Problemen führen kann. Hierbei ist auch zu beachten, dass wenn eine Kontrollfunktion an einen maschinellen Agenten ausgelagert wird, nicht nur absichtlicher Missbrauch passieren kann, sondern sich auch unabsichtliche Fehler (vgl. z. B. Overblocking) einschleichen können. Folgendes Unterkapitel gibt einen Überblick, an welchen Stellen KI-Systeme für das Risikomanagement bereits eingesetzt werden (können). Da bereits einige Risikomanagement-Strategien existieren (z. B. Transparenzanforderungen oder Sperrpflichten bei Falschinformationen), die ohne KI-Systeme funktionieren, ist im vorliegenden Fall konkret abzuwägen, wann KI-Systeme eingesetzt werden sollten.

2.3.1 Electoral Content Moderation

Im Zusammenhang mit Wahlen werden KI-Systeme zunehmend auch im Rahmen von Risikomanagement-Strategien, wie beispielsweise die „Electoral Content Moderation" (siehe Infobox, Seite 18), eingesetzt. Diese algorithmischen Content-Moderationssysteme stehen allerdings in der Kritik, da sie „oft undurchsichtig, nicht nachvollziehbar und schlecht verständlich sind" (vgl. Gorwa, Binns & Katzenbach 2020). So ist die Entscheidung, warum manche Inhalte – und andere wiederum nicht – entfernt werden, nicht transparent und nachvollziehbar. Eine empirische Untersuchung der Auswirkungen des Einsatzes von Algorithmen-basierten Content-Moderationssystemen auf die politische Meinungsbildung ist deshalb schwierig. Einen Ansatzpunkt stellen die Untersuchungen einer Forschungsgruppe der Universität Düsseldorf zur Frage dar, wie sich KI-Instrumente in unterschiedlichen Settings auf Inhalte, Prozesse und die Akzeptanz von politischen Entscheidungen auswirken (vgl. Vowe 2021).

Electoral Content Moderation
Unter Electoral Content Moderation wird eine Strategie zur Moderation von Inhalten in sozialen Netzwerken seitens der Social-Media-Plattformen und auch der Regulierungsbehörden verstanden. Ziel ist es, Fehlinformationen, Hatespeech, gewalttätige Inhalte oder auch Deepfakes zu entfernen.

Ein wesentliches Instrument der Electoral Content Moderation sind u. a. KI-basierte Uploadfilter (siehe Infobox). Viele verbinden mit Uploadfiltern vor allem die Urheberrechtsfilter, die mit der europäischen Urheberrechtsreform beschlossen wurden. Tatsächlich werden Uploadfilter bereits heute zur Identifikation und Sperrung unterschiedlicher Inhalte verwendet. So setzt Facebook zum Beispiel einen Filter ein, der verhindern soll, dass pornografisches Material den Weg auf die Plattform findet. Uploadfilter sind seit ihrer Einführung immer wieder kritisiert worden. Gründe hierfür liegen beispielsweise in der Gefahr von Kollateralschäden durch Fehlfilterungen, falschen Anreizen, die zur Zensur führen können (hohe Strafen, wenn etwas fälschlicherweise nicht gelöscht wird und kaum Strafen, wenn etwas fälschlicherweise gelöscht wird), sowie einseitigen Informationen, wenn viele Plattformen die gleichen Filter verwenden (vgl. Digitale Gesellschaft e. V. 2020).

Uploadfilter
Unter Uploadfiltern werden technische Systeme auf Plattformen verstanden, die hochzuladende Inhalte automatisch untersuchen und die Veröffentlichung gegebenenfalls ablehnen. Das Ziel ist, unerwünschte oder illegale Inhalte, wie Urheberrechtsverletzungen, Pornografie, Inhalte von Terrorgruppen, automatisiert zu erkennen und deren Veröffentlichung zu verhindern. Zur Inhaltserkennung kommen verschiedene Verfahren zum Einsatz: beispielsweise die Inhaltserkennung durch maschinelles Lernen (vgl. Digitale Gesellschaft e. V. 2020).

2.3.2 Detektion von Desinformation

Insbesondere in Krisenzeiten und im Vorfeld von Wahlen steigt die Gefahr von absichtlich eingesetzten Desinformationskampagnen im Internet, um irreführende Informationen zu verbreiten. Diese Kampagnen sind zumeist durch eine bestimmte politische Agenda motiviert (vgl. Oh et al. 2013; Stieglitz, Mirbabaie & Fromm 2018). Zur Verstärkung von solchen Desinformationskampagnen werden häufig soziale Bots (engl. Social Bots) eingesetzt, die auf KI-Verfahren basieren.[7] Die Verursachenden von digitalen Störkampagnen planen ihr Vorgehen strategisch häufig bereits im Hinblick auf den Schutz vor Entdeckung. So verteilen automatisierte soziale Bots ihre Beiträge über verschiedene Accounts gleich-

7 So zeigt das Projekt „Debater", dass KI-Systeme bereits überzeugend an Meinungsdebatten teilnehmen können (vgl. nature 2021).

zeitig (vgl. Zhang et al. 2013) oder tarnen sich durch eine menschenähnliche Kommunikation (vgl. Abokhodair, Yoo & McDonald 2015). Insofern ist die automatisierte Identifizierung und Bekämpfung von falschen Informationen häufig schwierig (vgl. Varol & Uluturk 2018). Beispielsweise stellten sich die Facebook-Warnhinweise auf Falschnachrichten häufig als ineffektiv heraus, weil sie das Verhalten der Nutzenden nicht nachhaltig beeinflussen (vgl. Ross et al. 2018).

Trotzdem können KI-Systeme einen ersten wertvollen Beitrag für die Detektion von Falschnachrichten leisten und die Bürgerinnen und Bürger bei ihrer demokratischen Meinungsbildung unterstützen. So nutzen Plattformen verstärkt KI-Systeme, um auffällige Muster in den Inhalten vor Wahlen zu erkennen oder Inhalte als Wahlwerbung zu detektieren. Für solche Inhalte gelten inzwischen auf allen bekannten Plattformen besondere Regeln.[8] Es ist zu erwarten, dass die Systeme sich verbessern und insbesondere im Bereich der Erkennung von tendenziöser oder gar falscher Berichterstattung Fortschritte gemacht werden. Deshalb ist anzunehmen, dass in Zukunft mehr und mehr Instrumente zur Erkennung von Fake News angeboten werden. Zumeist müssen solche Anwendungen wie Browser-Plugins von den Nutzenden jedoch aktiv installiert werden und setzen deshalb ein kritisches Bewusstsein für Fake News bereits voraus.

Plattformbetreiber behaupten, dass sie Inhalte nur sehr zurückhaltend mit Warnhinweisen versehen oder löschen, um möglichen Zensur-Vorwürfen vorzubeugen. Sie arbeiten bei der Identifikation und Löschung von Falschinformationen mit sogenannten „Fact Checkern" zusammen. Bei diesen Faktencheck-Organisationen überprüfen Fachleute mögliche Falschinformationen anhand einzelner Postings. Das Fact-Checking beginnt immer mit der Analyse, ob die vorliegende Information falsifizierbar ist. Ein Beispiel für eine einfache Falsifizierung wäre ein Foto, das eine Politikerin/einen Politiker ohne Maske zeigt. Ist dieses Foto bereits aus dem Jahr 2017, so handelt es sich um eine Falschinformation (vgl. Biederbeck & Bülow 2021). Ebenso versuchen die Fachleute Falschinformationen wie Deepfakes aufzufinden, teilweise bereits mithilfe von KI-Systemen.[9] Diese kommen aber meist erst im nächsten Schritt zum Einsatz. Sie können beispielsweise bei bereits als „Deepfake eingestuften" Dateien herausfinden, an welchen anderen Stellen die Datei noch auf der Plattform zu finden ist, und diese automatisiert mit einem Warnhinweis versehen oder auch löschen.

2.3.3 Ausgewogene Berichterstattung – Ausgleich von Media Bias

KI kann dabei unterstützen, einseitige Informationsangebote zu identifizieren und alternative Angebote zu machen. Generell, aber ganz besonders im Umfeld von Wahlen, sind tendenzielle, voreingenommene Berichterstattungen vorzufinden. Dieser sogenannte *Media Bias* entsteht durch eine bestimmte Wort- und Themenwahl („Framing"), die die

8 Informationen über diese Regeln bei den Plattform sind hier zu finden: Facebook; Twitter; TikTok; Instagram; Youtube; Snapchat; Xing; LinkedIn; Vimeo.

9 Darüber hinaus existieren zunehmend technische Möglichkeiten, Deepfakes mithilfe von KI-Systemen zu enttarnen (siehe z. B. Duck Duck Goose, Sensity oder auch 3D Universum).

gelieferte Information in einem bestimmten Licht erscheinen lassen (vgl. Hamborg et al. 2019). Extreme Formen dieser tendenziösen Berichterstattung zwingen mittlerweile auch die großen Plattformbetreiber zum Handeln, die auch hier zunehmend auf die Detektionsmöglichkeiten aus dem Bereich von KI und maschinellem Lernen setzen. In Anbetracht der immensen Menge von Nachrichten in den digitalen Medien ist eine menschliche Betrachtung und Analyse nicht mehr möglich. Deshalb setzen Internetkonzerne auf den Einsatz von automatisierten Analysen, die eine schnelle Erfassung und Bewertung von Media Bias ermöglichen und gegebenenfalls als Grundlage für Gegenangebote dienen können (siehe Electoral Content Moderation).

Zur Verbesserung der Kompetenzen, um Informationen zur politischen Meinungsbildung zu bewerten, können Hilfsmittel wie ergänzende Links verwendet werden (vgl. Wineburg & McGrew 2017). Bei verzerrenden Darstellungen, die durch algorithmische Filterblasen zustande kommen, können Gegenargumente bereitgestellt werden (vgl. Lorenz-Spreen et al. 2020) oder auch Pop-up-Informationen, die an einfache Schritte zur Bewertung von Informationen erinnern. Bei der Gegensteuerung zu falscher und tendenziöser Berichterstattung können auch soziale Bots helfen. Sie können automatisiert verifizierte Informationen verbreiten und gegebenenfalls interaktiv auf Fragen im Kontext von Wahlen und insgesamt zur politischen Meinungsbildung eingehen (vgl. Hofeditz et al. 2019).

3. Bedeutung der skizzierten Veränderungsprozesse

Im Folgenden wird die gesellschaftliche Bedeutung der in Kapitel 2 vorgestellten Einsatzmöglichkeiten von KI-Systemen im Zusammenhang mit Wahlen untersucht. Hierfür wird zuerst eine juristische Bewertung vorgenommen und anschließend werden die vorgestellten Änderungsprozesse in einen größeren Rahmen eingeordnet.

3.1 Juristische Perspektive

Wahlen sind für die Demokratie und die Rechtsordnung von entscheidender Bedeutung (siehe Infobox). Eine freie Wahl soll durch das Zusammenspiel mehrerer Grundrechte gewährleistet werden: Die Wahlberechtigten können sich nach Art. 5 Abs. 1 des Grundgesetzes (GG) aus öffentlich zugänglichen Quellen über Politikerinnen und Politiker, politische Ereignisse sowie politische Programme informieren. Sie können ungehindert auf Informationen aus Presse, Rundfunk und anderen Medien zugreifen. Presse und Rundfunk können gemäß Art. 5 Abs. 1 GG ungehindert ihren Aufgaben – unzensiert Nachrichten und Meinungen zu veröffentlichen – nachgehen, müssen aber hierbei Qualitätsanforderungen an Recherche und neutraler Berichterstattung erfüllen – online wie offline. Die Wahlberechtigten genießen nach Art. 5 Abs. 1 GG Meinungsfreiheit und können sich im Vorfeld der Wahl ungehindert argumentativ zu Vor- und Nachteilen der Wahlalternativen äußern. Versammlungs-, Vereinigungs- und Demonstrationsfreiheit nach Art. 8 und 9 GG ermöglichen, individuelle Meinungen zu bündeln. Die Alternativen, die gewählt werden können, werden nach Art. 21 GG durch Parteien strukturiert, die ungehindert für ihre Wahl sowie die Wahl ihrer Kandidatinnen und Kandidaten werben können. Die gleichen Rechte stehen auch den Kandidierenden selbst zu.

Die Bedeutung von freien und demokratischen Wahlen
Nach Art. 20 Abs. 2 GG geht alle Staatsgewalt vom Volk aus und wird von diesem in Wahlen ausgeübt. Die Wahl ist die einzige Handlung, in der das Volk direkt seinen Willen ausdrücken kann. Durch die Wahl überträgt es die staatliche Macht auf die gewählten Personen und Organe. Die Wahl ist die Grundlage der repräsentativen Demokratie. Sie bewirkt die demokratische Legitimität für die Abgeordneten und das Parlament und, von diesen abgeleitet, für die Regierung, Rechtsprechung und die gesamte Rechtsordnung. Aufgrund dieser außerordentlichen Bedeutung bestehen hohe Anforderungen an die Willensbildung des Volkes durch Wahlen und die freie und informierte Entscheidung der Wahlberechtigten in jedem Einzelfall.

Die Verwirklichungsbedingungen dieser Grundrechte können zum einen durch die Anwendung von KI verbessert werden (siehe Kapitel 2.1). Zum anderen können KI-Anwendungen aber auch die Verwirklichungsbedingungen der genannten Grundrechte verschlechtern und damit die Zielsetzung informierter und freier Wahlen gefährden (siehe Kapitel 2.2).

Das GG erwartet und ermöglicht einen argumentativen Wahlkampf, der die Vorzüge und Nachteile der Parteien, ihrer Wahlprogramme und ihrer Kandidierenden zum Gegenstand hat. Dabei geht es davon aus, dass die Wahlberechtigten jeweils selbst von ihren Grundrechten Gebrauch machen und ihre eigenen Argumente und Präferenzen in den Meinungsstreit einbringen. Die Fairness der Auseinandersetzung aber wird missbraucht, wenn einzelne Wahlkämpfende automatisierte Prozesse einsetzen, die vorgeben, eine natürliche Person zu sein, und millionenfach „Likes" vergeben, Meinungen teilen oder retweeten, Accounts generieren und als soziale Bots mit anderen Menschen kommunizieren. Dadurch werden die Gewichte von Meinungen und die Mehrheitsverhältnisse ihrer Unterstützung verfälscht und andere Wählende können in die Irre geführt werden (Löber & Roßnagel 2019b).

Das GG will eine freie und unbeeinflusste Wahl ermöglichen. Daher versucht es, den Wahlkampf von jeder rechtlichen und administrativen Einflussnahme freizuhalten. Jeder Wahlkampf trägt jedoch die Gefahr in sich, die Chancengleichheit der Parteien, die Persönlichkeitsrechte der Kandidierenden, die Meinungsfreiheit Andersdenkender und die informationelle Selbstbestimmung aller Wahlberechtigten einzuschränken. Für die gefährdeten Grundrechte hat der Staat jedoch eine Schutzpflicht. Recht und Staat müssen daher einen immer prekären und umstrittenen Ausgleich finden zwischen den Grundrechten, die eine freie und unbeeinflusste Wahl ermöglichen, und den Grundrechten, die ihres Schutzes bedürfen. Diese Grenzziehung wird durch die neuen Möglichkeiten der KI und die Vor- und Nachteile des Einsatzes von KI für die beschriebenen Rechtsziele noch schwieriger.

Bisher versucht das Recht diese Grenze dadurch zu bestimmen, dass es allgemeine Vorgaben beschreibt. Diese beziehen sich nicht auf bestimmte politische Inhalte und möglichst auch nicht auf die besondere Situation des Wahlkampfs, sondern sie stellen nur Regeln für ein verträgliches Zusammenleben allgemein auf. Hierzu gehören die allgemeinen Strafnormen, die Kommunikationsdelikte wie Beleidigung, üble Nachrede, Verleumdungen und Verletzung des höchstpersönlichen Lebensbereichs durch Bildaufnahmen oder das Verbreiten von Propagandamitteln oder Kennzeichen verfassungswidriger Organisationen, Verherrlichung von Nazisymbolen, die öffentliche Aufforderung zu Straftaten, die Störung des öffentlichen Friedens durch Androhung von Straftaten, Volksverhetzung oder Bedrohung unter Strafe stellen. Bei der Ausübung all dieser Delikte kann sich die Täterin oder der Täter nicht auf Meinungsfreiheit oder andere Grundrechte berufen (Löber & Roßnagel 2020). Das dürfte im Regelfall auch für Deepfakes und Hass-Postings gelten. Dagegen sind Fake News oft straffrei, weil es kein allgemeines Verbot der Lüge und keine allgemeine rechtliche Verpflichtung gibt, immer die Wahrheit zu sagen.

Allgemeine, auch im Wahlkampf, zu beachtende Regelungen sind das Datenschutzrecht, das Medienrecht und das Presserecht. Nach Datenschutzrecht sind beispielsweise Profilbildungen für Microtargeting verboten. Auch fordern Art. 13 und 14 der Datenschutzgrundverordnung (DSGVO) umfassende Informationen zur Datenverarbeitung für alle Personen, deren Daten verarbeitet werden. Das Medienrecht fordert in § 5 des Telemediengesetzes (TMG) für jeden im Internet veröffentlichten Inhalt die Angabe des Verantwortlichen in einem Impressum. Der neue Medienstaatsvertrag (MStV) verlangt unter anderem, automatisierte Prozesse kenntlich zu machen, um zu verhindern, dass die Empfangenden annehmen, die Nachricht stamme von einem Menschen. Nach § 18 Abs. 3 und 93 Abs. 4 MStV sind Anbieter von Social Networks verpflichtet, bei automatisiert erstellten Inhalten oder Mitteilungen den Umstand der Automatisierung kenntlich zu machen, sofern eine Gefahr der Verwechslung mit der Erstellung durch natürliche Personen besteht. Das Presserecht stellt spezifische Qualitätsansprüche an Recherche und Neutralität der Darstellung, denen Fake News oft nicht genügen.

Plattformbetreiber sind nach § 10 TMG verpflichtet, rechtswidrige Informationen, die sie für einen Nutzenden speichern, zu sperren oder zu entfernen, sobald sie davon Kenntnis erhalten. Da sie dieser Verpflichtung jedoch nicht ausreichend nachgekommen sind, erließ der Gesetzgeber am 1. September 2017 das Netzwerkdurchsetzungsgesetz (NetzDG). Dieses Gesetz verpflichtet die Betreiber von sozialen Netzwerken mit mehr als zwei Mio. Teilnehmenden in Deutschland dazu, ein Beschwerdemanagementsystem zu installieren, über das sie niedrigschwellig Kenntnis von rechtswidrigen Inhalten erlangen können. Als rechtswidrig gelten in diesem Sinn nur Inhalte, die gegen 22 explizit aufgezählte Straftatbestände verstoßen, für die keine Berufung auf die Meinungsfreiheit möglich ist. Diese Inhalte müssen die Plattformbetreiber in einfachen Fällen innerhalb von 24 Stunden entfernen und in komplizierten Fällen innerhalb von sieben Tagen. Verstößt ein Anbieter gegen seine Organisations- und Berichtspflichten, drohen ihm nach § 4 NetzDG Bußgelder in Höhe von bis zu fünf Millionen Euro.

Eine Evaluation des Gesetzes hat ergeben, dass die Plattformbetreiber ihre selbstgesetzten Regeln durchsetzen und sich nur an die gesetzlichen Vorgaben des NetzDG halten, wenn diese über die netzwerkeigenen Regeln hinausgehen. Ein Overblocking aufgrund des NetzDG war nicht festzustellen (Löber & Roßnagel 2019a). Die Plattformbetreiber befolgten die Vorgaben des Gesetzes bisher jedoch nur mit minimalen Anstrengungen, erschwerten NetzDG-Beschwerden und blockierten eine verbesserte Zusammenarbeit mit Strafverfolgungsbehörden (Löber & Roßnagel 2019a). Daher fordert eine Neufassung des Gesetzes 2021 ein verbessertes Beschwerdemanagement mit einem Verfahren zur Überprüfung von zu Unrecht blockierten Inhalten sowie Meldungen von Verstößen an die Staatsanwaltschaften.

Jenseits der vom NetzDG benannten rechtswidrigen und strafbaren Inhalte stehen Risikomanagement-Strategien der Plattformbetreiber immer in dem Spannungsfeld zwischen eigener Verantwortung für ihre bekanntgewordenen Inhalte, für die sie nach dem TMG Mitverantwortung tragen, und Beeinträchtigungen der Grundrechte der Nutzenden, die

Inhalte über die Plattformen veröffentlichen. Diese Risikomanagement-Strategien sind vor allem dann problematisch, wenn die Plattformbetreiber eigene Kriterien im Rahmen ihrer Community-Regeln für die Bewertungen der Inhalte verfolgen. Diese können nämlich leicht zu den Regeln der staatlichen Rechtsordnung in Widerspruch stehen.[10]

Gesetzliche Rahmenbedingungen

Europäische Vorgaben

- Datenschutz-Grundverordnung (DSGVO)
  EU-Verordnung 2016/679; 2016 in Kraft getreten und gültig seit dem 25.05.2018.
  Ziel: Vereinheitlichung der Verarbeitungsregeln personenbezogener Daten durch private und öffentliche Akteure
- Platforms for Business-Verordnung (P2B-VO)
  EU-Verordnung 2019/1150; seit 2020 in Kraft.
  Ziel: Beitrag „zur Förderung von Fairness und Transparenz für gewerbliche Nutzer von Online-Vermittlungsdiensten"
- Regulierungsvorschlag der Europäischen Kommission zu KI
  Entwurf der Europäischen Kommission, vorgelegt April 2021.
  Ziel: Definition verbindlicher Anforderungen an KI-Anwendungen mit hohem Risiko

Nationale Vorgaben

- Grundgesetz (GG)
  Verfassung Deutschlands; seit 1949 in Kraft.
- Bundesdatenschutzgesetz (BDSG-neu)
  Seit dem 25.05.2018 in Kraft (gleichzeitig mit der Geltung der DSGVO in Deutschland).
  Ziel: Ergänzung und/oder Präzisierung der DSGVO
- Telemediengesetz (TMG)
  Seit 2007 in Kraft.
  Ziel: Regelung der Bedingungen für das wirtschaftliche Angebot von Telemedien-diensten (Internetangeboten) und ihrer Verantwortung für Inhalte ihrer Nutzenden
- Medienstaatsvertrag (MStV)
  Staatsvertrag; seit 2020 wirksam.
  Ziel: Grundlegende Regelungen für die Veranstaltung und das Angebot, die Verbreitung und die Zugänglichmachung von Rundfunk und Telemedien
- Netzwerkdurchsetzungsgesetz (NetzDG)
  Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken; seit 2017 in Kraft, 2021 novelliert.
  Ziel: Bekämpfung von Hasskriminalität, strafbaren Falschnachrichten und anderen strafbaren Inhalten auf sozialen Netzwerken

10 Inwieweit die Betreiber die Grundrechte auch im Rahmen ihrer privatautonomen Rechtsordnung berücksichtigen müssen, wird derzeit in Rechtsprechung und Literatur neu vermessen (Löber & Roßnagel 2020). Viel deutet daraufhin, dass das Bundesverfassungsgericht (BVerfG) großen Kommunikationsplattformen aufgrund ihres wirkmächtigen Einflusses auf die öffentliche Kommunikation eine staatsgleiche Grundrechtsbindung auferlegt (BVerfG 2019).

3.2 Einordnung der skizzierten Veränderungsprozesse

Im Bereich der Wahlinformation und -organisation werden die Potenziale von KI aktuell noch nicht ausgeschöpft. Im Gegensatz dazu führen die zunehmend KI-gestützte Informationsverbreitung sowie die Nutzung von KI-Systemen für eine personalisierte Ansprache von Wahlberechtigten bereits gegenwärtig zu Einwirkungen auf die Wahlentscheidungen, ebenso wie die von Dritten mithilfe von KI-Systemen erstellten Deepfakes (siehe Infobox zu Desinformation und ihren Folgen). Zwar wird diesen Einwirkungen punktuell durch den Einsatz von KI-Systemen begegnet, so können zum Beispiel bereits verifizierte Deepfakes mithilfe von KI online schneller aufgefunden und gelöscht werden (siehe Kapitel 2.3.2). Dennoch gehört die Manipulation von Wahlen durch KI-basierte Systeme zu den gesellschaftlich immer wieder geäußerten Befürchtungen, auch das Bundesamt für Sicherheit in der Informationstechnik (BSI) wappnet sich gegen Desinformationskampagnen und IT-Angriffe im Rahmen der Bundestagswahl 2021.

Wahlen sind ein wesentlicher und sensibler Bestandteil von Demokratie. Das Wahlrecht, aber auch die konkrete organisatorische wie technische Durchführung von Wahlen stehen daher mit Recht immer wieder in hoher Aufmerksamkeit in Medien und Gesellschaft, wie dies etwa in den USA während und nach der Abwahl von Präsident Donald Trump zu beobachten war. Freie, geheime und faire Wahlen ohne Manipulationsmöglichkeiten sind ein Ideal, das immer nur annäherungsweise erreicht werden kann. Probleme im Zusammenhang mit Wahlen können durch den Einsatz von KI-Systemen auftreten, aber eben auch unabhängig von diesen, wie dies aus der Geschichte der Wahlen bekannt ist. Viele Probleme können aufgrund des jeweils besonderen sozialen und politischen Kontextes – anders als in anderen, rein technischen Anwendungsfällen – nicht so einfach oder möglicherweise gar nicht durch den Einsatz von KI-Systemen technisch gelöst werden. Ein Beispiel sind Uploadfilter (siehe Kapitel 2.3.1): Sie haben das Ziel, unerwünschte oder illegale Inhalte von Social-Media-Plattformen zu entfernen. Das übermäßige oder einseitige Löschen von Inhalten kann aber möglicherweise in Kombination mit der Verwendung der gleichen Uploadfilter auf unterschiedlichen Social-Media-Plattformen dazu führen, dass der demokratische Diskurs im Vorfeld von Wahlen eingeschränkt und verzerrt werden könnte und möglicherweise damit auch die Wahl an sich.

Diese Analyse ist jedoch nur eine Momentaufnahme. Die zunehmende Verfügbarkeit von Daten und deren Verarbeitung werden zukünftig immer mehr unser Verständnis von „Staat" und „Demokratie" infrage stellen. Dabei geht es um ein komplexes Zusammenspiel von Freiheitsrechten, Datenmanagement und Normen des demokratischen Zusammenlebens, das selbst Gegenstand partizipativer Verfahren und politischer Beratungsprozesse sein muss. Das untersuchte Feld befindet sich gerade im Wandel und der Einsatz von KI-Systemen bei Wahlen wird sich in ein paar Jahren mit großer Wahrscheinlichkeit anders darstellen – auch wenn aktuell noch keinerlei Hinweise auf große Motivationen unterschiedlicher Akteure zu sehen sind, KI-Systeme für den Bereich Wahlen zu entwickeln oder vermehrt einzusetzen. Es ist jedoch allein wegen der Bedeutung demokratischer Wahlen, aber auch, weil das Verhältnis von KI und Demokratie in der Öffentlich-

keit häufig kritisch gesehen wird, sinnvoll, diese Entwicklungen weiter zu beobachten, zu reflektieren und nach Möglichkeiten zu suchen, KI zur Stärkung der Demokratie einzusetzen.

Desinformation und ihre Folgen

Desinformation beeinflusst in ihrer Meinungsbildung offene Personen sowie Unentschlossene und Nichtwählende. Sie wirkt auf die jeweilige politische Anhängerschaft bestätigend und auch radikalisierend und untergräbt das Vertrauen in Institutionen, Verfahren oder Personen. Diese Wirkung steht der gesellschaftlichen Inklusion entgegen und fördert eine grundsätzliche politische und gesellschaftliche Lagerbildung. So entsteht eine Dynamik, die eine Fragmentierung von Öffentlichkeit begünstigt: Zum einen verliert die gruppenübergreifende Auseinandersetzung an Reichweite, da gesellschaftliche Gruppen sich in „Informationsblasen" einkapseln, in denen sie immer weniger mit dem Wissen, den Argumenten und Sichtweisen der „anderen Seite" konfrontiert werden. Dieser Trend wird durch Algorithmen-basierte Systeme zur Auswahl und Empfehlung von Informationen verstärkt. Zum anderen unterstützt Desinformation Verschwörungstheorien. Desinformation erzeugt Unsicherheit darüber, auf welche Wahrheit man sich innerhalb von Gemeinschaften von Kommunikationspartnern vernünftigerweise einigen kann. So entsteht aus der Gesamtperspektive eine Beliebigkeit verschiedener Wahrheiten.

4. Maßnahmen zur Demokratiestärkung durch KI-Systeme bei Wahlen

Damit Chancen von KI-Systemen bei Wahlen gestärkt und mögliche Risiken abgeschwächt werden können, sind unterschiedliche Maßnahmen denkbar. Hierzu bedarf es der Einbindung aller beteiligten Akteurinnen und Akteure.

4.1 Gesetzliche Rahmenbedingungen zur Reduzierung der Risiken durch KI-Systeme

Möglichen Risiken bei der Verwendung von KI-Systemen im Zusammenhang von Wahlen kann mit unterschiedlichen gesetzlichen und technischen Maßnahmen wie auch ethischen Vorgaben entgegengewirkt werden. Solche normativen Regulierungsansätze zeigen sich über die Kontrolle und Steuerung von Prozessen und der Verbreitung und Anwendung von KI-Technologien im Zusammenhang von Wahlinformation und Wahlkampf. Konkret geht es dabei insbesondere um die Verhinderung von Falschinformation beziehungsweise von manipulierter Information und um die Kuratierung bzw. automatisierte Auswahl von Medieninhalten, die für die politische Meinungsbildung von Bedeutung sind.

Electoral Content Moderation: Transparenz von Auswahlkriterien, Recht auf Begründung

Dem Grundsatz der Transparenz im Umgang mit KI und algorithmischen Systemen kommt eine große Bedeutung zu (z. B. AI HLEG 2019; Europäische Kommission 2021). In diesem Zusammenhang wird gleichzeitig zumeist nicht nur auf Transparenz, sondern auch auf Verständlichkeit, Erklärbarkeit bzw. die adressatengerechte Aufbereitung der Information verwiesen (vgl. z. B. Bundesministerium für Bildung und Forschung 2019). Diese Grundprinzipien der Transparenz und Verständlichkeit gelten genauso für den Einsatz von KI und algorithmischen Entscheidungssystemen zur Auswahl von Informationen, die sich auf demokratische Wahlen beziehen.

Die europäische Datenschutz-Grundverordnung (DSGVO) sowie die Platforms for Business (P2B)-Verordnung bieten in dieser Hinsicht bereits viele Vorgaben. Die DSGVO macht deutlich, dass Nutzende ein „Recht auf Begründung" automatisierter Entscheidungen haben, und die P2B-Verordnung verdeutlicht, dass die Betreiber von Online-Diensten zumindest die „Hauptparameter, die das (algorithmische) Ranking bestimmen" offenlegen (vgl. Richard et al. 2020).

Zudem wird im MStV festgelegt, dass journalistisch-redaktionelle Inhalte Dritter nicht ohne sachlichen Grund bei der Präsentation diskriminiert werden dürfen.

Der im April 2021 vorgelegte Regulierungsvorschlag der Europäischen Kommission zu KI sieht vor, dass ein System als risikoreich eingestuft wird, wenn es zu systemischen nachteiligen Auswirkungen für die Gesellschaft als Ganzes, einschließlich der Gefährdung des Funktionierens demokratischer Prozesse und Institutionen und des zivilgesellschaftlichen Diskurses, führen könnte. Hierunter könnten also einige der vorgestellten Systeme fallen. Damit risikoreiche Systeme in Verkehr gebracht werden dürfen, müssen diese eine Konformitätsbewertung durchlaufen und die Qualität der Datensätze muss dokumentiert werden (vgl. Europäische Kommission 2021). Hinzu kommt, dass Personen darauf hingewiesen werden müssen, dass sie mit einem KI-System interagieren. Wenn die Plattformbetreiber ihren Transparenzpflichten (noch) nicht hinreichend nachkommen, können insbesondere die Wissenschaft oder zivilgesellschaftliche Organisationen selbst tätig werden und eigene Versuche mit den Funktionsweisen von Algorithmen, die rekonstruieren, welchen Kriterien das Content Management unterliegt (sog. reverse engineering), durchführen. Diese Methode, die jedoch nur Expertinnen und Experten vorbehalten ist, ist zumeist zu aufwendig, sodass damit keine Transparenz für einen aktuellen Wahlkampf, sondern nur eine nachträgliche Aufklärung über die automatisierte Inhalteauswahl erfolgen kann.

Bekämpfung von Desinformation

Für den Bereich der Deepfakes existieren in Deutschland keine spezifischen gesetzlichen Regulierungen, sondern es finden generell-abstrakte Regelungen Anwendung (vgl. Antwort auf eine kleine Anfrage der FDP-Fraktion 2019 zu Deepfake). „Die Bundesregierung überprüft den Rechtsrahmen auf Bundesebene fortlaufend daraufhin, ob aufgrund von technologischen oder gesellschaftlichen Herausforderungen ein Anpassungsbedarf besteht" (vgl. Deutscher Bundestag 2019).

Im internationalen Kontext lassen sich jedoch Fallbeispiele für spezifische gesetzliche Regulierungen von KI-bezogenen Internetinhalten im Zusammenhang von Wahlen finden. Hier geht es in der Regel insbesondere um Deepfakes. Kalifornien und Texas haben ein zeitlich begrenztes und spezifisches Verbot politischer Deepfakes zur Wahlbeeinflussung erlassen (vgl. California Legislative Information 2019).[11] In China wurden Gesetze eingeführt, die die Erstellung, Ausstrahlung und Verwendung von Deepfakes insgesamt verbieten (vgl. Jing 2019; Kwok & Koh 2020). Auch Südkorea hat neue Gesetze verabschiedet, die versuchen, Deepfake-Videos unter Strafe zu stellen. Das Gesetz (seit Juni 2020 rechtskräftig) sieht Strafen von bis zu fünf Jahren Gefängnis oder eine empfindliche Geldstrafe vor (vgl. Ryall 2021).

11 Einige Rechtsexpertinnen und -experten haben in den USA jedoch die Durchsetzbarkeit der Gesetze infrage gestellt und argumentiert, dass die Bemühungen zum Verbot von Deepfakes dem Ersten Verfassungszusatz („Free Speech") widersprächen. Der Oberste Gerichtshof der USA hat mit Bezug zu diesem Verfassungszusatz bereits früher wissentlich gemachte falsche Äußerungen geschützt. Auch die California News Publishers Association und die American Civil Liberties Union sprachen sich gegen die kalifornische Gesetzgebung aus und unterstrichen damit die Bedenken hinsichtlich der Meinungsfreiheit (vgl. Toto & Keating 2020; Labbe 2020).

4.2 Möglichkeiten der Plattformbetreiber zur Reduzierung der Risiken durch KI-Systeme

Labelling

Ein weiteres Instrument gegen Falsch- und manipulierte Informationen ist das Labeln von Inhalten mit Warnhinweisen. Gelabelt werden Beiträge, die potenzielle Falschinformationen beinhalten. Diese Warnhinweise zeigen Nutzenden an, dass Faktenchecker die Behauptungen des Beitrags anzweifeln, und verweisen auf weitere verifizierte Quellen. So sollen Nutzende davon abgehalten werden, Falsch- und manipulierte Informationen für wahr zu halten und diese auch zu verbreiten.

Besonders im Vorfeld der US-Präsidentschaftswahlen im Jahr 2020 griffen viele Plattformen auf dieses Instrument zurück. Twitter beispielsweise verwendete Warnhinweise, wenn ein Tweet Fakten falsch wiedergab, zu Gewalt aufrief, Wahlergebnisse delegitimierte oder voreilig den Wahlsieg Trumps verkündete (siehe Abbildung 4). Enthielt ein Tweet eine Falschinformation, die potenziell einen Schaden hätte anrichten können, ging das Unternehmen sogar noch einen Schritt weiter und löschte den Beitrag. Facebook und Instagram dagegen löschen keine Informationen, sondern arbeiten nur mit Warnhinweisen. Jedoch sperren sie in bestimmten Fällen einzelne Personen wie beispielsweise Donald Trump (vgl. Rivero 2021).

Abbildung 4: Twitter, 23.08.2020

Die Effekte von Warnhinweisen wurden vielfach empirisch analysiert: Wird ein einzelner Beitrag mit einem Warnhinweis versehen, so sind Nutzende kritischer gegenüber dessen Inhalt und teilen den Beitrag auch seltener (vgl. Bode & Vraga 2015; Mena 2020; kritisch dazu vgl. Ross et al. 2018). Gleichzeitig kann es jedoch zu einem „implied truth"-Effekt kommen: Nutzende stufen weitere (nicht mit Warnhinweisen versehene) Beiträge als bestätigt und damit vertrauenswürdig ein. Dass diese Beiträge Falschinformationen enthalten könnten, aber (noch) nicht mit einem Warnhinweis versehen wurden, wird hierbei außer Acht gelassen (vgl. Pennycook et al. 2020). Dem „implied truth"-Effekt kann durch die Bereitstellung von (aufwendig umzusetzenden) Verifikationshinweisen begegnet werden (vgl. ebd.).

Plattformpolicies
Spätestens mit der Sperrung des Twitter Accounts im Januar 2021 des damaligen US-Präsidenten Donald Trump und seinem Ausschluss aus verschiedenen Plattformen (De-Platforming) sind auch der breiten Öffentlichkeit die Defizite der Plattformregulierung bewusst geworden. Die Vizepräsidentin der Europäischen Kommission, Věra Jourová, verweist in diesem Zusammenhang auf den Digital Services Act (DAS) und fordert einen neuen Pakt gegen Desinformation inklusive neuer berufsethischer Standards (vgl. Jourová 2021).

Ein Ansatzpunkt wären hier verbindliche Standards nach dem Modell einer regulierten Selbstregulierung, wie es in Deutschland und vielen anderen Ländern der EU für den Medienbereich üblich ist. Dafür könnten koordinierte und globale Standards für Unternehmen wie Twitter, Facebook, Instagram, Google, YouTube und TikTok etabliert werden, die sich über Content Moderation im Allgemeinen, aber auch insbesondere hinsichtlich der demokratischen Meinungsbildung im Zusammenhang von Wahlen verständigen. Eine offene und öffentliche Debatte über Fragen, was Meinungsfreiheit beziehungsweise falsche oder gefälschte Inhalte ausmacht, ist dringend geboten (vgl. Schaake & Reich 2020). Ein Ansatzpunkt könnte der im Jahr 2018 von Social-Media-Plattformen, Tech-Unternehmen und Verbänden der Werbeindustrie unterzeichnete EU-Verhaltenskodex zur Bekämpfung von Desinformation sein (vgl. Europäische Kommission). Dieser Verhaltenskodex wird jedoch für den Mangel an Transparenz sowie an klar messbaren Zielen kritisiert (vgl. Europäische Kommission).

4.3 Gestaltungsoptionen

Wie in diesem Papier dargestellt, weisen KI-Systeme ein gewisses Potenzial auf, im Kontext von demokratischen Wahlen auf medialer Ebene eine offene Meinungsbildung zu fördern. Eine Bedrohung von KI-Systemen für demokratische Wahlen ist eher als gering einzuschätzen. Trotzdem – oder vielleicht gerade deshalb – sollten Maßnahmen ergriffen werden, um Chancen von KI-Systemen für eine gelingende Meinungsbildung bei Wahlen zu realisieren und Risiken abzuschwächen. Hierbei können vielfältige Maßnahmen von unterschiedlichen Akteurinnen und Akteuren ergriffen werden. Da die KI-Systeme stets in ein soziotechnisches Gesamtsystem eingebettet sind, beziehen sich manche Gestaltungsoptionen nicht ausschließlich auf die Gestaltung der KI-Systeme direkt, sondern auch auf die Gestaltung des sie umgebenden Gesamtsystems. Gleichzeitig sind diese Optionen nicht als abschließend zu verstehen.

Betreiber von Social-Media-Plattformen

Die Betreiber von Social-Media-Plattformen haben einen direkten Zugang zu allen Inhalten der Plattformen und verfügen daher über einen großen Handlungsspielraum. Sie könnten:

- **Transparenz der Moderationsvorgänge schaffen:** Oft ist nicht bekannt, aus welchen Gründen ein Inhalt oder ein Konto gesperrt (oder auch nicht gesperrt) wurde – und ob diese Entscheidung Algorithmen-basiert oder durch einen Menschen erfolgte. Die Betreiber sind angehalten, hier für Transparenz zu sorgen und die Vorgaben öffentlich zu machen.

- **Effektive Beschwerdemechanismen bei vermuteten Fehlentscheidungen implementieren:** Die Plattformbetreiber sollten effektive und niedrigschwellige Beschwerdemechanismen implementieren (siehe § 3b ff. NetzDG), damit Nutzende (vermutete) Fehlentscheidungen, die sie selbst oder andere betreffen, beispielsweise im Rahmen eines Algorithmen-basierten Electoral Content Managements, anzeigen können.

- **Allgemeine Standards für Social-Media-Plattformen entwickeln und implementieren:** Die Betreiber sollten die Konstituierung unabhängiger, mit Expertinnen und Experten und Vertreterinnen und -vertretern zivilgesellschaftlicher Organisationen besetzter Räte vorantreiben. Diese beraten gemeinsam mit den Betreibern über Standards und Verhaltenskodizes zu Fragen des Algorithmen-basierten Electoral Content Managements wie auch zu allen anderen Themen einer pluralen und diskriminierungsfreien demokratischen Öffentlichkeit.

- **Community-Policing fördern:** Widerspruch durch andere Netznutzende zulassen und – beispielsweise mithilfe von Meldefunktionen – die gegenseitige soziale Kontrolle der Netznutzenden fördern.

Politische Entscheidungsträgerinnen und Entscheidungsträger

Die Verantwortlichen können beispielsweise rechtspolitisch, also durch den Beschluss von Gesetzen, als auch durch die Ausschreibung entsprechender Forschungsprojekte dazu beitragen, dass die Potenziale von KI-Systemen im Kontext der individuellen Wahlentscheidung gefördert und die Herausforderungen abgeschwächt werden. Sie könnten unter anderem:

- **Digitale Souveränität durch plurale Inhalte und Infrastrukturen fördern:** Dies umfasst die Förderung von alternativen kommerziellen Plattformen, deren Angebote an Zielen wie Gemeinwohlorientierung und Demokratieförderung ausgerichtet sind (vgl. Public-Value-Regelungen für die öffentlich-rechtlichen Rundfunkanstalten), ebenso wie die Weiterentwicklung des Qualitäts- und Datenjournalismus (für mögliche Kriterien vgl. Heesen et al. 2020). In Ergänzung dazu sollten sie Informationsangebote für Bürgerinnen und Bürger bereitstellen, wie – unter anderem mithilfe von KI erstellte –

Desinformationen erkannt werden können, und wie mit diesen umgegangen werden sollte. Darüber hinaus sollen durch KI erzeugte synthetische Medien gekennzeichnet werden.

- **Microtargeting einschränken:** Ähnlich der Regelung zur Kennzeichnungspflicht von Social Bots im MStV sollten die politischen Entscheidungsträgerinnen und Entscheidungsträger Transparenzpflichten für Microtargeting prüfen. Da Microtargeting, wenn es sich auf eine auf Persönlichkeitsmerkmalen beruhende Profilbildung stützt, aktuell bereits verboten ist (siehe Art. 6 DSGVO), sollte auch über weitere gesetzliche Einschränkungen nachgedacht werden.

- **Forschungs- und forschenden Nichtregierungsorganisationen Zugang zu relevanten Social-Media-Plattformdaten ermöglichen:** Anhand dieser Daten sollen Auswirkungen des oft mithilfe von KI-Systemen betriebenen Electoral Content Managements im Kontext von Wahlen der Social-Media-Plattformen und entsprechender Regulierungsvorschläge analysiert werden (vgl. auch den aktuellen Entwurf des Digital Services Act der Europäischen Kommission). Der DSA verlangt, dass Diensteanbieter transparent und verständlich darlegen, welche Maßnahmen ihr Hausrecht umfasst und wie sie wirken.

- **Forschungsförderung zur Bekämpfung von Desinformationen mittels KI-Systemen ausbauen:** Es ist wichtig, dass Forschungs- und forschende Nichtregierungsorganisationen über entsprechende Mittel verfügen, um KI-Methoden zur Bekämpfung von Desinformationen (weiter) zu entwickeln (siehe „Forschungs- und forschende Nichtregierungsorganisationen"). Hierzu ist auch regelmäßig zu überprüfen, ob die aktuellen Projekte und finanziellen Mittel ausreichen.

- **Datenjournalismus und netzpolitische Formate unterstützen:** Für einen lebendigen, öffentlichen, freien und diversen Diskurs zum Thema KI und Wahlen und zu Fragen von Demokratie und Technikentwicklung ist der Qualitäts- und Fachjournalismus ein unerlässlicher Baustein. Deren Bestand und Weiterentwicklung sollte in der journalistischen Ausbildung und am Markt unterstützt werden.

- **Konsequente Verfolgung von Straftaten (u. a. mithilfe von KI erstellte Deepfakes) in den sozialen Netzwerken fördern:** Hierunter fallen die bessere personelle Ausstattung der Strafverfolgungsbehörden und der Justiz sowie die Einführung von „Log-in-Fallen". Ziel ist die Identifikation der Verursachenden, ohne dass das Prinzip der Datenminimierung umgangen wird: Wenn eine Nutzende/ein Nutzender mit einem Pseudonym (z. B. „Highway1") in einem Netzwerk einen Verstoß begangen hat, würde der Netzwerkbetreiber verpflichtet werden, beim nächsten Login des „Highway1" deren/dessen aktuelle IP-Adresse zu speichern und der Polizei mitzuteilen, die daraus strafrechtliche Schritte ableiten kann. Diese Speicherung erfordert eine Vortat und ist damit anlassbezogen sowie örtlich und zeitlich beschränkt, sodass sie im Verhältnis zu der betroffenen Person verhältnismäßig sein dürfte.

Forschungs- und forschende Nichtregierungsorganisationen

Es besteht ein andauernder Wettbewerb zwischen denjenigen, die Social Bots und Deepfakes entwickeln, und denjenigen, die diese Phänomene eindämmen wollen. Letztere sollten unterstützt werden – auch mithilfe von KI-Systemen. Hier könnten Forschungs- und forschende NGOs unterstützen:

- **Verfahren zur Mobilisierung ethischer Leitlinien entwickeln:** Der Dual-Use-Charakter von KI-Systemen erfordert, dass ethische Leitlinien auch auf andere Kontexte übertragbar sein müssen. Wie diese Mobilisierung ethischer Leitlinien gelingen kann, sollte die Forschung analysieren.

- **Bestehende KI-Anwendungen zur Detektion, Kennzeichnung und Löschung von Desinformation und deren Verbreitungswegen verbessern und weitere KI-Anwendungen entwickeln:** Es ist wichtig, dass KI-Anwendungen zur Eindämmung von Desinformation weiterentwickelt und stetig verbessert werden. Idealerweise erfolgen diese Forschungsprojekte unter Einbindung unterschiedlicher Disziplinen (wie Technik, Journalistik, Psychologie, Ethik und Recht).

- **Analysieren, welche Organisationsstruktur zielführend ist, um die Entwicklung und Verbreitung von – u. a. mithilfe von KI erstellten – Desinformationen einzudämmen:** Hierbei ist eine Meta-Perspektive einzunehmen und zu fragen, welche (offiziellen) Stellen sich der Problematik annehmen sollten und ob noch neue Stellen zu schaffen sind. Es sollte darauf geachtet werden, dass Desinformationen wirksam bekämpft werden, ohne dass die Meinungsfreiheit oder die legitime demokratische Auseinandersetzung behindert werden.

Die (kritische) Öffentlichkeit

Die Entscheidung über das Für und Wider des Einsatzes von KI im Kontext von Wahlen obliegt dem demokratischen Souverän, also der Bürgerschaft. Umso wichtiger sind funktionierende Öffentlichkeiten und eine aktive, informierte Bürgerschaft für eine kritische und kritikfähige Begleitung der Entwicklung von KI-Systemen. Sie könnte:

- **(Digitale) Kompetenzen zur Bewertung von (Des-)Informationen aufbauen:** Hierfür müssen individuelle Anstrengungen, sich zu informieren, und allgemeine Angebote zum Aufbau von (digitalen) Kompetenzen ineinandergreifen (vgl. Heesen et al. 2020). Die Angebote zur Stärkung von Medienmündigkeit und demokratischer Teilhabe – vor allem im Kontext von Wahlen – müssen alle Ebenen der Gesellschaft adressieren (vorzugsweise über Bildungsinstitutionen).

- **Öffentlich-rechtliche Kommunikationsplattformen als neutrale Informationsquellen nutzen:** Kommunikationsplattformen der öffentlichen Hand bieten nichtkommerzielle Informationsangebote für Bürgerinnen und Bürger mit hohen Datenschutz- und Qualitätsstandards abseits von Werbeinteressen. Sie bieten damit eine Alternative zum teils intransparenten Handeln der Plattformbetreiber und den teilweise KI-basierten Methoden der Electoral Content Moderation.

KI-Entwickelnde

Entwickelnde tragen für die von ihnen gestalteten KI-Systeme eine moralische Verantwortung. Deshalb sollten sie sich dafür verantwortlich fühlen, dass KI-Systeme die offene Meinungsbildung im Kontext der individuellen Wahlentscheidung fördern. Hierfür könnten sie:

- **Möglichkeiten zu erklärbarer KI (engl. explainable AI, kurz XAI) sowie zu fairer KI nutzen und erforschen:** Hierbei ist es wichtig, diese Forschung von klassifizierender KI auf generative KI auszuweiten, damit vertrauenswürdige KI-Systeme gegen (Des-)Information zur Verfügung stehen.

- **Bei der Entwicklung der Produkte verantwortungsvoll vorgehen:** Wo schon möglich, sollten Entwickelnde Filter gegen Desinformation, manipulierte Inhalte, Deepfakes und Ähnliches einsetzen. Des Weiteren sollten auch aus Perspektive der Manipulationsmöglichkeiten heraus Aspekte wie Privacy by Design und Datensparsamkeit integriert werden.

- **Risikoanalyse ausweiten:** Entwickelnde sollten eine mögliche Manipulationsanfälligkeit beziehungsweise die Sicherheit vor Manipulationen für KI-Technologien, die den öffentlichen Diskurs beeinflussen, zum festen Bestandteil der Risikoanalyse/des Impact Assessments machen.

5. Fazit

Wahlen (und damit auch die individuelle Wahlentscheidung) sind ein wesentliches Element unserer Demokratie. Diese lebt davon, dass unterschiedliche Anliegen gegeneinander abgewogen und am Ende des Prozesses gut begründete Entscheidungen getroffen werden. Die Idee, dass dieser Prozess berechenbar wird, ist nicht neu (vgl. „Cybersyn" in Chile 1971–1973, Krauch 1972; Kevenhörster 1984). Inzwischen fallen jedoch auch in politischen Prozessen immer mehr Daten an (vgl. „Dataist State", Fourcade & Gordon 2020). Deshalb steigt die Nachfrage nach KI-basierten Auswertungsmethoden und belastbaren rechnerischen Vorhersagen sowie nach Methoden, auf Basis dieser Daten Entscheidungen zu beeinflussen (vgl. Heesen 2020). Es liegt nahe, dass sich diese Nachfrage auch auf Wahlen inklusive individueller Wahlentscheidungen und deren Beeinflussung bezieht.

Die Untersuchung hat gezeigt, dass der Einsatz von KI-Systemen das Potenzial hat, die Meinungsbildung hin zur individuellen Wahlentscheidung und damit das Funktionieren unserer Demokratie zu verbessern. Dieses Potenzial wird gegenwärtig allerdings noch nicht ausgeschöpft. Gleichzeitig werden Informationen zunehmend mithilfe von KI-Systemen verbreitet. Möglichen Risiken kann stellenweise mithilfe von KI-Systemen begegnet werden – allerdings können diese wiederum neue Risiken mit sich bringen.

Das Autorenteam plädiert dafür, die vorgestellten Potenziale und Herausforderungen, die der Einsatz von KI-Systemen im Kontext von Wahlen auf die individuelle Wahlentscheidung haben kann, weiter zu beobachten. Darüber hinaus ist nach Möglichkeiten zu suchen, KI-Systeme als Werkzeug zur Unterstützung der Meinungsbildung hin zur individuellen Wahlentscheidung und damit zur Stärkung der Demokratie einzusetzen. Hierzu ist das Schaffen von Vertrauen grundlegend. Dies kann über das Herstellen von Transparenz über die Funktionsweise und den Einsatz von KI-Systemen gelingen. Im Sinne eines überlegten Vorgehens sollte zuerst der Bedarf nach Verbesserungen ermittelt werden, anschließend die Optionen des Einsatzes für KI-Systeme analysiert und im dritten Schritt mögliche Konsequenzen des Einsatzes betrachtet werden. An dieser Stelle weist das Autorenteam auch darauf hin, dass KI-Systeme selbstverständlich nicht nur im Zusammenhang mit Wahlen auf die Demokratie einwirken können. Das breitere Verhältnis von KI und Demokratie ist im Zuge weiterer Arbeiten zu untersuchen.

Literatur

Aider, H. et al. (2019): The State of Deepfakes. Landscape, Threats, and Impact, Deeptrace.

AI High Level Expert Group (HILEG) (2019): Ethics Guidelines for trustworthy AI. https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai (abgerufen am 07.05.2021).

Berlant, L. (2012): Cruel Optimism, Duke University Press.

Biederbeck, M. & Bülow G. (2021): Wie funktionieren Fakten-Checks auf Facebook. https://wasmitmedien.de/2021/01/29/wie-funktionieren-fakten-checks-auf-facebook-max-biederbeck-und-guido-buelow/ (abgerufen am 25.06.2021).

Bischof, J. (2019): A surprising number of people trust AI to make better policy decisions than politicians. https://qz.com/1576057/could-ai-make-better-policy-than-politicians/ (abgerufen am 25.06.2021).

Bode, L. & Vraga, E. K. (2015): In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. In: J Commun, 65, S. 619-638.

Bond, R. et al. (2012): A 61-million-person experiment in social influence and political mobilization. In: nature, 489, S. 295-298.

Borgesius, F. et al. (2018): Online Political Microtargeting: Promises and Threats for Democracy. In: Utrecht Law Review, 1/2018, S. 82-96.

Bundesministerium für Bildung und Forschung (2019): KI-Erklärbarkeit und Transparenz. https://www.softwaresysteme.pt-dlr.de/de/ki-erkl-rbarkeit-und-transparenz.php (abgerufen am 24.06.2021).

California Legislative Information (2019): Assembly Bill No. 730 Elections: deceptive audio or visual media. https://www.leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730 (abgerufen am 24.06.2021).

Chaos Computer Club e.V. (Hrsg.) (2017): Analyse einer Wahlsoftware – Version 1.1. Online unter: Chaos Computer Club. https://www.ccc.de/system/uploads/230/original/PC-Wahl_Bericht_CCC.pdf (abgerufen am 07.05.2021).

Chester, J. & Montgomery, K. C. (2017): The Role of Digital Marketing in Political Campaigns. In: Internet Policy Review 4/2017. https://policyreview.info/archives/2017/issue-4 (abgerufen am 25.06.2021).

Deutscher Bundestag (2019): Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Manuel Höferlin, Frank Sitta, Grigorios Aggelidis, weiterer Abgeordneter und der Fraktion der FDP – Drucksache 19/15210. https://dip21.bundestag.de/dip21/btd/19/156/1915657.pdf (abgerufen am 25.06.2021).

Digitale Gesellschaft e.V. (Hrsg.) (2021): Was sind Uploadfilter? https://digitalegesellschaft.de/2020/07/was-sind-uploadfilter-neue-broschuere-der-digitalen-gesellschaft-klaert-auf/ (abgerufen am 24.06.2021).

D64 (2019): Politische Kommunikation im digitalen Raum. https://d-64.org/wp-content/uploads/2019/05/d64-Thesenpapier-Politische-Kommunikation-Online.pdf (abgerufen am 28.05.2021).

Europäische Kommission (2021): Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain union legislative acts. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence (abgerufen am 07.05.2021).

Fourcade, M. & Gordon, J. (2020): Learning Like a State: Statecraft in the Digital Age. In: Journal of Law and Political Economy, 1(1).

Gerlitz, C. & Helmond, A. (2013): The like Economy: Social Buttons and the Data-Intensive Web. In: New Media & Society, 15(8), S. 1348-1365.

Gorwa, R., Binns, R. & Katzenbach, C. (2020): Algorithmic content moderation: Technical and political challenges in the automation of platform governance. https://journals.sagepub.com/doi/10.1177/2053951719897945 (abgerufen am 07.05.2021).

Hamborg, F., Donnay, K. & Gipp, B. (2019): Automated identification of media bias in news articles: an interdisciplinary literature review. In: International Journal on Digital Libraries, S. 391-415.

Heesen, J. (2020): Verantwortlich Forschen mit und zu Big Data und Künstlicher Intelligenz. In: Anja Seibert-Fohr (Hrsg.): Entgrenzte Verantwortung. Zur Reichweite und Regulierung von Verantwortung in Wirtschaft, Medien, Technik und Umwelt, New York/Heidelberg, S. 285-303.

Heesen, J. et al. (2020): Ethik-Briefing. Whitepaper aus der Plattform Lernende Systeme, München. https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen/AG3_Whitepaper_EB_200831.pdf (abgerufen am 07.05.2021).

Helbing D. (2015): The Automation of Society is Next: How to Survive the Digital Revolution. CreateSpace Independent Publishing Platform: South Carolina.

Helbing D. (2021): Digital Democracy (Democracy 2.0, 3.0, 4.0). In: Next Civilization. Springer, Cham.

Hochschild, A. R. (2016): Strangers in their own land: anger and mourning on the American right. New York: New Press.

Hofeditz, L. et al. (2019): Meaningful Use Of Social Bots? Possible Applications In Crisis Communication During Disasters. European Conference on Information Systems. https://aisel.aisnet.org/ecis2019_rp/138 (abgerufen am 07.05.2021).

Johnson, K. (2020): How AI can empower communities and strengthen democracy. https://venturebeat.com/2020/07/04/how-ai-can-empower-communities-and-strengthen-democracy/ (abgerufen am 30.07.2021).

Jonsson O., De Tena C. L. (2021): European Tech Insights: Part II Embracing and Governing Technological Disruption. https://www.ie.edu/cgc/research/european-tech-insights/?submissionGuid=cfd2c5db-0cc7-449c-9e4f-d1224edad93d#download-cgc (abgerufen am 09.07.2021).

Jungherr, A., Rivero, G. & Gayo-Avello, D. (2020): Retooling Politics. In: Retooling Politics: How Digital Media Are Shaping Democracy. Cambridge: Cambridge University Press.

Kevenhörster, H. (1984): Politik im elektronischen Zeitalter. Politische Wirkungen der Informationstechnik. Baden-Baden: Nomos.

Kind, S. et al. (2017): Social Bots. TA-Vorstudie. https://www.tab-beim-bundestag.de/de/pdf/publikationen/berichte/TAB-Arbeitsbericht-hs003.pdf (abgerufen am 07.07.2021).

Krauch, H. (1972): Die Computer-Demokratie. Hilft uns die Technik entscheiden? München: Goldmann.

Kwok, A. O. J. & Koh, S. (2020): Deepfake: a social construction of technology perspective. Current Issues in Tourism. https://www.tandfonline.com/doi/full/10.1080/13683500.2020.1738357 (abgerufen am 07.05.2021).

Labbe, M. (2020): The deepfake 2020 election threat is real, but containable. https://searchenterpriseai.techtarget.com/feature/The-deepfake-2020-election-threat-is-real-but-containable (abgerufen am 24.06.21).

Löber, L. I. & Roßnagel, A. (2019a): Das Netzwerkdurchsetzungsgesetz in der Umsetzung. Bilanz nach den ersten Transparenzberichten, Multimedia und Recht (MMR), 21. Jg. (2019), Heft 2, S. 71-76.

Löber, L. I. & Roßnagel, A. (2019b): Kennzeichnung von Social Bots – Transparenzpflichten zum Schutz integrer Kommunikation, Multimedia und Recht (MMR), 21. Jg. (2019), Heft 8, S. 493-498.

Löber, L. I. & Roßnagel, A. (2020): Desinformation aus der Perspektive des Rechts. In: Steinebach, M./Bader, K./Rinsdorf, L./Krämer, N./Roßnagel, A. (Hrsg.), Desinformation aufdecken und bekämpfen. Interdisziplinäre Ansätze gegen Desinformationskampagnen und für Meinungspluralität, Nomos Verlag, Baden-Baden 2020, S. 149-194.

Makse, H. A. & Zhou, Z. (2019): Artificial intelligence for elections: the case of 2019 Argentina primary and presidential election. https://arxiv.org/abs/1910.11227 (abgerufen am 24.06.21).

Mena, P. (2020): Cleaning Up Social Media: The Effect of Warning Labels on Likelihood of Sharing False News on Facebook. Policy & Internet, 12, S. 165-183.

nature (2021): Am I arguing with a machine? AI debaters highlight need for transparency. https://www.nature.com/articles/d41586-021-00867-6 (abgerufen am 07.06.2021).

Oh, O., Agrawal, M. & Rao, H. (2013): Community intelligence and social media services: a rumor theoretic analysis of tweets during social crises. In: MIS Quarterly, 37(2), S. 407-426.

Papacharissi, Z. (2015): Affective publics and structures of storytelling: sentiment, events and mediality. Information. In: Communication & Society, 19(3), S. 307-324.

Pennycook, G. et al. (2020): The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. In: Management Science, 66(11), S. 4944-4957.

Polonski, S. (2017): Artificial intelligence can save democracy, unless it destroys it first. https://medium.com/@drpolonski/artificial-intelligence-can-save-democracy-unless-it-destroys-it-first-7b1257cb4285 (abgerufen am 24.06.2021).

Richard, É. et al. (2020): Instagram-Algorithmus: Wer gesehen werden will, muss Haut zeigen. https://algorithmwatch.org/de/haut-zeigen-auf-instagram/ (abgerufen am 25.06.2021).

Rivero, N. (2021): The risk of putting warning labels on election misinformation. https://qz.com/1925272/the-risk-of-putting-warning-labels-on-election-misinformation/ (abgerufen am 03.05.2021).

Ross, B. et al. (2018): Fake News on Social Media: The (In)Effectiveness of Warning Messages. International Conference on Information Systems.

Ryall, J. (2021): ‚Deepfakes' rattle South Korea's tech. https://www.dw.com/en/deepfakes-rattle-south-koreas-tech-culture/a-56310213 (abgerufen am 03.05.2021).

Schaake M. & Reich R. (2020): Election 2020: Content Moderation and Accountability. https://hai.stanford.edu/sites/default/files/2020-10/HAI_CyberPolicy_IssueBrief_3.pdf (abgerufen am 03.05.2021).

Stieglitz, S., Mirbabaie, M. & Fromm, J. (2018): Understanding Sense-Making on Social Media during Crises: Categorization of Sense-Making Barriers and Strategies. In: International Journal of Information Systems for Crisis Response and Management, 9(4), S. 49-69.

Toto, C. S. & Keating T. (2020): Protecting Elections: Regulating Deepfakes in Politics. https://www.internetandtechnologylaw.com/elections-deepfakes-politics-regulation/ (abgerufen am 24.06.21).

Twitter (2021): Die Twitter Regeln. https://help.twitter.com/de/rules-and-policies/twitter-rules (abgerufen am 28.06.21).

Varol, O. & Uluturk, I. (2018): Deception strategies and threats for online discussions. First Monday, 23, S. 5-7.

von Ungern-Sternberg, A. (2018): Demokratische Meinungsbildung und künstliche Intelligenz (Democracy, Public Opinion Formation, and Artificial Intelligence). https://ssrn.com/abstract=3400756 (abgerufen am 07.05.2021).

Vowe, G. (2021): Use Case Politics. https://www.heicad.hhu.de/en/research/manchot-research-group-decision-making-with-the-help-of-artificial-intelligence/use-case-politics (abgerufen am 24.06.2021).

Wittmann, L. (2021): Wenn die CDU ihren Wahlkampf digitalisiert. https://lilithwittmann.medium.com/wenn-die-cdu-ihren-wahlkampf-digitalisiert-a3e9a0398b4d (abgerufen am 17.05.2021).

Zhang et al. (2013): On the impact of social botnets for spam distribution and digital-influence manipulation. In: 2013 IEEE Conference on Communications and Network Security (CNS), S. 46-54.

Zhou, Z. et al. (2021): Why polls fail to predict elections. https://arxiv.org/abs/2101.11389 (abgerufen am 17.05.2021).

Über dieses Whitepaper

Die Autorin 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 der Plattform Lernende Systeme.

Autorinnen und Autoren
PD Dr. Jessica Heesen, Universität Tübingen
Prof. Dr. Christoph Bieber, Universität Duisburg-Essen
Prof. Dr. Armin Grunwald, Universität Karlsruhe, TAB
Prof. Dr. Tobias Matzner, Universität Paderborn
Prof. Dr. Alexander Roßnagel, Universität Kassel, Hessischer Beauftragter für Datenschutz und Informationsfreiheit

Redaktion
Stephanie Dachsberger, Geschäftsstelle der Plattform Lernende Systeme
Christine Wirth, Geschäftsstelle der Plattform Lernende Systeme

Über die Plattform Lernende Systeme

Lernende Systeme im Sinne der Gesellschaft zu gestalten – mit diesem Anspruch wurde die Plattform Lernende Systeme im Jahr 2017 vom Bundesministerium für Bildung und Forschung (BMBF) auf Anregung des Fachforums Autonome Systeme des Hightech-Forums und acatech – Deutsche Akademie der Technikwissenschaften initiiert. Die Plattform bündelt die vorhandene Expertise im Bereich Künstliche Intelligenz und unterstützt den weiteren Weg Deutschlands zu einem international führenden Technologieanbieter. Die rund 200 Mitglieder der Plattform sind in Arbeitsgruppen und einem Lenkungskreis organisiert. Sie zeigen den persönlichen, gesellschaftlichen und wirtschaftlichen Nutzen von Lernenden Systemen auf und benennen Herausforderungen und Gestaltungsoptionen.

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
September 2021

Bildnachweis
scyther5/iStock/Titel

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

Folgen Sie uns auf Twitter: @LernendeSysteme

Empfohlene Zitierweise
Jessica Heesen et al. (Hrsg.): KI-Systeme und die individuelle Wahlentscheidung – Chancen und Herausforderungen für die Demokratie. Whitepaper aus der Plattform Lernende Systeme, München 2021.
DOI: https://doi.org/10.48669/pls_2021-1