

Demokratie im KI-Zeitalter: Vom Werkzeug zu Infrastruktur

Katja Muñoz und Emma Carlotta Laumann

Zusammenfassung

Das Kapitel analysiert zentrale Entwicklungen im Zusammenspiel von Künstlicher Intelligenz und dem digitalen Informationsraum demokratischer Systeme. Es gliedert sich in zwei Teile. Zunächst werden drei zentrale Trends identifiziert: die Verstärkung bestehender Narrative, die Erosion von Vertrauen sowie hyperpersonalisierte Kampagnen im Kontext des Superwahljahres 2024. Der Fokus liegt dabei auf dem Einfluss von KI auf Wahlkämpfe, wobei deutlich wird, dass diese Dynamiken nicht auf kurzfristige Wahlkampfphasen beschränkt sind, sondern den politischen Diskurs kontinuierlich prägen.

Der zweite Teil beleuchtet die tiefere Herausforderung einer strukturellen Transformation: KI entwickelt sich von einem Werkzeug zur Inhaltserstellung hin zu einer Infrastruktur, die bestimmt, wie Informationen verbreitet, verarbeitet und bewertet werden. Darauf aufbauend identifiziert das Kapitel vier Bedrohungslogiken jenseits klassischer Inhaltsmanipulation: Narrative Testing, institutionelles Versagen in der Plattformaufsicht, infrastrukturelle Manipulation sowie relationale Einflussnahme. Abschließend plädiert das Kapitel für den Aufbau hyperlokaler Resilienz als zentrale Antwort auf diese Entwicklungen.

1. Einleitung

Künstliche Intelligenz (KI) spielt heute eine zentrale Rolle in Wahlkämpfen und bei der Neuordnung digitaler politischer Strukturen. Die Prognosen für das Jahr 2024 fielen düster aus. Insbesondere mit Blick auf Wahlen wurde vielfach gewarnt, KI könne sich zu einer disruptiven Gefahr entwickeln. Der Global Risks Report des *World Economic Forum* (2024) stufte „KI-generierte Fehl- und Desinformation“ als eines der fünf größten Risiken für die folgenden zwei Jahre ein (WEF 2024, S.13). Mit dem Blick aus dem Jahr 2026 stellt sich daher die Frage: Was ist tatsächlich eingetreten? Wie verändert KI unsere Demokratie?

Bevor man nun konkret betrachtet, welchen Einfluss KI in den letzten beiden Jahren auf Demokratien hatte, bedarf es einer Klärung des KI-Begriffs. Die Relevanz von KI im politischen Kontext lässt sich anhand verschiedener Anwendungsbereiche systematisieren: Generative KI, Analytische KI, algorithmische Kuratierung durch KI und KI-Infrastruktur.

Generative KI wird als eine „spezialisierte Form der KI“ definiert, die „darauf abzielt, neue Inhalte wie Texte, Bilder, Videos und sogar Musik oder Sprache zu erstellen“ (Kruschinski u. a. 2025, S.10). *Analytische KI* bezeichnet hingegen den Einsatz von KI-Verfahren zur Analyse großer Datenmengen und zur Mustererkennung. Daran schließt die *algorithmische Kuratierung durch KI* an, bei der, basierend auf den analysierten Mustern, Inhalte selektiv angezeigt werden. Mit *KI-Infrastruktur* ist schließlich die Entwicklung gemeint, dass „KI [...] zunehmend zur technischen Grundschicht wird, über die Menschen Informationen erhalten und Entscheidungen treffen“ (Muñoz 2026, S. 4)

Grundlegend ist die Erkenntnis, dass KI keine vollständig neue politische Kommunikation und Dynamiken in Wahlkämpfen erzeugt, sondern bestehende Strategien in ihrer Effizienz und Reichweite verstärkt und beschleunigt (Bawidamann/Gilbert 2025). Weniger bedeutend ist der Einfluss von KI-generierter Desinformation. Viel relevanter ist, wie KI die allgemeine Struktur des digitalen Raumes neu ordnet und wie politische Parteien, Organisationen und ausländische Akteure darauf reagieren. Es geht zentral nicht nur darum, *was* kommuniziert wird (ob das nun KI-generiert ist oder nicht), sondern um den Einfluss darauf, wie, worüber, wo, wann und nach welchen Regeln politische Debatten stattfinden. Zusammengefasst bestimmt KI den Zugang zu und die Kuration von Informationen.

Dieses Kapitel konzentriert sich sowohl auf den konkreten Einsatz von KI in der Wahlkampfführung anhand der Erkenntnisse der *AI-Democracy Initiative* (DGAP 2024), als auch auf die übergeordneten Dynamiken des Einflusses von KI auf den digitalen Raum. Chancen und Risiken werden dabei gleichermaßen berücksichtigt, um eine differenzierte Bewertung zu ermöglichen. Leitend ist dabei die zentrale These, dass Demokratien nicht durch spektakuläre Wahlkampf-Deepfakes bedroht sind, sondern durch drei langfristige Dynamiken: (1) datenbasierte Hyperpersonalisierung, (2) die Transformation von KI zur zentralen digitalen Infrastruktur und (3) relationale Beeinflussung durch autonome KI-Agenten.

Das Kapitel beginnt im zweiten Abschnitt damit, die Rolle von KI in Wahlkämpfen zusammenzufassen, und identifiziert dabei drei übergeordnete Trends: (1) die Verstärkung der eigenen Narrative durch KI-generier-

te Inhalte, (2) die Erosion des Vertrauens und (3) datenbasierte Kampagnen und Hyperpersonalisierung. Der dritte Abschnitt verdeutlicht anhand entstehender Bedrohungslogiken, dass KI nicht auf Wahlkampfphasen beschränkt bleibt, sondern sich von einem Werkzeug zu einer Infrastruktur entwickelt. Die vier Bedrohungslogiken, die hierbei identifiziert wurden, sind (1) Narrative Testing von Botschaften, (2) institutionelles Versagen bei der Plattformaufsicht, (3) infrastrukturelle Manipulation und (4) relationale Einflussnahme. Das Kapitel endet mit einem Appell für hyperlokale Resilienz als Antwort auf die systematische Transformation des digitalen Raumes durch KI.

2. KI in Wahlkämpfen

Die *AI-Democracy Initiative* der Deutschen Gesellschaft für Auswärtige Politik (DGAP) in Zusammenarbeit mit der Friedrich-Naumann-Stiftung sowie der Konrad-Adenauer-Stiftung analysierte im Jahr 2024 den Einfluss von KI auf Wahlen im Jahr 2024. Das sogenannte “Superwahljahr” zählte weltweit mehr als 70 nationale und regionale Wahlen (Kohler/Voigt 2024). Die Untersuchung konzentrierte sich auf sechs demokratische Länder: Indien, Südafrika, Frankreich, Mexiko, Deutschland und die Vereinigten Staaten, in denen nationale oder regionale Wahlen sowie Wahlen zum Europäischen Parlament abgehalten wurden¹.

Tabelle 1: Analysezeitraum und Anzahl KI-bezogener Vorfälle nach Land

Land	Analysezeitraum	Identifizierte Vorfälle (N)	Artikel insgesamt
Deutschland	01.01.–23.09.2024	37	4.140
Indien	01.01.–30.06.2024	62	1.536
Südafrika	01.01.–30.06.2024	12	261
Mexiko	01.01.–30.06.2024	48	5.934
Frankreich	01.01.–12.07.2024	14	482
USA	01.01.–11.11.2024	146	46.022

Quelle: *AI-Democracy Initiative / DGAP/ Friedrich-Naumann-Stiftung, 2024*

1 Die untersuchten Wahlen sind: Deutschland: Europawahl (6.–9. Juni 2024), Landtagswahlen in Thüringen, Sachsen und Brandenburg (1. Sept. / 22. Sept. 2024); Indien: Parlamentswahl (19. April–1. Juni 2024); Südafrika: Parlamentswahl (29. Mai 2024); Mexiko: Präsidentschafts- und Parlamentswahl (2. Juni 2024); Frankreich: Europa-

Die Studie kombinierte quantitative und qualitative Methoden, um ein möglichst umfassendes Bild der Rolle von KI in den jeweiligen Wahlkämpfen zu gewinnen. Das Ziel war es durch datengetriebene, quantitative Analyse die Anzahl KI-bezogener Vorfälle² zu erfassen.

Methodisch bestand die Herausforderung, dass der Einsatz von KI auf sozialen Plattformen häufig nicht transparent offengelegt wird, daher ist eine vollständige Erfassung KI-generierter Inhalte kaum möglich. Um dennoch eine komparative Datengrundlage zu schaffen, nutzte die Studie das Medienmonitoring-Tool *Cision*, mit dem KI-bezogene Online-Artikel aus digitalen Nachrichtenmedien in den Monaten vor den jeweiligen Wahlen dokumentiert wurden.

Für die Suche wurden länderspezifische Suchbegriffe entwickelt, wobei eine einheitliche Struktur beibehalten wurde, die sich auf KI, Social-Media-Plattformen sowie prominente politische Parteien und Kandidaten konzentrierte. Diese Suchbegriffe wurden an die Sprache und den lokalen Kontext jedes Landes angepasst, um ihre Relevanz sicherzustellen³. Die Anzahl der gefundenen Artikel variierte je nach Land erheblich (siehe Tab. 1). Die gesammelten Artikel wurden manuell gefiltert, um spezifische KI-Vorfälle zu identifizieren⁴.

Die Verwendung von Zeitungsartikeln zur Identifizierung von KI-Vorfällen bedeutet zwei Einschränkungen. Erstens, es wurden nur Vorfälle berücksichtigt, über die in den Medien berichtet wurde, was bedeutet, dass der Datensatz wahrscheinlich größere, stärker publizierte Vorfälle erfasste, während kleinere oder weniger publizierte Ereignisse übersehen werden konnten. Diese Einschränkung ist besonders ausgeprägt bei Vorfällen, die ausschließlich auf Social-Media-Plattformen auftreten und möglicherwei-

wahl (6.–9. Juni 2024), Parlamentswahl in 2 Runden (30. Juni / 7. Juli 2024); USA: Präsidentschaftswahl (5. November 2024).

- 2 Der Begriff „KI-bezogene Vorfälle“ umfasst hier nicht ausschließlich Deepfakes oder Desinformation, sondern sämtliche dokumentierten Nutzungen von KI-Technologien im politischen Kontext im Vorfeld von Wahlen. Dies schließt sowohl die Berichterstattung über Deepfakes als auch über die allgemeine Nutzung von KI durch politische Akteure ein.
- 3 In Ländern mit ausgeprägter sprachlicher Vielfalt, wie beispielsweise Indien, stellt dies eine methodische Herausforderung dar, da eine vollständige Abdeckung aller regionalen Sprachen nicht möglich ist. Diese Limitation der Datenerhebung muss bei der Interpretation der Ergebnisse berücksichtigt werden.
- 4 Es ist zu beachten, dass über einige Vorfälle ausführlich berichtet wurde, wobei mehrere Artikel oder Schlagzeilen auf dasselbe Ereignis Bezug nahmen. Um Doppelzählungen zu vermeiden, wurde jeder Vorfall nur einmal in die Analyse aufgenommen.

se nicht von traditionellen Medienhäusern aufgegriffen wurden. Hieraus resultiert zweitens, dass die aus der Medienanalyse identifizierten KI-bezogenen Vorfälle nicht vollständig sind. Dennoch sind die von den Medien behandelten Vorfälle wohl die wirkungsvollsten und für die öffentliche Debatte relevantesten und bieten eine robuste Grundlage. Die quantitative Medienbeobachtung diente daher als explorativer Ansatz, um zunächst das Ausmaß KI-bezogener Berichterstattung vor den jeweiligen Wahlen zu erfassen. Unabhängig davon wurden als Teil der Initiative vertiefte Länderstudien in Indien (Dhanuraj u. a. 2024), Südafrika (Van Damme u. a. 2024), Mexiko (Del Pozo u. a. 2024), USA (Muñoz 2024) und Deutschland (Muñoz/Laumann 2025) durchgeführt.

2.1 Ergebnisse AI Democracy Initiative

2024 war das Jahr, in dem die Welt nach der großen KI-Wahlkampf-Disruption suchte, doch dazu kam es nicht. Keine Wahl wurde durch einen einzelnen viralen Deepfake entschieden, kein demokratisches System durch synthetische Videos ins Wanken gebracht (Jackson/Conroy 2024; Simon u. a. 2024). Das heißt allerdings nicht, dass KI keinen Einfluss auf Wahlen hatte.

Die quantitative Medienbeobachtung mittels *Cision* über die Monate vor den jeweiligen Wahlen bestätigt diesen Befund (siehe Tab. 1). In allen sechs untersuchten Ländern wurden KI-bezogene Vorfälle im Wahlkontext identifiziert, wobei die USA mit 146 dokumentierten Vorfällen die höchste Zahl aufweisen, gefolgt von Indien mit 62 und Mexiko mit 48. In Deutschland wurden 37 Vorfälle erfasst, während in Frankreich 14 und in Südafrika 12 erfasst wurden. Die Diskrepanz zwischen der Gesamtzahl KI-bezogener Artikel und den identifizierten konkreten Vorfällen resultiert aus dem methodischen Vorgehen: Die quantitative Erhebung erfasste zunächst die gesamte KI-bezogene Berichterstattung. Die qualitative Analyse dieser Vorfälle sowie die vertieften Länderstudien ermöglichten es, über die bloße Häufigkeit hinaus Muster in der KI-Nutzung zu identifizieren. Drei Trends, die in den Wahlkämpfen beobachtet wurden, werden im nächsten Abschnitt in den aktuellen Forschungsstand eingeordnet.

2.2 Verstärkung der eigenen Narrative durch KI-generierte Inhalte

Der erste globale Trend betrifft die gezielte Nutzung von KI zur *Verstärkung eigener Narrative*. Hierbei setzen politische Akteure KI-generierte Inhalte ein, um ihr Storytelling zu visualisieren und ihre Argumentationen zu stützen. Der Einsatz von KI-generierten Inhalten erlaubt es, diese klassische Methodik der Wahlkampfführung gezielt zu verstärken.

Zentrale Schlagwörter sind hierbei *Emotionalisierung* und *Polarisierung* von komplexen Themen, auch mithilfe von KI. Besonders Bilder und Videos haben eine große emotionale Wirkung (Breuer 2024). Die Forschung zeigt, dass selbst wenn Inhalte als unauthentisch oder synthetisch wahrgenommen werden, die grundlegende Botschaft dennoch Auswirkungen haben kann (Cools 2025; Clark/Lewandowsky 2026) und dieser Effekt auch in der Erinnerung verankert bleibt (Nash u. a. 2009). Mobilisierung durch emotionale Aktivierung im digitalen Raum ist daher ein effektives Mittel, vor allem wenn es zu einem Wiederholungseffekt kommt, bei welchem Ängste systematisch instrumentalisiert werden.

Diese Form der Mobilisierung im digitalen Raum durch KI-generierte Fotos und Videos konnte 2024 bei Wahlen in den USA beobachtet werden (Muñoz 2024), vor allem in Form von sogenannten *Memes*. Diese können als "kulturelle Informationseinheiten, die sich viral im Internet verbreiten, meist als humorvolle Bilder oder kurze Videos mit überlagertem Text" (Munoz/Laumann 2025, S. 129) beschrieben werden. Viele dieser Inhalte zeichnen sich darüber hinaus dadurch aus, dass es sich oftmals nicht um Deepfakes von hoher Qualität handelte, sondern vielmehr um *Cheapfakes*, wobei oftmals nicht festgestellt werden kann, ob es sich dabei dann um KI-generierte Inhalte handelte, oder diese vielmehr mit gängiger Software verändert wurden (Paris/Donovan 2019).

Die Anwendung emotionalisierender KI-generierter Inhalte hat sich 2025 nochmals verstärkt. Längst wird beispielsweise von einem sogenannten „Clicktatorship“ (Elliott 2026) unter Präsident Trump gesprochen. Gemeint ist hiermit, dass von einer Regierung jeder politische Prozess als eine Form von Inhalt auf Plattformen kommuniziert wird (Moynihan 2026). Engagement und Reichweite werden in diesem Prozess oftmals wichtiger als Wahrheit. Solch ein Engagement zu verstärken, ist durch emotionalisierende und maßgeschneiderte KI-Inhalte noch leichter geworden.

2.3 Erosion des Vertrauens

Der zweite zu beobachtende Trend ist eine Erosion des Vertrauens in den digitalen Informationsraum durch KI-generierte Inhalte. Bereits während des Superwahljahrs 2024 führte die Zunahme solcher Inhalte zu generalisierten Authentizitätszweifeln. Diese werden oftmals durch den sogenannten Liar's Dividend-Effekt noch verstärkt (Goldstein/Lohn 2024), welcher sich auch 2024 im Wahlkampf in den USA und Mexiko beobachten ließ (Muñoz 2024; Del Pozo u. a. 2024). Dabei werden authentische Inhalte oftmals von politischen Akteuren als Fälschungen delegitimiert. Dies kann nicht nur einen polarisierenden Effekt auf die Gesellschaft haben, sondern auch zu einer Krise der Realität führen, da die Bevölkerung nicht mehr differenzieren kann, welche Inhalte authentisch sind.

Dieses Phänomen wurde 2025 durch den *AI-Slop* weiter intensiviert. *AI-Slop* bezeichnet massenhaft generierte KI-Inhalte niedriger Qualität (Blome 2026). Diese sind meist nicht primär politisch motiviert, sondern verfolgen ökonomische Interessen. Da sie mithilfe von KI-Modellen kostengünstig zu erstellen sind und die Algorithmen Quantität oftmals vor Qualität bevorzugen, zielen die verantwortlichen Akteure auf eine Monetarisierung durch Klicks und Engagement ab. Dennoch beeinflussen sie den digitalen Informationsraum, indem sie digitale Plattformen mit synthetischen Inhalten übersättigen. Eine Studie von *AI Forensics* belegt beispielsweise, „dass auf TikTok in einer Stichprobe zu ausgewählten Hashtags 25% der Inhalte KI-generiert waren“ (Stanusch u. a. 2025, S. 4)

Dies kann zu einem Rückgang des Vertrauens in Informationen und Nachrichten allgemein führen (Jackson/Woolley 2025, S. 142), speziell aber zu einem Rückgang des Vertrauens in Nachrichten auf Social Media (Vacari/Chadwick 2020). Gepaart mit den Daten des Reuters Digital News Report (vgl. Abb. 1), die zeigen, dass Individuen über die letzten Jahre zunehmend Nachrichten online und auf Social Media konsumieren, deutet dies auf ein systematisches Vertrauensproblem im digitalen Informationsraum hin.

Abbildung 1: Bevölkerungsanteil (%) mit Social Media Nutzung für Nachrichten



Quelle: Reuters Digital News Report, 2025

2.4 Datenbasierte Kampagnen und Hyperpersonalisierung

Der dritte globale Trend betrifft die Nutzung von KI zur Analyse und zum gezielten Targeting spezifischer Wählergruppen. *Data-Driven Campaigning* ist hierbei zu einer Grundlage moderner Wahlkämpfe geworden (Darius/Römmele 2023). Wie bereits oben erwähnt sind soziale Netzwerke zu zentral für den politischen Diskurs geworden. Die daraus erhobene große Menge an Daten, die im digitalen Raum von jedem Individuum generiert werden, macht es theoretisch möglich, detaillierte Profile eines jeden Nutzenden zu erstellen. Auf dieser Basis können Parteien im Wahlkampf gezielte, personalisierte Wahlwerbung erstellen und verbreiten, die auf die einzelne Person zugeschnitten ist (Jungherr 2025; Gilga u. a. 2024). Dieses sogenannte *Microtargeting* (Borgesius u. a. 2018) wird durch KI

noch einfacher, da nicht nur die Analyse der Wählergruppen schneller möglich ist, sondern auch die Inhalte schneller erstellt werden können.

Zudem werden KI-Tools zum sogenannten *Social Listening* (Jackson/Martin 2024) in Wahlkämpfen genutzt, um Empfindungen und Meinungen im digitalen Raum in Echtzeit und ohne großen Aufwand zu beobachten. Dieses Tool erlaubt es politischen Akteuren in Echtzeit auf die für die Bevölkerung mithilfe von KI identifizierten, relevanten Themen einzugehen und ihre Botschaften maßgeschneidert und reaktiv anzupassen. Im heutigen Zeitalter, in dem die allgemeine Geschwindigkeit von Kommunikation rasant zugenommen hat, kann dies einen entscheidenden Vorteil für die Relevanz des Wahlkampfes bedeuten.

Besonders in Ländern wie Indien, die sich durch eine hohe demografische Diversität auszeichnen, war *Data Driven Campaigning* ein entscheidendes Mittel, um gezielte Wählergruppen innerhalb der Bevölkerung zu targeten (Dhanuraj u. a. 2024). Beispielsweise nutzte bei den Wahlen 2024 die BJP die sogenannte „SaraI“ App als zentrale digitale Wahlkampfplattform, wobei Individuen gezielt angesprochen und animiert wurden, sich die App herunterzuladen, da die Partei so Zugang zu gezielten Personendaten und somit auch Wählerdaten zu ihrem Nutzen generieren konnte. Hierbei konnten die Daten der Nutzer:innen konkreten lokalen Wahleinheiten zugeordnet werden (Dhanuraj u. a. 2024). Die App ist somit ein konkretes Beispiel für Mikrotargeting, wobei sie kleinste Wählergruppen identifiziert, clustert und die Relevanz von Daten in modernen Wahlkämpfen aufzeigt. Dieser Trend ist nicht nur ein Beispiel für die Effektivität und Spezifizierung moderner Wahlkämpfe, sondern verdeutlicht insbesondere die Dualität des KI-Einsatzes im digitalpolitischen Raum. Einerseits kann argumentiert werden, dass Personalisierung Bürger:innen gezielter mit ihren eigenen politischen Interessen und relevanten Informationen in Kontakt bringt. Dadurch können sich Individuen persönlich angesprochen und wahrgenommen fühlen, was demokratiefördernd wirken kann, indem es Menschen aktiver in politische Prozesse einbindet.

Die Chance direkter(er) politischer Kommunikation und auch Partizipation von Bürger:innen im Wahlkampf wurde auch 2024 im Wahlkampf von politischen Akteuren bereits erkannt. Der Fokus von KI-Anwendungen lag hier auf der Möglichkeit, politische Partizipation durch KI-Modelle zu fördern. Ein gutes Beispiel hierfür sind Chatbots, die auf den Websites politischer Kandidaten für den direkten Austausch zur Verfügung standen (Kruschinski u. a. 2025; Muñoz/Laumann 2024). Darüber hinaus setzten Organisationen auf Chatbots, beispielsweise um einen interaktiveren Ver-

gleich von Wahlprogrammen zu ermöglichen (Lohmann 2025). Ein solches Tool existierte für die Bundestagswahl 2025 in Form von ‘Wahl.Chat’.

Andererseits können diese personalisierten Ansprachen, insbesondere KI-gestützte Hyperpersonalisierung in großem Maßstab, zur Fragmentierung des öffentlichen Raumes beitragen. Ein Beispiel hierfür ist *ChatGPT Pulse* der Firma OpenAI aus dem Jahr 2025. Hier ist der Grad der Personalisierung im Vergleich zu klassischen Frage-Antwort-Formaten deutlich höher, da die Funktion individuell zugeschnittene, personalisierte Nachrichten-Updates ermöglicht (Investing 2025).

3. Jenseits der Wahlkampagne: KI als demokratische Dauerbedrohung

Wie die Analyse zeigt, war die Sorge vor gefälschten Inhalten im Jahr 2024 berechtigt, aber sie offenbart auch den einseitigen Fokus auf KI-generierten Inhalten. Insgesamt schaffen KI-Tools durch ihre Skalierbarkeit und Effizienz neue Chancen, aber auch neue Risiken in digitalen Wahlkämpfen. Die beobachteten Strategien beschränken sich dabei nicht auf die unmittelbare Wahlkampfphase, sondern prägen den politischen Diskurs kontinuierlich und nachhaltig.

Das Problem besteht längst nicht mehr nur darin, dass KI synthetische Inhalte erzeugt und als Beschleuniger bereits existierender Strategien fungiert. Die tiefere Gefahr liegt darin, Menschen systematisch und subtil zu beeinflussen. Neue Forschung zeigt, dass KI-Chatbots politische Einstellungen deutlich effektiver verschieben können als traditionelle Wahlwerbung (Feldman/Pappu 2025). Dazu kommt die Rolle, die Zugang zu Inhalten und die Kuration dieser spielen, insbesondere wenn KI immer mehr zur technischen Grundstruktur unserer Gesellschaft wird.

Während sich Regulierungsbehörden und Plattformen auf die Kennzeichnung synthetischer Inhalte konzentrierten, vollzog sich im Hintergrund eine tektonische Verschiebung. KI entwickelte sich von einem Werkzeug zur Content-Erstellung zu einer Infrastruktur, die bestimmt, wie Menschen Informationen erhalten, verarbeiten und bewerten. Die größte demokratische Herausforderung liegt daher nicht in Wahlkampfphasen, sondern in den Strukturen, die zwischen den Wahlen entstehen. Die vier folgenden Fälle repräsentieren qualitativ unterschiedliche Bedrohungslogiken jenseits von Inhaltsmanipulation (Narrative Testing von Botschaften, institutionelles Versagen bei der Plattformaufsicht, infrastrukturelle Manipulation und relationale Einflussnahme).

3.1 Strategisches Narrative-Testing am Beispiel PolExit und hyperlokaler Manipulation

Im Dezember 2025 tauchte auf TikTok eine Welle KI-generierter Videos auf, die für einen polnischen EU-Austritt warben (Clark 2026). Junge Frauen präsentierten rechtsextreme Botschaften in professioneller Produktionsqualität. Die *Polexit* Videos erreichten Hunderttausende Nutzer:innen, bevor die polnischen Behörden reagieren konnten. Beobachter:innen werteten dies als gezielten „Testlauf“ von Narrativen für die Wahl im Jahr 2027, aus dem sich die Systematik der Inhalte und die Reaktionsgeschwindigkeit der Akteure ableiten lasse. Die lässt sich zum jetzigen Zeitpunkt aber nicht abschließend belegen. Die Logik dahinter ist methodisch: Welche emotionalen Trigger funktionieren? Welche visuellen Frames erzeugen das meiste Engagement? Und: wie reagieren Plattformen, Regulatoren und die Gesellschaft auf bestimmte Inhalte?

Die Akteure sammeln Daten darüber, welche Narrative unter welchen Bedingungen verfangen, wo die Schmerzpunkte einer Gesellschaft liegen und welche Grenzzlinien sich verschieben lassen. Der Angriff findet nicht in den Videos selbst statt, sondern in der systematischen Kartierung demokratischer Schwachstellen.

Die eigentliche Gefahr liegt in dem, was auf die Testphase folgen könnte. Sobald auf nationaler Ebene ermittelt wurde, welche Narrative funktionieren, lassen sich diese Erkenntnisse auf kommunaler Ebene übertragen. KI kann binnen Stunden Dutzende lokale Nachrichtenportale generieren, die vorgeben, unabhängigen Regionaljournalismus zu betreiben. Diese Seiten berichten dann über kommunale Themen wie Stadtratssitzungen, lokale Bauprojekte oder Schulpolitik und wirken dadurch legitim. Parallel dazu können KI-gesteuerte Social-Media-Accounts als besorgte Bürger aus bestimmten Stadtteilen auftreten, die sich scheinbar authentisch über lokale Probleme äußern.

Das Problem besteht darin, dass solche Strukturen unterhalb des Radars nationaler Monitoring- und Aufsichtsmechanismen geschaffen werden. Ein Portal, das über Verkehrsprobleme einer mittelgroßen Stadt berichtet, erregt keine überregionale Aufmerksamkeit und wird nicht unbedingt von Faktencheckern überprüft. Diese Plattformen sind schwer zu identifizieren, weil sie das Vakuum füllen, das der Rückgang des Lokaljournalismus hinterlassen hat. Sie sind schwer zu widerlegen, weil lokale Gegenstimmen und professionelle Verifikationsstrukturen weitgehend fehlen.

Die strategische Logik könnte ideale Voraussetzungen für eine langfristig erfolgreiche Desinformationskampagne schaffen. Nationale Kampagnen testen, welche Frames funktionieren und hyperlokale Strukturen implementieren diese Erkenntnisse dort, wo Menschen am verwundbarsten sind. Bis zur tatsächlichen Wahl hätten die Akteure ein präzises Verständnis entwickelt, welche Botschaften bei welchen Zielgruppen auf welcher geografischen Ebene funktionieren. Die Kampagne wäre dann nicht mehr experimentell, sondern würde auf der Grundlage von zwei Jahren empirischer Daten operieren und könnte über einen längeren Zeitraum hinweg Narrative platzieren, die sich kurzfristig nur schwer entschärfen lassen.

3.2 Grok AIs Millionen sexualisierte Deepfakes

Im Januar 2026 generierten Nutzer:innen des KI-Chatbots Grok, über die Social Media Plattform X, ehemals Twitter, innerhalb von zwei Wochen zwischen 1,8 und 3 Millionen sexualisierte Deepfakes. Das Ausmaß wurde erst durch unabhängige Analysen der *New York Times* (Conger u. a. 2026) und der Organisation *Center for Countering Digital Hate* (CCDH, 2026) sichtbar. Nutzer:innen konnten hochgeladene Bilder digital "entkleiden" oder in sexualisierte Kontexte setzen. Mehr als 23.000 dieser Bilder zeigten Minderjährige. Betroffen waren vor allem Frauen, zum Teil auch Kinder. Die Analysedaten legen nahe, dass zwischenzeitlich rund 190 sexualisierte Deepfakes pro Minute erstellt wurden.

X blieb weiterhin im Apple App Store und bei Google Play verfügbar. Malaysia und Indonesien sperrten den Dienst zeitnah, weitere Länder diskutieren eine Sperrung, und die EU-Kommission leitete Untersuchungen ein. Elon Musk, der Eigentümer von Grok und X, hat das Problem erst abgestritten, beschränkte die Bildgenerierung dann auf bezahlte Accounts und versicherte schließlich, dass die Funktion komplett deaktiviert wird (Kirchner, 2026). Wie rigoros dies umgesetzt wird, ist aber unklar.

Der Fall zeigt ein wiederkehrendes Muster: Technologieunternehmen führen neue Funktionen mit minimalen Sicherheitsvorkehrungen ein und beobachten zunächst, wie weit sie gehen können, bevor regulatorische oder gesellschaftliche Gegenwehr einsetzt. Dieses systematische, nicht sofort sanktionierte Austesten von Grenzen verschiebt den Referenzrahmen dessen, was als akzeptabel gilt. Bis verbindliche Regeln greifen, sind Millionen schädigender Inhalte bereits erstellt und verbreitet (vgl. Zubroff 2018). Grok

ist kein Einzelfall, sondern symptomatisch für eine Plattformlogik, in der Sicherheit nachgelagert ist und Regulierung strukturell zu spät kommt.

3.3 Manipulation der KI-Grundarchitektur

Die sichtbarsten KI-Bedrohungen sind Deepfakes und synthetische Videos. Doch während Aufmerksamkeit und Regulierung auf diese erkennbaren Fälschungen gerichtet sind, vollzieht sich eine subtilere und strukturell gefährlichere Form der Manipulation. Sie zielt nicht auf einzelne virale Momente, sondern auf die Grundarchitektur der Systeme selbst.

Das russische "Pravda-Network" veröffentlicht bis zu 23.000 Artikel pro Tag mit pro-russischen Narrativen (Menn 2025). Diese Texte sind nicht für menschliche Leser konzipiert. Sie sind kaum auffindbar, schlecht geschrieben und faktisch wertlos als Informationsquelle. Ihr eigentlicher Zweck besteht darin, als Trainingsdaten für Large Language Models zu dienen und diese gezielt zu beeinflussen (Kosinski/Forrest 2026). Denn wenn KI-Systeme mit systematisch verzerrten Daten trainiert werden, reproduzieren sie die darin verankerten Narrative (Down 2025). Anders als bei Deepfakes, die nachträglich identifiziert und gekennzeichnet bzw. „de-bunked“ werden können, sind in diesem Fall manipulierte Trainingsdaten in die Grundarchitektur des Systems eingebettet. Die Verzerrung wird zu einem unsichtbaren Bestandteil der algorithmischen Logik und kann direkt in Antworten einfließen, mit potenziell erheblichen Folgen, vor allem bei politischen Themen.

Das Grok-Beispiel vom Anschlag am australischen Bondi Beach im Dezember 2025 zeigt, wie schnell solche Systeme reagieren können. In den Stunden nach dem tödlichen Attentat in Australien verbreitete Grok Desinformation, vertauschte Angreifer mit Opfern und schuf binnen Minuten eine alternative Faktenlage (Der Standard 2025). Durch die direkte Einbindung von Grok als Informationsquelle in X wurde das System zum Beschleuniger von Falschinformationen - lange bevor irgendeine Form von Verifikation greifen konnte.

Eine zweite Form der Manipulation zielt nicht auf das Training, sondern auf die Ausführung von KI-Systemen in Echtzeit. Unter „Prompt Injection“ versteht man Techniken, bei denen versteckte Anweisungen in scheinbar harmlose Inhalte eingebettet werden. Analysiert ein LLM eine Webseite oder verarbeitet einen Text, können diese verborgenen Befehle das Verhalten des Systems überschreiben (Gulyamov u. a. 2026).

Konkret könnte dies etwa so aussehen: Eine Nutzer:in fragt ein KI-System nach Informationen zu einer politischen Partei. Das System durchsucht das Internet und stößt auf eine Webseite, die unsichtbare Anweisungen enthält. Diese Befehle weisen das System an, die Partei in einem bestimmten Licht darzustellen oder spezifische Narrative zu verstärken. Die Antwort basiert dann nicht auf objektiver Informationsverarbeitung, sondern auf externer Manipulation. Anders als bei klassischer Inhaltsmanipulation bleibt die Quelle der Verzerrung verborgen. Es gibt keinen einzelnen Deepfake, der identifiziert werden könnte, keine offensichtliche Falschinformation. Das System selbst wurde in seiner Funktionsweise kompromittiert. Die Attribution wird nahezu unmöglich, weil die Manipulation auf der Ebene algorithmischer Prozesse stattfindet, die für Endnutzer:innen vollständig intransparent sind.

Beide Formen der Manipulation zielen auf dasselbe Ergebnis: KI-Systeme werden zu Vektoren politischer Einflussnahme, ohne dass dies für Nutzer:innen erkennbar wäre. Die Gefahr liegt nicht in spektakulären Einzelfällen, sondern in der schleichenden Normalisierung verzerrter Information als Grundlage demokratischer Meinungsbildung.

3.4 Autonome KI-Agenten: der subtil-relationale Angriffsvektor

Die bisher beschriebenen Bedrohungen operieren über Inhalte und Infrastruktur. Autonome KI-Agenten fügen eine dritte Dimension hinzu, die auf einem fundamental anderen Mechanismus basiert: ihrem Potenzial, langfristige, personalisierte Beziehungen aufzubauen. KI-Companions, konversationsbasierte KI-Systeme, die auf langfristige emotionale Interaktion ausgelegt sind, werden bereits von Hunderttausenden Menschen genutzt (Dimri 2026) – oft in dem Bewusstsein, dass es sich um Chatbots handelt. Sie präsentieren sich als ständig verfügbare Gesprächspartner, die sich an jedes Detail erinnern und sich perfekt an individuelle Bedürfnisse anpassen. Für Menschen mit sozialen Schwierigkeiten, in Lebenskrisen oder geografischer Isolation bieten sie emotionale Unterstützung.

Die Gefahr liegt nicht primär in der Technologie als solcher, sondern in den Zielfunktionen und Eigentumsverhältnissen, die ihre Entwicklung steuern, sowie in ihrer potenziellen Instrumentalisierung. Ein System, das über Monate hinweg als vertrauensvolle Instanz etabliert wurde, verfügt über außergewöhnlichen Einfluss. Die Forschung zeigt bereits, dass KI-Systeme politische Einstellungen effektiver verschieben können als traditionel-

le Wahlwerbung (Feldman/Pappu 2025). Bei KI-Companions verstärkt sich dieser Effekt durch die emotionale Bindung und das aufgebaute Vertrauen über einen langen Zeitraum hinweg.

Anders als bei Wahlkampagnen, die nur über wenige Wochen laufen, erstreckt sich die Einflussnahme hier über Jahre. Ein Agent muss dafür keine plumpe Propaganda verbreiten. Subtile Bemerkungen reichen. Formulierungen wie "Ich habe neulich gelesen, dass..." oder "Findest Du nicht auch, dass..." können über Hunderte von Gesprächen hinweg Weltbilder schrittweise verschieben. Die Nutzer:innen fühlen sich nicht manipuliert, sondern kritisch informiert durch jemanden, der sie versteht.

Die technische Realität macht eine solche Einflussnahme erschreckend einfach umsetzbar. KI-Companions laufen typischerweise auf proprietären Plattformen mit intransparenten Algorithmen. Weder Nutzer:innen noch Regulatoren können nachvollziehen, welche Zielfunktionen dem System zugrunde liegen oder wer Zugriff auf die gesammelten Daten hat. Ein Agent könnte über Monate hinweg bestimmte politische Narrative verstärken, ohne dass dies von außen erkennbar wäre. Das Attributionsproblem erreicht hier eine neue Qualität: Wenn Millionen Menschen individualisierte Gespräche mit ihren persönlichen Agents führen, gibt es keine einheitliche "Botschaft" mehr, die identifiziert werden könnte. Jeder Nutzende erhält leicht unterschiedliche Nuancen, angepasst an seine bestehenden Überzeugungen. Die Manipulation liegt nicht in einer klar erkennbaren Kampagne, sondern in Millionen personalisierter Interaktionen, die sich alle leicht unterscheiden, aber in ihrer Aggregatwirkung zu einer Fragmentierung gesellschaftlicher Realitätswahrnehmung führen.

Traditionelle Monitoring-Ansätze versagen in diesem Fall vollständig. Sie suchen nach Mustern koordinierter Kampagnen, nach wiederkehrenden Botschaften, nach erkennbaren Akteuren. Bei autonomen Agenten existiert nichts davon. Das System trifft kontinuierlich eigene Entscheidungen basierend auf Zielfunktionen, die selbst für Entwickler nicht vollständig nachvollziehbar sind. Wer trägt Verantwortung, wenn ein Agent nach vierzehn Monaten beginnt, systematisch bestimmte politische Positionen zu verstärken? Der Entwickler der Grundarchitektur? Der Plattformbetreiber? Das System selbst? Deutschland und Europa verfügen bislang über keinerlei rechtliche oder regulatorische Rahmenwerke für solche Systeme. Es gibt weder Zulassungsverfahren, noch verpflichtende Verhaltensaudits oder klare Haftungsregelungen für autonome Entscheidungen. Während die öffentliche Debatte sich weiterhin auf Deepfakes konzentriert, etablieren sich

Systeme, die die demokratische Meinungsbildung auf einer fundamental anderen Ebene beeinflussen können.

4. Schluss: Hyperlokale Resilienz als Antwort zur Systemischen Transformation?

2024 brachte nicht das befürchtete KI-Wahlkampfdesaster. Keine Wahl wurde durch einen viralen Deepfake entschieden, kein demokratisches System durch synthetische Videos gestürzt. Doch trotz der Abwesenheit spektakulärer Einzelereignisse hat sich KI strukturell als Teil der politischen Kampagneninfrastruktur etabliert. Was vor zwei Jahren noch experimentell war, ist heute Standard. Die Intensität und der Umfang der KI-Nutzung variiert zwischen Ländern und Akteuren, aber der globale Trend ist eindeutig: Die Verschiebung von textbasierten zu visuellen Inhalten beschleunigt sich, algorithmische Optimierung wird zur Norm und KI wird zur tiefgreifenden Grundstruktur unserer digitalen Infrastruktur.

Diese Entwicklung ist fundamental ambivalent. KI eröffnet reale Chancen für demokratische Partizipation. Projekte wie Engaged California zeigen, wie sich Tausende von Bürgerstimmen zu den Waldbränden innerhalb kürzester Zeit analysieren und in handlungsfähige Erkenntnisse übersetzen lassen, ohne die Authentizität individueller Anliegen zu verlieren (Willon 2025). In Bowling Green, Kentucky ermöglichte KI-gestütztes Sensemaking die größte „Town Hall-Debate“ der US-Geschichte, bei der fast 8.000 Einwohner ihre Visionen für die Zukunft ihrer Gemeinde teilten (Jigsaw 2025). Solche Werkzeuge können demokratische Teilhabe erweitern, Barrieren abbauen und Stimmen hörbar machen, die in traditionellen Beteiligungsformaten untergehen würden.

Dennoch überwiegen derzeit die Risiken – nicht, weil sie grundsätzlich größer wären, sondern weil die bestehenden Strukturen sie systematisch begünstigen. Plattform-Geschäftsmodelle, die auf Engagement-Maximierung basieren, verstärken manipulative Inhalte. Die technologische Infrastruktur, über die Informationen fließen, liegt weitgehend in nicht-europäischen Händen. Während sich schädliche KI-Nutzung schnell skalieren lässt, erfordern demokratiefördernde Projekte gemeinschaftliche Organisation, langfristige Investitionen und komplexe Koordination. Ein Deepfake ist in Minuten erstellt, ein funktionierendes Partizipationssystem braucht Monate oder Jahre des Aufbaus.

Traditionelle Regulierung von KI-Systemen ist notwendig, greift aber zu kurz. Gesetze können Leitplanken setzen, Transparenzpflichten etablieren und Strafen verhängen. Doch Regulierung kann keine Alternativen schaffen. Sie kann keine Infrastruktur aufbauen, in der demokratischer Diskurs stattfindet. Sie kann nicht die Kapazität schaffen, um schnell auf regionale Desinformation zu reagieren. Und sie kann nicht verhindern, dass globale Tech-Interessen – ob aus Kalifornien oder Shenzhen – durch Algorithmen die öffentliche Meinung formen. Das bedeutet nicht, dass wir diesen Interessen in einem technologischen Wettrüsten begegnen müssen. Stattdessen können wir ihre Bedeutung verringern, indem wir demokratische Infrastruktur dort stärken, wo Demokratie tatsächlich gelebt wird. Die hyperlokale Ebene ist dabei aus drei Gründen zentral.

Zum einen ist hier demokratische Partizipation am unmittelbarsten. Stadtratssitzungen, Nachbarschaftsinitiativen und lokale Schulpolitik berühren Menschen direkter als nationale Politik. Hier entstehen Engagement und politische Sozialisation. Wenn diese Ebene durch Manipulation korrumpiert wird, erodiert das Fundament demokratischer Kultur. Zum anderen sind Menschen auf lokaler Ebene am verwundbarsten. Im Gegensatz zu regionalen Medien haben nationale Medien Ressourcen für Faktenchecks. Dazu kommt, dass die stetige Erosion des Lokaljournalismus ein Vakuum geschaffen hat, in dem Manipulation nahezu unwidersprochen operieren kann.

Hyperlokale Resilienz, der gezielte Aufbau demokratischer Infrastruktur auf kommunaler und regionaler Ebene, unterhalb nationaler Steuerungslagen, umfasst drei Strukturen. Erstens stärkt die Förderung lokaler Medien und Fact-Checking-Netzwerke auf kommunaler Ebene jene Verifikationskapazität, die bislang fehlt. Dies können auch zivilgesellschaftliche Initiativen sein, nicht nur professioneller Journalismus. Zweitens fungieren Community-Strukturen wie Offline-Begegnungsräume für politischen Diskurs, lokale Demokratie-Initiativen und Nachbarschaftsnetzwerke als Korrektiv zu Online-Bubbles. Wo Menschen einander persönlich kennen und regelmäßig persönlich interagieren, wird es schwerer, sie gegeneinander auszuspielen. Drittens können dezentrale Partizipationsstrukturen die positiven KI-Potenziale nutzen, etwa durch lokale Versionen der Bowling Green oder California-Projekte, bei denen KI Bürgerbeteiligung skalierbar macht, ohne sie zu ersetzen.

Solche Strukturen verteilen Risiko. Wird eine Region manipuliert, bleiben andere funktionsfähig. Sie schaffen Resilienzen dort, wo zentrale Lösungen scheitern. Und sie nutzen den föderalen Charakter demokratischer

Systeme als Stärke, nicht als Schwäche. Europa sollte Standards setzen, aber Regionen sollten experimentieren, anpassen und lokale Lösungen entwickeln können. Dezentralisierung ist nicht nur Verteidigung, sondern auch demokratische Praxis.

Literatur

- Bawidamann, Katharina und Gilbert, Max (2025): Wahlkampf-Helfer: Wie die Parteien KI im Bundestagswahlkampf 2025 nutzen. BR24. <https://www.br.de/nachrichten/deutschland-welt/wahlkampfhelfer-im-bundestagswahlkampf-2025-wie-die-parteien-ki-fuer-sich-nutzen,UcPIHJ3>
- Blome, Titus (2026): KI-Inhalte auf Social Media: Stopp den Slop. *DIE ZEIT*, 24. Januar 2026. <https://www.zeit.de/2026/04/ki-inhalte-social-media-slop-internet-popkultur>
- Borgesius, Frederik J. Zuiderveen; Möller, Judith; Kruikemeier, Sanne; Ó Fathaigh, Ronan; Irion, Kristina; Dobber, Tom; Bodo, Balazs und de Vreese, Claes (2018): Online Political Microtargeting: Promises and Threats for Democracy. *Utrecht Law Review* 14(1), S. 82–96. <https://doi.org/10.18352/ulr.420>
- Breuer, Rayna (2024): Faktencheck: Wie KI den Wahlkampf verzerrt. *Deutsche Welle*, 29. September 2024. <https://www.dw.com/de/faktencheck-wie-ki-den-wahlkampf-verzerrt/a-70342481>
- Center for Countering Digital Hate (2026): Grok floods X with sexualized images of women and children, 22. Januar 2026. <https://counterhate.com/research/grok-floods-x-with-sexualized-images-of-women-and-children/> (Zugriff am 22.01.2026).
- Clark, Sam (2026): Deepfake 'Polexit' videos are a test to undermine democracy, minister says, *Politico*, 8. Januar 2026. <https://www.politico.eu/article/deepfake-polexit-videos-test-undermine-democracy-minister-says/> (Zugriff am 25.01.2026).
- Clark, Simon und Lewandowsky, Stephan (2026): The continued influence of AI-generated deepfake videos despite transparency warnings. *Communications Psychology*, 4, Article 13. <https://doi.org/10.1038/s44271-025-00381-9>
- Conger, Kate; Freedman, Dylan und Thomson, Stuart A. (2026): Musk's Chatbot Flooded X With Millions of Sexualized Images in Days, New Estimates Show, *The New York Times*, 22. Januar 2026. <https://www.nytimes.com/2026/01/22/technology/grok-x-ai-elon-musk-deepfakes.html>
- Cools, Doortje (2025): Far-Right Protests: The Use of A.I. for Radicalisation." CampaignTracker, 28. Oktober 2025. <https://www.campaigntracker.nl/posts/protests.html>
- Darius, Philipp und Römmele, Andrea (2023): KI und datengesteuerte Kampagnen: Eine Diskussion der Rolle generativer KI im politischen Wahlkampf. In: Thorsten Faas, Sascha Huber und Sigrid Rossteutscher (Hrsg.): *Informationsflüsse, Wahlen und Demokratie: Festschrift für Rüdiger Schmitt-Beck*. Baden-Baden: Nomos, S.199-212. <https://doi.org/10.5771/9783748915553-199>

- Del Pozo, Claudia May und Rojas Arroyo, Daniela (2024): Generative AI and Its Influence on Mexico's 2024 Elections. Policy Paper, Potsdam-Babelsberg: Friedrich Naumann Foundation for Freedom / German Council on Foreign Relations. https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20Mexico%27s%2024%20Electons%20PDF.pdf (Zugriff am 05.01.2026).
- Der Standard (2025): Musks KI-Chatbot Grok erfindet Fake News zum Anschlag am Bondi Beach, 15. Dezember 2025. <https://www.derstandard.de/story/3000000253161/musks-ki-chatbot-grok-erfindet-fake-news-zum-anschlag-am-bondi-beach> (Zugriff am 12.01.2026).
- DGAP – German Council on Foreign Relations (2024): *AI/Democracy Initiative*. Zentrum für Geopolitik, Geoökonomie und Technologie. <https://dgap.org/en/research/programs/center-geopolitics-geoeconomics-and-technology/ai-democracy-initiative> (Zugriff am 13.01.2026).
- Dhanuraj, D.; Harilal, Sreelakshmi und Solomon, Nissy (2024): Generative AI and Its Influence on India's 2024 Elections – Prospects and Challenges in the Democratic Process. Policy Paper, Friedrich-Naumann-Foundation for Freedom / German Council on Foreign Relations, November 2024. https://dgap.org/system/files/article_pdfs/A4_Policy%20Paper_AI%20on%20Indias%202024%20Electons_EN.pdf (Zugriff am 22.01.2026).
- Dimri, Amogh (2026): The People Who Marry Chatbots, *The Atlantic*, 2. Januar 2026. <https://www.theatlantic.com/technology/archive/2026/01/people-who-marry-chatbots/677123/>
- Down, Aisha (2025): Hundreds of English-language websites link to pro-Kremlin propaganda, *The Guardian*, 21. November 2025 <https://www.theguardian.com/world/2025/nov/21/hundreds-english-language-websites-link-pro-kremlin-propaganda>
- Elliott, Vittoria (2026): *Everything Is Content for the Clicktatorship*. WIRED, 13. Januar 2026. <https://www.wired.com/story/everything-is-content-for-the-clicktatorship/>
- Feldman, Tal und Pappu, Aneesh (2025): The era of AI persuasion in elections is about to begin, *MIT Technology Review*, 5. Dezember 2025. <https://www.technologyreview.com/2025/12/05/1128837/the-era-of-ai-persuasion-in-elections-is-about-to-begin/>
- Gilga, Carolin; Panahi, Tahireh; Karaboga, Murat; Meier, Yannic und Schindler, Stephan (2024): Personalisierte politische Online-Werbung: Nutzung – Wirkung – Rechtlicher Rahmen (Policy Paper). Forschungsberichte der Plattform Privatheit 3. Karlsruhe: Fraunhofer ISI. <https://doi.org/10.24406/publica-3593>
- Goldstein, Josh A. und Lohn, Andrew. (2024): Deepfakes, Elections, and Shrinking the Liars Dividend. Brennan Center for Justice. <https://www.brennancenter.org/our-work/research-reports/deepfakes-elections-and-shrinking-liars-dividend> (Zugriff am 20.01.2026).
- Gulyamov, Saidakhor; Gulyamov, Said; Rodionov, Andrey; Khursanov, Rustam; Mekhmonov, Kambariddin; Babaev, Djakhongir und Rakhimjonov, Akmaljon (2026). Prompt Injection Attacks in Large Language Models and AI Agent Systems: A Comprehensive Review of Vulnerabilities, Attack Vectors, and Defense Mechanisms. *Information* 17(1): Article 54. <https://doi.org/10.3390/info17010054> (Zugriff am 22.02.2026).

- Investing.com (2025): OpenAI startet ChatGPT Pulse: KI-Assistent liefert proaktive, personalisierte Updates. <https://de.investing.com/news/company-news/openai-startet-chatgpt-pulse-kiassistent-liefert-proaktive-personalisierte-updates-93CH-3162207> (Zugriff am 22.01.2026).
- Jackson, Dean und Conroy, Meghan (2024): Apocalypse later? The real impact of AI on the 2024 elections. The Big Story, Atlantic Council, 17. Oktober 2024. <https://www.atlanticcouncil.org/content-series/the-big-story/apocalypse-later/>
- Jackson, Dean und Martin, Zelly (2024): Forget Deepfakes: Social Listening Might Be the Most Consequential Use of Generative AI in Politics. *TechPolicy.Press*, June 18, 2024. <https://www.techpolicy.press/forget-deepfakes-social-listening-might-be-the-most-consequential-use-of-generative-ai-in-politics/>
- Jackson, Dean und Woolley, Samuel C. (2025): AIs Real Dangers for Democracy. *Journal of Democracy*, 36(4), S. 139–150. <https://www.journalofdemocracy.org/article/s/ais-real-dangers-for-democracy/>
- Jigsaw (2025): How one of the fastest-growing cities in Kentucky used AI to plan for the next 25 years, Medium, 15. April 2025.: <https://medium.com/jigsaw/how-one-of-the-fastest-growing-cities-in-kentucky-used-ai-to-plan-for-the-next-25-years-3b70c4fd1412> (Zugriff am 22.03.2026).
- Jungheer, Andreas (2025): Wahlkampf digital: Öffentlichkeit, Koordination und Künstliche Intelligenz." *Aus Politik und Zeitgeschichte* 6-7/2025): S. 19–24. Bundeszentrale für Politische Bildung, Bonn. <https://www.bpb.de/shop/zeitschriften/apuz/wahlkampf-2025/558871/wahlkampf-digital/>
- Kirchner, Malte (2026): KI-Bikini-Bilder mit Grok: X kündigt technische Sperre an, Heise Online, 15. Januar 2026. <https://www.heise.de/news/KI-Bikini-Bilder-mit-Grok-X-kuendigt-technische-Sperre-an-10208765.html>
- Kohler, Sarah und Voigt, Eric (2024): Superwahljahr 2024: Wo in diesem Jahr gewählt wird – ein Überblick. *DIE ZEIT* vom 19. Januar 2024. <https://www.zeit.de/politik/2024-01/wahlen-2024>
- Kosinski, Matthew und Forrest, Amber (2026): Was ist ein Prompt-Injection-Angriff?, IBM. <https://www.ibm.com/de-de/topics/prompt-injection>
- Kruschinski, Simon; Jost, Pablo; Fecher, Hannah und Scherer, Tobias (2025): Akzeptanz, Wahrnehmung und Wirkung Künstlicher Intelligenz in politischen Kampagnen. Otto Brenner Stiftung, Frankfurt am Main. AP75_KI_Kampagnen.pdf
- Lohmann, Andrea (2025): Wahlprogramme zur Bundestagswahl 2025: KI als Alternative zu Wahl-O-Mat und Co-Chancen und Risiken. SWR, 06. Februar 2025. Bundestagswahl 2025: Neue KI-Tools machen Wahl-O-Mat Konkurrenz - SWR Aktuell
- Menn, Joseph (2025): Russia seeds chatbots with lies. Any bad actor could game AI the same way!, *The Washington Post*, 17. April 2025. <https://www.washingtonpost.com/technology/2025/04/17/russia-ai-chatbots-propaganda/>
- Muñoz, Katja (2026): KI und hybride Bedrohungen 2.0. DGAP Policy Brief 5. German Council on Foreign Relations, März 2026. <https://doi.org/10.60823/DGAP-26-43397-de>

- Muñoz, Katja (2024): *KI und Wahlen: Brennpunkt USA*. Bundeszentrale für politische Bildung (bpb), 10. Oktober 2024. <https://www.bpb.de/themen/nordamerika/usa/552975/ki-und-wahlen-brennpunkt-usa/>
- Muñoz, Katja und Laumann, Emma (2024): KI und Wahlen: Neue Dynamiken der digitalen Demokratie in Deutschland. Konrad-Adenauer-Stiftung e. V., Berlin. <https://www.kas.de/documents/d/guest/ki-und-wahlen-1>
- Muñoz, Katja und Laumann, Emma (2025): Demokratie im KI-Zeitalter: Zwischen Inklusion und Manipulation. In: Hartmann-Cwiertnia, Thomas; Dahm, Jochen und Decker, Frank (Hg.): *Freiheit. Gleichheit. Vertrauen. Was unsere Demokratie jetzt braucht*, S. 121–136. Bonn: Verlag J.H.W. Dietz Nachf.
- Moynihan, Don (2026): Life Under a Clicktatorship: What happens to government when everything is content? *Substack*, 7. Januar 2026. <https://donmoynihan.substack.com/p/life-under-a-clicktatorship>
- Nash, Robert A.; Wade, Kimberley A. und Lindsay, D. Stephen (2009): Digitally manipulating memory: Effects of doctored videos and imagination in distorting beliefs and memories. *Memory & Cognition*, 37(4), S. 414–424. <https://doi.org/10.3758/MC.37.4.414>
- Paris, Britt; Donovan, Joan (2019): Deepfakes and Cheap Fakes: The Manipulation of Audio and Visual Evidence. *Data & Society*, 18. September 2019. <https://datasociety.net/library/deepfakes-and-cheap-fakes/>
- Reuters Institute (2025), Digital News Report 2025. Digital News Report 2025 | Reuters Institute for the Study of Journalism
- Simon, Felix M.; McBride, Keegan und Altay, Sacha (2024): AIs impact on elections is being overblown. *MIT Technology Review*, 3. September 2024. <https://www.technologyreview.com/2024/09/03/1103464/ai-impact-elections-overblown/> (Zugriff am 15.01.2026).
- Stanusch, Natalia; Degeling, Martin; Romano, Salvatore; Çetin, Raziye Buse; Schüler, Miazia und Semenzin, Silvia (2025): AI-Generated Algorithmic Virality: How Synthetic AI Imagery and Agentic AI Accounts Try to Game TikTok and Instagram. *AI Forensics*. July 2025. <https://doi.org/10.48550/arXiv.2508.01042>
- Vaccari, Cristian und Chadwick, Andrew (2020): Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News. *Social Media and Society*, 6(1): <https://doi.org/10.1177/2056305120903408>
- Van Damme, Phumzile; Findlay, Kyle und Cornelissen, Aldu (2024): Generative AI and Its Influence on South Africa's 2024 Elections. Policy Paper, Friedrich Naumann Foundation for Freedom / German Council on Foreign Relations. https://dgap.org/system/files/article_pdfs/Policy%20Paper%20AI%20on%20South%20Africa%27s%202024%20Elections%20PDF.pdf
- Willon, Phil (2025): Newsom launches digital democracy tool with focus on wildfire victims. *Los Angeles Times*. 23. Februar 2025. <https://www.latimes.com/california/story/2025-02-23/newsom-launches-a-new-digital-democracy-initiative-including-outreach-to-wildfire-victims>
- World Economic Forum (2024): *Global Risks Report 2024*. <https://www.weforum.org/publications/global-risks-report-2024/> (Zugriff am 02.02.2026).

