

Skeptisches Sehen

KI-generierte Bilder fordern unser Verhältnis zur Wahrheit heraus

Mittels Künstlicher Intelligenz erzeugte Fotos und Videos werden zunehmend als Bedrohung für Individuen und demokratische Systeme wahrgenommen. Sie nähren den Verdacht, dass echt aussehendes Bildmaterial gefälscht sein könnte – oder aber für falsch gehaltenes echt. Trotzdem führen sie nicht unvermeidlich in eine postfaktische Gesellschaft, in der Wahrheit und Lüge ununterscheidbar werden.

Florian Irgmaier und Florian Eyert

Wie der Technikphilosoph Andreas Kaminski gezeigt hat, muss eine Technologie noch nicht flächendeckend zum Einsatz kommen, um gesellschaftlich wirksam zu werden. Bereits das Wissen um ihre Funktionsweise und die Erwartung ihrer Effekte können politisch, rechtlich, wirtschaftlich oder anderweitig einen Unterschied machen. Beobachten kann man das aktuell an Verfahren zur Bearbeitung und Erzeugung von Bildmaterial, die auf Künstlicher Intelligenz beruhen.

Jüngere Technologien in diesem Bereich machen es möglich, Fotos und Videos herzustellen, die auf den ersten Blick wie wirklichkeitsgetreue Aufnahmen erscheinen – von „Deepfakes“ ist dann oft die Rede, die wissenschaftliche Literatur spricht, weniger wertend, von „synthetischen Medien“. Mit der zunehmenden Verbreitung dieser Verfahren gewinnt die Erwartung an Plausibilität, dass gefälschtes Bildmaterial als echt und echtes als gefälscht ausgegeben werden kann. Dies verunsichert, wie auch die mediale und wissenschaftliche Diskussion zeigt, das gesellschaftliche Vertrauen in Fotos und Videos als Mittel der neutralen Abbildung der Wirklichkeit. Im Folgenden beleuchten wir aus einer wissenssoziologischen Perspektive, woher

diese Verunsicherung rührt und wie Gesellschaften darauf reagieren.

Der Erfolg KI-gestützter synthetischer Medien beruht auf sogenannten künstlichen neuronalen Netzwerken, die mit großen Mengen von Bild- oder Videomaterial versorgt werden, an die sich die Parameter des Netzwerks anpassen. Sie erlauben es, komplexe Regelmäßigkeiten in diesen Daten zu identifizieren und so beispielsweise Kerneigenschaften dessen zu isolieren, was ein menschliches Gesicht ausmacht. Auf dieser Basis können dann Netzwerkarchitekturen konstruiert werden, die nicht nur Gesichter erkennen, sondern auch neue generieren. Autoencoder etwa reduzieren ein Bild auf seine Kernmerkmale und rekonstruieren aus diesen mit einem Decoder wieder das ursprüngliche Bild. Codiert man das Gesicht einer Person und verwendet bei der Rekonstruktion den Decoder einer anderen Person, ergibt sich ein sogenannter „Face Swap“. Eine weitere Architektur sind Generative Adversarial Networks (GAN), bei denen ein erstes Netzwerk Bilder generiert, die ein zweites Netzwerk dann als echt oder falsch einzustufen versucht. Durch mehrfaches Wiederholen dieses Prozesses „lernt“ das erste Netzwerk, immer echter wirkende Bilder oder Videos her-

zustellen – so etwa die Gesichter auf der Website thispersondoesnotexist.com. An Popularität gewinnen derzeit Diffusionsmodelle, die zunächst lernen, zufälliges Rauschen aus Bildern zu entfernen, und dann genutzt werden können, um auch aus künstlich generiertem Rauschen realistische Bilder zu erzeugen. Anwendungen wie Midjourney, Stable Diffusion oder DALL•E 2 sind zudem in der Lage, eine Sequenz von beschreibenden Worten in ein Bild zu verwandeln, also gewissermaßen die Strategie des automatisierten Erstellens einer Bildbeschreibung umzudrehen.

Auf die so generierten Bilder und Videos reagieren einzelne Gesellschaftsbereiche unterschiedlich. In der Kunst beispielsweise entstehen neue Ausdrucksmöglichkeiten, aber auch neue Konflikte um Urhebererschaft und Vorstellungen von künstlerischem Schaffen. Im Bildungssektor ergeben sich neuartige Lehrformate, etwa durch künstlich generierte Videos historischer Persönlichkeiten. Wir interessieren uns hier jedoch für die Bedeutung synthetischer Medien für die Felder, in denen Wahrheitsfragen verhandelt werden, und dabei vor allem für die politische Öffentlichkeit.

Traditionell fungieren Fotos und Videos für die Öffentlichkeit als Mittel zur Herstellung von Konsens über Tatsachen. Mit ihnen kann man die Behauptung, dass etwas geschehen ist, gegenüber jenen stützen, die nicht dabei waren. Erst hinreichendes Einvernehmen über das, was der Fall ist, ermöglicht koordiniertes Han-

„Erst hinreichendes Einvernehmen über das, was der Fall ist, ermöglicht koordiniertes Handeln“

deln: Dass Menschen gemeinsam auf Kriegsverbrechen, Polizeigewalt oder Umweltkatastrophen reagieren können, die sie nicht mit eigenen Augen beobachtet haben, beruht nicht unwesentlich auf der bildlichen Dokumentation dieser Ereignisse. Wer die Wahrhaftigkeit der Bilder abstreitet, muss überzeugende Gründe vorbringen können oder riskiert, aus dem Kreis der ernst zu nehmenden Diskursteilnehmer*innen ausgeschlossen zu werden.

Der Vertrauensvorschuss für das statische oder bewegte fotografische Bild stützt sich auf die



Florian Irgmaier ist wissenschaftlicher Mitarbeiter in den Forschungsgruppen Politik der Digitalisierung am WZB und Technik, Macht und Herrschaft am Weizenbaum-Institut für die vernetzte Gesellschaft. In seiner Dissertation untersucht er den Einfluss von Wissenschaft und Technik auf Vorstellungen des Möglichen und Machbaren und deren normative Implikationen.

florian.irgmaier@wzb.eu

Foto: © WZB/David Ausserhofer, alle Rechte vorbehalten.

wahrgenommene Passivität des Mediums, die eine unverfälschte Wiedergabe der Wirklichkeit verspricht. Auch ohne ausgeprägtes technisches Verständnis weiß man, dass Fotos und Videos das Resultat eines Prozesses sind, bei dem sich ein optisches Geschehen einem hochempfindlichen Medium einprägt – sei es ein chemisch präparierter Film, sei es, vermittelt über einen Bildsensor, ein digitaler Datenspeicher. Was das Medium festhält, sobald der Auslöser betätigt wird, ist unabhängig vom Willen der Beteiligten. Wer sich jemals eines unvoreilhaftigen Gesichtsausdrucks auf einem Gruppenbild geschämt hat, weiß das nur zu gut. Deshalb kann man davon ausgehen, dass Fotos und Videos in etwa das zeigen, was man mit eigenen Augen hätte sehen können, wäre man zum entsprechenden Zeitpunkt am selben Ort gewesen. Wie die Wissenschaftshistoriker*innen Lorraine Daston und Peter Galison zeigen, hat gerade diese unterstellte Passivität der Fotografie seit dem 19. Jahrhundert zu ihrem besonderen Stellenwert als Medium der Darstellung wissenschaftlicher Gegenstände verholfen. So konnte das Foto zum Archetyp „mechanischer Objektivität“ werden. Damit meinen Daston und Galison die historisch keineswegs konkurrenzlose Vorstellung, dass sich wahre Beobachtungen am besten durch größtmögliche Ausschaltung der Beobachtenden, ihres Willens und ihrer Vorannahmen erzeugen lassen. Standardisierte, insbesondere maschi-



Florian Eyert ist wissenschaftlicher Mitarbeiter in den Forschungsgruppen Politik der Digitalisierung am WZB und Technik, Macht und Herrschaft am Weizenbaum-Institut für die vernetzte Gesellschaft. In seiner Dissertation untersucht er die Rolle von Computermodellen in politischen Prozessen und die Praktiken ihrer Gestaltung.
florian.eyert@wzb.eu

Foto: © WZB/Thu-Ha Nguyen, alle Rechte vorbehalten.

nelle Verfahren gelten dafür als besonders geeignet. Solange Foto und Video im Sinne dieses Paradigmas plausibel versprechen, das tatsächlich Geschehene unverzerrt wiederzugeben, fungieren sie wissenschaftlich wie außerwissenschaftlich als neutrale Instanzen zur Entscheidung von Wahrheitsfragen, etwa über Geschwindigkeitsübertretungen oder Fouls im Sport.

Da Fotos und Videos ihre Glaubwürdigkeit aus der erwarteten Passivität des Mediums beziehen, lassen sie sich am einfachsten durch den Vorwurf in Zweifel ziehen, es sei verborgene Aktivität im Spiel: Man unterstellt einen Unterschied zwischen dem, was wirklich passiert ist, und dem, was das Foto oder Video zeigt, und führt ihn auf das aktive Eingreifen irgendeiner Instanz, einen intervenierenden Willen, zurück. Dieser Zweifel ist so alt wie die Fotografie selbst. Daston und Galison weisen darauf hin, dass die Manipulierbarkeit von Fotos schon im 19. Jahrhundert weithin bekannt war.

Die Plausibilität eines konkreten Fälschungsverdachts hängt auch von den Erwartungen des Publikums über die Leistungsfähigkeit aktueller Bildbearbeitungstechniken ab. Was die durchschnittlich informierte Bürgerin für technisch machbar hält, wird zum einen durch Massen- und soziale Medien beeinflusst, die über Fälschungen und die Verfahren ihrer Herstellung berichten. Dass sich nicht nur Fotos, sondern

auch Videos immer einfacher nachahmen lassen, konnte die Öffentlichkeit etwa an dem im März 2022 verbreiteten, noch deutlich als Fälschung erkennbaren Video beobachten, in dem ein vermeintlicher ukrainischer Präsident Selenskyj seine Landsleute zur Kapitulation auffordert. Zum anderen prägen eigene Erfahrungen im Umgang mit Technik die Vorstellungen des Machbaren. Bereits das Herumspielen mit Smartphone-Apps, die ein fotografiertes Gesicht auf einen anderen Kopf übertragen, oder mit Websites, die aus Worten Bilder generieren, macht die Leistungsfähigkeit gegenwärtiger Techniken der Bilderzeugung erlebbar.

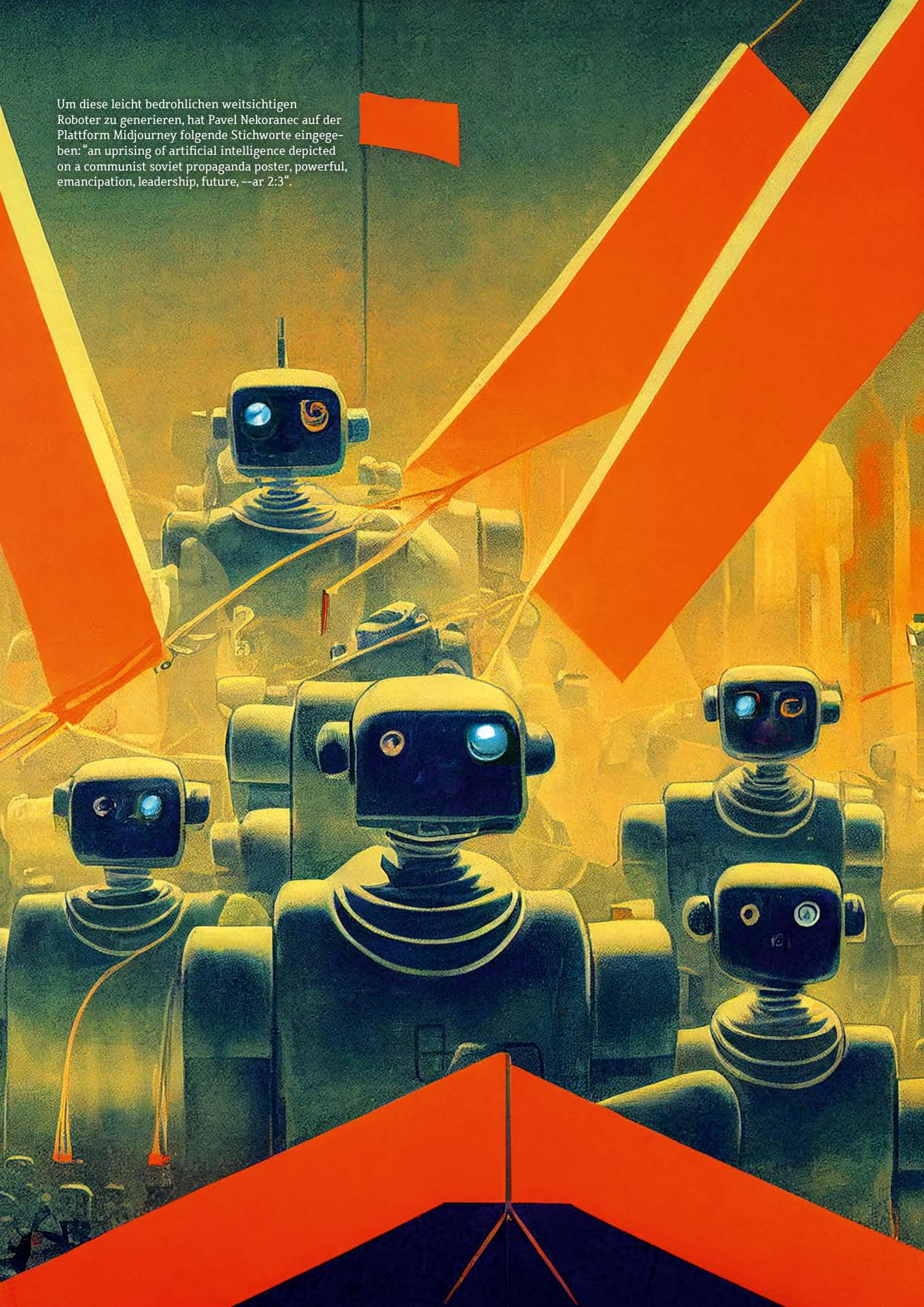
Die Erwartung, dass im Prinzip jedes Foto und jedes Video gefälscht sein könnte, schwächt deren Gewicht als Mittel zur Herstellung von Konsens über Tatsachen. Sie erleichtert es Diskursteilnehmer*innen, eine missliebige, aber fotografisch belegte Tatsachenbehauptung anzuzweifeln, ohne sich damit völlig zu disqualifizieren. Das macht es bei politischen Auseinandersetzungen schwieriger, die Beteiligten auf eine geteilte Faktenbasis zu verpflichten. Ohne wenigstens relative Einigkeit über die Wirklichkeit werden Kompromisse über den Umgang mit ihr jedoch zunehmend unwahrscheinlich. Auch wenn die Zahl glaubwürdiger Deepfakes gering sein mag, können gesellschaftliche Verständigungsprozesse also bereits durch den Verdacht gestört werden, dass, was wie ein Abdruck der Wirklichkeit in einem passiven Medium aussieht, auch das Produkt aktiven Eingreifens sein könnte.

Daher ist es kein Wunder, dass die Fortschritte der bildgenerierenden Verfahren in den vergangenen Jahren vor allem Sorgen ausgelöst haben. Berechtigterweise werden Gefahren für

„Auf die Erwartbarkeit von Deepfakes wird mit der Schulung des eigenen Blicks reagiert“

Individuen und Demokratien befürchtet. Demgegenüber möchten wir den Blick auf die sich herausbildenden Formen des Umgangs mit dieser neuen Situation lenken, die wir unter dem Stichwort des skeptischen Sehens beschreiben wollen. Wird ein Medium, das früher durch seine schiere physische Existenz überzeugte, von der Möglichkeit eingeholt, beliebige Inhalte in

Um diese leicht bedrohlichen weitsichtigen Roboter zu generieren, hat Pavel Nekoranec auf der Plattform Midjourney folgende Stichworte eingegeben: "an uprising of artificial intelligence depicted on a communist soviet propaganda poster, powerful, emancipation, leadership, future, --ar 2:3".



ihm zu formen, kommt es zur Problematisierung bisher unhinterfragten Vertrauens. In der Folge lässt sich auf verschiedenen Ebenen beobachten, wie neue, skeptischere Weisen des Sehens eingeübt werden. Zunächst einmal beginnen viele Menschen, Bildmaterial mit dem prinzipiellen Verdacht der Gefälschtheit zu begegnen, nach erkennbaren Hinweisen auf aktives Eingreifen zu suchen und nach der Plausibilität des Gezeigten zu fragen. Auf die Erwartbarkeit von Deepfakes wird mit der Schulung des eigenen Blicks reagiert. In diesem Lernprozess spielen auch Organisationen eine wichtige Rolle. Das Bundesamt für Sicherheit in der Informationstechnik etwa gibt konkrete Hinweise zur Überprüfung bildlicher Quellen und empfiehlt beispielsweise, nach verwaschenen Konturen oder unplausibler Belichtung Ausschau zu halten. Auch an Schulen wird derartige Wissen zunehmend vermittelt, unterstützt von Einrichtungen wie der Bundeszentrale für politische Bildung. Darüber hinaus arbeiten verschiedene Unternehmen der Digitalwirtschaft an der technischen Einbettung eines solchen skeptischen Sehens in ihre Produkte. Auf den Plattformen der sozialen Medien etwa sollen gefälschte Bilder mithilfe Künstlicher Intelligenz automatisiert erkannt und gegebenenfalls gekennzeichnet oder entfernt werden. Auch in der Medienforensik entwickelt sich die Erkennung von Deepfakes zu einem wichtigen Feld, und zivilgesellschaftliche Organisationen wie Mimikama und Volksverpetzer sowie verschiedene Unternehmen beteiligen sich an der Aufklärung über und der Identifikation von Deepfakes. Schließlich werden gesetzliche Maßnahmen, wie eine Pflicht zur Kennzeichnung KI-generierter Bilder, verhandelt.

Bilder büßen also einen Teil des Vertrauensvorschlusses ein, den das Versprechen mecha-

nischer Objektivität ihnen verschafft hatte. Zugleich bilden sich aber gesellschaftliche Institutionen heraus, die diesen Verlust zu kompensieren versuchen. Bei der Beurteilung von Wahrheit und Falschheit vollzieht sich gewissermaßen eine Gewichtsverschiebung von der Sach- in die Sozialdimension. Wie diese konkret aussehen wird, ist noch unklar, aber sicherlich werden Vertrauensbeziehungen eine wichtige Rolle spielen: Die Glaubwürdigkeit der das Bildmedium präsentierenden oder seine Wahrheit verbürgenden Person oder Organisation rückt in den Fokus. Allgemeinen Konsens über Tatsachen wird freilich auch diese Gewichtsverschiebung nicht stiften können. Denn der Vorwurf strategischer Selektivität – jemand stütze sich nur auf das, was eigenen Interessen dient – lässt sich nicht nur gegen vorgebrachtes Bildmaterial richten, sondern auch gegen Institutionen des skeptischen Sehens. Das Ringen um die Grenze zwischen wahr und falsch wird also weitergehen.

Die Diagnose einer postfaktischen Gesellschaft ist weit verbreitet – greift aber zu kurz. Dagegen lenkt eine Perspektive, die auf die Institutionen des skeptischen Sehens abzielt, den Blick auf die Formenvielfalt und Gestaltbarkeit des Umgangs mit Technik. Diese komplexere Betrachtungsweise berücksichtigt, dass die Ermittlung von Wahrheit schon immer ein gesellschaftlicher Prozess war und durch Erwartungen über Technologien und durch organisationale Umgänge mit ihnen geprägt wurde. Diese unausgesprochenen Voraussetzungen geraten jetzt in den Fokus der gesellschaftlichen Aufmerksamkeit. Wie die darauf reagierenden sozialen Formen künftig aussehen und gesellschaftlich strukturiert werden, wird in den kommenden Jahren zu beobachten – und mitzugestalten – sein. ●

Literatur

Chesney, Bobby/Citron, Danielle: „Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security“. In: California Law Review, 2019, Jg. 107, S. 1753-1819.

Daston, Lorraine/Galison, Peter: Objektivität. Frankfurt a. M.: Suhrkamp 2007.

Heider, Fritz: Ding und Medium. Neuedition mit Vorwort von Dirk Baecker. Berlin: Kadmos 2005 [1926].

Kaminski, Andreas: Technik als Erwartung: Grundzüge einer allgemeinen Technikphilosophie. Bielefeld: transcript 2010.

Pawelec, Maria: „Deepfakes als Chance für die Demokratie?“. In: Alexander Bogner/Michael Decker/Michael Nentwich/Constanze Scherz (Hg.): Digitalisierung und die Zukunft der Demokratie: Beiträge aus der Technikfolgenabschätzung. Baden-Baden: Nomos, 2022, S. 89-101.