

PROJECT IMMERSIVE DEMOCRACY

Part of the European Metaverse Research Network

Ein sicherer Raum für alle – Plädoyer für ein demokratisches und partizipatives Metaverse

Dr. Octavia Madeira, Dr. Georg Plattner | Juli 2023

ÜBER DIE AUTOR:INNEN

Dr. Octavia Madeira, Institut für
Technikfolgenabschätzung und Systemanalyse
(ITAS), Karlsruher Institut für Technologie (KIT)

Dr. Georg Plattner, Institut für
Technikfolgenabschätzung und Systemanalyse
(ITAS), Karlsruher Institut für Technologie (KIT)

1 MALEVOLENTE AKTEUR:INNEN IM METAVERSE

Die Vorstellung der Vision eines Metaverses durch den Meta-CEO Mark Zuckerberg im Oktober 2021 stellte für die Gesellschaft sowie die Tech-Welt eine deutliche Zäsur dar. Zwar ist die Idee eines zweiten digitalen Lebens inklusive digitaler Identität weder neu noch META-exklusiv (siehe bspw. Second Life), jedoch war dies bis dahin die erste nahezu vollständige Darstellung der möglichen Funktionsweisen des Metaverses. Hierbei liegt ein besonderes Augenmerk auf der Immersion über virtuelle Realität, die Nutzer:innen in Erweiterung heutiger Internetapplikationen ein völlig neues Gefühl der Beteiligung geben und sie multimodal bzw. -sinnlich am Metaverse teilhaben lassen soll. Des Weiteren lag in der Visionspräsentation ein besonderer Fokus auf der technologischen Permeabilität, der Diffusion sozialer Medien in alle menschlichen Lebensbereiche hinein und folglich der Ablösung sozialer Medien als reine Unterhaltungsplattform.

Sollte das Metaverse also kommen, so wie es sich Zuckerberg und andere Proponent:innen ausmalen, dann würde es einen radikalen Wandel des gesellschaftlichen Umgangs mit dem digitalen Raum, aber auch eine radikale Veränderung unseres alltäglichen Lebens bedeuten. Der Einkauf könnte sich als immersives Shopping-Erlebnis zusehends in das Metaverse verlagern, Sportkurse könnten in einer virtuellen Umgebung abgehalten werden oder virtuelle Gottesdienste mit Gläubigen aus aller Welt stattfinden. Die Arbeitswelt hat sich, unter anderem durch die Corona-Pandemie, bereits nachhaltig verändert – demnächst könnte der Schritt vom Home- ins Meta-Office anstehen.

Doch diese Innovationen werden nicht nur unseren Alltag verändern – sie werden auch dazu führen, dass Extremismus und Radikalisierung neue Wege gehen werden und sich transformieren, um sich den neuen Umgebungen anzupassen. Im Allgemeinen nutzen Extremist:innen Technologien, die günstig, leicht verfügbar, einfach zu bedienen und für ihre Zwecke, wie Propaganda, Kommunikation und Rekrutierung, weithin zugänglich sind. Die Nutzung von Technologie in einer anderen als von den Entwickler:innen beabsichtigten Funktion mit der Absicht, anderen zu schaden, ist ein inhärent kreativer Prozess. Cropley, Kaufman und Cropley (2008) nennen dies „malevolente Kreativität“. Sie definieren es als eine Form der

Kreativität, die „von einer Gesellschaft, einer Gruppe oder einem Individuum als notwendig erachtet wird, um Ziele zu erreichen, die sie für wünschenswert halten, die aber für eine andere Gruppe schwerwiegende negative Folgen hat, wobei diese negativen Folgen von der ersten Gruppe voll beabsichtigt sind“ (106). Akteur:innen, die malevolent kreativ sind (wie zum Beispiel Extremist:innen oder Verbreiter:innen von Fake News usw.), bezeichnen wir als malevolente Akteur:innen.

Malevolente Akteur:innen waren in der Vergangenheit vor allem dann sehr kreativ, wenn es um die Neuausrichtung der eigenen Organisation und die Distribution der eigenen Ideologie ging. Die digitale Revolution hat sie mit einer noch nie dagewesenen Anzahl von Werkzeugen ausgestattet, um ihre Sache voranzutreiben: von (verschlüsselter und sofortiger) Massenkommunikation für Propaganda und Rekrutierung über alternative Instrumente zur Finanzierung von Operationen und Logistik bis hin zu neuen Mitteln der Zerstörung und des Terrors. Die jüngsten technologischen Fortschritte haben malevolenten Akteur:innen viele neue Möglichkeiten eröffnet. So haben beispielsweise das Web 2.0, das Aufkommen der sozialen Medien und die Verfügbarkeit fast aller Inhalte im Internet es diesen Akteur:innen ermöglicht, leicht mit anderen Gleichgesinnten in Kontakt zu treten und nahezu geschlossene Gemeinschaften zu bilden, die sich in ihren Ansichten bestärken.

Die Erforschung des Metaverse als Nachfolge der sozialen Medien und des mobilen Internets kann wichtige Erkenntnisse darüber liefern, wie malevolente Akteur:innen das Metaverse kreativ nutzen könnten. Während wir im Allgemeinen mit Joe Whittaker und anderen (Whittaker 2022; Valentini, Lorusso und Stephan 2020) darin übereinstimmen, dass die Unterscheidung zwischen „Offline“- und „Online“-Radikalisierung analytisch nicht sinnvoll ist, könnte die Art und Weise, wie malevolente Akteur:innen aktuell soziale Medien nutzen, einen Ausblick auf das zukünftige Metaverse erlauben:

Es ist allgemein anerkannt, dass malevolente Akteur:innen (mit unterschiedlichen ideologischen Hintergründen) das Internet und seine Möglichkeiten früh genutzt haben (Feldman 2020; Fisher 2015; Stewart 2021; Lehmann und Schröder 2021). Sie nutzten neue Technologien in kreativer Weise, um sich der Überwachung und Entdeckung zu entziehen, aber auch um ihre eigenen Operationen zu verbessern. Das Internet als anonymer Ort der unzähligen Möglichkeiten, in dem man jede Information finden kann, die auf die eigenen Interessen zugeschnitten ist, macht es zu einer Goldgrube für Extremist:innen (Bertram 2016, S. 232).

Während Forschungen zu Radikalisierungsmustern verurteilter Dschihad-Terrorist:innen gezeigt haben, dass Offline-Netzwerke eine viel größere Rolle bei ihrer Radikalisierung spielten als Online-Netzwerke (Hamid und Ariza 2022), deuten andere Forschungen darauf hin, dass bei rechtsextremen Täter:innen das Internet eine bedeutendere Rolle einnimmt. Dies gilt vor allem für die Planungsphase ihrer Angriffe und Aktionen (von Behr et al. 2013; Gill et al. 2017). „Das Internet ist in erster Linie ein Hilfsmittel, das größere Möglichkeiten für Radikalisierung und Anschlagplanung bietet. Dennoch sind Radikalisierung und Anschlagplanung nicht vom Internet abhängig [...]“ (Gill et al. 2017, S. 113)¹

Insbesondere Social Media wurde von malevolenten Gruppen genutzt, um selbst erstellte Inhalte gezielt einzusetzen und zu verbreiten, ohne die traditionellen Prüfverfahren der Medienunternehmen zu durchlaufen und sich somit der Kontrolle und Zensur durch die Nationalstaaten zu entziehen“ (Droogan et al. 2018, S. 171)². Darüber hinaus sind die sozialen Medien auch zu einem Instrument der sozialen Interaktion für diejenigen geworden, die bereits radikalisiert sind, und diejenigen, die sie überzeugen wollen bzw. die sich für ihre Aktivitäten interessieren (Conway 2017).

Die Einführung des Metaverse könnte diese Dynamik noch verstärken. Indem die Kluft zwischen „offline“ und „online“ weiter überbrückt wird, könnte die Unterscheidung zwischen den beiden Sphären der Radikalisierung (und des Extremismus und Terrorismus) noch schwieriger zu halten sein. Während heute Offline-Netzwerke Vertrautheit und ein enges Umfeld bieten und sich den Sicherheitsbehörden mit größerer Wahrscheinlichkeit entziehen als Online-Extremist:innen (Hamid und Ariza 2022), könnte das künftige Metaverse diese Vorteile der Offline-Welt in einer umfassenden und immersiven digitalen Erfahrung vereinen. Zusammen mit den Vorteilen der Online-Welt – sofortige Massenkommunikation und

¹ Übersetzung mit DeepL.

² Übersetzung mit DeepL.

Propaganda – könnte das Metaverse ein noch größerer Gamechanger werden, als es das Internet und die sozialen Medien waren.

2 DAS METAVERSE ALS DEMOKRATISCHER RAUM

Das Metaverse befindet sich noch im Anfangsstadium seiner Entwicklung und es hat noch einen weiten Weg vor sich, bis es ein gewisses Reifestadium erreicht hat, in dem Versprechen und tatsächliche Funktionalitäten erfüllt werden. Bereits jetzt ist sichtbar, dass dessen Risiken mit denen sozialer Medien vergleichbar sind und in der Vergangenheit oft zu spät reagiert wurde. Freiheit und Sicherheit werden wahrscheinlich die entscheidenden Variablen in dieser Zukunftstechnologie sein, was die Auseinandersetzung mit malevolenten Akteur:innen umso zentraler macht (Neuberger 2023). Bereits in der Anfangsphase des Metaverses zeichnet sich ab, dass malevolente Akteur:innen hier durchaus fruchtbaren Boden vorfinden – veranschaulicht etwa durch die Fälle sexueller Belästigung, die in den heute verfügbaren Testversionen des Metaverses verbreitet sind (Bazu 2021; Bovermann 2022; Diaz 2022; Wiederhold 2022).

Wie kann diesen Entwicklungen begegnet werden? Wie lassen sie sich gar verhindern, bevor sie Schaden anrichten? Ein wichtiger Aspekt wird die demokratische Einbindung von Akteur:innen und marginalisierten Gruppen in Entscheidungs- und Entwicklungsprozesse sein. Während dies bei sozialen Medien bereits ein genuin reaktiver Prozess wäre, haben die Entwickler:innen des Metaverses noch die Möglichkeit, förderliche Strukturen aufzubauen. Eine Demokratisierung der sozialen Medien ist aus gesellschaftspolitischer Sicht wünschenswert, da sie aufgrund ihrer Verbreitung, ihrer wirtschaftlichen und kulturellen Bedeutung mit einer großen Machtfülle ausgestattet sind. Diese Macht sollte demokratisch legitimiert und kontrolliert werden (Engelmann et al. 2020). Dabei sollte eine demokratische Absicherung jedoch keinen parteipolitischen Mustern folgen.

Soziale Medien sollten bei der Entwicklung des Metaverses in verschiedenen Aspekten lehrreich sein – von der Kreativität, mit der malevolente Akteur:innen neue Medien und Technologien nutzen (s.o.), bis zur demokratischen Einbindung von Nutzer:innen. Social-Media-Betreiber:innen versuchten bereits dem Aspekt der Partizipation Rechnung zu tragen:

- **META** hat im Jahr 2018 die Idee eines Oversight Boards³ vorgestellt, eine unabhängige Instanz, die das Unternehmen bei schwierigen Entscheidungen zu Inhalten unterstützen soll. Dieses Gremium soll unabhängig, zugänglich und transparent handeln und dabei von META mit voller Handlungsfähigkeit in Bezug auf die Entscheidung zur Zulassung oder Entfernung von Inhalten ausgestattet sein.
- **Twitter** wurde in der Vergangenheit von einem Trust and Safety Council beraten. Dieser setzte sich aus verschiedenen NGOs und Forscher:innen zusammen und beriet das Unternehmen zu Themen der Online-Sicherheit. Elon Musk löste den Council nach seiner Übernahme auf (The Associated Press 2022).
- Google hat auf seiner Videoplattform **YouTube** das „Priority Flagger Program“⁴ eingeführt. Dieses bietet für NGOs und staatliche Behörden die Möglichkeit, besonders effektive Tools zu nutzen, um Inhalte zu melden, die gegen die Community-Richtlinien verstoßen. Diese werden dann von Inhaltsmoderator:innen priorisiert überprüft. Die Löschkriterien sind jedoch dieselben wie bei „normalen“ Meldungen. Dieses Programm wurde 2021 von YouTube umgearbeitet, was zu großer Kritik aus der Community führte (Meineck 2021).

Generell scheint eine besorgniserregende Tendenz in den sozialen Medien zu bestehen, diese partizipativen Modelle in Moderation und Sicherheit zugunsten von Anwendungen Künstlicher Intelligenz (KI) zurückzufahren (Gorwa et al. 2020; Llansó 2020). KI-Lösungen können und sollten jedoch die Einbindung der Zivilgesellschaft in Entscheidungsprozesse und Fragen der demokratischen Kultur keinesfalls ersetzen, schon allein deshalb, weil KI-gestützte Lösungen zur Content-Moderation nach wie vor fehleranfällig und intransparent agieren (Gillespie 2020; Gorwa et al. 2020).

³ Siehe <https://www.oversightboard.com/>.

⁴ Siehe <https://support.google.com/youtube/answer/7554338?hl=de>.

3 PARTIZIPATION UND DEMOKRATISCHE TEILHABE WAGEN

In der Social-Media-Forschung und insbesondere in der Plattform-Governance-Forschung lassen sich wichtige Ansätze finden, die dazu beitragen können, ein demokratisches und inklusives Metaverse zu ermöglichen. Neben der zentralen Kooperation von Betreiber:innen mit staatlichen und nicht staatlichen Akteur:innen in Fragen der Transparenz und Forschung wird vor allem die aktive Stärkung demokratischer Akteur:innen und Narrative betont (Bundtzen und Schwieter 2023; Engelmann et al. 2020; Rau et al. 2022).

Diese Strategie ist von größter Wichtigkeit, um sicherzustellen, dass der staatliche Repressionsapparat tatsächlich nur als Ultima Ratio eingesetzt wird, um malevolente Akteur:innen zu stoppen. Der demokratische Streit und Diskurs muss in einem inklusiven Metaverse möglich sein, ohne beständig Repression und Einschränkung fürchten zu müssen. Stattdessen können Plattformbetreiber:innen auch im Metaverse Schritte setzen, um demokratische Akteur:innen und Narrative bewusst und aktiv zu fördern und somit demokratische Resilienz im Metaverse aufzubauen.

Auch hier kann sich das Metaverse an bestehenden Versuchen in den sozialen Medien orientieren, zum Beispiel an YouTube's „Trusted Flagging“-Programm: Demokratische Akteur:innen, zum Beispiel nicht staatliche und staatliche Organisationen, die sich auf Thematiken wie Hassrede, gruppenbezogene Menschenfeindlichkeit oder Demokratiestärkung spezialisiert haben, könnten Zugang zu speziellen Meldetools bekommen. Des Weiteren könnten sie mit erweiterten Befugnissen zur Kontextualisierung von zweifelhaften Inhalten ausgestattet werden.

Neben der Stärkung demokratischer Narrative ist jedoch auch die Demokratisierung der Plattform selbst ein wichtiger Faktor für Inklusivität. Die Einbindung von Nutzer:innen in Entscheidungs- und Designprozesse kann von großem Mehrwert für eine Plattform sein, die an einem demokratischen Miteinander interessiert ist. Marginalisierte Gruppen und deren Vertreter:innen wissen sehr genau, wo im digitalen Raum Hass und Hetze lauern können, und mit einer frühzeitigen Einbindung solcher Stakeholder:innen könnten einige der Fehler, die in den sozialen Medien begangen wurden, bereits von vornherein minimiert werden.

In der politischen Praxis haben sich bereits „**Mini-Publics**“ als Instrument der Nutzer:innenbeteiligung bewährt (Escobar und Elstub 2017; Smith und Setälä 2018). Mini-Publics bezeichnen eine Gruppe von (zufällig oder systematisch) ausgewählten Bürger:innen, die sich über einen längeren Zeitraum und unter Einbeziehung externer Quellen, z.B. wissenschaftlicher Expertise, mit gesellschaftlich relevanten Themen auseinandersetzen. Dabei werden Themenfelder multiperspektivisch beleuchtet, diskutiert, bewertet und darauf aufbauend Empfehlungen an politische Entscheidungsträger:innen weitergegeben (Escobar und Elstub 2017; Pek et al. 2023). Als Beispiel sei hier der virtuelle Bürger:innenrat in Deutschland genannt, der im Juni 2022 über die Folgen des Einsatzes Künstlicher Intelligenz diskutierte (Buergerrat.de 2022). Mit solchen Räten könnten plattformspezifische Themen diskutiert werden, um somit zu einer demokratischeren Entscheidungsfindung zu kommen.

Obwohl durchaus umstritten (s.o.), können **Plattformräte** auch demokratieförderndes Potenzial entwickeln, wenn sie unabhängig, objektiv und transparent agieren können (Haggart und Keller 2021; Rau et al. 2022). Um dies zu gewährleisten, könnten sich derartige Plattformräte entsprechend den Vorschlägen von Kettemann und Fertmann (2021) an dem Status der bereits in Deutschland etablierten Presse- und Rundfunkräte orientieren. Hier bleibt jedoch anzumerken, dass Zuständigkeiten (geografisch, sachlich), Besetzung (Bürger:innen, Expert:innen, NGOs, politische Entscheidungsträger:innen) und nicht zuletzt Befugnisse (quasi-judikativ, beratend) Teil des gesellschaftlichen Diskurses sein müssen und noch nicht abschließend geklärt werden können (COWLS et al. 2022; Kettemann und Fertmann 2021). Außerdem könnten solche Räte das öffentliche Vertrauen fördern – je diverser und transparenter deren Besetzung ist und auch je mehr die Auswirkungen ihrer Empfehlungen öffentlich sichtbar sind.

Nicht zuletzt muss es darum gehen, die Medien- und Politikkompetenz zu stärken, indem verschiedene **Fortbildungsmöglichkeiten** angeboten werden. Diese sollten so gestaltet sein, dass auch Personen, die nicht (mehr) im Bildungssystem angebunden sind, von ihnen profitieren können. Hier gilt es, das nötige „Rüstzeug“ im Umgang mit Fake News, anderweitig manipulierten bzw. extremistischen Inhalten oder auch Hate Speech im Internet zu vermitteln. Exemplarisch ist das Projekt „Good Gaming – Well Played

Democracy⁵ der Amadeu-Antonio-Stiftung zu nennen, welches sich u.a. mit der Sensibilisierung der Gaming-Community bezüglich extremistischer Inhalte auseinandersetzt.

Daneben muss angemerkt werden, dass der Aufbau eines demokratischen Metaverse nicht allein die Aufgabe der Bürger:innen ist. Der Aufbau eines digitalen Zwillinges im Sinne einer „wehrhaften Demokratie“ ist ebenfalls von Bedeutung. Dies bedeutet aber laut Rau et al. (2022), nicht ausschließlich repressive Maßnahmen wie Löschung oder Unterdrückung problematischer Inhalte (siehe beispielsweise Bellanova und De Goede 2022), sondern damit einhergehend auch die Stärkung demokratischer Akteur:innen, beispielsweise durch algorithmisch erhöhte Sichtbarkeit. In diesem Zusammenhang kommt dem Empowerment marginalisierter demokratischer Akteursgruppen eine besondere Bedeutung zu, um die gesellschaftliche Vielfalt adäquat abzubilden. Diese sind u.a. ausreichend geschult, um problematische Inhalte rechtzeitig zu erkennen und können somit auch beratend hinzugezogen werden (Rau et al. 2022). Auch der Einsatz von Gegenrede (Counter Speech) könnte eine weitere Handlungsstrategie gegen extremistische Metaverse-Inhalte sein (Clever et al. 2022; Hangartner et al. 2021; Kunst et al. 2021; Morten et al. 2020). Unter (digitaler) Gegenrede versteht man die Erstellung von Kommentaren oder anderweitigen Inhalten als Antwort auf hasserfüllte Inhalte, um deren Effekt zu minimieren und zu entkräften oder um potenzielle Opfer zu unterstützen (Ernst et al. 2022; Garland et al. 2022). Diesbezüglich konnte in Studien gezeigt werden, dass Gegenrede ein wirksames Mittel im Umgang mit extremistischen Inhalten sein kann und diese effektiv reduziert (Garland et al. 2022; Hangartner et al. 2021). Im Kontext neuerer Technologiekomplexe, z.B. KI, bestehen inzwischen sogar Überlegungen, Gegenrede unter bestimmten Bedingungen automatisiert zu implementieren, auch wenn finale Konzepte und Zuständigkeiten bislang noch Gegenstand intensiver Diskussion sind (Clever et al. 2022).

Zusätzlich zu partizipierenden Methoden bietet schließlich auch die Gesetzgebung Möglichkeiten, extremistische Inhalte zu unterbinden. In Deutschland ist die Verbreitung verfassungswidriger Symbole oder Zeichen per se verboten und kann somit strafrechtlich verfolgt werden. Ebenso gibt das Netzwerkdurchsetzungsgesetz (NetzDG) in Deutschland rechtliche Rahmenbedingungen für den Umgang mit Hasskriminalität in den sozialen Medien. Dementsprechend verpflichtet die Terrorist Content Online-Verordnung (TCO-VO; Europäische Union 2021) auf EU-Ebene Plattformbetreibende dazu, als terroristisch gemeldete Inhalte innerhalb von einer Stunde zu entfernen respektive zu sperren. Mit Blick auf neuere Forschungsergebnisse der Extremismusforschung ist jedoch zu verzeichnen, dass sogenannte „Legal-but-Harmful“-Inhalte bereits jetzt eine große Herausforderung darstellen und vermutlich auch im Metaverse von Bedeutung sein werden (Jiang et al. 2021; Rau et al. 2022). Hierzu zählt beispielsweise digitaler Content, der subtil radikalierend wirken kann und gleichzeitig nicht strafrechtlich relevant ist. Diesbezüglich ist jedoch anzumerken, dass Content-Moderation sich der freien Meinungsäußerung als verfassungsrechtlich verankerter Grundsatz unterwerfen muss. Folglich ist zu vermuten, dass auch die anhaltende Diskussion zum Verhältnis aus Freiheit und Sicherheit maßgeblich die Gestaltung des Metaverse beeinflussen wird und das Ergebnis eines gesamtgesellschaftlichen Verhandlungsprozesses sein wird bzw. muss, um die demokratische Ausgestaltung gewährleisten zu können.

4 DISKUSSION

Sollte das Metaverse in seiner Immersivität die Vision von Mark Zuckerberg erreichen, dann ist es sehr wahrscheinlich, dass es unser Alltagsleben und das gesellschaftliche Miteinander massiv beeinflussen wird. Diese Immersivität würde bedeuten, dass sich die Betreiber:innen des Metaverses (oder der Metaversen) intensiv mit Fragen der Demokratisierung auseinandersetzen müssen. Nicht nur würde wahrscheinlich der Staat in einem Metaverse eine (noch zu definierende) Rolle spielen, sondern auch seinen Nutzer:innen muss demokratische Teilhabe ermöglicht werden. Diese würde dazu beitragen, die Plattform inklusiv und möglichst sicher vor malevolenten Akteur:innen zu machen.

Aufbauend auf der Social-Media-Forschung der letzten Jahrzehnte finden sich viele Anknüpfungspunkte, welche die Ausgestaltung eines demokratischen Metaverse unterstützen und lenken können. Wie oben angemerkt, befindet sich das Metaverse noch in einem frühen Entwicklungsstadium. Dabei ist es mit Blick auf die rasante Dynamik jedoch wichtig, diesen Prozess zu begleiten, „am Ball“ zu bleiben und eine aktive Rolle in der Diskussion einzunehmen. Nicht zuletzt ist die multiperspektivische Brille im Sinne aller

⁵ <https://www.amadeu-antonio-stiftung.de/projekte/good-gaming-well-played-democracy/>.

beteiligten Stakeholder:innen relevant, um die Balance zwischen Sicherheit und Freiheit für alle Nutzer:innen gewährleisten zu können. Die hier vorgestellten Möglichkeiten zum Aufbau eines Metaverse stellen einige Lösungen zur Implementierung demokratischer Grundpfeiler dar. Zusammengefasst sollten folgende Lösungsmöglichkeiten von den Plattformbetreibern eingesetzt werden:

- frühe Implementierung von Methoden zur Nutzer:innenbeteiligung, z.B. Mini Publics oder unabhängige Plattformräte
- Stärkung demokratischer Akteur:innen und Einbezug marginalisierter Gruppen
- Bezug auf bisherige wissenschaftliche Forschungsergebnisse zu Social Media, Hassrede und (digitalem) Extremismus sowie offene Zusammenarbeit mit Forschungseinrichtungen
- Angebot von Bildungsmöglichkeiten in Zusammenarbeit mit demokratischen Akteur:innen

Die konkrete und finale Umsetzung ist aktuell noch Gegenstand einer lebhaften Diskussion. Allerdings animiert der Status der frühen Entwicklungsphase des Metaverse zur aktiven Beteiligung, was sich auch in dem hiesigen Project Immersive Democracy widerspiegelt und als Einladung zu diesem Prozess verstanden werden kann. Partizipation ist kein Allheilmittel für die Gefahren, die im digitalen Raum lauern. Sie stellt aber eine wichtige Stütze dar, die dabei helfen kann, marginalisierte Gruppen oder Individuen gezielt zu empowern und ihnen so das Werkzeug in die Hand zu geben, um gemeinsam mit Betreiber:innen gegen Diskriminierung und Hass im Metaverse vorzugehen. Jetzt ist die Zeit, um diese Werkzeuge zu entwickeln und dafür zu sorgen, dass ein zukünftiges Metaverse ein möglichst sicherer Ort für alle sein wird.

LITERATUR

Bazu, Tanya (2021): The metaverse has a groping problem already. MIT Technology Review. Online verfügbar unter <https://www.technologyreview.com/2021/12/16/1042516/the-metaverse-has-a-groping-problem/>, zuletzt geprüft am 28.09.2022.

von Behr, Ines; Reding, Anaïs; Edwards, Charlie; Gribbon, Luke (2013): Radicalisation in the Digital Era: The Use of the Internet in 15 Cases of Terrorism and Extremism | Office of Justice Programs. In: RAND Europe.

Bellanova, Rocco; De Goede, Marieke (2022): Co-Producing Security: Platform Content Moderation and European Security Integration. In: JCMS: Journal of Common Market Studies 60 (5), S. 1316–1334. <https://doi.org/10.1111/jcms.13306>

Bertram, Luke (2016): Terrorism, the Internet and the Social Media Advantage: Exploring how terrorist organizations exploit aspects of the internet, social media and how these same platforms could be used to counter-violent extremism. In: Journal for Deradicalization 2016 (7), S. 225–252.

Bovermann, Philipp (2022): Online-Belästigungen im Metaverse – Am eigenen Leib. Süddeutsche Zeitung. Online verfügbar unter <https://www.sueddeutsche.de/kultur/metaverse-vr-virtual-reality-microsoft-sexuelle-belaestigung-1.5519527?print=true>, zuletzt geprüft am 04.02.2022.

Buergerrat.de (2022): Bürgerrat diskutierte über künstliche Intelligenz. Buergerrat.de. Online verfügbar unter <https://www.buergerrat.de/aktuelles/buergerrat-diskutierte-ueber-kuenstliche-intelligenz/>, zuletzt geprüft am 22.06.2023.

Bundtzen, Sara; Schwieter, Christian (2023): Datenzugang zu Social-Media-Plattformen für die Forschung: Lehren aus bisherigen Maßnahmen und Empfehlungen zur Stärkung von Initiativen inner- und außerhalb der EU. Berlin: Institute for Strategic Dialogue (ISD).

Clever, Lena; Klapproth, Johanna; Frischlich, Lena (2022): Automatisierte (Gegen-)Rede? Social Bots als digitales Sprachrohr ihrer Nutzer*innen. In: Julian Ernst, Michalina Trompeta und Hans-Joachim Roth (Hg.): Gegenrede digital: Neue und alte Herausforderungen interkultureller Bildungsarbeit in Zeiten der Digitalisierung. Wiesbaden: Springer Fachmedien, (Interkulturelle Studien), S. 11–26. https://doi.org/10.1007/978-3-658-36540-0_2

Conway, Maura (2017): Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. In: Studies in Conflict & Terrorism Routledge, 40 (1), S. 77–98. <https://doi.org/10.1080/1057610X.2016.1157408>

- Cowls, Josh; Darius, Philipp; Santistevan, Dominiquo; Schramm, Moritz (2022): Constitutional metaphors: Facebook's "supreme court" and the legitimation of platform governance. In: *New Media & Society*.
<https://doi.org/10.1177/14614448221085559>
- Cropley, David H.; Kaufman, James C.; Cropley, Arthur J. (2008): Malevolent Creativity: A Functional Model of Creativity in Terrorism and Crime. In: *Creativity Research Journal* 20 (2), S. 105–115.
<https://doi.org/10.1080/10400410802059424>
- Diaz, Adriana (2022): Disturbing reports of sexual assaults in the metaverse: 'It's a free show'. *New York Post*. Online verfügbar unter <https://nypost.com/2022/05/27/women-are-being-sexually-assaulted-in-the-metaverse/>, zuletzt geprüft am 28.09.2022.
- Droogan, Julian; Waldek, Lise; Blackhall, Ryan (2018): Innovation and terror: an analysis of the use of social media by terror-related groups in the Asia Pacific. In: *Journal of Policing, Intelligence and Counter Terrorism* Routledge, 13 (2), S. 170–184. <https://doi.org/10.1080/18335330.2018.1476773>
- Engelmann, Severin; Grossklags, Jens; Herzog, Lisa (2020): Should users participate in governing social media? Philosophical and technical considerations of democratic social media. In: *First Monday*.
<https://doi.org/10.5210/fm.v25i12.10525>
- Ernst, Julian; Trompeta, Michalina; Roth, Hans-Joachim (2022): Gegenrede digital – Einleitung in den Band. In: Julian Ernst, Michalina Trompeta und Hans-Joachim Roth (Hg.): *Gegenrede digital*. Wiesbaden: Springer Fachmedien Wiesbaden, (Interkulturelle Studien), S. 1–7. https://doi.org/10.1007/978-3-658-36540-0_1
- Escobar, Oliver; Elstub, Stephen (2017): *Forms of Mini-Publics: An introduction to deliberative innovations in democratic practice*. (Research and Development Note) New Democracy.
- Europäische Union (2021): Verordnung (EU) 2021/784 des Europäischen Parlaments und des Rates vom 29. April 2021 zur Bekämpfung der Verbreitung terroristischer Online-Inhalte (Text von Bedeutung für den EWR). OJ L.
- Garland, Joshua; Ghazi-Zahedi, Keyan; Young, Jean-Gabriel; Hébert-Dufresne, Laurent; Galesic, Mirta (2022): Impact and dynamics of hate and counter speech online. In: *EPJ Data Science* 11 (1), S. 3. <https://doi.org/10.1140/epjds/s13688-021-00314-6>
- Gill, Paul; Corner, Emily; Conway, Maura; Thornton, Amy; Bloom, Mia; Horgan, John (2017): Terrorist Use of the Internet by the Numbers: Quantifying Behaviors, Patterns, and Processes. In: *Criminology & Public Policy* 16 (1), S. 99–117.
<https://doi.org/10.1111/1745-9133.12249>
- Gillespie, Tarleton (2020): Content moderation, AI, and the question of scale. In: *Big Data & Society* 7 (2).
<https://doi.org/10.1177/2053951720943234>
- Gorwa, Robert; Binns, Reuben; Katzenbach, Christian (2020): Algorithmic content moderation: Technical and political challenges in the automation of platform governance. In: *Big Data & Society* 7 (1).
<https://doi.org/10.1177/2053951719897945>
- Haggart, Blayne; Keller, Clara Iglesias (2021): Democratic legitimacy in global platform governance. In: *Telecommunications Policy* 45 (6). <https://doi.org/10.1016/j.telpol.2021.102152>
- Hamid, Nafees; Ariza, Cristina (2022): *Offline Versus Online Radicalisation: Which is the Bigger Threat?* London: Global Network on Extremism&Technology.
- Hangartner, Dominik et al. (2021): Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. In: *Proceedings of the National Academy of Sciences* 118 (50). <https://doi.org/10.1073/pnas.2116310118>
- Jiang, Jialun Aaron; Scheuerman, Morgan Klaus; Fiesler, Casey; Brubaker, Jed R.; Alexandre Bovet (Hg.) (2021): Understanding international perceptions of the severity of harmful content online. In: *PLOS ONE* 16 (8).
<https://doi.org/10.1371/journal.pone.0256762>
- Kettemann, Matthias C.; Fertmann, Martin (2021): *Die Demokratie Plattformfest Machen: Social Media Councils als Werkzeug zur gesellschaftlichen Rückbindung der privaten Ordnungen digitaler Plattformen*. Potsdam-Babelsberg: Friedrich-Naumann-Stiftung.

Kunst, Marlene; Porten-Cheé, Pablo; Emmer, Martin; Eilders, Christiane (2021): Do "Good Citizens" fight hate speech online? Effects of solidarity citizenship norms on user responses to hate comments. In: *Journal of Information Technology & Politics* 18 (3), S. 258–273. <https://doi.org/10.1080/19331681.2020.1871149>

Llansó, Emma J (2020): No amount of "AI" in content moderation will solve filtering's prior-restraint problem. In: *Big Data & Society* 7 (1). <https://doi.org/10.1177/2053951720920686>

Meineck, Sebastian (2021): Trusted Flagger: YouTube serviert freiwillige Helfer:innen ab. [netzpolitik.org](https://netzpolitik.org/2021/trusted-flagger-youtube-serviert-freiwillige-helferinnen-ab/). Online verfügbar unter <https://netzpolitik.org/2021/trusted-flagger-youtube-serviert-freiwillige-helferinnen-ab/>, zuletzt geprüft am 27.06.2023.

Morten, Anna; Frischlich, Lena; Rieger, Diana (2020): Gegenbotschaften als Baustein der Extremismusprävention. In: Josephine B. Schmitt, Julian Ernst, Diana Rieger und Hans-Joachim Roth (Hg.): *Propaganda und Prävention*. Wiesbaden: Springer Fachmedien Wiesbaden, S. 581–589. https://doi.org/10.1007/978-3-658-28538-8_32

Neuberger, Christoph (2023): Sicherheit und Freiheit in der digitalen Öffentlichkeit. In: Nicole J. Saam und Heiner Bielefeldt (Hg.): *Sozialtheorie*. Bielefeld: transcript Verlag, S. 297–308.

Pek, Simon; Mena, Sébastien; Lyons, Brent (2023): The Role of Deliberative Mini-Publics in Improving the Deliberative Capacity of Multi-Stakeholder Initiatives. In: *Business Ethics Quarterly* 33 (1), S. 102–145. <https://doi.org/10.1017/beq.2022.20>

Rau, Jan; Kero, Sandra; Hofmann, Vincent; Dinar, Christina; Heldt, Amélie P. (2022): Rechtsextreme Online-Kommunikation in Krisenzeiten: Herausforderungen und Interventionsmöglichkeiten aus Sicht der Rechtsextremismus- und Plattform-Governance-Forschung. In: *Arbeitspapiere des Hans-Bredow-Instituts SSOAR - GESIS Leibniz Institute for the Social Sciences*. <https://doi.org/10.21241/SSOAR.78072>

Smith, Graham; Setälä, Maija (2018): Mini-Publics and Deliberative Democracy. In: Andre Bächtiger, John S. Dryzek, Jane Mansbridge und Mark Warren (Hg.): *The Oxford Handbook of Deliberative Democracy*. Oxford University Press, S. 299–314. <https://doi.org/10.1093/oxfordhb/9780198747369.013.27>

The Associated Press (2022): Musk's Twitter has dissolved its Trust and Safety Council. *National Public Radio (NPR)*. Washington, DC, 12.12.2022.

Wiederhold, Brenda K. (2022): Sexual Harassment in the Metaverse. In: *Cyberpsychology, Behavior, and Social Networking* 25 (8), S. 479–480. <https://doi.org/10.1089/cyber.2022.29253.editorial>