

DIE ZENTRALE ROLLE DER SOCIAL-MEDIA-FORSCHUNG BEI DER BEKÄMPFUNG VON HASS: EIN VIROLOGISCHER ANSATZ

Dr. Matthias J. Becker

Die zentrale Rolle der Social-Media-Forschung bei der Bekämpfung von Hass: Ein virologischer Ansatz

Dr. Matthias J. Becker
Leiter, Decoding Antisemitism

Zusammenfassung

Social-Media-Plattformen sind zum zentralen Marktplatz der Ideen in modernen Demokratien geworden – und bleiben doch eine Black Box. Ohne systematische Instrumente, um die Dynamiken zu entschlüsseln, die sich in diesen digitalen Räumen entfalten, oder ihre Entwicklung in Echtzeit nachzuverfolgen, fehlt uns die grundlegende Wissensbasis, die notwendig ist, um die gesellschaftlichen und politischen Trends zu verstehen, welche unsere Gesellschaften prägen. Dieser Artikel plädiert dafür, Social-Media-Forschung – verstanden als Verbindung von Diskursanalyse und akteurszentrierter Forschung – dringend in jeden Versuch zu integrieren, Polarisierung, Radikalisierung und die Erosion demokratischer Normen zu begreifen.

Ausgehend von einem **virologischen Ansatz** behandelt dieser Artikel Hassideologien – Antisemitismus, Rassismus, Misogynie, Verschwörungstheorien und antidemokratische Ressentiments – als adaptive Krankheitserreger, die mutieren, sich ausbreiten und Schwachstellen innerhalb der digital umstrukturierten Öffentlichkeit ausnutzen. So wie Virologen virale Mechanismen auf molekularer Ebene entschlüsseln müssen, bevor sie Impfstoffe entwickeln können, müssen Forschende systematisch erfassen, wie sich Hassdiskurse durch strategische Mehrdeutigkeit, algorithmische Verstärkung und partizipative Radikalisierung manifestieren, verändern und über Plattformen hinweg zirkulieren. Dies erfordert eine transdisziplinäre Synthese: computergestützte Sozialwissenschaft zur Erkennung großskaliger Muster, qualitative Diskursanalyse zur Interpretation impliziter Bedeutungen und kultureller Codes sowie KI-gestützte Methoden, um Entwicklungen in Echtzeit zu verfolgen.

Für demokratische Gesellschaften steht dabei existenziell viel auf dem Spiel. Social-Media-Plattformen haben die politische Kommunikation grundlegend umgestaltet und dadurch günstige „politisch-kulturelle Opportunitätsstrukturen“ für extremistische Narrative geschaffen, während sie zugleich traditionelle Gatekeeper und Wahrheitsinstanzen demontiert haben. Neue Intermediäre – Influencer, Algorithmen, milliarden schwere Plattformbesitzer – prägen heute den öffentlichen Diskurs, ohne institutioneller

Rechenschaftspflicht zu unterliegen, und verstärken grenzüberschreitende Inhalte, die Aufmerksamkeit erzeugen, ungeachtet des sozialen Schadens. Das Ergebnis ist eine sich beschleunigende Polarisierung, die Normalisierung vormals marginaler Ideologien und die Erosion jener gemeinsamen epistemischen Grundlagen, die für demokratische Deliberation notwendig sind.

Dennoch haben politische Entscheidungsträger diesen quasi unregulierten „öffentlichen Raum“ weitgehend vernachlässigt und zugelassen, dass er das gesellschaftliche Bewusstsein auf unkontrollierte und zunehmend gefährliche Weise beeinflusst.

Ohne eine rigorose Social-Media-Forschung, die in die Black Box des digitalen Diskurses eindringen kann – und offenlegt, wer was zu wem sagt, wie sich Botschaften während ihrer Zirkulation verändern, welche Akteure und Algorithmen welche Narrative verstärken und welche Offline-Folgen daraus hervorgehen –, agieren wir blind. Dieser Artikel skizziert einen Fahrplan von der Forschung zum Handeln: den Aufbau fortgeschrittener Infrastrukturen mit Mixed Methods, das Verständnis von Phänomenen und Trends durch systematische Beobachtung, die Übersetzung von Wissen in Politik und Bildung sowie die Entwicklung kontextsensibler Interventionen. Nur durch nachhaltige Investitionen in die Social-Media-Forschung als zentrale Säule demokratischer Governance können wir hoffen, der viralen Ausbreitung von Hass entgegenzuwirken und die Integrität der Öffentlichkeit zu bewahren.

Einleitung

In den vergangenen Jahren haben wir in westlichen Demokratien ein alarmierendes Wiederaufleben von Hassideologien erlebt – Antisemitismus, Rassismus, Misogynie, Verschwörungsdenken und antidemokratische Ressentiments –, begleitet von einer breiteren Renaissance autoritärer Bewegungen und systematischer Desinformationskampagnen. Diese Phänomene sind nicht bloß vorübergehende Ausschläge von Vorurteilen, sondern Ausdruck des Wiederauftauchens und der Transformation uralter Formen des Hasses, die nun durch digitale Kommunikationstechnologien verstärkt werden. Besonders besorgniserregend ist die sich beschleunigende gesellschaftliche Polarisierung und Radikalisierung, die diese Dynamiken hervorbringen und die den Zusammenhalt demokratischer Gesellschaften grundlegend bedrohen.

Diese Entwicklung spiegelt wider, was Forscher als grundlegenden Wandel der „politisch-kulturellen Opportunitätsstrukturen“ bezeichnet haben – also jener Bedingungen, unter denen Hassrede, Verschwörungstheorien und antidemokratische Ideologien sich ausbreiten und Legitimität gewinnen können. ¹ Die digital umstrukturierte Öffentlichkeit, verbunden mit einer sogenannten „epistemischen Krise“, die aus dem schwindenden Vertrauen in traditionelle Gatekeeper und Wahrheitsinstanzen wie Leitmedien und

demokratische Institutionen hervorgeht, hat äußerst günstige Bedingungen für die Verbreitung extremistischer Narrative über das gesamte ideologische Spektrum hinweg geschaffen. 2 Diese Dynamiken werden zusätzlich verschärft durch das, was manche als „globale Welle der Autokratisierung“ beschreiben, in der demokratische Normen und Institutionen weltweit systematisch erodieren. 3 Was dieses Wiederaufleben besonders alarmierend macht, ist die weit verbreitete moralische Desorientierung, die darin sichtbar wird: Teile der Zivilgesellschaft, die sich traditionell als Verteidiger progressiver Werte verstehen, haben sich – oft ungewollt – mit Bewegungen verbündet, die genau jenen Autoritarismus, Sexismus, Rassismus und jene Homophobie verkörpern, die sie eigentlich bekämpfen wollen.

Bis vor kurzem glaubten viele, Hochschulbildung und kritisches Denken würden als Bollwerke gegen extremistische Narrative wirken. Doch die Vorgänge an Universitäten in Nordamerika und Europa stellen diese Annahme infrage. Diese Demonstrationen, die von autoritären Regimen gelobt werden, die der liberalen Demokratie feindlich gegenüberstehen, werfen grundlegende Fragen darüber auf, wie gebildete Menschen dazu gelangen, Ideologien zu übernehmen, die den demokratischen Prinzipien fundamental widersprechen.

Mehrere Erklärungen drängen sich auf. Diese Personen könnten tatsächlich uninformiert sein und die Glaubenssysteme, die sie offenbar unterstützen, nie ernsthaft geprüft haben. Ebenso denkbar ist, dass die gegenwärtig im Westen vorherrschende Desorientierung und Unsicherheit eine Rolle spielen. Im Lauf der Geschichte hat kollektive Angst Menschen häufig dazu gebracht, entgegen ihrem besseren Urteil Positionen zu übernehmen, die scheinbar umfassende Lösungen für komplexe Probleme bieten. Das relativistische Denken, das den gegenwärtigen Diskurs dominiert, verstärkt diese Tendenz zusätzlich, stiftet Verwirrung über grundlegende Werte und treibt Menschen in ideologische Extreme.

Doch selbst diese Faktoren reichen nicht aus, um die gegenwärtige Erosion moralischer und intellektueller Klarheit vollständig zu erklären. Tatsächlich entsteht das Phänomen aus einer Gemengelage all dieser Elemente, die durch Social-Media-Plattformen massiv verstärkt wird – Plattformen, die ein Umfeld gefördert haben, das von Materialismus, Hyperindividualismus und dem unerbittlichen Streben nach Selbstbestätigung geprägt ist. Social Media bündelt damit alle Voraussetzungen für einen perfekten Sturm aus Aggression, Angst und Isolation.

Der jüngste Anstieg von Antisemitismus, Rassismus, Misogynie und verschwörungsideologischer Desinformation, verbunden mit einer weit verbreiteten Ablehnung demokratischer Werte, geht auf multiple, ineinandergreifende Faktoren zurück. Teilweise lassen sich diese Phänomene auf böswillige Akteure zurückführen, die den öffentlichen Diskurs bewusst manipulieren, um eigene Interessen durchzusetzen. Zahlreiche Berichte haben dieses Phänomen belegt und dokumentiert, wie staatliche und nichtstaatliche Akteure die öffentliche Meinung orchestrieren, um demokratische Gesellschaften zu destabilisieren. 4

Hinzu kommt, dass digitale Kommunikation sowohl Bottom-up- als auch Top-down-Dynamiken umfasst, die Hass und Desinformation verstärken. Einerseits ermöglichen soziale Medien horizontale, bottom-up verlaufende Informationsströme, die traditionelle Gatekeeper umgehen. Andererseits greifen diese Dynamiken in neue vertikale, top-down strukturierte Mechanismen ein, die durch intransparente algorithmische Kuratierung geprägt sind. Neue Gatekeeper – Influencer, milliardenschwere Medienmogule wie Elon Musk und Plattformalgorithmen – operieren ohne jene rechtlichen und institutionellen Kontrollstandards, die etablierte Medien begrenzen. Algorithmen privilegieren aufmerksamkeitsstarke, grenzverletzende, polarisierende und emotionalisierende Inhalte, fördern den kommerziellen Traffic und verstärken Stimmen, die Desinformation und Hass in einem „postfaktischen Zeitalter“ verbreiten. 5 Die Nutzung dieser algorithmischen Systeme durch Social-Media-Plattformen – vermeintlich zu kommerziellen Zwecken – verschärft damit Tendenzen, die einem konstruktiven und empathischen Dialog in Online-Gemeinschaften entgegenstehen. 6

Trotz alledem bleibt der wichtigste Faktor die inhärente Natur der Online-Kommunikation selbst – gekennzeichnet durch Anonymität, wechselseitige Bestärkung unter Nutzern und die rasche Verbreitung extremistischer Ideologien. Herausgebildet hat sich eine komplexe und weitreichende Toxizität, für die es in der Menschheitsgeschichte kein Beispiel gibt. Das interaktive Web, das nach eigenen Regeln funktioniert, formt das Bewusstsein der Gesellschaft insgesamt.

Bottom-up- und Top-down-Dynamiken wirken zusammen, um digitalen Hass auszuweiten und zu legitimieren, und fördern damit, was als kommunikativ-autoritär geprägte „Politik der Grenzüberschreitung“ im Netz und in der Gesellschaft insgesamt beschrieben worden ist. 7 Diese Politik der Grenzüberschreitung wirkt über alle Formen von Hass und antidemokratischem Diskurs hinweg – von Antisemitismus und Rassismus bis zu Misogynie und Verschwörungstheorien – und verschiebt schrittweise die Grenzen des sozial Sagbaren. Auf diese Weise werden extremistische Deutungsmuster als legitime Meinung normalisiert, und Hassideologien wandern aus subkulturellen Randmilieus in die gesellschaftliche Mitte.

Von politischen Entscheidungsträgern weitgehend ignoriert, beeinflusst dieser quasi unregulierte „öffentliche Platz“, wie Cal Newport treffend formuliert, das gesellschaftliche Denken auf unkontrollierte Weise. 8 Alles, was hinter einer ernsthaften Analyse dieser Verkettung von Entwicklungen zurückbleibt, läuft praktisch auf eine Einladung zu schädlichen Folgen für die demokratische Stabilität hinaus.

Vor diesem Hintergrund werden sprachwissenschaftliche und computergestützte Analysen zu unverzichtbaren Instrumenten. Wie bei jeder sozialen Ansteckung setzt die Bekämpfung hassbasierter Ideologien – ob Antisemitismus, Rassismus, Misogynie oder Verschwörungsdenken – Wissen über ihre gegenwärtigen Ausdrucksformen und Übertragungsmechanismen voraus. Auch wenn die in diesem Artikel beschriebenen

Dynamiken auf verschiedene Formen von Hass und Desinformation zutreffen, konzentrieren wir uns in erster Linie auf den Antisemitismus als paradigmatischen Fall, an dem sich umfassendere Muster digitalen Hasses und digitaler Radikalisierung besonders gut untersuchen lassen.

Antisemitismus ist deshalb ein besonders aufschlussreicher Fokus, weil er diskursiv außerordentlich wandlungsfähig ist und als verbindende Ideologie fungiert, die ansonsten unterschiedliche Formen von Hass und illiberalem Ressentiment miteinander verknüpft. Darüber hinaus wird das gegenwärtige Wiederaufleben des Antisemitismus als Teil einer „globalen antisemitischen Welle“ beschrieben, die sich online wie offline entfaltet und mit umfassenderen Krisen demokratischer Ordnung sowie dem weltweiten Aufstieg autoritärer Bewegungen verknüpft ist. 9 Diese Welle veranschaulicht, wie Hassideologien innerhalb günstiger politisch-kultureller Opportunitätsstrukturen operieren, die durch die digital umstrukturierte Öffentlichkeit hervorgebracht wurden.

Die Antisemitismusforschung hat zwei teils konkurrierende, teils komplementäre Dynamiken identifiziert, über die sich Hassrede in digitalen Umgebungen ausweitet. Erstens schaffen soziale Medien günstige Voraussetzungen für transgressive Dynamiken, indem sie die Grenzen des sozial Akzeptablen allmählich verschieben – wodurch ungefilterter Hass, Verschwörungsmythen und extremistische Narrative Raum und Legitimität gewinnen. Zweitens erweitern Hassideologien ihre Reichweite, indem sie sich im Laufe der Zeit transformieren und an soziale Codes der Akzeptabilität anpassen, sodass sie in camouflierten Formen zirkulieren, die plausible Abstreitbarkeit ermöglichen. 10 Beide Dynamiken – Grenzüberschreitung und Transformation – wirken gleichzeitig in unterschiedlichen Erscheinungsformen von Hassrede, von offenem Rassismus über codierten Antisemitismus bis hin zu verschwörungsideologisch geprägter antidemokratischer Rhetorik.

Wie Rob Williams von der USC Shoah Foundation in Bezug auf Antisemitismus betont: „We have to name it.“ Das Sammeln von Wissen über seine unterschiedlichen Erscheinungsformen sei „the first steps we must take in the fight against antisemitism“. 11 Doch die methodischen und analytischen Ansätze, die zur Erforschung des Antisemitismus entwickelt wurden, reichen weit darüber hinaus: Dieselben Techniken zur Entschlüsselung impliziter Vorurteile, zur Nachverfolgung algorithmischer Verstärkung und zur Identifikation von Radikalisierungspfaden sind auch für die Auseinandersetzung mit Rassismus, Misogynie, Verschwörungstheorien und antidemokratischen Desinformationskampagnen unverzichtbar. Wenn wir die gegenwärtig häufigsten Erscheinungsformen des Judenhasses verstehen und bekämpfen wollen, müssen wir unseren Blick auf das richten, was inzwischen mit Abstand zum wichtigsten Ort politischer Meinungsbildung und Debattenkultur geworden ist: soziale Medien.

Noch immer wissen wir viel zu wenig über den gegenwärtigen Stand des antisemitischen Diskurses in der Online-Welt. Das betrifft konventionellere Plattformen wie YouTube und Facebook ebenso wie Instagram, TikTok, Reddit und verschiedene Fringe-Plattformen wie

4chan. Diese digitalen Räume spielen eine zentrale Rolle bei der Prägung der politischen Perspektiven junger Menschen und bleiben doch weitgehend unerforscht – sie fungieren im Wesentlichen als Black Box.

Forschungsprojekte und zivilgesellschaftliche Initiativen, die versuchen, dieses Problem anzugehen, verdienen Anerkennung. Das Projekt „Decoding Antisemitism“¹² ist ein solches Beispiel. Zwar hat dieses Projekt versucht, die Vielzahl offener und codierter Formen von Antisemitismus im Netz freizulegen,¹³ doch war auch es durch technologische, administrative und finanzielle Hürden begrenzt. Folglich konzentrieren sich die meisten Initiativen auf einzelne Plattformen, bestimmte Erhebungszeiträume und spezifische Forschungsdesigns. Von einem umfassenden Verständnis des Online-Ökosystems, wie Sozialwissenschaften es für Phänomene der Offline-Welt beanspruchen, sind sie noch weit entfernt.

Neuere empirische Arbeiten haben begonnen, Ausmaß und Komplexität dieser Herausforderung zu dokumentieren. Die Analyse von YouTube-Kommentarspalten nach den Anschlägen vom 7. Oktober 2023 zeigte, dass antisemitischer Diskurs bei großen britischen Nachrichtenkanälen auf 36 bis 38 Prozent der Kommentare anstieg – nahezu doppelt so hoch wie der Vorkrisenwert von 15 bis 25 Prozent.¹⁴ Ebenso ergaben die Reaktionen auf die Schüsse im Museum in Washington, D.C. im Mai 2025 im Durchschnitt 43 Prozent antisemitische Inhalte über acht große englischsprachige Nachrichtenkanäle hinweg, wobei einzelne Kanäle sogar 66 Prozent erreichten.¹⁵ Diese Zahlen verweisen nicht bloß auf Einzelfälle, sondern auf eine systematische Normalisierung antisemitischer Äußerungen in digitalen Mainstream-Räumen – und damit auf umfassendere Muster gesellschaftlicher Polarisierung und Radikalisierung.

Die Vielzahl dieser Erscheinungsformen – Leugnung, Verharmlosung, Rechtfertigung oder Feier antisemitischer Gewalt sowie stereotype und verschwörungsideologische Narrative auf unterschiedlichen Plattformen und in verschiedenen Online-Umgebungen – ist bislang nicht hinreichend verstanden. Ein solches Verständnis ist jedoch eine grundlegende Voraussetzung für jede wirksame Gegenstrategie, ob sie nun von Bildungsakteuren oder aus der Politik ausgeht.

Das Spektrum antisemitischer Ausdrucksformen

Um Antisemitismus in digitalen Räumen zu verstehen, muss man seine diskursive Vielgestaltigkeit erkennen. Antisemitische Kommunikation bewegt sich auf einem Kontinuum von expliziter Hetze bis hin zu impliziten und codierten Ausdrucksformen, deren Interpretation von Kontextwissen und kulturellem Wissen abhängt. Dieses Spektrum existiert, weil Antisemitismus von dem lebt, was Forscher als „Kommunikationslatenz“ bezeichnet haben – dem Fortbestehen antisemitischer Bedeutung in codierten oder camouflierten Formen, die sozial verständlich bleiben und zugleich ihre Abstreitbarkeit bewahren.¹⁶ Solche latenten Formen arbeiten mit Ironie,

Aussagen und intertextuellen Signalen statt mit offen feindseligen Aussagen; dadurch wird antisemitische Bedeutung kontextabhängig und diskursiv vermittelt. Digitale Plattformen verstärken diese Dynamik durch das, was sich als „Politik der Grenzüberschreitung“ beschreiben lässt: Implizite Ausdrucksformen testen und erweitern durch Wiederholung und algorithmische Verstärkung die Grenzen des Akzeptablen, während sie zugleich plausible Abstreitbarkeit aufrechterhalten. 17

Diese Fragen werden umso dringlicher, wenn man Fälle physischer Gewalt einschließlich tödlicher Angriffe betrachtet. Historisch wurden Vorfälle wie Pittsburgh, Christchurch und Halle mit den Online-Umfeldern der Täter in Verbindung gebracht, in denen Hassrede, Verschwörungstheorien und Gewaltaufrufe kontinuierlich verbreitet wurden. Jüngst hat die Schießerei in Washington, D.C. im Mai 2025 – bei der der 30-jährige Elias Rodriguez im Capital Jewish Museum zwei jüdische Botschaftsmitarbeiter, Yaron Lischinsky und Sarah Milgrim, ermordete und dabei „free, free Palestine“ rief – erneut gezeigt, wie Online-Radikalisierungspfadwege der Offline-Gewalt vorausgehen. Rodriguez, der Berichten zufolge sagte, er habe „did it for Gaza“, war zuvor im radikalen antiisraelischen Aktivismus aktiv. Wenn Menschen kontinuierlich einem realitätsverzerrenden Diskurs ausgesetzt sind, steigt die Wahrscheinlichkeit erheblich, dass sich ihre Überzeugungen radikalieren. Das unterstreicht, welche zentrale Rolle Sprache und Diskurs bei der Prägung von Einstellungen spielen – Einstellungen, die sich letztlich in gewaltsamen Handlungen niederschlagen können.

Die multimodale Herausforderung

Zeitgenössischer Antisemitismus beschränkt sich längst nicht mehr auf Text. Er manifestiert sich zunehmend in multimodalen Assemblagen, in denen Bilder, Memes, Emojis und Videos ebenso viel semantisches Gewicht tragen wie Worte. Nach dem 7. Oktober fungierten Emojis wie Gleitschirme (🏂), palästinensische Flaggen (🇵🇸) und Wassermelonen als Chiffren für Solidarität mit der Hamas oder für die Billigung von Gewalt. Das Wassermelonen-Emoji (🍉) erfüllt dabei eine besonders komplexe semiotische Funktion: Es recontextualisiert die Farben der palästinensischen Flagge und impliziert zugleich, dass Solidaritätsbekundungen aufgrund vermeintlicher Zensur in codierter Form erfolgen müssten – eine Anspielung, die mit verschwörungsideologischen Vorstellungen jüdischer Kontrolle über den öffentlichen Diskurs korrespondiert. 18

Diese multimodalen Signale entziehen sich keywordbasierten Erkennungssystemen, bleiben für In-Group-Publika jedoch problemlos lesbar. Memes greifen populäre Formate auf und fügen antisemitische Analogien ein, die Juden oder Israel als betrügerisch, gierig oder manipulativ darstellen. Profilbilder mit Hakenkreuzen oder Hitler-Porträts signalisieren antisemitische Affinitäten, ohne dass es dafür überhaupt textlicher Inhalte bedarf. Diese Komplexität stellt computergestützte Systeme vor Herausforderungen höherer Ordnung:

Modelle müssen visuelle, textuelle und symbolische Hinweise aufeinander beziehen und zugleich kulturelle Literalität integrieren, um Bedeutung korrekt zu erfassen.

Die Analyse dieser Erscheinungsformen und der Zusammenhänge zwischen Auslösern, Diskursereignissen – sowohl Offline-Vorfällen als auch Online-Ereignissen mit dem Potenzial, Antisemitismus anzufachen – und den darauf folgenden Wellen von Verzerrung, Verschwörungsnarrativen und Desinformation kann uns in die Lage versetzen, die Motivationen der Nutzer zu rekonstruieren und damit auch Offline-Verhalten besser zu verstehen. Dadurch wird sichtbar, was auf bestimmte Publika anziehend wirkt und welche zugrunde liegenden Bedürfnisse und Ängste dabei eine Rolle spielen. Diese Art von Social-Media-Forschung ist die wirksamste – und wahrscheinlich die einzige – Möglichkeit, handlungsrelevante Einsichten in gesellschaftliche Stimmungen zu gewinnen.

Ein virologischer Ansatz: Fahrplan von der Forschung zum Handeln

Um Hassideologien im digitalen Zeitalter zu verstehen, müssen wir eine grundlegend neue Perspektive einnehmen. Wie Virologen, die Krankheitserreger untersuchen, müssen Forscher im Bereich Social Media und Hass erfassen, wie Antisemitismus, Rassismus, Misogynie, Verschwörungsdiskurs und antidemokratische Ressentiments in gegenwärtigen Kommunikationsnetzwerken existieren, sich verändern und verbreiten. So wie Virologen Impfstoffe entwickeln, indem sie virale Mechanismen auf molekularer Ebene verstehen, müssen wir die Struktur, die Übertragungsmuster und die Mutationsdynamik von Hassdiskursen entschlüsseln, um wirksame Gegenmittel zu entwickeln.

Neuere Forschung hat einen zentralen Übertragungsmechanismus offengelegt: Strategische Mehrdeutigkeit fungiert als Katalysator für das, was sich als „kaskadierende Radikalisierung“ über digitale Plattformen hinweg beschreiben lässt. 19 Dieser Prozess folgt einem Drei-Phasen-Modell, in dem mehrdeutiger Elitendiskurs zunächst von digitalen Intermediären umgerahmt wird und dann in der Publikumsbeteiligung in explizite Hassrede umkippt – eine Abfolge, die sich als „Ambiguität → Reframing → Kollaps“ zusammenfassen lässt. 20 Das Verständnis dieser Übertragungsdynamiken ist entscheidend für die Bekämpfung aller Formen von Hassrede, von Antisemitismus und Rassismus bis hin zu Misogynie und verschwörungsideologisch getriebener Desinformation.

Dieser virologische Ansatz verlangt nach ausgefeilten analytischen Instrumenten, die nur aus der systematischen Integration von Geisteswissenschaften, Sozialwissenschaften und Data Science hervorgehen können. Die Herausforderung, vor der wir stehen, ist nicht einfach interdisziplinär – sie erfordert eine transdisziplinäre Synthese, in der Informatik, quantitative Sozialforschung und qualitative interpretative Methoden in jeder Phase der Untersuchung zusammenwirken. Auch wenn sich dieser Artikel zur Veranschaulichung

methodischer Prinzipien vor allem auf die Antisemitismusforschung stützt, lässt sich das Rahmenmodell gleichermaßen auf andere Formen von Hassrede und Desinformation anwenden.

Schritt 1: Fortschrittliche Forschungsinfrastruktur – Mixed Methods und LLM-gestützte Methodologien

Die Grundlage wirksamer Intervention beginnt mit belastbaren Forschungskapazitäten, die die Grenzen traditioneller Disziplinen überschreiten. Die gegenwärtige Computational Social Science verbindet meist Data Science mit quantitativen Methoden, übersieht dabei jedoch die unverzichtbare Rolle qualitativer, interpretativer Forschung im gesamten Forschungsprozess – ein Versäumnis, das erhebliche blinde Flecken und Verzerrungen erzeugt. 21

Ein umfassender Ansatz integriert drei komplementäre Bereiche: **Informatik** – also technische Infrastruktur für groß angelegte Datenerhebung, algorithmische Verarbeitung und maschinelles Lernen; **quantitative Sozialwissenschaft** – also theoriegeleitete Modelle, statistische Modellierung und strenge Hypothesentests; und qualitative Sozialwissenschaft – also explorative Zugänge zum Verständnis neu entstehender Phänomene, zur Entwicklung differenzierter theoretischer Rahmen und zur Bereitstellung jenes Kontextwissens, das für sinnvolle Interpretation notwendig ist. Diese Integration führt zu einer Computational Social Science, die sowohl statistische Analyse als auch interpretative Tiefe ermöglicht – einer Wissenschaft also, die neue Phänomene explorieren und zugleich Hypothesen in großem Maßstab testen, komplexe kulturelle Bedeutungen interpretieren und Millionen von Datenpunkten verarbeiten sowie Evaluationskriterien etablieren kann, die tatsächliches Verstehen statt bloßes Pattern Matching widerspiegeln.

Neuere Entwicklungen zeigen die Notwendigkeit systematischer Integration über fünf Forschungsstufen hinweg: 21

(1) Konzeptdefinition und Operationalisierung übersetzen theoretische Konstrukte wie die IHRA-Arbeitsdefinition in sprachlich prüfbare Indikatoren, die Ko-Text, Kontext und multimodalen Ausdruck berücksichtigen. 22 **(2) Datenerhebung und Sampling** balancieren die Breite der Erfassung mit der Tiefe des Kontextverständnisses aus – das Projekt „Decoding Antisemitism“ hat über 300.000 Kommentare annotiert, doch gemessen an den Milliarden Beiträgen, die online zirkulieren, ist das weiterhin bescheiden. 23 **(3) Annotation und Modelltraining** bilden die entscheidende Schnittstelle, an der menschliche Interpretation das maschinelle Lernen anleitet. Expertenannotation integriert drei Wissensbereiche: Sprachwissen (linguistische Strukturen, pragmatische Bedeutungen), Ko-Text-Wissen (unmittelbarer Diskurskontext) und Weltwissen (historisches Verständnis antisemitischer Tropen und kultureller Codes). 24 Entscheidend ist dabei ein konservatives Zuschreibungsprinzip: Wenn eine Aussage mehrere

Interpretationen zulässt – eine vorurteilsbehaftete und eine nicht-vorurteilsbehaftete –, wird sie als nicht-vorurteilsbehaftet codiert, sofern nicht substanzielle Evidenz dafür spricht, dass die vorurteilsbehaftete Deutung wahrscheinlicher ist. Dieses Prinzip schützt vor Überinterpretation, nimmt aber zugleich in Kauf, dass die tatsächliche Verbreitung von Hass möglicherweise unterschätzt wird. 25

(4) Die computergestützte Analyse nutzt jüngste Fortschritte bei Large Language Models mittels Prompt Engineering und Chain-of-Thought-Reasoning. Aufkommende RAG-Architekturen (Retrieval-Augmented Generation) koppeln LLMs an externe Wissensbasen – annotierte Korpora, Lexika, Stereotyp-Taxonomien – und reduzieren Halluzinationen durch die Einbindung relevanter Evidenz. Perspektivisch könnten kontext-engineerte Architekturen Modelle mit konversationellem Ko-Text, historischem Ereigniskontext, Named-Entity-Resolution und situativen Metadaten versorgen und so menschlichem Expertenurteil näherkommen. 26 **(5) Modellevaluation und Validierung** stellen sicher, dass automatisierte Systeme tatsächlich sinnvolle Erkenntnisse liefern, statt bloß Verzerrungen zu reproduzieren. Vorläufige Experimente mit neueren LLMs – Gemini-2.5-Flash, Llama-3.3-70B, MoonshotAI Kimi-K2 – erreichten bei Antisemitismuserkennung F1-Werte von über 0,88. 27 Dennoch bestehen erhebliche Grenzen fort: feinjustierte BERT-Modelle übersehen ungefähr ein Drittel der Fälle, 28 Antisemitismus bleibt statistisch selten – bei nur 1 bis 2 Prozent vieler Datensätze –, und der Mangel an Daten wird durch Plattformbeschränkungen und die ressourcenintensive Expertenannotation zusätzlich verschärft. 29

Schritt 2: Phänomene und Trends verstehen

Ausgestattet mit diesen anspruchsvollen Mixed-Methods-Instrumenten können Forschende über bloße statische Beschreibung hinaus zu einem dynamischen Verständnis gelangen. In dieser Phase geht es darum zu erkennen, wie Hassnarrative – antisemitische, rassistische, misogyne, verschwörungsideologische und antidemokratische – in digitalen Netzwerken entstehen, sich entwickeln und verbreiten. Zu den zentralen Aufgaben gehören die Kartierung der Mutation von Hasstropen bei ihrer Anpassung an gegenwärtige Diskurse, die Nachverfolgung von Übertragungswegen über Plattformen und Communities hinweg, die Identifikation von Superspreader-Accounts und Verstärkungsmechanismen, die Erkennung von Frühwarnsignalen eskalierender Radikalisierung und gesellschaftlicher Polarisierung sowie das Verständnis der Wechselwirkungen zwischen Offline-Ereignissen und der Intensivierung von Online-Diskursen.

Jüngere Forschung hat ein systematisches Drei-Phasen-Muster identifiziert, über das sich Hassrede und Desinformation in digitalen Ökosystemen ausbreiten: 30

(1) Primärdiskurs bezeichnet Äußerungen von Eliten, die durch strategische Mehrdeutigkeit gekennzeichnet sind – eine Rhetorik, die mehrere Lesarten zulässt, unterschiedliche Zielgruppen anspricht und zugleich plausible Abstreitbarkeit wahrt. 31

(2) Sekundärdiskurs entsteht dann, wenn digitale Intermediäre – YouTuber, TikTok, Podcaster, Influencer – diese Botschaften umrahmen und weiterverarbeiten, dabei Mehrdeutigkeit typischerweise reduzieren, ihre Sprache zuspitzen und sie in explizitere ideologische Raster einfügen. Anders als klassische Journalisten operieren diese Akteure in Aufmerksamkeitsökonomien, die Provokation und emotionale Intensität belohnen, sodass sie die nachgelagerte Interpretation oft stärker prägen als die ursprüngliche Aussage selbst. **(3) Tertiärdiskurs** entfaltet sich in Kommentarspalten, in denen das Publikum durch partizipative Mitwirkung aktiv an der Bedeutungsproduktion beteiligt ist. Hier kollabiert die Mehrdeutigkeit: Implizite Bedeutungen werden explizit, politische Kritik geht in Verschwörungstheorie über, und extreme Positionen gewinnen ihre Legitimität eher durch soziale Bestätigung als durch autoritative Billigung.

Diese dreistufige Zirkulation ermöglicht, was Forscher als „kaskadierende Radikalisierung“ bezeichnen: mehrdeutige Elitenrhetorik wird durch semantische Kristallisierung – Formulierungen werden zu Bekenntnissen –, emotionale Intensivierung – abgewogene Rhetorik wird affektiv aufgeladen – und die Normalisierung des Extremen – schockierende Aussagen werden alltäglich – zunehmend expliziter und radikaler. 32 Hassideologien funktionieren als lebendige, adaptive Systeme, die auf Interventionen reagieren, Schwachstellen von Plattformen ausnutzen und fortlaufend neue Ausdrucksformen entwickeln, um sich durch orthografische Manipulation, zusammengesetzte Neologismen und rhetorische Indirektheit der Erkennung zu entziehen. 33 Die Zeit nach dem 7. Oktober 2023 veranschaulichte diese Dynamik in Form offener Gewaltfeier, während die Reaktionen auf die Schüsse im Museum in Washington im Mai 2025 stärker auf Verschwörungsnarrative, Leugnung und codierte Zurückweisung setzten – ein Beispiel dafür, wie unterschiedliche Auslöser unterschiedliche Muster innerhalb des breiteren Spektrums von Hasskommunikation aktivieren. 34

Das Verständnis dieser Dynamiken in Echtzeit liefert die Wissensgrundlage, die für wirksame Reaktionen auf alle Kategorien von Hassrede und Desinformation notwendig ist.

Schritt 3: Wissenstransfer in Politik, Recht und Bildung

Forschungserkenntnisse bleiben akademische Übungen, wenn sie nicht in handlungsfähige Rahmen für gesellschaftliche Akteure übersetzt werden. Diese entscheidende Phase umfasst mehrere Dimensionen des Wissenstransfers.

Politikentwicklung bedeutet, politischen Entscheidungsträgern evidenzbasierte Empfehlungen für Plattformregulierung, Standards der Inhaltsmoderation und gesetzliche Rahmenbedingungen an die Hand zu geben, die die Freiheit der Meinungsäußerung mit dem Schutz vor Aufstachelung ins Gleichgewicht bringen. Der Digital Services Act der Europäischen Union stellt einen ersten Versuch dar, mehr Transparenz bei Moderationsentscheidungen und algorithmischen Prozessen verbindlich zu machen; seine wirksame Umsetzung erfordert jedoch die Zusammenarbeit mit wissenschaftlichen Experten und den Zugang zu hochwertigen Datensätzen. 35

Rechtliche Anwendung bedeutet, Strafverfolgungsbehörden und Justizsysteme mit dem Wissen auszustatten, Hassverbrechen zu erkennen, zu untersuchen und zu verfolgen, und zugleich die Online-Radikalisierungspfade zu verstehen, die der Offline-Gewalt vorausgehen. Rechtliche Rahmen müssen dabei grenzüberschreitende Unterschiede berücksichtigen: Was in den Vereinigten Staaten als geschützte Rede gilt, kann nach EU- oder deutschem Recht bereits als Volksverhetzung, gruppenbezogene Verleumdung oder Hassrede qualifizieren.

Bildungsintervention bedeutet die Entwicklung von Curricula und pädagogischen Ansätzen, die auf einem realitätsnahen Verständnis beruhen, wie junge Menschen im Netz auf antisemitische Narrative stoßen und diese internalisieren. Dazu gehört auch die Entwicklung von Medienkompetenzprogrammen, die gezielt auf die rhetorischen Strategien und emotionalen Appelle eingehen, welche extremistische Inhalte attraktiv machen – also darauf, Kommunikationslatenz, codierte Sprache und multimodale Signale erkennen zu lernen. Solche Bildungsarbeit ist entscheidend, um Radikalisierung entgegenzuwirken und gesellschaftliche Polarisierung unter jüngeren Generationen zu verringern.

Plattformverantwortung bedeutet, Technologieunternehmen mit konkreten, evidenzbasierten Einsichten darüber zu konfrontieren, wie ihre algorithmischen Systeme schädliche Inhalte verstärken, und ihnen architektonische Änderungen zu empfehlen, die diese Verstärkung reduzieren würden, ohne ein unrealistisches Maß an Moderation zu verlangen. Derzeitige Moderationsansätze konzentrieren sich häufig auf Schimpfwörter oder direkte Drohungen und übersehen implizite oder multimodale Formen von Hassrede; zugleich bleibt ihre Durchsetzung über Sprachen und Regionen hinweg inkonsistent. 36

Zusätzliche Komplexität entsteht durch das, was als „context collapse“ bezeichnet worden ist – die Zusammenfaltung heterogener Publika zu einer einzigen Öffentlichkeit. 37 Aussagen, die für eine bestimmte Interpretationsgemeinschaft formuliert wurden, können von anderen vollkommen anders gelesen werden, wenn sie plattformübergreifend zirkulieren. Nutzer sozialer Medien betreiben Impression Management über kollabierte Kontexte hinweg, was ihre Abhängigkeit von mehrdeutigen Formulierungen verstärken kann, die mehrere Lesarten zulassen. Hinzu kommt, dass Algorithmen Verbreitungsmuster entscheidend prägen: Inhalte, die Engagement erzeugen – auch in Form von Kontroversen –, werden verstärkt, unabhängig davon, ob dieses Engagement auf korrekter Interpretation beruht. Das Verständnis dieser Dynamiken ist essenziell für wirksame Strategien der Inhaltsmoderation und Plattform-Governance, die nicht nur einzelne schädliche Posts adressieren, sondern die systemischen Verstärkungsmechanismen, die Hassrede erst viral werden lassen.

Eine bleibende Herausforderung ist dabei die Spannung zwischen Overblocking und Underblocking. Werden Inhalte zu aggressiv entfernt, besteht die Gefahr, dass legitime politische Kritik oder neutrale Bezüge auf jüdische Themen zensiert werden. Studien haben gezeigt, dass weit verbreitete Moderationsdienste Toxizitätswerte für Kommentare,

in denen Wörter wie „Jude“ oder „Israel“ vorkommen, systematisch überhöhen – unabhängig von der inhaltlichen Positionierung. 38 Umgekehrt drohen Normalisierung und emotionale Schädigung jüdischer Nutzer, wenn antisemitische Inhalte online bleiben. Automatisierte Moderationssysteme müssen daher zu einer kontextsensitiven Klassifikation fähig sein, die etwa zwischen „Israel begeht Genozid“ als umstrittener politischer Kritik und antisemitischer Holocaust-Inversion unterscheiden kann.

Schritt 4: Voraussetzungen für wirksame Prävention und Intervention schaffen

Das letztliche Ziel besteht nicht nur darin, Antisemitismus zu verstehen, sondern seine Ausbreitung zu verhindern und wirksam einzugreifen, wenn er auftritt. Das verlangt mehrere miteinander verbundene Fähigkeiten.

Frühwarnsysteme setzen die in Schritt 1 entwickelten computergestützten Werkzeuge ein, um Online-Umgebungen kontinuierlich zu beobachten und beunruhigende Trends zu erkennen, bevor sie sich in Offline-Schaden übersetzen. Was sich in der nächsten Pandemie, globalen Wirtschaftskrise oder klimabedingten Ressourcenknappheit entladen könnte, beginnt oft bereits vorher in den Tiefen des anonymen Netzes zu gären. Mit den richtigen Analyseinstrumenten lassen sich solche Frühwarnsignale erkennen.

Schnelle Reaktionsfähigkeit bedeutet, institutionelle Strukturen zu schaffen, die bei aufkommenden antisemitischen Desinformationskampagnen zügig Gegen narrative, Faktenchecks und strategische Kommunikation mobilisieren können. Dafür muss man nicht nur verstehen, welche Narrative zirkulieren, sondern auch, welche Superspreader-Accounts und Verstärkungsmechanismen ihre Verbreitung antreiben.

Gesellschaftliche Resilienz nutzt Forschungserkenntnisse, um die psychologischen und sozialen Ressourcen jener Gemeinschaften zu stärken, die Ziel von Hass werden, damit sie Manipulationstaktiken erkennen und ihren Zusammenhalt auch unter Druck bewahren können. Jüdische Gemeinschaften erwarten Schutz vor Online-Antisemitismus, doch schlecht konzipierte Interventionen bergen das Risiko erneuter Traumatisierung oder weiterer Marginalisierung.

Rehabilitationspfade entwickeln Interventionsstrategien für Individuen, die Anzeichen von Radikalisierung zeigen – auf der Grundlage eines Verständnisses dafür, was Menschen in extremistische Ideologien hineinzieht und was sie wieder herausführen könnte. Dazu gehört nicht nur die Erfassung expliziter Äußerungen, sondern auch der schleichende Radikalisierungsprozess, durch den sich verzerrte Vorstellungen zu festen Weltbildern verfestigen. Wirksame Deradikalisierungsprogramme müssen deshalb auch die gesellschaftliche Polarisierung adressieren, die den Nährboden für extremistische Rekrutierung bildet.

Dieser Fahrplan stellt eine umfassende Antwort auf Hassideologien dar, die der Komplexität der Bedrohung gerecht wird. So wie moderne Medizin auf Molekularbiologie angewiesen ist, um gezielte Therapien zu entwickeln, braucht die moderne Demokratie die Computational Social Science, um gezielte Interventionen gegen Hass in all seinen Formen zu entwickeln – Antisemitismus, Rassismus, Misogynie, Verschwörungsdenken und antidemokratische Desinformation. Diese Viren des Hasses haben in sozialen Medien neue Übertragungsvektoren gefunden; wir müssen zu den Virologen werden, die ihre Mechanismen entschlüsseln und die Gegenmittel entwickeln, die unsere Gesellschaften so dringend brauchen.

So sehr soziale Medien zu einem der bedeutendsten Destabilisierungsfaktoren unserer Zeit geworden sind, so sehr bergen sie zugleich eine enorme Chance. Im Laufe des vergangenen Jahrhunderts haben unsere Gesellschaften in unterschiedlichem Ausmaß versucht, den tief verwurzelten Hass zu überwinden, der Kulturen seit Jahrtausenden durchzieht. Gerade diese dauerhafte Herausforderung hat viele daran zweifeln lassen, ob sich eine solche Ideologie des Hasses jemals vollständig aus der Welt schaffen lässt.

Traditionell waren Umfragen unser wichtigstes Mittel, um die Verbreitung des Antisemitismus in der Gesellschaft zu erfassen. Doch im Vergleich zu den Einsichten, die sich aus einer gründlichen Analyse von Social-Media-Diskursen gewinnen lassen, verblassen sie. Interaktionen auf sozialen Medien erfassen freiwillige Äußerungen, Dialoge zwischen Tausenden von Menschen und das gesamte Spektrum menschlicher Interaktion. Online-Diskurse zu verstehen heißt, die Bildung von Einstellungen in unserer Gesellschaft zu verstehen – ein Grad an Einsicht, über den die Menschheit zuvor nie verfügt hat.

Die gründliche Identifikation dominanter Kommunikationsmuster im Netz geht über die bloße Dokumentation der Gegenwart hinaus; sie kann die Zukunft erhellen. Wer versteht, wie Trends in unserer Gesellschaft entstehen und auf welche Weise Hass sich manifestiert, über welche Kanäle er sich verbreitet und mit welchen Botschaftsstrategien er operiert, kann besser erkennen, wo künftig Schaden entstehen wird.

Wenn wir begreifen wollen, wohin sich unsere Gesellschaft bewegt – sowohl in Bezug auf Online-Trends als auch auf ihre Offline-Folgen –, dann ist eine rigorose interdisziplinäre Forschung zu Antisemitismus, Rassismus, Misogynie, Verschwörungsdenken und anderen hassbasierten Ideologien, die unsere freien Gesellschaften, demokratischen politischen Systeme und den sozialen Zusammenhalt bedrohen, unabdingbar. Diese Forschung muss alle verfügbaren methodischen Zugänge nutzen. Sozialwissenschaftler und Data-Science-Experten müssen in einen fortlaufenden Dialog treten und voneinander lernen. Weder politische Interventionen noch Plattformregulierung allein werden diese Gefahren für die Demokratie lösen; ebenso wenig kann künstliche Intelligenz dies isoliert leisten. Nur durch gemeinsame Anstrengungen können wir in Echtzeit verstehen, wie sich Online-Debatten entwickeln, welche Resonanz und Verbreitung bestimmte extremistische Narrative und Desinformationskampagnen entfalten und welche Personen oder Gruppen die Deutung zentraler Fragen beeinflussen.

Dieser Text ist daher nicht nur eine Bestandsaufnahme hassbezogener Social-Media-Forschung, sondern auch ein Appell. Als Gesellschaft müssen wir anerkennen, dass die traditionellen Methoden zur Sicherung des sozialen Friedens, wie sie in politischen, rechtlichen, bildungsbezogenen und sicherheitspolitischen Institutionen verankert sind, für die Navigation in diesem neuen digitalen Raum nicht ausreichen. Wir dürfen uns nicht damit begnügen, diese Dynamiken mit ein paar Scheinwerfern punktuell auszuleuchten; wir müssen das Phänomen in seiner Gesamtheit verstehen, um sowohl die Symptome als auch ihre tieferen Ursachen und möglichen Folgen zu erkennen. Nur so wird es möglich sein, einen wirklich informierten Dialog mit jenen Teilen der Gesellschaft zu führen, die zunehmend abdriften.

Dafür brauchen wir belastbare Forschung und fruchtbare Kooperationen zwischen Wissenschaft, Politik und Zivilgesellschaft – und die Entwicklung von Methoden, die für alle Formen von Hassrede funktionieren, von Antisemitismus und Rassismus bis hin zu Misogynie und antidemokratischer Desinformation. Nur wenn wir die Diskurse entschlüsseln, die unser digitales Zeitalter prägen, können wir beginnen, die moralische Grammatik der demokratischen Gesellschaft wiederaufzubauen.

Fußnoten

1. Lars Rensmann, "Politischer Antisemitismus im postfaktischen Zeitalter: Formen und Ursachen in Demokratien des 21. Jahrhunderts" (Baden-Baden: Nomos, 2025); Heiko Beyer et al., "Antisemitismus in der Gesamtgesellschaft von Nordrhein-Westfalen im Jahr 2024," Antisemitismusbeauftragte des Landes Nordrhein-Westfalen (2024). ↩
2. Gabriele Cosentino, "Social Media and the Post-Truth World Order: The Global Dynamics of Disinformation" (Cham: Palgrave Macmillan, 2020). ↩
3. Anna Lührmann and Staffan I. Lindberg, "A Third Wave of Autocratization is Here: What is New About it?" *Democratization* 26, no. 7 (2019): 1095–1113. ↩
4. Institute for Strategic Dialogue, "Information Laundromat Detects Banned Russian Propaganda Across Hundreds of Websites," accessed October 2024, <https://www.isdglobal.org/isd-in-the-news/information-laundromat-detects-banned-russian-propaganda-across-hundreds-of-websites>; Institute for Strategic Dialogue, "Russia-Ukraine War," accessed October 2024, <https://www.isdglobal.org/tag/russia-ukraine-war>; Institute for Strategic Dialogue, "China, Russia and Iran Are Exploiting the Israel-Hamas Conflict for Their Advantage," accessed October 2024, <https://www.isdglobal.org/isd-in-the-news/china-russia-and-iran-are-exploiting-the-israel-hamas-conflict-for-their-advantage>; see also Deen Freelon and Chris Wells, "Disinformation as Political Communication," *Frontiers in Political Science* 4 (2022), <https://doi.org/10.3389/fpos.2022.885362>. ↩
5. "Facebook Whistleblower Frances Haugen Testified That the Company's Algorithms Are Dangerous—Here's How They Can Manipulate You," *The Conversation*, October 5, 2021, <https://theconversation.com/facebook-whistleblower-frances-haugen-testified->

that-the-companys-algorithms-are-dangerous-heres-how-they-can-manipulate-you-169420. ↩

6. Ibid. ↩
7. Lars Rensmann, "The Noisy Counter-Revolution: Understanding the Cultural Conditions and Dynamics of Populist Politics in Europe in the Digital Age," *Politics and Governance* 5, no. 4 (2017): 123–135. ↩
8. Cal Newport, "The Real Problem with Twitter," accessed October 2024, <https://calnewport.com/the-real-problem-with-twitter>. ↩
9. Matthias J. Becker and Lars Rensmann, "The Dynamics of Digital Antisemitism: Anti-Jewish Narratives in the Aftermath of October 7th," in *Antisemitism Online: An Ancient Hatred in the Modern World*, ed. Todd Pittinsky (Oxford: Oxford University Press, in press). ↩
10. Ibid. ↩
11. USC Shoah Foundation, "Facing Antisemitism," accessed October 2024, <https://sfi.usc.edu/video/facing-antisemitism>. ↩
12. Decoding Antisemitism, accessed October 2024, <https://decoding-antisemitism.eu>. ↩
13. Matthias J. Becker et al., eds., *Decoding Antisemitism: An AI-Driven Study on Hate Speech and Imagery Online* (Cham: Springer, 2024), <https://doi.org/10.1007/978-3-031-49237-2>. ↩
14. Matthias J. Becker et al., *Celebrating Terror: Antisemitism Online After the Hamas Attacks on Israel. Preliminary Results I* (Berlin: Technische Universität Berlin, Center for Research on Antisemitism, 2023), <https://doi.org/10.14279/depositonce-19143>. ↩
15. Matthias J. Becker, Jordan Blatter, and Oksana Stanevich, "Decoding Antisemitism Online: Linguistic and Multimodal Challenges in the Age of AI" (working paper, 2025). ↩
16. Werner Bergmann and Rainer Erb, "Kommunikationslatenz, Moral und öffentliche Meinung. Theoretische Überlegungen zum Antisemitismus in der Bundesrepublik Deutschland," *Kölner Zeitschrift für Soziologie und Sozialpsychologie* 38 (1986): 223–246. ↩
17. Becker and Rensmann, "The Dynamics of Digital Antisemitism." ↩
18. Becker, Blatter, and Stanevich, "Decoding Antisemitism Online." ↩
19. Parker Bach, Carolyn E. Schmitt, and Shannon C. McGregor, "Let Me Be Perfectly Unclear: Strategic Ambiguity in Political Communication," *Communication Theory* 35, no. 2 (2025): 96–106; Matthias J. Becker, Gabrielle Beacken, and Liora Sabra, "How Political Ambiguity Transforms Across Digital Platforms: Mechanisms of Discourse Radicalization" (working paper, 2025). ↩
20. Ibid. ↩
21. Katharina Soemer, Daniela Gurschow, and Steffen Egert, "Social Science and AI Joining Forces: Towards New Approaches for Computational Social Science" (paper presented at the interdisciplinary conference on computational methods, 2025). ↩ ↩2
22. International Holocaust Remembrance Alliance, "Working Definition of Antisemitism" (2016), <https://holocaustremembrance.com/resources/working-definition-antisemitism>; Becker et al., *Decoding Antisemitism: An AI-Driven Study*. ↩

23. Becker, Blatter, and Stanevich, "Decoding Antisemitism Online." ↩
24. Ibid. ↩
25. Matthias J. Becker and Hagen Troschke, "Decoding Implicit Hate Speech: The Example of Antisemitism," in *Challenges and Perspectives of Hate Speech Research*, ed. C. Strippel et al., *Digital Communication Research* 12 (2023): 335–352, <https://doi.org/10.48541/dcr.v12.20>. ↩
26. Becker, Blatter, and Stanevich, "Decoding Antisemitism Online." ↩
27. Jordan Blatter and Blue Square Alliance Against Hate, "Testing Next-Generation AI to Better Detect Antisemitism," Blue Square Alliance Command Center Insights (2025), <https://www.bluesquarealliance.org/command-center-insights/testing-next-generation-ai-to-better-detect-antisemitism>. ↩
28. Milena Pustet and Helena Mihaljević, "Automated Detection of Antisemitic Texts: Is Context All We Need?" in *Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online. Discourse Report 6*, ed. Matthias J. Becker et al. (Berlin: Technische Universität Berlin, Center for Research on Antisemitism, 2024), <https://decoding-antisemitism.eu/publications/sixth-discourse-report>. ↩
29. Becker, Blatter, and Stanevich, "Decoding Antisemitism Online"; Eva Steffen, Milena Pustet, and Helena Mihaljević, "Algorithms Against Antisemitism? Towards the Automated Detection of Antisemitic Content Online," in *Antisemitism in Online Communication: Transdisciplinary Approaches to Hate Speech in the Twenty-First Century*, ed. Matthias J. Becker et al. (London: Open Book Publishers, 2024), <https://doi.org/10.11647/OBP.0406.08>. ↩
30. Becker, Beacken, and Sabra, "How Political Ambiguity Transforms Across Digital Platforms." ↩
31. Bach, Schmitt, and McGregor, "Let Me Be Perfectly Unclear." ↩
32. Becker, Blatter, and Stanevich, "Decoding Antisemitism Online." ↩
33. Ibid. ↩
34. Becker and Rensmann, "The Dynamics of Digital Antisemitism." ↩
35. Jash Patel, Heer Mehta, and Jeremy Blackburn, "Evaluating Large Language Models for Detecting Antisemitism," arXiv preprint arXiv:2509.18293 (2025), accepted to EMNLP 2025 Main Conference. ↩
36. Ibid. ↩
37. danah boyd, "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications," in *A Networked Self: Identity, Community, and Culture on Social Network Sites*, ed. Zizi Papacharissi (New York: Routledge, 2010), 39–58. ↩
38. Alexis Chapelan et al., *Decoding Antisemitism: An AI-driven Study on Hate Speech and Imagery Online. Discourse Report 5* (Berlin: Technische Universität Berlin, Center for Research on Antisemitism, 2023), <https://doi.org/10.14279/depositonce-17105>; Lucas Dixon et al., "Measuring and Mitigating Unintended Bias in Text Classification," in *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society (AIES)* (2018), 67–73. ↩