

# Misogynie als Vektor der Radikalisierung:

## Empfohlene Maßnahmen für Plattformen zur Erkennung und Eindämmung von geschlechts- spezifischer Online-Gewalt

Sara Bundtzen

## Über das Digital Policy Lab

Als zwischenstaatliche Arbeitsgruppe engagiert sich das Digital Policy Lab (DPL) dafür, politische Lösungen zur Verhinderung und Bekämpfung der Verbreitung von Desinformation, Hassrede sowie extremistischen und terroristischen Inhalten im Internet aufzuzeigen. Die Arbeitsgruppe besteht aus Vertreter:innen der zuständigen Ministerien und Aufsichtsbehörden ausgewählter liberal-demokratischer Länder. Die Arbeit des DPL zielt darauf ab, den regierungsübergreifenden Dialog zu fördern, politischen Entscheidungsträger:innen und Aufsichtsbehörden Zugang zu einschlägigem Fachwissen und Forschungsergebnissen zu verschaffen sowie eine internationale Arbeitsgemeinschaft zur Bewältigung der wichtigsten digitalpolitischen Herausforderungen aufzubauen. Wir danken dem Auswärtigen Amt für die Unterstützung des Projekts.

## Über diesen Bericht

Im Rahmen des DPL organisierte das ISD zwischen Mai und Juni 2023 mehrere Arbeitsgruppentreffen zum Thema geschlechtsspezifischer Online-Gewalt. Die Arbeitsgruppe bestand aus Vertreter:innen nationaler Ministerien und Aufsichtsbehörden aus Australien, Deutschland, Großbritannien, Kanada, Neuseeland und den USA. Dieser Bericht baut auf den Diskussionen auf und vertieft diese. Mitglieder der Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership) und der Multi-Stakeholder-Gemeinschaft des Christchurch Call trugen ebenfalls zu diesem Bericht bei. Zu der Arbeitsgruppe gehörten zudem Vertreter:innen der Zivilgesellschaft und der Wissenschaft. Auch wenn die Teilnehmer:innen an diesem Bericht mitgewirkt haben, spiegeln die darin geäußerten Ansichten nicht unbedingt die Ansichten aller Teilnehmer:innen oder der an diesem Projekt beteiligten Regierungen wider.



Copyright © Institute for Strategic Dialogue (2023).  
Das Institute for Strategic Dialogue (gGmbH) ist beim  
Amtsgericht Berlin-Charlottenburg registriert (HRB 207 328B).  
Die Geschäftsführerin ist Huberta von Voss. Die Anschrift lautet:  
Postfach 80647, 10006 Berlin. Alle Rechte vorbehalten.

[www.isdgermany.org](http://www.isdgermany.org)

## Über die Autorin

**Sara Bundtzen** ist Analytistin beim ISD Germany. Sie erforscht die Verbreitung von Desinformation und Informationsmanipulation durch staatliche und nichtstaatliche Akteur:innen im deutschen und englischen Sprachraum. Im Rahmen des DPL unterstützt sie die digitalpolitische Arbeit des ISD und untersucht Vorschläge zur Bekämpfung von Desinformation, ausländischer Einflussnahme, Hassrede und extremistischen Inhalten. Zuvor arbeitete Sara Bundtzen im Bundesministerium der Verteidigung (BMVg) in Berlin und im NATO-Hauptquartier in Brüssel.

## Mitwirkung

Helena Schwertheim ist Senior Digital Policy & Research Manager beim ISD Germany, wo sie das DPL leitet. Zuvor leitete sie die digitalpolitische Arbeit und Forschungsprojekte bei Democracy Reporting International. Sie verfügt über Erfahrungen in der Risiko- und Politikanalyse bei internationalen Organisationen und Denkfabriken, unter anderem beim UN-Welternährungsprogramm in Rom und bei International IDEA in Stockholm.

## Herausgeberische Verantwortung:

Huberta von Voss, Executive Director, ISD Germany & Henry Tuck, Head of Digital Policy, ISD

## Danksagungen

Wir bedanken uns bei allen Teilnehmer:innen der Arbeitsgruppe, einschließlich der Expert:innen aus Wissenschaft und Zivilgesellschaft, für ihre Beiträge. Besonderer Dank gilt allen Mitwirkenden an diesem Bericht für ihre wertvollen Einblicke, ihre Kritik und ihr Feedback, darunter: Elisabeth Brown (Christchurch Call Unit, Department of the Prime Minister and Cabinet/DPMC, Neuseeland), Cailin Crockett (National Security Council & Gender Policy Council, The White House, USA), Stella Ivory (Christchurch Call Unit, DPMC, Neuseeland), Clara Martiny (ISD), David Reid (Christchurch Call Unit, DPMC, Neuseeland), Sara Soleymani (Ministry for Women, Neuseeland), Caroline Sinders (Convocation Research + Design) und Professorin Lorna Woods (Universität Essex, Großbritannien). Wir danken zudem den ISD-Kolleg:innen Milo Comerford, Cooper Gatewood, Melanie Smith, Dr. Tim Squirrel und Henry Tuck für die Durchsicht dieses Berichts.

---

# Inhaltsverzeichnis

<b>Executive Summary</b>	<b>4</b>
<b>Glossar</b>	<b>9</b>
<b>Einleitung</b>	<b>11</b>
<b>Phänomene der geschlechtsspezifischen Online-Gewalt (OGBV)</b>	<b>12</b>
Misogyne Gruppen und (Sub-)Kulturen: OGBV als Vektor für Radikalisierung	12
Taktiken und Verhaltensweisen: OGBV als ein Kontinuum von Gewalt	14
Schäden: Auswirkungen von OGBV auf Individuen und Gesellschaft	16
<b>Bewertung von Community-Richtlinien und deren Durchsetzung</b>	<b>18</b>
<b>Plattformdesign und -systeme: Risiken der Reproduktion und Verstärkung von OGBV</b>	<b>20</b>
<b>Empfohlene Maßnahmen: Risikobewertung und Eindämmung von OGBV</b>	<b>22</b>
Ermöglichung des API-Zugangs zu Plattformdaten und Entwicklung standardisierter Transparenzberichte	22
Anwendung eines traumainformierten »Safety and Privacy by Design«-Ansatzes	23
Verstärkung der plattformübergreifenden Zusammenarbeit und des Informationsaustauschs	25
Überprüfung und Aktualisierung von Richtlinien, Verfahren und Systemen zur Inhaltsmoderation	26
Risikobewertung und -minderung von Misogynie in KI-basierten Systemen	26
<b>Fazit</b>	<b>28</b>
<b>Endnoten</b>	<b>29</b>

---

## Executive Summary

In diesem Bericht wird geschlechtsspezifische Online-Gewalt (*online gender-based violence, OGBV*) als Teil eines Kontinuums von (On- und Offline-) Gewalt untersucht, wobei die Verbindung zu extremistischen Ideologien, einschließlich der Verbreitung von terroristischen und gewalttätigen extremistischen Inhalten (*terrorist and violent extremist content, TVEC*), in den Fokus genommen wird. Ziel des Berichts ist, die Geschlechterperspektive bei der Prävention und Intervention in der Verbreitung von TVEC zu priorisieren. So sollten Social-Media-Plattformen Misogynie als einen Vektor für Radikalisierung, die oftmals online beginnt (oder verstärkt wird), erkennen und diese eindämmen.

In Anbetracht des Umfangs dieser Untersuchung, die sich auf Plattformen und nicht auf breitere Formen digitaler Technologien konzentriert, wird im vorliegenden Bericht der Begriff OGBV anstelle von technologiegestützter geschlechtsspezifischer Gewalt (*technology-facilitated gender-based violence, TFGBV*) verwendet. Die Diskussion macht deutlich, dass die Ursachen von geschlechtsspezifischem Hass, misogynen Gewalt und anderen sich überschneidenden Diskriminierungsformen eine breitere gesellschaftliche Herausforderung widerspiegeln, die nicht allein durch Plattformen angegangen oder gelöst werden kann. Somit wird anerkannt, dass die Eindämmung von OGBV einen gesamtgesellschaftlichen und gesamtstaatlichen Ansatz erfordert. Zwar bestehen Maßnahmen, die Regierungen und die Zivilgesellschaft ergreifen können und sollten, wie beispielsweise die Durchsetzung neuer regulatorischer Rahmenbedingungen und freiwilliger Verpflichtungen der Unternehmen, doch der vorliegende Bericht und seine Empfehlungen betonen dabei vor allem die verantwortliche Rolle von Plattformen.

In diesem Bericht werden die Auswirkungen von OGBV auf der Mikro- (Individuen) und Makroebene (Gesellschaft) dargelegt. Es wird untersucht, inwiefern OGBV ein Vektor für Radikalisierung sein kann, und durch Misogynie motiviert ist, die wiederum terroristische und gewalttätige extremistische Ideologien durchdringt. Der Bericht konzentriert sich auf die Rolle, die Plattformen bei der Verschärfung der Risiken von

OGBV spielen können, und bewertet die Community-Richtlinien der Plattformen, die Praktiken der Inhaltsmoderation, die Gestaltung der Nutzungsoberfläche und die algorithmischen Empfehlungssysteme.

Die Diskussion betrachtet OGBV als jede Form von Gewalt, einschließlich entmenschlichender Sprache, die sich gegen Personen aufgrund ihrer Geschlechtsidentität oder ihrem Geschlechtsausdruck richtet, wobei intersektionale Diskriminierung aufgrund einer Kombination oder Akkumulation struktureller Diskriminierungsformen wie Rassismus, Antisemitismus, Islamophobie, Homophobie, Transphobie und Ableismus das Risiko erhöht, Ziel von OGBV zu werden. In dem Bericht wird anerkannt, dass Frauen und LGBTQ+-Personen in unverhältnismäßig hohem Maße von OGBV betroffen sind. Zudem wird OGBV im Kontext von strukturellen, patriarchalischen Geschlechternormen untersucht, wobei Misogynie als ideologisches Bindeglied in einem Kontinuum von Gewalt (*continuum of violence*) und als Vektor für unterschiedliche extremistische Ideologien fungieren kann.

In diesem Zusammenhang betont der Bericht, dass die Erforschung und Eindämmung der Risiken von OGBV eine frühere Warnung vor und ein Eingreifen in misogynen Wegbereiter zu verschiedenen Formen des gewalttätigen Extremismus ermöglichen kann. Der vorliegende Bericht und seine Empfehlungen unterstreichen zudem, dass jegliche Risikominderung mit der Wahrung der Grundrechte der Nutzer:innen, einschließlich ihres Rechts auf Privatsphäre und freie Meinungsäußerung und Informationsfreiheit, einhergehen muss.

### Empfehlungen

#### **Ermöglichung des API-Zugangs zu öffentlich verfügbaren Daten für im öffentlichen Interesse liegende Forschung:**

- Die systematische Sammlung öffentlich zugänglicher Daten über den Zugang zu Application Programming Interfaces (APIs) kann zu einer Ergänzung digitaler ethnografischer (und anderer) Methoden beitragen, indem sie Datenlücken zum Zweck der im öffentlichen Interesse liegenden Forschung schließt.

- Plattformen sollten den Zugang zu kontinuierlichen Echtzeit- oder echtzeitnahen und durchsuchbaren APIs ermöglichen, damit zugelassene Forscher:innen die Taktiken und Formen von OGBV sowie die Verbindung zwischen Misogynie, Radikalisierung und gewalttätigem Extremismus untersuchen können. Ein solcher Zugang könnte zum Beispiel Verlaufs- oder Longitudinalstudien zu geschlechtsspezifischen Normen und Verhaltensweisen einer »Eigengruppe« im Laufe der Zeit, über verschiedene extremistische Ideologien und über verschiedene Plattformen hinweg unterstützen.
  - Die Zulassung von Forscher:innen sollte so erfolgen, dass interdisziplinäre Forschung möglich ist, die eine Reihe von Disziplinen wie Computerlinguistik, Kritische Terrorismusforschung und Kritische Männer- und Männlichkeitsforschung einbezieht sowie den Vorteil der vergleichenden Forschung über verschiedene Ideologien, Länder und Sprachen hinweg anerkennt.
  - Der API-Zugang sollte zwar in einer gewissen Form zulassungsbeschränkt sein, um eine schädliche oder kommerzielle Nutzung zu verhindern, doch sollte der Zugang für Forscher:innen kostenlos oder zu einem geringen Kostenbeitrag möglich sein. Bei höheren Kosten besteht die Gefahr, dass der Zugang de facto nicht möglich ist oder dass Forscher:innen mit geringeren finanziellen Möglichkeiten benachteiligt werden.
- sollten die Plattformen bei der Berichterstattung über Hassrede auf ihren Diensten Daten darüber einbeziehen, ob ein Verstoß auf der Grundlage des Geschlechts und anderer geschützter Merkmale erfolgte, um eine intersektionale Analyse von OGBV zu ermöglichen.
- Plattformen sollten auf eine Standardisierung der Berichterstattung hinarbeiten, indem sie nach Möglichkeit eine Reihe gemeinsamer Metriken und Inhaltskategorien entwickeln, um einen Vergleich und eine Nachverfolgung von Richtlinienverstößen auf verschiedenen Plattformen zu ermöglichen. So könnte beispielsweise eine standardisierte Berichterstattung den Anteil der bildbasierten gegenüber den textbasierten Inhalten offenlegen, die gegen Hassrede-Richtlinien verstoßen.
  - Eine plattformübergreifende Bemühung zur Standardisierung könnte mit der Arbeit von UN Women zur Entwicklung eines statistischen Rahmens für TFGBV und mit den Ansätzen der UN-Sonderberichterstatterin für Meinungsfreiheit zur Entwicklung einer gemeinsamen Definition für geschlechtsspezifische Desinformation (*gendered disinformation*) koordiniert und abgestimmt werden.
  - Plattformen sollten bei der Entwicklung der Methodiken von Transparenzberichten und internen Untersuchungen (beispielsweise bei der Durchführung von Umfragen) mit Feminismus-Forscher:innen, Frauenrechtler:innen, GBV-Expert:innen sowie Opfern und Überlebenden von OGBV zusammenarbeiten. Zumindest sollten die Plattformen die Methodik ihrer Berichterstattung (und etwaige Änderungen daran) transparent machen.

### **Entwicklung von nach Geschlecht aufgeschlüsselten und standardisierten Transparenzberichten:**

- Die Transparenzberichterstattung der Plattformen sollte es Forscher:innen ermöglichen, den Umfang und das Ausmaß von OGBV und die Durchsetzung von Community-Richtlinien im Laufe der Zeit zu verfolgen und zu bewerten.
- Plattformen sollten Transparenzberichte erstellen, die nach Geschlecht aufgeschlüsselte Daten enthalten, die sich auf Statistiken zu Verstößen gegen die Community-Richtlinien beziehen. Beispielsweise

### **Anwendung eines traumainformierten »Safety and Privacy by Design«-Ansatzes:**

- Bei der Entwicklung von Nutzungsoberflächen sollte in allen Phasen des Designs eine traumainformierte Perspektive eingenommen werden, die die Erfahrungen von Opfern und Überlebenden von geschlechtsspezifischer Gewalt in den Mittelpunkt stellt.

- Plattformen sollten proaktive Maßnahmen ergreifen, um die Handlungskompetenz der Nutzer:innen (*user agency*) mithilfe von Einstellungen zu unterstützen, die ihre Privatsphäre schützen und die Gefährdung durch OGBV verringern; reaktive Maßnahmen, die eine effiziente Meldung von Verstößen durch die Nutzer:innen ermöglichen (nach Möglichkeit auch plattformübergreifend); und Maßnahmen zur Rechenschaftspflicht, die Täter in angemessenem Maße abschrecken und bestrafen.
- Datenschutz und Sicherheit sollten nicht nur über zugängliche und transparente Einstellungen, sondern auch in den Richtlinien fest verankert werden, um die Verwendung personenbezogener Daten für OGBV einzudämmen (beispielsweise, um Doxing oder die Weitergabe von intimen Bildern ohne Einwilligung zu verhindern).
- Von der Industrie entwickelte Werkzeuge zur Inhaltsmoderation wie beispielsweise die Perspective API von Google sollten unter Berücksichtigung einer traumainformierten und geschlechtsspezifischen Perspektive kontinuierlich getestet und geprüft werden. Diese Bemühungen sollten Teil eines sektorübergreifenden Dialogs sein, an dem alle betroffenen Stakeholder beteiligt sind.

#### **Verbesserung der plattformübergreifenden Zusammenarbeit und des Informationsaustauschs über OGBV-Vorfälle:**

- Plattformen sollten Kanäle des Informationsaustausches (*exchange channels*) zwischen den zuständigen Teams, einschließlich der Moderationsteams, einrichten, um proaktiv Informationen über OGBV-Vorfälle, einschließlich plattformübergreifender Belästigung (beispielsweise Informationen über Täter, die plattformübergreifende Konten nutzen), auszutauschen. Dies ist wichtig, um den Umfang und das Ausmaß von OGBV zu verstehen, aber auch, um gegebenenfalls notwendige plattformübergreifende Abhilfemaßnahmen zu koordinieren.
- Plattformen sollten, wo sinnvoll, Informationen über Meldungen von Nutzer:innen austauschen und nach Möglichkeit interoperable Meldemechanismen entwickeln, um die Handlungskompetenz der Nutzer:innen zu stärken.
- Bestehende plattformübergreifende Koordinierungsstellen wie das Christchurch Call Crisis Response Protocol und das Global Internet Forum to Counter Terrorism sollten prüfen, inwieweit OGBV für ihre Mandate relevant ist, und ihren Anwendungsbereich und ihre Mechanismen entsprechend anpassen.
- Ein plattformübergreifender Wissensaustausch sollte auf die bestehenden Prozesse zur Inhaltsmoderation aufbauen und diese verbessern, unter anderem durch regelmäßige Bewertungen und Berichte über die Effizienz und Wirksamkeit dieser Werkzeuge.

#### **Überprüfung und Anpassung der Richtlinien, Prozesse und Systeme für die Inhaltsmoderation:**

- Die Inhaltsmoderation sollte das Kontinuum der Gewalt berücksichtigen und Misogynie als möglichen Ausgangspunkt und Frühwarnzeichen für verschiedene gewalttätige extremistische Ideologien anerkennen.
- Plattformen sollten ihre Community-Richtlinien im Zusammenhang mit Hassrede und gewalttätigen extremistischen und terroristischen Inhalten überprüfen und anpassen, um zu erkennen, wie misogynen Einstellungen oftmals ideologische Wege zum Extremismus durchdringen. Dazu gehören auch Überlegungen darüber, inwiefern Misogynie eine Ideologie sein kann, die gewalttätigen Extremismus begünstigt und Gewalt gegen Frauen und LGBTQ+-Personen rechtfertigt.
- Community-Richtlinien sollten mehrdimensionale Perspektiven, darunter die Erfahrungen von Opfern und Überlebenden von OGBV, einbeziehen. Richtlinien sollten weitere geschützte Merkmale, wie etwa ethnische Herkunft und Religion, einbeziehen, da geschlechtsspezifische Gewalt gegen rassifizierte Gemeinschaften oft durch rassistische und entmenslichende Sprache zum Ausdruck gebracht wird.

- Inhaltsmoderation sollte subtile und verschlüsselte misogynie Inhalte, einschließlich kontextbezogener bildbasierter Inhalte, sowie den vielsprachigen, kulturübergreifenden Kontext von Online-Räumen berücksichtigen. Dies kann durch die Entwicklung von Lexika von Wörtern und Ausdrücken in Zusammenarbeit mit lokalen Organisationen geschehen. Diese Bemühungen sollten auf Grundlage der Traumaforschung erfolgen.
- Der Einsatz von Systemen, die auf künstlicher Intelligenz (KI) basieren, um misogynie Inhalte zu erkennen und zu moderieren, muss durch menschliche Aufsicht ergänzt werden, um nuancierte Ansätze zu ermöglichen. Diese sollten subtile und verschlüsselte Misogynie erkennen und gleichzeitig falschpositive Meldungen verhindern.

#### **Anwendung von intersektionalen feministischen Ansätzen bei der Risikobewertung von KI-basierten Systemen:**

- Plattformen sollten bei der Bewertung der Risiken von Algorithmen und Modellen des maschinellen Lernens (*machine learning, ML*), die in ihre Dienste eingebettet sind, Gender-Analysen und feministische Methoden einbeziehen. Dies hilft zu verstehen, wie strukturelle Ungleichheiten und patriarchalische Geschlechternormen durch KI-basierte Systeme reproduziert und verstärkt werden können.
- Plattformen sollten dabei ihre Empfehlungsrichtlinien (beispielsweise Richtlinien für die in Feeds herabgestuften Inhalte) in Abstimmung mit ihren Community-Richtlinien überprüfen und anpassen.
- Plattformen sollten bei der Ideenfindung, der Konzeption, der Entwicklung, dem Testen und der Skalierung neuer Funktionen oder bei der Änderung bereits bestehender Funktionen Designprozesse anwenden, bei denen die mehrdimensionalen Diskriminierungserfahrungen von Opfern und Überlebenden von OGBV einbezogen werden.
- Plattformen sollten sicherstellen, dass die zuständigen Teams (beispielsweise diejenigen, die Algorithmen entwerfen, testen und bewerten) divers sind und darin geschult werden, wie sie Gender-Analysen durchführen können, um diskriminierende Normen und Muster in ihren Systemen zu erkennen und einzudämmen.

#### **Stärkung und Förderung eines Multi-Stakeholder-Dialogs:**

- Als Teil einer breit angelegten Bemühung sollten Plattformen zu einem vertrauenswürdigen Umfeld beitragen, das den Austausch zwischen den Stakeholdern, einschließlich politischen Entscheidungsträger:innen, Regierungsbehörden, der Zivilgesellschaft, der Wissenschaft, Produktentwickler:innen und Softwareingenieur:innen, unterstützt.
- Ein Austausch zwischen den Stakeholdern sollte sich um die Einbeziehung von mehrdimensionalen Perspektiven bemühen, einschließlich der Diskriminierungserfahrungen von Opfern und Überlebenden von OGBV.
- Ein regelmäßiger Austausch könnte sich auf das Testen und die Verbesserung von Methoden zur Bewertung und Eindämmung von OGBV konzentrieren. So könnten Stakeholder über interoperable Melde-mechanismen für Nutzer:innen, KI-basierte Werkzeuge zur Inhaltsmoderation und Auswirkungen von algorithmischen Empfehlungssystemen diskutieren.
- Die Zusammenarbeit zwischen den verschiedenen Stakeholdern sollte den Austausch von bestehenden Untersuchungen fördern. Plattformen können beispielsweise die folgenden Ressourcen für zusätzliche Erkenntnisse und Empfehlungen zu Rate ziehen:
  - [»Technology-facilitated gender-based violence: preliminary landscape analysis«](#) der Global Partnership for Action on Gender-Based Online Harassment and Abuse;
  - [»Technology Companies Must Make Platforms Safer for Women in Politics«](#), [»Interventions to End Online Violence Against Women in Politics«](#) und [»Landscape Tracker«](#) des National Democratic Institute (NDI);

- »[The Chilling: A global study of online violence against women journalists](#)« des International Center for Journalists (ICF);
  - »[Guidance on the Safe and Ethical Use of Technology to Address Gender-based Violence and Harmful Practices](#)« des UN-Bevölkerungsfonds (UNFPA);
  - »[Measuring technology-facilitated gender-based violence. A discussion paper](#)« des UN-Bevölkerungsfonds (UNFPA);
  - »[Report on freedom of expression and the gender dimensions of disinformation](#)« der UN-Sonderberichterstatterin für Meinungsfreiheit;
  - »[Technology-facilitated violence against women: Taking stock of evidence and data collection](#)« von UN Women;
  - »[Violence Against Women and Girls \(VAWG\) Code of Practice](#)« von The End Violence Against Women Coalition, Glitch, Refuge, Carnegie UK, NSPCC, 5Rights, Professorin Clare McGlynn und Professorin Lorna Woods.
-



## Glossar

**Application Programming Interfaces (APIs)** sind Softwareschnittstellen, die eine Kommunikation zwischen zwei Anwendungen ermöglichen. APIs haben eine riesige Vielzahl an Verwendungszwecken, aber im Zusammenhang mit diesem Bericht ermöglichen sie Forscher:innen den Zugriff auf bestimmte Daten von Online-Plattformen über Datenanfragen. Als zwischen-geschaltete Instanz stellen APIs eine zusätzliche Sicherheitsebene bereit, indem sie keinen direkten Zugriff auf Daten zulassen und das Volumen und die Häufigkeit der Anfragen protokollieren, verwalten und überwachen.

**Extremismus** beschreibt die Befürwortung einer Weltanschauung, die die Überlegenheit und Dominanz einer »Eigengruppe« (*in-group*) über alle »Fremdgruppen« (*out-groups*) fordert. Extremismus propagiert somit eine Entmenschlichung der »Anderen« (*othering*), die mit Pluralismus und dem universellen Geltungsanspruch der Menschenrechte unvereinbar ist. Nach der Definition des ISD kann sich Extremismus durch gruppenbezogene Gewalt manifestieren, aber auch durch sukzessive, extreme soziale oder politische Absichten, die die Menschenrechte, demokratische Institutionen und die Zivilgesellschaft untergraben. Es ist wichtig, OGBV im Spektrum des Extremismus einzuordnen, da misogynen Inhalte in Radikalisierungsprozessen verwendet werden und zu Gewalt anstiften und diese auslösen können. **Gewalttätiger Extremismus** gilt in diesem Kontext als eine spezifische gewaltsame Erscheinungsform des Phänomens des Extremismus.

**Gender bzw. das soziale Geschlecht** ist das innere Gefühl einer Person, eine Frau, ein Mann, keines von beiden, beides oder irgendwo dazwischen in einem Spektrum zu sein.<sup>1</sup> Der Begriff beschreibt gesellschaftlich konstruierte Rollen für Frauen und Männer und ist eine erworbene Identität, die erlernt wird, sich im Laufe der Zeit verändert und innerhalb und zwischen Kulturen stark variiert. **Geschlechternormen** oder **Geschlechterstereotypen** sind »verallgemeinerte Ansichten oder Vorurteile über Eigenschaften oder Merkmale bzw. die Rollen, die Frauen und Männer innehaben oder einnehmen sollten«.<sup>2</sup> Sie werden oft in einen binären Rahmen gestellt, der die gelebten Erfahrungen diversgeschlechtlicher

Menschen außer Acht lässt und gleichzeitig Transpersonen ausschließt. Im Gegensatz dazu wird das **biologische Geschlecht** bei der Geburt auf der Grundlage der körperlichen Merkmale, die mit einer weiblichen oder männlichen Person assoziiert werden, zugewiesen.

**Geschlechtsspezifische Gewalt (gender-based violence, GBV)** – im Deutschen oft synonym mit geschlechtsbezogener oder geschlechtsbasierter Gewalt verwendet – bezieht sich auf »Gewalt, die sich gegen eine Person aufgrund ihres Geschlechts richtet«. <sup>3</sup> Frauen und die LGBTQ+-Community, einschließlich trans und diversgeschlechtlicher Personen, sind überproportional häufig von GBV betroffen.

**Geschlechtsspezifische Online-Gewalt (online gender-based violence, OGBV)** kann als Unterkategorie der **technologiegestützten geschlechtsspezifischen Gewalt (technology-facilitated gender-based violence, TFGBV)** beschrieben werden, die sich auf jede »Handlung bezieht, die durch den Einsatz von Informations- und Kommunikationstechnologien oder anderen digitalen Werkzeugen begangen, unterstützt, verschlimmert oder verstärkt wird und die zu körperlichem, sexuellem, psychologischem, sozialem, politischem oder wirtschaftlichem Schaden oder anderen Verletzungen von Rechten und Freiheiten führt oder führen kann«. <sup>4</sup> Das Phänomen wird auch als **technologiegestützte Gewalt gegen Frauen (technology-facilitated violence against women, TFVAW)** bezeichnet, wobei darauf hingewiesen wird, dass VAW (*violence against women*) durch GBV (*gender-based violence*) ersetzt werden kann, während die gemeinsame Definition zur Beschreibung des Phänomens beibehalten wird.

Die **Manosphere** ist ein Überbegriff, der sich auf mehrere, lose miteinander verbundene misogynen und antifeministische Online-Gemeinschaften bezieht. Sie umfasst verschiedene Arten von Misogynie mit unterschiedlichen Ausprägungen von Gewalt. Dazu gehören sogenannte Pick Up Artists, Männerrechtsaktivisten (*Men's rights activists, MRAs*), »Männer, die ihren eigenen Weg gehen« (*Men Going Their Own Way, MGTOW*) und Incels (»unfreiwillig Zölibatäre«). <sup>5</sup>

**Männliche Vorherrschaft (*male supremacy*)** ist eine »hasserfüllte Ideologie, die im Glauben an eine angeborene Überlegenheit von cis Männern und deren Recht, Frauen [und trans und diversgeschlechtliche Personen] zu unterdrücken, begründet ist«. <sup>6</sup> Die Vorstellung der männlichen Vorherrschaft richtet sich generell gegen Menschen, die als unmännlich oder nicht männlich wahrgenommen werden. Sie steht in Verbindung mit **hegemonialer Männlichkeit**, die das Patriarchat strukturiert und die »Legitimation ungleicher Geschlechterverhältnisse« beschreibt. <sup>7</sup>

**Misogynie** dient der Aufrechterhaltung einer patriarchalischen Gesellschaftsordnung, indem sie die Geschlechternormen und Genderidentitäten kontrolliert. <sup>8</sup> Sie dient dazu, Gewalt zu rechtfertigen, wenn von diesen Normen abgewichen wird. <sup>9</sup> Misogynie ist eng mit anderen Formen der Diskriminierung verbunden, einschließlich der gegen Transfrauen gerichteten Misogynie (Transmisogynie) und der gegen Frauen, die im rassistischen oder antisemitischen Weltbild als ungleichwertig imaginiert werden, wie beispielsweise der spezifische Hass gegen Schwarze Frauen (Misogynoir <sup>10</sup>). Sie versteckt sich oft in gewalttätigen extremistischen Ideologien, wobei Misogynie selbst eine motivierende gewalttätige Ideologie darstellen kann. <sup>11</sup>

**Radikalisierung** ist ein Begriff, der in diesem Kontext den Prozess beschreibt, durch den eine Person eine extremistische Ideologie annimmt, welche unter Umständen zu gewalttätigen extremistischen oder terroristischen Handlungen führen kann. In der Literatur über Terrorismus und gewalttätigen Extremismus wird häufig zwischen kognitiver Radikalisierung (Übernahme extremistischer Überzeugungen) und Verhaltensradikalisierung (der Prozess, der zu gewalttätigem Verhalten führt) unterschieden.

---

## Einleitung

Geschlechtsspezifische Gewalt (*gender-based violence, GBV*) in der Öffentlichkeit und im Privatleben ist eine globale Herausforderung, die zunehmend mit den Online-Räumen der Social-Media-Plattformen, Messaging-Diensten und anderen Kommunikationstechnologien verbunden ist und durch diese verstärkt wird. GBV spiegelt die Manifestation und Verstärkung ungleicher Machtverhältnisse auf Grundlage patriarchalischer Geschlechternormen wider,<sup>12</sup> die sich gegen alle Geschlechter richten können, aber am häufigsten gegen Frauen und LGBTQ+-Personen. OGBV überschneidet sich mit weiteren Formen der Diskriminierung und Hierarchisierung, beispielsweise haben muslimische Frauen und Schwarze Frauen ein erhöhtes Risiko aus rassistischen und misogynen Motiven angegriffen zu werden.

Während Plattformen zur Stärkung feministischer Bewegungen beitragen können, indem sie beispielsweise geschlechtliche und sexuelle Vielfalt sichtbar machen, ermöglicht das derzeitige Online-Umfeld jedoch auch die Verbreitung von misogynen und antifeministischen Inhalten. Zudem sind Frauen und LGBTQ+-Personen im öffentlichen Leben, darunter Aktivist:innen und Menschenrechtsverteidiger:innen,<sup>13</sup> Politiker:innen,<sup>14</sup> und Journalist:innen, unverhältnismäßig stark von OGBV betroffen.<sup>15</sup> Dies hat eine abschreckende Wirkung (*chilling effect*) auf die gleichberechtigte Teilhabe, wobei geschlechtsspezifische und sexualisierte Fehl- und Desinformationen von (sowohl nichtstaatlichen als auch staatlichen) antidemokratischen Akteur:innen als gezielte Taktik eingesetzt werden.<sup>16</sup> Einige der Plattformen haben bereits vereinzelte Maßnahmen ergriffen. Bislang hat jedoch noch keine Plattform ausreichende Schritte unternommen, um die individuellen und gesellschaftlichen Risiken, die von OGBV ausgehen, wirksam anzugehen.

Online-Manifestationen von GBV weisen zwar unterschiedliche Merkmale auf, sind jedoch Teil eines »Kontinuums vielfältiger, wiederkehrender und miteinander verbundener Formen von GBV«.<sup>17</sup> Die Täter<sup>18</sup> misogynen Gewalt setzen so eine patriarchalische, geschlechternormative Ordnung durch und verstärken sie mit Mitteln

aus dem gesamten taktischen Spektrum, das von legalem, aber schädlichem Verhalten bis hin zu gewalttätigem Extremismus reicht.<sup>19</sup> Sie manifestiert sich online, während die Reproduktion und Verstärkung von Misogynie im Internet zu Offline-Gewalt führen kann – von Gewalt in der Partnerschaft über physische Angriffe auf Journalist:innen bis hin zu Terrorismus.<sup>20</sup> In diesem Zusammenhang ist es wichtig anzuerkennen, dass OGBV in einem Ökosystem stattfindet, das durch eine digitale Kluft zwischen den Geschlechtern (*digital gender divide*) gekennzeichnet ist. Diese ist in strukturellen Ungleichheiten der Geschlechter begründet. So bestehen beispielsweise die Produktteams der Plattformen (und in einigen Ländern auch die Gruppen, die auf diese Plattformen zugreifen und sie nutzen) in einem unverhältnismäßig hohen Maß aus Männern.<sup>21</sup> In Anbetracht der vielschichtigen Dynamik von On- und Offline-Gewalt wird im vorliegenden Bericht die Verbindung von OGBV mit extremistischen Ideologien und gewalttätigem Extremismus erläutert, wobei der Schwerpunkt auf der Verantwortung und Rolle von Plattformen liegt.

Eine Reihe multilateraler Foren, darunter UN Women, die Weltgesundheitsorganisation (WHO) und der UN-Bevölkerungsfonds (UNFPA), sowie Multi-Stakeholder-Initiativen wie die von den USA einberufene Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership) erkennen die Dringlichkeit an, Plattformen mit Blick auf die Risiken von OGBV in die Verantwortung zu ziehen. Die in dem Christchurch Call involvierten Stakeholder beabsichtigen zudem, die Erforschung der Verbindung zwischen misogynen und gewalttätigen extremistischen Inhalten zu vertiefen. Auf der Grundlage eines regierungs- und ressortübergreifenden Dialogs und einer Konsultierung zivilgesellschaftlicher Expertise gibt dieser Bericht einen Überblick über die gegenwärtigen Phänomene und die vielschichtigen Auswirkungen von OGBV, wobei die Verbindung zu gewalttätigem Extremismus hervorgehoben wird. Der Bericht formuliert zudem mögliche Maßnahmen für Plattformen zur Risikobewertung und -minderung.

## Phänomene der geschlechtsspezifischen Online-Gewalt (OGBV)

OGBV hat greifbare und messbare Auswirkungen, und Offline-Schäden können sich wiederum online ausweiten und verstärken. In vielen Fällen kennen die Opfer und Überlebenden die Täter, bei denen es sich oft um derzeitige oder ehemalige Partner:innen, Verwandte, Mitarbeitende oder Personen aus dem Freundeskreis handelt.<sup>22</sup>

In Anbetracht der Tatsache, dass patriarchalische Geschlechternormen und Machtverhältnisse sowohl bei Gewalt in der Partnerschaft als auch bei extremistischen Ideologien eine Rolle spielen, wird in diesem Abschnitt dargelegt, dass OGBV ein Vektor für Radikalisierung darstellen kann. Es wird untersucht, welche Rolle Misogynie, einschließlich geschlechtsspezifischer und sexualisierter Motive, im gewalttätigen Extremismus spielt. Dabei werden wiederum die Überschneidungen von OGBV mit anderen Formen von Gewalt und Diskriminierungsformen wie Rassismus, Antisemitismus und Islamophobie betont, sowie der Zusammenhang zwischen On- und Offline-Manifestationen von Gewalt berücksichtigt.

### Misogyne Gruppen und (Sub-)Kulturen: OGBV als Vektor für Radikalisierung

In den vergangenen Jahren rückte die Untersuchung der Geschlechterdynamik im Kontext der Erforschung von Online-Radikalisierung und der Verbreitung von extremistischen und terroristischen Inhalten über verschiedene Ideologien hinweg immer stärker in den Mittelpunkt.<sup>23</sup> Forscher:innen weisen darauf hin, dass Misogynie als ideologischer Vektor für Radikalisierung in Praktiken zur Verhinderung und Bekämpfung von gewalttätigem Extremismus (*Preventing and Countering Violent Extremism, P/CVE*) stärker anerkannt werden muss.<sup>24</sup>

Ein wachsender Teil der Forschung befasst sich verstärkt mit den Beziehungen zwischen Rechtsextremismus, misogynen Weltanschauungen und der Manosphere, einem losen Netzwerk misogynen Online-Gemeinschaften, die männliche Vorherrschaft (*male supremacy*) und patriarchalische Geschlechternormen durchsetzen wollen.<sup>25</sup> So erklärt das Southern Poverty Law Center (SPLC), dass die Anhänger männlicher Vorherrschaftsideologien »auf starre Geschlechterrollen fixiert sind und jegliche Abweichung von ihrer strikten Geschlechterdichotomie

abwerten«. SPLC beschreibt männliche Vorherrschaft zudem als eine »starke Unterströmung der weißen Vorherrschaft (*white supremacy*), und ihre Denkansätze liegen einem Großteil der heutigen extremen Rechten zugrunde.«<sup>26</sup> Misogyne Gewalt hat sich in physischen Angriffen auf Frauen manifestiert, wobei sich misogynen Motive oft mit rassistischen und fremdenfeindlichen Beweggründen überschneiden. Der Täter der Schießerei in einem Spa in Atlanta, Georgia, bei der 2021 sechs Frauen asiatischer Herkunft getötet wurden, verbreitete beispielsweise Botschaften, denen eine Hypersexualisierung und Fetischisierung asiatischer Frauen zugrunde lagen.<sup>27</sup>

Die Beziehungen zwischen misogynen Gruppen, wie den »unfreiwillig Zölibatären« (Incels) – einer Untergruppe der Manosphere, die Frauen und die Gesellschaft für ihren mangelnden romantischen Erfolg verantwortlich machen<sup>28</sup> – und anderen Ideologien der Vorherrschaft (*supremacy*) sind jedoch komplex und vielschichtig. Zum Beispiel haben Incels eine bestimmte Perspektive auf *Race* und ethnische Herkunft, die sich von rechtsextremen Gruppen unterscheidet. Incels glauben an eine vermeintliche Rassenhierarchie in der Dating-Sphäre – laut ihrer Ansicht werden beim Dating weiße Männer bevorzugt, was sie mit den Entscheidungen der Frauen bei der Auswahl von Sexualpartnern begründen, anstatt diese aktiv zu befürworten.<sup>29</sup> Rechtsextreme Gruppen verfolgen dagegen eine rassistische Vorherrschaft und versuchen, rassifizierte sexuelle Grenzen durchzusetzen, um eine Homogenität der »Eigengruppe« aufrechtzuerhalten. In Anbetracht dieser Nuancen findet eine gegenseitige Inspiration der Gruppen statt, da sowohl Incels als auch rechtsextreme Gruppen misogynen und antifeministischen Haltungen teilen. Forscher:innen des ISD stellten zudem fest, dass sich einige Incels ausdrücklich mit rassistisch oder ethnisch motiviertem gewalttätigem Extremismus (*racially or ethnically motivated violent extremism, REMVE*) identifizieren, indem sie sich beispielsweise als »Stormcels« bezeichnen – eine Anlehnung an die Website »Stormfront«, die rassistische Inhalte verbreitete.<sup>30</sup>

In diesem Zusammenhang wurde bislang in P/CVE-Praktiken und -Maßnahmen der Schwerpunkt nicht ausreichend auf die potenziell gewalttätigen Folgen misogynen Ideologie gelegt. Jüngste wissenschaftliche

Debatten drehen sich auch um die Frage, ob Gewalt im Zusammenhang mit Incels als Terrorismus identifiziert werden sollten.<sup>31</sup> Während Incels politische Einstellungen vertreten und einen Grundsatz verfolgen, der auf gruppenbezogene Unterdrückung ausgerichtet ist, gibt es keinen Konsens darüber, ob die Gewalt dieser Gruppe als primär ideologisch oder alternativ als nihilistisch zu betrachten ist. In diesem Zusammenhang werden die ersten juristischen Präzedenzfälle geschaffen: Der Oberste Gerichtshof von Ontario (Kanada) hat beispielsweise entschieden, dass ein Incel-motivierter Mord als terroristische Aktivität einzustufen ist.<sup>32</sup> In solchen Fällen gilt die Bezeichnung »Einsamer Wolf« (*lone wolf*) unter Umständen nicht als ein treffender Begriff. Auch wenn die Täter von geschlechtsspezifischer Gewalt, sowohl online als auch offline, nicht zwingend einer bestimmten ideologischen Gruppe angehören, kann dies auf die misogynen Ideologie selbst zurückzuführen sein, welche zugleich diffus und vernetzt ist – ein Phänomen, das mit der weiter gefassten Herausforderung des »post-organisatorischen« (*post-organisational*) Extremismus zusammenhängt.<sup>33</sup>

Misogynie – wie auch Antisemitismus – dient oft als verbindendes Kernmerkmal verschiedener extremistischer Ideologien. Im Zusammenhang mit der Förderung heteronormativer Geschlechternormen und -identitäten zeigen die Verbindungen zwischen gewalttätigem Extremismus und Misogynie Parallelen zwischen militanter Männlichkeit in verschiedenen Ideologien auf, vor allem im Rechtsextremismus und islamistischem Extremismus. Die Forschung betont, dass »beide Gruppen den Begriff der Männlichkeit mit der Bereitschaft zur Verteidigung gleichsetzen, wobei es ist nicht ungewöhnlich ist, dass offene oder subtile Misogynie als Motiv für die Hinwendung zur jeweiligen Ideologie dient«.<sup>34</sup> Bei der Untersuchung extremistisch-islamistischer Gruppen stellten Forscher:innen fest, dass ISIS »einen militarisierten, maskulinisierten, religiösen und genozidalen Nationalismus innerhalb ihres »Islamischen Staates« praktiziert, indem jesidische Frauen und Mädchen und andere marginalisierte Gruppen geschlechtsspezifischer Gewalt ausgesetzt werden«.<sup>35</sup> Sowohl islamistischer Extremismus als auch Rechtsextremismus »setzen patriarchalische geschlechtsspezifische Rollen, Binaritäten, Hierarchien und Normen durch«,<sup>36</sup> und zeigen, dass männliche Vorherrschaft und misogynen

Glaubenssysteme in einem breiten ideologischen Spektrum zu finden sind. Allen gemeinsam ist die missbräuchliche Nutzung etablierter und »alternativer« Plattformen und Messaging-Dienste, und zwar über geografische Grenzen und verschiedene Sprachen hinweg.

Darüber hinaus heben Forscher:innen hervor, wie sich Antisemitismus mit misogynen Einstellungen überschneidet. Evelyn Torton Beck erklärt, dass das Stereotyp der »jüdischen Prinzessin« (*Jewish Princess*) »die traditionellen antisemitischen Tropen auf eine weibliche Form überträgt: Sie ist materialistisch, geldgierig, manipulativ, oberflächlich, hinterlistig und prahlerisch«.<sup>37</sup> Blyth Crawford stellt fest, dass die neofaschistische militante Akzelerationismus-Bewegung jüdische Personen »als Einflussnehmer:innen auf die Sexualpolitik betrachtet, die als familienfeindlich angesehen wird und so eine Bedrohung für die weiße Rasse darstellt«.<sup>38</sup>

Als Teil des Verschwörungsmythos eines »Völkermords an der weißen Rasse« (*white genocide*) gibt es Verbindungen zwischen antisemitischen Ausdrücken und anti-LGBTQ+-Hetze, die geschlechtsspezifische Hassrede einschließt. So stellt die Anti-Defamation League (ADL) fest, dass die »angebliche gezielte Propaganda von LGBTQ+-Identitäten als ein Schlüsselement der Versuche des Judentum dargestellt wird, mit dem Ziel Reproduktionsraten unter heterosexuellen, weißen Cisgender-Personen zu senken«.<sup>39</sup> Anti-LGBTQ+-Ausdrücke, die die Community als »Pädophile« oder »Groomer«<sup>40</sup> bezeichnen, decken sich oft mit der antisemitischen Behauptung, dass sich Juden an Nicht-Juden, insbesondere an nicht-jüdischen Kindern, vergreifen würden.<sup>41</sup> Diese Verbindungen zwischen Antisemitismus, Misogynie und Hass gegen LGBTQ+-Personen wird von einem breiten extremistischen Spektrum mit unterschiedlichem ideologischem Hintergrund vorangetrieben.

Strukturelle Geschlechterungleichheit und Geschlechternormen können zudem zu einer verinnerlichten Misogynie führen. Ein extremes Beispiel dafür ist die Entstehung der sogenannten Tradwives, einer Online-Gemeinschaft, in der die Ansichten von rechtsextremem Ideologie, christlichem Nationalismus, weißer Vorherrschaft (*white supremacy*) und patriarchalischer Geschlechternormen zusammenkommen. Tradwives vertreten ein stark heteronormatives und

traditionalistisches Rollenverständnis von Frauen als »Ehefrau und Mutter«, das gegen die Gleichstellung der Geschlechter, Feminismus, reproduktive Rechte, und LGBTQ+-Rechte steht. Forscher:innen heben hervor, dass Tradwives »ihre Präsenz in den sozialen Medien nutzen, um eine starke weibliche Eigengruppen-Assoziation zu bieten« und »mit ihrer anti-globalistischen, anti-modernen Lebenseinstellung erfolgreich die sozialen Mainstream-Medien infiltriert haben«.42 Das Rollenverständnis der Frau in gewalttätigen extremistischen Ideologien und Gruppen ist zwar unterschiedlich, doch die Forschung legt nahe, dass reine Frauenforen auch »als geschlechtsspezifische Orte der ideologischen Anfechtung« dienen können, in denen Frauen »in ihren Alltagspraktiken trotz anderweitig einschränkender geschlechtsspezifischer ideologischer Konstrukte Handlungsfähigkeit beweisen«.43

Während misogynie Online-Gemeinschaften oft subtile und verschlüsselte Sprache und Rhetorik verwenden, um eine Kultur der »Eigengruppe« zu verdecken und zu vermitteln,44 ist Online-Misogynie gleichzeitig weit verbreitet und populär. Das ISD stellt fest, dass die Verbreitung misogynen Inhalte am wirkungsvollsten und erfolgreichsten ist, wenn sich verschiedene ideologische Gruppen an der Verbreitung beteiligen. Anti-Drag- und anti-LGBTQ+-Aktivitäten beschränken sich beispielsweise nicht auf Randgruppen, sondern sind zu einer verbindenden Motivation der extremen Rechten und lokalisierter Aktivist:innen geworden, darunter bestimmte Elternrechtsgruppen, Anti-Impfstoff- oder Anti-Lockdown-Gruppen bis hin zu christlichen Nationalist:innen.45 Darüber hinaus wird misogynie Gewalt vermehrt mit gewalttätigen Verschwörungsbewegungen in Verbindung gebracht, die infolge der Covid-19-Pandemie stark zugenommen haben. So zeigt die Erforschung der Online-Aktivitäten von QAnon-Anhänger:innen, wie gezielter Hass, einschließlich gewalttätiger misogynen, rassistischer und anti-LGBTQ+-Rhetorik, zu einem Problem insbesondere für Frauen im öffentlichen Leben geworden ist, die häufig Opfer von koordinierter Belästigung werden.46 Zudem hat oft eine kleine Anzahl an Akteur:innen einen überproportionalen Einfluss auf die Verbreitung misogynen Inhalte wie beispielsweise der Fall des misogynen Influencers Andrew Tate, der gewaltverherrlichende Vorstellungen von Männlichkeit propagiert, zeigt.47

Online-Misogynie kann als ideologischer Türöffner in verschiedene extremistische und radikale Gruppen agieren, wobei die Weiterverbreitung und Amplifizierung misogynen Inhalte das Risiko einer Normalisierung unter den Nutzer:innen birgt, insbesondere unter männlichen Nutzern, die sich in Online-Räumen aufhalten, um Kontakte zu knüpfen, sich zu vernetzen und Anschluss zu finden.

### Taktiken und Verhaltensweisen: OGBV als ein Kontinuum von Gewalt

Wie bereits erwähnt, findet OGBV im Rahmen eines Gewaltkontinuums (*continuum of violence*) statt,48 das die komplexen und miteinander verknüpften Erfahrungen mit verschiedenen Formen und Verhaltensweisen der geschlechtsspezifischen Gewalt anerkennt.49

In diesem Abschnitt werden weitverbreitete Formen von OGBV beschrieben. Dabei werden eine im Jahr 2020 von der Economist Intelligence Unit (EIU)<sup>50</sup> durchgeführte Umfrage zur Messung der Prävalenz von Online-Gewalt gegen Frauen sowie weitere Forschungsergebnisse herangezogen, darunter Recherchen von PEN America,<sup>51</sup> der Global Partnership<sup>52</sup> und des International Center for Journalists (ICJ).<sup>53</sup> Dieser Abschnitt soll keineswegs als eine vollständige Übersicht fungieren, sondern vielmehr die vielfältigen Formen von OGBV skizzieren. Wichtig ist, dass diese Formen und Verhaltensweisen häufig in Kombination und über mehrere Plattformen hinweg zu beobachten sind.

- **Geschlechtsspezifische oder sexualisierte Fehl- und Desinformation** bezieht sich auf »eine Untergruppe des geschlechtsspezifischen Online-Missbrauchs, die falsche oder irreführende geschlechtsspezifische und sexualisierte Narrative gegen Frauen [und trans und diversgeschlechtliche Personen] verwendet, oft mit einem gewissen Grad an Koordination und mit dem Ziel, Frauen [und trans und diversgeschlechtliche Personen] von der Teilnahme am öffentlichen Leben abzuhalten«.54 Dabei kann es sich um verleumderische Kommentare handeln, die den Ruf einer Person schädigen sollen. Eine Kombination aus Falschinformationen und der Veröffentlichung von sachlichen, dekontextualisierten und irreführenden Informationen ist oft am schädlichsten.

Geschlechtsspezifische und sexualisierte Fehl- und Desinformationen bedienen sich häufig einer kodierten und subtilen Sprache sowie sich wiederholender, kontextbezogener Memes.

- **Online-Belästigung** umfasst ein breites Spektrum unerwünschter oder negativer Kontaktaufnahmen, die dazu dienen, ein einschüchterndes, lästiges, beängstigendes oder feindliches Umfeld zu schaffen.<sup>55</sup> Sie kann eine langanhaltende koordinierte Narrativkampagne, die Verbreitung von Namenslisten und plattformübergreifende Belästigung beinhalten.<sup>56</sup> Es kann sich dabei auch um eine einzelne Bemerkung oder einen einmaligen Vorfall handeln. Diese Art von Belästigung ist oft geschlechtsspezifischer oder sexualisierter Natur.
- **Geschlechtsspezifische Hassrede (*gender-based hate speech*)** erniedrigt Personen aufgrund ihrer Geschlechtsidentität oder ihres Geschlechtsausdrucks, wobei sich überschneidende geschützte Merkmale wie beispielsweise ethnische Herkunft oder Religion das Risiko erhöhen, Ziel von Hassrede zu werden.<sup>57</sup> Geschlechtsspezifische Hassrede umfasst entmenschlichende und abwertende Sprache als auch Drohungen und Aufforderungen zur Gewalt.<sup>58</sup>
- **Online-Identitätsdiebstahl** bezeichnet die unrechtmäßige Erlangung und Nutzung der persönlichen Daten einer anderen Person in einer Weise, die Betrug oder Täuschung beinhaltet. Geschlechtsspezifische Beispiele umfassen die Erstellung gefälschter Konten zur Anwerbung von Mädchen und Frauen für den Sexhandel.<sup>59</sup>
- **Stalking und Überwachung** bezeichnen den Missbrauch von Technologie, beispielsweise die Installation von kommerzieller Stalker-Software auf einem Gerät. Stalking und Überwachung finden oft wiederholt statt und können eine Erweiterung von Gewalt in der Partnerschaft sein.<sup>60</sup>
- **Astroturfing** bezieht sich auf die betrügerische Praxis der Verbreitung oder Verstärkung von Inhalten, die scheinbar organisch entstehen und sich verbreiten, in Wirklichkeit aber von einer Einzelperson, einer Interessengruppe, einer politischen Partei oder einer Organisation koordiniert werden.<sup>61</sup> Astroturfing kann Teil der **vernetzten Belästigung (*networked harassment*)** sein, die Taktiken wie **Trolling** (absichtliche Störung oder Unterbrechung von Online-Veranstaltungen, Debatten oder Hashtags)<sup>62</sup> und **koordiniertes Flagging** (unberechtigte Meldung von Nutzer:innen, um diese von der Plattform zu entfernen) umfasst.<sup>63</sup>
- **Bildbasierter sexueller Missbrauch** beinhaltet die Erstellung, Verbreitung, Weitergabe oder Androhung der Weitergabe intimer Bilder oder Videos einer Person ohne deren Einwilligung.<sup>64</sup> Der Begriff umfasst eine Vielzahl von Verhaltensweisen, wie beispielsweise **sexuelle Erpressung** (wenn eine Person ein intimes Bild einer anderen Person hat oder behauptet, ein solches zu haben, und es benutzt, um sie zu zwingen, etwas zu tun, was sie nicht tun will);<sup>65</sup> **Dokumentation oder Verbreitung von sexueller Gewalt**, die in sozialen Medien veröffentlicht, an andere gesendet, verkauft oder gehandelt wird, was zu einer zusätzlichen Form sexueller Gewalt gegen Opfer und Überlebende führt;<sup>66</sup> und die **Verwendung von generativer KI zur Erstellung von Deepfakes**, einschließlich künstlicher Bilder oder Videos, die tatsächlichen Aufnahmen ähneln.<sup>67</sup>
- Beim **Doxing** werden persönliche oder identifizierende Informationen (beispielsweise Adressen, Telefonnummern, E-Mails, Namen von Partner:innen oder Kindern) ohne Einwilligung abgerufen und veröffentlicht – oft in der böswilligen Absicht, die Betroffenen am Arbeitsplatz oder zu Hause aufzusuchen oder negative oder unerwünschte Kontakte herzustellen.<sup>68</sup>
- **Androhung von Offline-Gewalt** wie Vergewaltigung und Todesdrohungen oder Anstiftung zu körperlicher Gewalt. Journalist:innen,<sup>69</sup> Akademiker:innen,<sup>70</sup> Politiker:innen<sup>71</sup> und Menschenrechtsverteidiger:innen<sup>72</sup> sind häufig gewalttätigen Drohungen ausgesetzt, die geschlechtsspezifisch und sexualisiert sind, insbesondere wenn sie über Geschlechtergerechtigkeit und Gleichberechtigung oder von Männern dominier-

te Themen sprechen oder schreiben. In einer Studie des International Center for Journalists (ICFJ) wurde der Kreislauf der Gewalt aufgezeigt und hervorgehoben, dass »digitale Angriffe die Offline-Gewalt anheizen können, während der Offline-Missbrauch durch prominente Personen wiederum Online-Angriffe auslösen kann«.73

Die vorgestellten Taktiken und Verhaltensweisen treten plattformübergreifend auf, oft gleichzeitig und in koordinierter Weise. Astroturfing und vernetzte Belästigung werden eingesetzt, um eine größere Reichweite sowie die Vernetzung und Gruppenbildung von Tätern zu ermöglichen, die ansonsten voneinander isoliert wären. So sind beispielsweise Incel-Foren auf Reddit und 4chan sowie in Gaming-Foren wie Discord oder auf speziellen Webseiten zu finden. Diese Foren stärken die Zugehörigkeit zu einer »Eigengruppe« oder einer Gemeinschaft.74 Während diese Art von plattformübergreifendem misogynen Verhalten und Networking zusätzliche Risiken für die Nutzer:innen mit sich bringt, unterstreicht sie auch die Herausforderung, zu verstehen, wie Personen radikalisiert werden und wie ein gesundes Online-Umfeld dazu beitragen kann, die Radikalisierung zu verhindern, einschließlich zielorientierter Interventionen, die auf individuelle Missstände eingehen.

Darüber hinaus nutzen koordinierte Belästigungskampagnen häufig Online-Unterhaltungen über aktuelle Themen, die beispielsweise die Verwendung missbräuchlicher Hashtags beinhalten können, um misogynen Inhalte zu amplifizieren.75 Dies kann sich auch in der koordinierten Belästigung einer Person über mehrere Plattformen hinweg äußern. Die koordinierte Online-Belästigung wiederum stellt eine Herausforderung für die Verfolgung und Meldung von OGBV dar, da sie die Verantwortung häufig auf die Opfer und Überlebenden überträgt und die Notwendigkeit einer traumainformierten Koordinierung und Zusammenarbeit zwischen den Plattformen verdeutlicht. Schließlich vergrößert die globale Reichweite vieler Plattformen die misogynen Online-Einflüsse auf Radikalisierungswege und trägt zu einer Aufrechterhaltung von Online-Kulturen mit extremistischen Ansichten bei. So zeigen Studien, dass 40% bis 50 % der jüngeren Menschen gelegentlich mit extremistischen Inhalten in Berührung kommen.76 So ergeben sich immer wieder Möglichkeiten für Online-Radikalisierungsprozesse mit Misogynie als Vektor.

## Schäden: Auswirkungen von OGBV auf Individuen und Gesellschaft

In diesem Abschnitt werden die Auswirkungen von OGBV sowohl auf der Mikroebene (Individuen) als auch auf der Makroebene (Gesellschaft) dargestellt, wobei die Bandbreite der Schäden, einschließlich der Risiken für die private und öffentliche Sicherheit, berücksichtigt wird.

- **Psychische Schäden:** OGBV kann bei Opfern und Überlebenden schwerwiegende psychische Schäden, mentalen oder emotionalen Stress sowie Symptome einer posttraumatischen Belastungsstörung hervorrufen, insbesondere wenn der Missbrauch häufig auftritt. Die kumulativen Auswirkungen von Off- und Online-Gewalt können auch zu Selbstverletzungen, Depressionen und Selbstmord führen.77
- **Bedrohungen für die reproduktive Gesundheit:** Fehl- und Desinformationen über Abtreibung und reproduktive Rechte können geschlechtsspezifischen Schaden anrichten, da sie den Zugang zu faktischen Informationen über die Gesundheitsversorgung untergraben und unsichere Alternativen oder ungetestete Medikamente vorantreiben.78
- **Eingriffe in die Privatsphäre:** Sobald personenbezogene Daten online veröffentlicht wurden (beispielsweise durch Doxing), kann es sich schwierig, wenn nicht gar unmöglich gestalten, diese abzurufen oder zu entfernen. Dies birgt auch die Gefahr künftiger Angriffe, da die Informationen dauerhaft im Internet oder auf dem Gerät einer anderen Person gespeichert bleiben. Somit wirkt sich OGBV negativ auf das Recht auf Privatsphäre der betroffenen Person aus.
- **Wirtschaftlicher und materieller Schaden:** Der Begriff »wirtschaftlicher Vandalismus« (*economic vandalism*) hebt die wirtschaftlichen und beruflichen Konsequenzen hervor, die durch OGBV verursacht werden können – beispielsweise durch verpasste Arbeitsmöglichkeiten, verringerte Produktivität und den Rückzug aus dem Internet.79
- **Verstärkung der strukturellen Ungleichheit der Geschlechter:** OGBV normalisiert Misogynie und fördert eine Kultur der patriarchalischen Gewalt, die Vergewaltigungskultur (*rape culture*), Opfer-Be-



schuldigung (*victim blaming*) und die Verharmlosung sexueller Übergriffe beinhaltet.<sup>80</sup> Die Normalisierung von OGBV verstärkt das »Verstummen« (*silencing*) von Frauen und LGBTQ+-Personen, wodurch Opfer und Überlebende von der Teilnahme am öffentlichen Leben abgehalten werden. OGBV verschärft somit die Ungleichheit zwischen den Geschlechtern, die Frauen und die LGBTQ+-Community an der Ausübung ihrer Freiheiten und Menschenrechte hindert. Es gibt zudem generationenübergreifende Auswirkungen, da OGBV junge Frauen und Mädchen sowie LGBTQ+-Personen davon abhält, Berufe wie beispielsweise in der Politik und im Journalismus zu ergreifen, aus Angst vor ähnlichem Missbrauch, was wiederum die digitale Kluft zwischen den Geschlechtern verstärkt.<sup>81</sup>

- **Bedrohung der privaten und öffentlichen Sicherheit:** Es gilt erneut zu betonen, dass OGBV in einem Kontinuum stattfindet und misogynen Verhalten, das im Online-Raum beginnt, auch zu Offline-Gewalt führen kann – sowohl im privaten als auch im öffentlichen Raum.<sup>82</sup> In einem Bericht des U.S. Secret Service National Threat Assessment Center (NTAC) aus dem Jahr 2023 wird die Bedrohung der öffentlichen Sicherheit durch Personen beschrieben, die gezielte Gewalttaten begehen, wobei die Täter vor einem Angriff häusliche Gewalt, misogynen Verhaltensweisen oder beides aufzeigten. In dem Bericht wird erläutert, dass Männer, die misogynen Gewalttaten begangen haben (in der Regel Schießereien und Messerstechereien), bereits eine Vorgeschichte mit bedrohlicher Online-Kommunikation sowie anderer Risikofaktoren aufzeigten (beispielsweise eine Vorgeschichte mit Mobbing, finanzieller Instabilität und zwischenmenschlichen Schwierigkeiten).<sup>83</sup>

- **Bedrohung der Demokratie:** Auf gesellschaftlicher und globaler Ebene nutzen antidemokratische Kräfte – sowohl staatliche als auch nichtstaatliche Akteure – Online-Räume, um Frauen und LGBTQ+-Personen im öffentlichen Leben anzugreifen.<sup>84</sup> Ein Bericht von #ShePersisted stellt fest, dass geschlechtsspezifische Desinformation als Frühwarnsystem für »Rückschritte bei den Frauenrechten und die Erosion demokratischer Prinzipien und Institutionen« dienen kann.<sup>85</sup> Eine vom International Center for Journalists (ICJ) durchgeführte globale Studie stellt einen »alarmierenden Trend« fest, wonach politische Akteure und extreme Gruppierungen am politischen Rand, als »Anstifter und Täter von Online-Gewalt gegen Journalist:innen« fungieren.<sup>86</sup> Ein Bericht der USA, Kanadas, des Europäischen Auswärtigen Dienstes, Deutschlands, der Slowakei und des Vereinigten Königreichs über geschlechtsspezifische Desinformation (*gendered disinformation*) unterstreicht, dass ausländische staatliche Akteure wie Russland und die Volksrepublik China strategisch auf Frauen und Menschen mit sich überschneidenden Identitäten abzielen, um Einzelpersonen und Personengruppen davon abzuhalten, ihre Grundrechte auszuüben. In dem Bericht wird zudem erläutert, dass identitätsbasierte Desinformation (*identity-based disinformation*) die »Fähigkeit untergräbt, auf unparteiische, faktenbasierte Informationen zuzugreifen, und sich negativ auf die Zusammensetzung der demokratischen Vertretung auswirkt«.<sup>87</sup>

## Bewertung von Community-Richtlinien und deren Durchsetzung

Ein Schlüsselfaktor bei der Bekämpfung von OGBV ist die Formulierung als auch wirksame Durchsetzung umfassender Community-Richtlinien, die festlegen, was auf einer Plattform erlaubt ist und was nicht. Dabei handelt es sich in der Regel um Verträge, die den Nutzer:innen auf einer »Take-it-or-leave-it«-Basis („Nimm es oder lass es sein“) vorgelegt werden und eine Reihe von Richtlinien enthalten, die von den Plattformen regelmäßig aktualisiert werden.

Während die meisten Plattformen<sup>88</sup> im Allgemeinen einige Formen von OGBV in ihren Richtlinien zu Hassrede oder Belästigung und Missbrauch berücksichtigen (einschließlich der geschützten Merkmale der Nutzer:innen), weisen sowohl die Richtlinien an sich als auch deren Durchsetzung Lücken auf. In Bezug auf die Richtlinien an sich befasst sich keine der Plattformen ausdrücklich mit geschlechtsspezifischer oder sexualisierter Fehlinformation. Derartige Inhalte kommen oft in Form von verschlüsselter und subtiler Sprache, kontextbasierten visuellen Memes vor oder durch die absichtliche Unkenntlichmachung bestimmter Wörter.<sup>89</sup> Darüber hinaus hat das Oversight Board, das die von Meta getroffenen Moderationsentscheidungen überprüft, kürzlich die Entscheidung von Meta aufgehoben, einen Facebook-Post, der eine Überlebende von GBV verhöhnt, nicht zu löschen. Konkret stellte das Oversight Board fest, dass der Beitrag gegen die Richtlinien von Meta im Zusammenhang mit Mobbing und Belästigung verstößt, da dieser sich über die schwere Körperverletzung der abgebildeten Frau lustig macht. Das Oversight Board erklärte jedoch, dass »dieser Beitrag nicht gegen die Richtlinien von Meta verstoßen hätte, wenn die abgebildete Frau nicht identifizierbar gewesen wäre oder wenn dieselbe Bildunterschrift einem Bild einer fiktiven Frau beigefügt gewesen wäre«, was auf eine Lücke in den Richtlinien hinweist, die Inhalte zuzulassen scheint, welche geschlechtsspezifische Gewalt normalisieren.<sup>90</sup>

Abgesehen von Lücken in der Richtlinienformulierung haben ISD-Untersuchungen ergeben, dass auch die Durchsetzung der bestehenden Richtlinien lückenhaft ist. Auf X (ehemals Twitter), das seinen Nutzer:innen verbietet, »andere mit wiederholten Verunglimpfungen, bildlichen Ausdrücken oder anderen Inhalten anzusprechen, die darauf abzielen, eine geschützte Kategorie zu erniedrigen oder negative oder schädliche Stereotype zu verstärken«,<sup>91</sup> fand das ISD mehrere Fälle,

in denen diese Art von Inhalten nicht moderiert wurde, darunter Tweets mit sexistischen Ausdrücken gegen die Schauspielerin Amber Heard sowie allgemeine Angriffe auf das Erscheinungsbild von Frauen.<sup>92</sup> Zudem wurden auch offen abwertende Beleidigungen wie »Hure« (*whore*), »Fotze« (*cunt*) oder »Schlampe« (*bitch*) im Kommentarbereich von YouTube festgestellt. Diese tauchten häufig nicht nur unter Videos auf, die zu hass erfüllten Kommentaren einzuladen schienen, sondern auch unter unverdächtigen Videos (solche Kommentare wurden beispielsweise sowohl unter dem Video »Andrew Tate zerstört moderne Frauen« als auch unter dem Video »Sprecherin Nancy Pelosi hält ihre letzte Pressekonferenz« gefunden).<sup>93</sup> Auf TikTok stellte das ISD fest, dass misogynen Inhalte immer noch offen gepostet werden und dabei hohes Engagement erhalten. Beispielsweise fanden sich auf der Plattform Videos von Accounts, die für Men Going Their Own Way (MGTOW) werben, eine Gemeinschaft in der Manosphere, die unter anderem alleinerziehende Mütter herabwürdigt und entmenschlicht.<sup>94</sup>

Die Plattform mit den laxesten (oder gar fehlenden) Richtlinien und Durchsetzungsmaßnahmen (d. h. wo ISD die meisten misogynen Inhalte fand) ist Telegram. Insbesondere private Kanäle (welche einen Einladungslink für einen Beitritt erfordern, denen jedoch in der Praxis oft problemlos beitreten werden kann) sind nicht durch die Nutzungsbedingungen abgedeckt. Zum Zeitpunkt der Erstellung dieses Berichts liegen bei diesem Dienst keine Richtlinien vor, die sich mit Hassrede befassen, noch verbietet Telegram Doxing, obwohl der Dienst dafür kritisiert wurde, dass er »eine Epidemie von politisch motiviertem Doxing beherbergt«. <sup>95</sup>

Forscher:innen stellen zudem fest, dass Telegram zwar als Messenger-Dienst konzipiert wurde, sich jedoch im Lauf der Zeit zu einem Hybrid aus einem Messenger-Dienst und einer Social-Media-Plattform entwickelt hat, da Nachrichten in öffentlichen Kanälen Hunderttausende (oder gar Millionen) Aufrufe erreichen können, und einige Kanäle die Möglichkeit bieten, Beiträge zu kommentieren oder mit Emojis darauf zu reagieren.<sup>96</sup> Darüber hinaus stellte das ISD fest, dass das Teilen von Links zu anderen (etablierten) audiovisuellen Plattformen, darunter Video- und Livestreaming-Websites, bei Rechtsextremist:innen und Verschwörungsideolog:innen auf Telegram besonders beliebt ist.<sup>97</sup> Dies

verdeutlicht erneut die Herausforderung der plattformübergreifenden Verbreitung von misogynen Inhalten, einschließlich der Relevanz der Berücksichtigung von Risiken, die von kleineren und »alternativen« Plattformen ausgehen.

Zuletzt ist ein Mangel an ausreichenden Ressourcen für die Moderation von Inhalten, die in unterschiedlichen lokalen und sprachlichen Kontexten veröffentlicht werden, insbesondere auch in nicht-westlichen Regionen oder nicht-englischen Sprachen ein bekanntes Problem. Ein Bericht der slowakischen Aufsichtsbehörde (*Slovak Council for Media Services*) und der Organisation Reset untersuchte beispielsweise die Rolle der Plattformen im Fall einer Schießerei vor einer LGBTQ+-Bar in Bratislava im Jahr 2022 und stellte fest, dass Facebook die gemeldeten Inhalte an einen beauftragten externen Faktenprüfer weitergeleitet hatte. Allerdings gab es zu diesem Zeitpunkt nur einen einzigen von Facebook beauftragten Faktenprüfer für die gesamte Slowakei. Dies zeigt, wie begrenzte Ressourcen die schnelle und effiziente Überprüfung von Inhalten, die gegen die Community-Richtlinien verstoßen, erschweren.<sup>98</sup>

---

## Plattformdesign und -systeme: Risiken der Reproduktion und Verstärkung von OGBV

Die Untersuchung der Auswirkungen von Plattformdesign und -systemen auf die Verstärkung des Risikos von geschlechtsspezifischer Online-Gewalt stellt nach wie vor eine Herausforderung dar, da sich die Nutzer:innen in personalisierten Online-Räumen bewegen. In diesem Abschnitt wird untersucht, wie Plattformdesign und -systeme das Risiko bergen, OGBV zu reproduzieren und zu erhöhen. Dabei werden das Design von Nutzungsoberflächen sowie auf künstlicher Intelligenz (KI) basierende Systeme bewertet, einschließlich der Verwendung von Algorithmen und Modellen des maschinellen Lernens (*machine learning, ML*) für die Erstellung von personalisierten Feeds und (Such-)Empfehlungen.

**Gestaltung der Nutzungsoberfläche:** Ein relativ gut erforschtes Phänomen, das zeigt, wie Design den Nutzer:innen schaden kann, sind sogenannte täuschende Muster, auch bekannt als »dark patterns«. Es gibt viele Formen von täuschenden Designmustern<sup>99</sup>, aber im Wesentlichen handelt es sich um »Entscheidungen, die Nutzer:innen [unabsichtlich oder absichtlich] beeinflussen oder dazu bringen können, unbeabsichtigte Entscheidungen zu treffen«. <sup>100</sup> Im Jahr 2021 untersuchten Caroline Sindors, Vandinika Shukla und Elyse Voegeli in einer interdisziplinären Studie die Interaktionen und Beziehungen von Journalist:innen mit den Nutzungsoberflächen von Plattformen, da sie eine Gruppe darstellen, die online einer Reihe von Belästigungen und Schäden ausgesetzt ist, darunter geschlechtsspezifische oder rassistische Beleidigungen, Doxing und Vergewaltigungsdrohungen. <sup>101</sup> Ihre Forschung unterstreicht, dass »Technologie ein geplanter Raum ist, in dem die Nutzer:innen nur bestimmte Handlungen ausführen können, die von der Software, Anwendung oder Plattform, die sie nutzen, vorgesehen und erlaubt sind«. <sup>102</sup> Das Design von Online-Räumen ist dabei eng mit Diskriminierungserfahrungen der Nutzer:innen und Risiken von OGBV verknüpft. Täuschendes Design kann sich demnach negativ auf die Privatsphäre und die Sicherheit der Nutzer:innen auswirken, dazu zählen:

- Einstellungen, die standardmäßig auf die am wenigsten datenschutzfreundliche Option eingestellt sind;
- Belohnungen und Einschränkungen, wenn Nutzer:innen die Einstellungen ablehnen oder sich dagegen entscheiden, beispielsweise Verlust von bestimmten Funktionen;

- Erzwungene Durchführung der Einstellungsüberprüfung zu einem von der Plattform festgelegten Zeitpunkt, wodurch die Nutzer:innen unter Druck gesetzt werden, ohne dass sie eine klare Möglichkeit haben, den Prozess auf einen späteren Zeitpunkt zu verschieben;
- Eine Illusion von Kontrolle, da die Plattform den Nutzer:innen übermäßig detaillierte Auswahlmöglichkeiten bietet, die sie letztlich davon abhalten, ihre Einstellungen selbst zu ändern oder zu kontrollieren.

Täuschendes Design unterstreicht die Notwendigkeit eines »Safety by Design«-Ansatzes, der Plattformen dazu ermutigt, die Sicherheit bereits bei der Gestaltung, der Entwicklung und der Bereitstellung ihrer Funktionen zu integrieren, anstatt Sicherheitslösungen erst nachträglich einzubauen, nachdem Schäden bereits entstanden sind. <sup>103</sup>

Die Studie zeigt zudem auf, dass sowohl Meldemechanismen, bei denen die Opfer und Überlebenden im Mittelpunkt stehen als auch transparente Kommunikation von Community-Richtlinien fehlen. Die befragten Journalist:innen erlebten eine Vielzahl von Belästigungen, wobei es sich häufig nicht um Einzelfälle handelte, sondern um Muster von Belästigungen. Die Betroffenen waren jedoch nur in der Lage, einzelne Fälle zu melden, nicht aber mehrere Fälle als Teil einer Belästigungskampagne. Die Journalist:innen äußerten zudem »Frustration und Verwirrung darüber, wie die Plattformen auf belästigende Inhalte reagierten«. Es ist wichtig zu bedenken, dass eine eingeschränkte Handlungsfähigkeit der Nutzer:innen den erlebten Verlust von Macht, Kontrolle und Rechten verstärken kann – eine häufige Erfahrung für Opfer und Überlebende von OGBV. Die Koordinierung von Belästigungskampagnen über verschiedene Plattformen hinweg und das Fehlen von interoperablen Meldemechanismen schwächen die Handlungsfähigkeit der Nutzer:innen und die Risikominderung weiter ab.

**KI-basierte Systeme:** Personalisierte Feeds und andere KI-gestützte Funktionen wie Suchempfehlungen beeinflussen die Erfahrungen der Nutzer:innen und bergen besondere Risiken für OGBV. Algorithmische Ranking-Systeme treffen automatisierte Entscheidungen, die Inhalte und Vorschläge, mit wem sich Nutzer:innen verbinden sollten (oder welchen Seiten oder Gruppen sie folgen sollten), ordnen. Dabei verwenden die eingesetzten ML-Modelle bestimmte Signale wie das historische Verhalten der Nutzer:innen und Wahrscheinlichkeiten,

die aus dem Verhalten ähnlicher Nutzer:innen abgeleitet werden (unter Verwendung von Techniken wie dem kollaborativen Filtern).<sup>104</sup> ML-Modelle bewerten Inhalte üblicherweise nach dem zu erwartenden Engagement (also die Wahrscheinlichkeit, dass Nutzer:innen Inhalte liken, teilen, ansehen, etc.) und weisen ihnen auf dieser Grundlage ein entsprechendes Ranking zu. So soll die Reihenfolge, in der die Inhalte in den Feeds der Nutzer:innen angezeigt werden, optimiert werden. Das Ziel ist, ein von dem Unternehmen festgelegtes Interaktionsziel (Metrik) zu maximieren, oft auf der Ebene individueller Indikatoren (beispielsweise die Zeit, die Nutzer:innen auf einer Plattform verbringen).

Auch wenn die Untersuchung der Auswirkungen algorithmischer Feeds auf Diskurse und Nutzungsverhalten nach wie vor eine Herausforderung darstellt, gibt es Hinweise auf ein »Engagement-Problem«, das die Tendenz beschreibt, mehr mit Inhalten zu interagieren, die von geringer Qualität sind (beispielsweise Clickbait-Schlagzeilen) oder sich einem »Grenzwert« (*cut-off point*) dessen nähern, was gemäß den Community-Richtlinien erlaubt ist.<sup>105</sup> Daher besteht bei Feeds mit einem auf Engagement basierten Ranking oft die Gefahr, dass eine »Anziehungskraft« auf grenzwertige Inhalte (*borderline content*) entsteht, was auch das Risiko erhöht, dass misogynen Inhalte amplifiziert werden.<sup>106</sup> Eine Untersuchung des ISD legt beispielsweise nahe, dass Plattformen missbräuchlichen Hashtags eine größere Sichtbarkeit einräumen als nicht missbräuchlichen Hashtags. Auf Instagram lag der transphobische Hashtag #rachellevineisaman (gelesen als rachel-levine-is-a-man) mit 25 Beiträgen an dritter Stelle der Empfehlungen, vor nicht missbräuchlichen Hashtags, die in mehr Beiträgen vorkamen (beispielsweise #rachellevinephotography mit 183 Beiträgen).<sup>107</sup> Zudem zeigen Untersuchungen, dass Tradwives in der Lage sind, »ihre Inhalte anzupassen«, um algorithmische Feeds auszunutzen, indem sie Self-Branding-Strategien anwenden, die »#tradlife« präsentieren und »Inhalte zu den Themen Hausarbeit und Beauty« teilen, um ihr Publikum anzusprechen, während sie gleichzeitig antifeministische und anti-LGBTQ+-Ansichten verbreiten. Tradwives nutzen auch das »Engagement-Problem« aus, indem sie ihre extremistische Ideologie durch Werbung, Markenkooperationen oder Werbeaktionen kommerzialisieren.<sup>108</sup>

Bei der Untersuchung von KI-basierten Systemen im Allgemeinen gehen Forscher:innen davon aus, dass

Algorithmen bereits in der Gesellschaft vorhandene Geschlechternormen widerspiegeln und verschärfen. Geschlechterstereotypen und geschlechtsspezifische Diskriminierung wurden unter anderem beim Einsatz von Algorithmen für Personaleinstellungsentscheidungen,<sup>109</sup> für die Verurteilung von Straftäter:innen,<sup>110</sup> und für die Zuweisung von Gesundheitsdiensten<sup>111</sup> festgestellt. Forscher:innen der New York University (NYU) zeigen auf, dass gesellschaftliche Ungleichheiten in Suchalgorithmen ersichtlich sind, und stellen einen »Kreislauf der Ausbreitung von Vorurteilen zwischen Gesellschaft und KI« fest.<sup>112</sup> Dies unterstreicht die Notwendigkeit, sich mit den gesellschaftlichen Auswirkungen der KI auseinanderzusetzen, einschließlich der Frage, wie OGBV durch KI-basierte Systeme reproduziert und verstärkt wird und wie dies eingedämmt werden kann. Dies erfordert interdisziplinäre Forschung und die Entwicklung von On- und Offline-Maßnahmen, die von allen betroffenen Stakeholdern getragen werden.

Wie erwähnt, setzen Plattformen Algorithmen und ML-Modelle ein, um anhand großer Mengen von Daten Vorhersagen über die Interaktionen ihrer Nutzer:innen zu treffen. Bei automatisierten Schlussfolgerungen besteht die Gefahr, dass die Ergebnisse verzerrt werden, einschließlich geschlechtsspezifischer Stereotypen.<sup>113</sup> So haben Forscher:innen beispielsweise die Berechnungen von X (ehemals Twitter) mit Blick auf die Geschlechtsidentität der Nutzer:innen getestet. Sie stellten fest, dass LGBTQ+-Personen und heterosexuelle Frauen häufiger falsch gegendert wurden als heterosexuelle Männer. Die Forscher:innen betonen dabei, dass das falsch gendern der Nutzer:innen »nicht nur tief verwurzelte Stereotypen widerspiegelt, sondern auch zu Problemen mit dem Datenschutz und der Diskriminierung führen kann«.<sup>114</sup>

Rebekah Tromble, Direktorin des Institute for Data, Democracy and Politics an der George Washington University, beschreibt das Problem wie folgt: »Die Art und Weise, wie wir Inhalte konsumieren, ist ein inhärent menschliches Konstrukt. Und wenn es Probleme mit diesem Konsum gibt, dann liegt das an aktiven Entscheidungen von Manager:innen und Ingenieur:innen der Plattformen – und nicht an einem natürlichen Phänomen, auf das niemand Einfluss hat«.<sup>115</sup> Wie Plattformen ihre Algorithmen gestalten und bewerten, wirkt sich unmittelbar auf die Erfahrungen und Sicherheit der Nutzer:innen aus, zumal die Gefahr besteht, dass sie die Verbreitung misogynen Inhalte verstärken.

## Empfohlene Maßnahmen: Risikobewertung und Eindämmung von OGBV

Auf der Grundlage der Diskussion über die Phänomene geschlechtsspezifischer Online-Gewalt und der Bewertung der Richtlinien, des Design und der Systeme der Plattformen wird in den folgenden Abschnitten erläutert, wie Plattformen auf eine bessere Risikobewertung und Eindämmung von OGBV, einschließlich gewalttätiger misogynen Inhalte, hinarbeiten können. In Anbetracht der Notwendigkeit von Ansätzen und Lösungen, an denen alle betroffenen Stakeholder beteiligt sind, werden Vorschläge für Maßnahmen auf multilateraler und staatlicher Ebene sowie von der Industrie, der Zivilgesellschaft und der Wissenschaft erörtert.

### Ermöglichung des API-Zugangs zu Plattformdaten und Entwicklung standardisierter Transparenzberichte

Transparenzberichte der Plattformen, die geschlechtsspezifische Online-Gewalt explizit einbeziehen, sind Voraussetzung, um die Art, das Ausmaß und den Umfang des Phänomens nachzuvollziehen und zu erklären. Zusätzlich sollte ein gesicherter API-Zugriff auf öffentlich zugängliche Daten<sup>116</sup> die im öffentlichen Interesse liegende Forschung<sup>117</sup> unterstützen und somit eine faktengestützte Entscheidungsfindung ermöglichen.<sup>118</sup>

Die von den USA einberufene Global Partnership hat gemeinsam mit UN Women, der WHO, dem UNFPA und UNICEF Anstrengungen unternommen, um die Erstellung genauer, zuverlässiger und vergleichbarer Daten und Kenntnisse über OGBV zu ermöglichen.<sup>119</sup> Im Jahr 2023 veröffentlichte das gemeinsame Programm von UN Women und WHO einen Bericht zu den vorliegenden Daten im Zusammenhang mit Gewalt gegen Frauen (*Violence against Women, VAW*), in dem Methoden und Empfehlungen zu den Ansätzen für die Erhebung von Daten über technologiegestützte Gewalt gegen Frauen (*Technology-Facilitated Violence against Women, TFVAW*) zusammengefasst sind.<sup>120</sup> Der Bericht beleuchtet existierende Methoden<sup>121</sup> sowie methodische, ethische und gesellschaftspolitische Herausforderungen bei der Datensammlung. Dazu gehören das Fehlen einer »allgemeinen Problematisierung und eines Bewusstseins« rund um das Thema TFVAW aufgrund mangelnder Daten und der fehlenden Verbreitung von Forschungsergebnissen sowie die einseitige Ausrichtung der Daten auf den globalen Norden, wodurch

die unterschiedlichen Auswirkungen in vielfältigen und verschiedenen Zusammenhängen vernachlässigt werden. Der Bericht unterstreicht zudem die Notwendigkeit einer gemeinsamen operativen Definition und Methodik für die Messung und Analyse von TFVAW. Ferner wird darauf hingewiesen, wie wichtig die Einbeziehung von Plattformdaten ist und dass eine »Vielfalt von Methoden« in Betracht gezogen werden muss, um verschiedene Datenquellen zu berücksichtigen.

In diesem Zusammenhang sollte die Sammlung und Analyse von Plattformdaten Lücken in der Forschung schließen, um die Taktiken und Formen von OGBV sowie die Verbindung zwischen misogynen und verschiedenen extremistischen Ideologien zu untersuchen. Ein verbesserter und gesicherter API-Zugang könnte beispielsweise Forschung ermöglichen, die die geschlechtsspezifischen Dimensionen gewalttätiger extremistischer und terroristischer Online-Aktivitäten vergleichend – über verschiedene Ideologien hinweg – nachverfolgt und untersucht.<sup>122</sup> Zugelassene Forscher:innen aus verschiedenen Disziplinen wie etwa der Computerlinguistik, Kritischen Terrorismusforschung oder der Kritischen Männer- und Männlichkeitsforschung sollten über einen API-Zugang verfügen, um systematisch relevante Plattformdaten zu sammeln und zu analysieren.

Die rechtlichen Rahmenbedingungen tragen der Notwendigkeit eines solchen Datenzugangs bereits Rechnung. Artikel 40 des Gesetzes über digitale Dienste (*Digital Services Act, DSA*) der Europäischen Union schreibt beispielsweise vor, dass der Zugang zu Daten, die »in ihrer Online-Nutzungsoberfläche öffentlich zugänglich sind«, Forscher:innen, einschließlich derjenigen, die gemeinnützigen Einrichtungen, Organisationen und Verbänden angehören, nach Möglichkeit in Echtzeit zur Verfügung gestellt werden sollte. Parallel dazu verpflichteten sich die Unterzeichnenden des gestärkten Verhaltenskodex für Desinformation (*Strengthened Code of Practice on Disinformation*) der EU zu freiwilligen Standards, die als Ko-regulierungsmaßnahmen für den DSA dienen werden. Der Kodex beinhaltet unter anderem die Verpflichtung zu »kontinuierlichem, Echtzeit- oder echtzeitnahem, durchsuchbarem, stabilem Zugang zu nicht-personenbezogenen Daten und anonymisierten, aggregierten oder offenkundig öffentlich gemachten Daten zu Forschungszwecken über Desinformation

durch automatisierte Mittel wie APIs«. <sup>123</sup> Die verfügbaren Plattformdaten sollten mit Daten aus anderen Quellen wie Verwaltungsdaten, Statistiken oder Umfragen verglichen und trianguliert werden, um eine umfassende Abbildung der Phänomene zu gewährleisten.

Darüber hinaus sollten die Plattformen eine standardisierte Transparenzberichterstattung entwickeln, die nach Geschlecht aufgeschlüsselte Daten (*gender-disaggregated data*) mit Blick auf Richtlinienverstöße enthält, damit Forscher:innen die Durchsetzung von Richtlinien prüfen und verfolgen können. So sollten Transparenzberichte beispielsweise Statistiken über die Verbreitung von und die Interaktionen der Nutzer:innen mit Inhalten einbeziehen, die als geschlechtsspezifische Hassrede eingestuft wurden. Zudem sollten Plattformen Informationen über den Anteil an bildbasierten Inhalten, die gegen diese Richtlinien verstoßen, sowie Daten darüber, wie auf Meldungen von Nutzer:innen reagiert wurde (beispielsweise welche spezifischen Maßnahmen ergriffen wurden) in Transparenzberichte aufnehmen. <sup>124</sup> Die Berichte sollten Hassrede berücksichtigen, die sich gegen andere geschützte Gruppen richten, um den sich überschneidenden identitätsbasierten Hass zu messen und eine intersektionale Analyse der Beweggründe für OGBV zu unterstützen.

Die Plattformen sollten bei der Entwicklung von Methoden für Transparenzberichte (einschließlich Inhaltskategorien und Metriken) und bei der Durchführung interner Forschung (beispielsweise Umfragen zu Erfahrungen mit OGBV) mit Feminismus-Forscher:innen, Frauenrechtler:innen, GBV-Expert:innen und Opfern und Überlebenden von OGBV zusammenarbeiten. Da derzeit keine allgemein gültige Definition von geschlechtsspezifischer Online-Gewalt vorliegt und die Notwendigkeit einer kohärenteren Transparenzberichterstattung (und damit der Vergleichbarkeit von Maßnahmen der unterschiedlichen Plattformen) besteht, könnten sektorübergreifende Bemühungen zu den Arbeiten von UN Women <sup>125</sup> zur Entwicklung einer Terminologie und eines statistischen Rahmens für TFGBV sowie der UN-Sonderberichterstatteerin für Meinungsfreiheit zur Entwicklung einer gemeinsamen Definition für geschlechtsspezifische Desinformation beitragen und sich an diesen beteiligen. <sup>126</sup>

### Anwendung eines traumainformierten »Safety and Privacy by Design«-Ansatzes

In wissenschaftlichen und politischen Debatten wird anerkannt, dass Maßnahmen ergriffen werden müssen, um kurzfristig Abhilfe zu schaffen und die Risiken von OGBV einzudämmen. Die Gestaltung von Nutzungsoberflächen kann zwar die Handlungsfähigkeit und Sicherheit der Nutzer:innen untergraben (durch die beschriebenen täuschenden Muster), aber sie kann Nutzer:innen auch befähigen, sich gegen Misogynie und geschlechtsspezifischen Hass zu schützen.

Forschung von Sinders, Shukla und Voegeli (2021) <sup>127</sup> und PEN America <sup>128</sup> unterstreicht bei der Überprüfung und Bewertung der Sofortmaßnahmen die Notwendigkeit, dass die Plattformen verbesserte Werkzeuge für Nutzer:innen einführen, um die Risiken von OGBV zu verringern. Die Forscher:innen empfehlen proaktive Maßnahmen, die es den Nutzer:innen ermöglichen, Risiken und Gefährdungen zu verringern; reaktive Handlungen, die wirksamere Sofortmaßnahmen ermöglichen, wenn Nutzer:innen mit OGBV konfrontiert werden; sowie Maßnahmen zur Durchsetzung von Rechenschaftspflichten, die darauf abzielen, missbräuchliches Verhalten zu verhindern und die Täter davon abzuhalten, Plattformen zur Vernetzung und koordinierten Belästigung auszunutzen.

Ein »Safety and Privacy by Design«-Ansatz stellt die Handlungsfähigkeit der Nutzer:innen bei der Entwicklung und Gestaltung von Plattformprodukten und -diensten in den Mittelpunkt. Die vorgeschlagenen Maßnahmen sind nicht in ihrer Vollständigkeit aufgeführt, und weitere Untersuchungen sind erforderlich, um ihre potenzielle Wirksamkeit zu bewerten.

### Proaktive Maßnahmen könnten Folgendes beinhalten:

- Tools zur Inhaltsmoderation wie »Schutzschilder«, die es den Nutzer:innen ermöglichen, missbräuchliche Inhalte (in Feeds, Kommentaren, Direktnachrichten etc.) proaktiv zu filtern und sie in einem Dashboard unter Quarantäne zu stellen, wo sie diese gemeinsam mit vertrauenswürdigen Verbündeten überprüfen und auf sie geeignet reagieren können;
- Robuste, intuitive und nutzungsfreundliche Funktionen, die eine Feinabstimmung der Datenschutz- und Sicherheitseinstellungen ermöglichen, einschließlich »Sichtbarkeits-Snapshots«, die in Echtzeit zeigen, wie sich die geänderten Einstellungen auf die Reichweite auswirken;
- Strukturen, die es den Nutzer:innen ermöglichen, Teams aus vertrauenswürdigen Verbündeten zusammenzustellen, die bei Bedarf schnell reagieren, einschließlich der Delegation des Kontozugangs.

### Reaktive Maßnahmen könnten Folgendes beinhalten:

- Notfall-Hotlines, über die die Nutzer:innen in Echtzeit traumainformierte Unterstützung erhalten können;
- Dokumentationsfunktionen, die es den Nutzer:innen ermöglichen, Beweise für OGBV schnell und einfach festzuhalten (beispielsweise mithilfe einer sofortigen Erfassung von Screenshots, Hyperlinks und anderen öffentlich zugänglichen Daten), die interoperabel gemacht werden sollten, um eine plattformübergreifende Beweisaufnahme zu ermöglichen;
- Verbesserte und standardisierte Funktionen zum Sperren von Kontakten, Stummschalten von Inhalten und Einschränken oder Ausblenden von Inhalten;
- Verbesserte Meldemechanismen, einschließlich Massenmeldungen in Anerkennung des koordinierten Charakters von Belästigungskampagnen, sowie zirkuläre Meldungen, die es ermöglichen, eine Meldung erneut zu öffnen und zu bearbeiten, und zwar plattformübergreifend.

### Maßnahmen der Rechenschaftspflicht könnten Folgendes beinhalten:

- Ein transparentes System »eskalierender Strafen« für missbräuchliches Verhalten, einschließlich Verwarnungen, »Strikes«, vorübergehender Funktionseinschränkungen, Kontosperrungen, der Entfernung von Inhalten und der dauerhaften Sperrung von Konten. In Bezug auf Letzteres haben Forscher:innen festgestellt, dass »die Löschung von Konten das Problem der Rechenschaftspflicht unter Umständen nicht an der Wurzel packt«, während gleichzeitig betont wird, dass Sperrungen Metadaten und Kontoinformationen für die Beweisaufnahme und für Zwecke der Rechenschaftspflicht aufbewahren sollten;<sup>129</sup>
- Testen sogenannter »proaktiver Anstöße (nudges)«, die Nutzer:innen dazu bewegen sollen, missbräuchliche Inhalte zu überarbeiten, bevor sie sie veröffentlichen (sowie Forschung zur Messung der Wirksamkeit von diesen Anstößen);
- Beschwerdeverfahren mit ausreichenden Mitteln, um eine klare und zeitnahe Prüfung von Beschwerden zu gewährleisten.

Gesetzliche Rahmenbedingungen haben bereits die Notwendigkeit für mehr Handlungsfähigkeit der Nutzer:innen erkannt, und verlangen von den Plattformen mehr Rechenschaft und Transparenz in Bezug auf ihr Design und ihre Richtlinien. Das australische Gesetz zur Online-Sicherheit (*Online Safety Act*) aus dem Jahr 2021 verweist beispielsweise auf die »grundlegenden Erwartungen an die Online-Sicherheit« (*Basic Online Safety Expectations*). Diese verlangen von den Plattformen, klare und leicht erkennbare Mechanismen zur Meldung von Inhalten und zum Einlegen von Beschwerden sowie Nutzungsbedingungen, Richtlinien und Verfahren für die Bearbeitung von Beschwerden und Meldungen einzuführen. Diese Anforderungen verlangen zudem, dass die Plattformen Daten über Meldungen und Beschwerden der Nutzer:innen für einen Zeitraum von fünf Jahren aufbewahren.<sup>130</sup>



## Harassment Manager von Googles Jigsaw

Der von der Google-Tochter Jigsaw entwickelte Harassment Manager ist eine »eine Open-Source-Webanwendung, die es Nutzer:innen ermöglicht, gegen sie gerichtete Belästigung in sozialen Medien zu dokumentieren und zu verwalten«, angefangen bei X (ehemals Twitter), welches sich an dem Projekt beteiligte. Das Tool soll Nutzer:innen dabei helfen, »schädliche Beiträge zu identifizieren und zu dokumentieren, Täter stumm zu schalten oder zu blockieren und belästigende Antworten auf ihre eigenen Tweets zu verbergen«. Nutzer:innen können Tweets auf der Grundlage von Hashtags, Profilnamen, Stichwörtern oder Datum überprüfen und die Perspective API nutzen, um Kommentare zu erkennen, die mit hoher Wahrscheinlichkeit toxisch sind (dies wird weiter unten diskutiert). Der Code des Harassment Manager ist auf Github verfügbar<sup>131</sup> und kann von Entwickler:innen und Organisationen kostenlos eingebaut und angepasst werden. Dieses Werkzeug sollte von Feminismus-Forscher:innen, Frauenrechtler:innen, GBV-Expert:innen sowie von Opfern und Überlebenden von mehrdimensionaler Diskriminierung getestet und geprüft werden, um die weitere Verbesserung und Entwicklung zu ermöglichen.

## Verstärkung der plattformübergreifenden Zusammenarbeit und des Informationsaustauschs

Die Plattformen sollten sich darüber im Klaren sein, dass das, was auf anderen Plattformen geschieht, sich auch auf ihre eigenen Dienste (und umgekehrt) ausweiten kann. Dies gilt nicht nur für TVEC, sondern auch für Akteur:innen und Taktiken von OGBV. Online-Belästigungskampagnen, die auf eine Einzelperson abzielen, können auf einer Plattform koordiniert werden, wobei die Inhalte oder URLs zu diesen Inhalten daraufhin auf anderen Plattformen geteilt werden, unabhängig davon, ob die betroffenen Nutzer:innen auf diesen Plattformen ein Konto haben. Wie bereits erwähnt, kann die Belästigung oder die Koordinierung der Belästigung auch auf kleineren und »alternativen« Plattformen stattfinden.

Plattformen sollten Kanäle für einen Austausch zwischen den zuständigen Teams, einschließlich der

Moderationsteams, entwickeln und betreiben, um proaktiv Informationen über plattformübergreifende Belästigung (beispielsweise Täter, die mehrere Konten auf verschiedenen Plattformen nutzen) auszutauschen – hierzu sollten gegebenenfalls auch Meldungen von Nutzer:innen über plattformübergreifende Belästigung gehören. Plattformen sollten auch interoperable Melde-mechanismen entwickeln, um den Nutzer:innen eine effiziente Reaktionsmöglichkeit zur Verfügung zu stellen. Austauschkanäle können ein schnelleres Handeln gewährleisten, wenn beispielsweise festgestellt wird, dass ein prominenter, misogyner Account mit wiederholtem schädlichem Verhalten in Verbindung gebracht wird, etwa wenn dieser Account plattformübergreifend gegen Community-Richtlinien verstößt. Bemühungen dieser Art sind wichtig, um das Ausmaß und die Art von OGBV zu verstehen, aber auch, um die Abhilfemaßnahmen von Plattformen und anderen Stakeholdern zu koordinieren, einschließlich der Regierungen, der Zivilgesellschaft oder gegebenenfalls der Strafverfolgungsbehörden.

Bereits bestehende plattformübergreifende Krisenprotokolle wie das Content Incident Protocol des Global Internet Forum to Counter Terrorism<sup>132</sup> und das Christchurch Call Crisis Response Protocol<sup>133</sup> sollten prüfen, inwiefern OGBV für ihren jeweiligen Anwendungsbereich und ihr Mandat relevant ist und wie die Mechanismen gestärkt werden können, indem Misogynie als ein Radikalisierungsvektor für gewalttätigen Extremismus anerkannt wird. Einschlägige freiwillige Verpflichtungen oder Ko-regulierungsrahmen könnten ebenfalls überprüft werden. So enthält der Verhaltenskodex für Desinformation der EU die Verpflichtung, »Kanäle für den Austausch zwischen den zuständigen Teams einzurichten, um proaktiv Informationen über plattformübergreifende Einflussoperationen (*influence operations*), ausländische Eingriffe in den Informationsraum und relevante Vorfälle miteinander zu kommunizieren, um somit deren Verbreitung zu verhindern«. <sup>134</sup> Die entsprechenden Kommunikationskanäle könnten sich zum Beispiel auf Vorfälle von geschlechtsspezifischen Belästigungskampagnen erstrecken. Bemühungen dieser Art könnten auch für die Durchsetzung des Gesetzes über digitale Dienste der EU von Vorteil sein, der die Plattformen dazu verpflichtet, systemische Risiken im Zusammenhang mit OGBV zu bewerten und zu mindern.

## Überprüfung und Aktualisierung von Richtlinien, Verfahren und Systemen zur Inhaltsmoderation

Plattformen sollten prüfen, inwiefern ihre Moderationsrichtlinien und -verfahren von patriarchalischen Geschlechternormen beeinflusst werden (und diese reproduzieren). Community-Richtlinien und Inhaltsmoderation sollten eine geschlechtsspezifische und traumainformierte Sichtweise in ihre Entwicklung, Verfahren und Systeme integrieren. Dabei sollten Richtlinien (und deren Durchsetzung) für das Kontinuum von Gewalt und für die Verbindungen zwischen Misogynie und gewalttätigem Extremismus sensibilisiert werden.

Die zuständigen Teams der Plattformen sollten die Durchführung von Interviews und Fokusgruppen mit Opfern und Überlebenden in Betracht ziehen, damit diese Sichtweisen bei der Formulierung ihrer Richtlinien und Durchsetzungsprozesse berücksichtigt werden. Das mangelnde Verständnis der Moderator:innen für lokale Sprachen und regionale Zusammenhänge sollte zudem durch die aktive Einbeziehung verschiedener Bevölkerungsgruppen behoben werden. Zivilgesellschaftliche Organisationen schlagen zudem vor, dass Plattformen lokale Organisationen finanziell fördern, die Hassrede überprüfen und lokale Lexika für misogynen Wörter und Ausdrücke entwickeln, einschließlich der Unterstützung bei Traumata.<sup>135</sup>

Ein Beispiel dafür, wie KI-basierte Systeme zur Unterstützung der Inhaltsmoderation eingesetzt werden könnten, ist die Perspective API<sup>136</sup> von Google's Jigsaw. Diese verwendet ML-Modelle, um missbräuchliche Online-Kommentare zu erkennen. Perspective API sagt die wahrgenommene Auswirkung eines Kommentars auf ein Gespräch voraus, indem es diesen Kommentar anhand einer Reihe von emotionalen Konzepten (Attributen) bewertet. Derzeit kann Perspective API Bewertungen für die Attribute »Toxizität«, »Schwere Toxizität«, »Beleidigung«, »Obszönität«, »Identitätsangriff«, »Bedrohung« und »Sexuell explizit« liefern. Dieses Tool soll Moderator:innen dabei helfen, gemeldete Kommentare schnell zu priorisieren und zu überprüfen, den Kommentator:innen Feedback zu geben und den Nutzer:innen die Kontrolle darüber zu ermöglichen, welche Kommentare sie sehen.

Perspective API gibt einen Ausblick auf den Einsatz von KI-gestützten Systemen zur Verbesserung der Inhaltsmoderation. Da sich die Entwicklung und Anwendung

solcher Werkzeuge noch in der Entwicklungsphase befindet, haben Forscher:innen Perspective API getestet, um den Grad der vermeintlichen Toxizität von Tweets prominenter Drag Queens in den USA zu messen. Die Untersuchung legt nahe, dass Perspective eine signifikante Anzahl von Drag-Queen-Accounts als toxischer einstufte als Accounts von weißen Nationalisten. Somit war das Tool nicht in der Lage, den sozialen Kontext bei der Messung der Toxizität zu berücksichtigen – beispielsweise wurden Fälle nicht erkannt, in denen Wörter, die konventionell als beleidigend angesehen werden könnten, in der LGBTQ+-Sprache eine andere Bedeutung haben.<sup>137</sup> Diese Untersuchungen betonen wiederum die Notwendigkeit einer kontinuierlichen und verstärkten Zusammenarbeit zwischen den verschiedenen Stakeholdern, um auf Brancheninstrumenten wie Perspective API aufzubauen und diese weiterzuentwickeln.

Sowohl die KI-gestützte als auch die menschliche Inhaltsmoderation erfordern umfassende und regelmäßige Überprüfung der Richtlinien, einschließlich traumainformierter Prozesse, um die nuancierten Formen von OGBV zu berücksichtigen und kontraproduktive Ergebnisse zu verhindern. Diese Bemühungen erfordern einen politischen Willen zur Verbesserung der Systeme und sollten nicht durch die Zielvorgaben (Metriken) der Unternehmen, die das Engagement der Nutzer:innen priorisieren, abgeschwächt werden. Stattdessen sollte bei den Risikobewertungen und Maßnahmen der Plattformen das Prinzip »richte keinen Schaden an« (*do no harm*) im Vordergrund stehen, um mögliche negative Auswirkungen ihrer Systeme zu vermeiden.

## Risikobewertung und -minderung von Misogynie in KI-basierten Systemen

Rechtsvorschriften wie der DSA der EU und vorgeschlagene (nicht verbindliche) Richtlinien wie der Verhaltenskodex für Gewalt gegen Frauen und Mädchen (*Violence Against Women and Girls, VAWG*), der im Zusammenhang mit dem britischen Gesetz zur Online-Sicherheit (*Online Safety Bill, OSB*)<sup>138</sup> erarbeitet wurde, zielen darauf ab, die geschlechtsspezifischen Auswirkungen des Plattfordesigns, einschließlich ihrer algorithmischen Feeds, zu bewerten und einzudämmen. Artikel 34 und 35 des DSA fordern die Plattformen ausdrücklich auf, die von ihren Diensten ausgehenden sogenannten systemischen Risiken zu bewerten und einzudämmen, einschließlich jeglicher negativen

Auswirkungen im Zusammenhang mit OGBV. Dabei sollten die Risikobewertungen auch das »Design ihrer Empfehlungssysteme und anderer relevanter algorithmischer Systeme« berücksichtigen.<sup>139</sup> Der VAWG-Kodex betont ebenso einen »Safety by Design«-Ansatz, um sicherzustellen, dass »die im Dienst verwendeten Algorithmen keine hasserfüllten Inhalte fördern und somit vorhersehbaren Schaden verursachen, indem sie beispielsweise misogynen Influencer mit größerer Reichweite belohnen«. Der Kodex argumentiert, dass »Präventivmaßnahmen die Rolle von algorithmischen Produktentscheidungen berücksichtigen müssen« und bekräftigt, dass die Entscheidungsprozesse hinsichtlich ihrer Entwicklung und ihres Einsatzes genau geprüft werden müssen.<sup>140</sup> Die algorithmische Rechenschaftspflicht und Prüfung sollten einen auf die Opfer und Überlebenden ausgerichteten Ansatz verfolgen. Dieser Ansatz involviert beispielsweise die Durchführung von Sicherheitstests sowie eine geschlechtsspezifische und intersektionale Analyse der Funktionalitäten, um zu prüfen, wie Nutzer:innen verstärkt gruppenbezogene Diskriminierung aufgrund der Systeme erleben.

Die Überprüfung und Bewertung der Auswirkungen von Algorithmen bleibt eine Herausforderung, selbst bei direktem Zugang zu proprietärem Code, da algorithmische Feeds personalisiert sind und von vielen Faktoren abhängen, einschließlich der historischen Daten der Nutzer:innen. Darüber hinaus müssen Forscher:innen ein kontrafaktisches Szenario zum Vergleich heranziehen, beispielsweise bei der Durchführung randomisierter kontrollierter Experimente. In einer kürzlich durchgeführten Studie wurden einer Gruppe von Nutzer:innen mit deren Zustimmung umgekehrt chronologisch geordnete Feeds zugewiesen, um die Auswirkungen algorithmischer Feeds zu bewerten, einschließlich der Frage, inwieweit sie negative Einstellungen gegenüber Personengruppen beeinflussen.<sup>141</sup> Im Zusammenhang mit OGBV kann diese Art der Untersuchung der algorithmischen Empfehlungen bei der Entwicklung evidenzbasierter Interventionen in Online-Radikalisierung helfen. So haben beispielsweise männliche Nutzer, die in die Manosphere eintreten, mit einer größeren Wahrscheinlichkeit zunächst weniger extreme Ansichten, und bilden diese erst über neue Verbindungen und »Eigengruppe«-Dynamiken.<sup>142</sup> Hier könnte die Erforschung von KI-basierten Systemen den potenziell algorithmischen Pfad zu misogynen Online-Räumen untersuchen und nachvollziehen.

### Christchurch Call Initiative on Algorithmic Outcomes

Die Christchurch Call Initiative on Algorithmic Outcomes, die von Neuseeland, den USA, X (ehemals Twitter) und Microsoft geleitet wird, zielt darauf ab, Software-Tools zu entwickeln, die eine unabhängige Forschung über die Auswirkungen von algorithmischen Systemen erleichtern.<sup>143</sup> In Zusammenarbeit mit OpenMined, DailyMotion und LinkedIn wird eine neue Software-Infrastruktur Technologien zur Verbesserung des Datenschutzes integrieren, die es externen Forscher:innen und Datenwissenschaftler:innen ermöglicht, Algorithmen sicher aus der Ferne zu untersuchen. Bemühungen dieser Art sind von entscheidender Bedeutung, um eine unabhängige Forschung über die Auswirkungen algorithmischer Feeds zu ermöglichen. Die unabhängige Prüfung von Algorithmen und ML-Modellen über eine solche Software-Infrastruktur sollte sich darauf konzentrieren, die Rolle von algorithmischen Empfehlungen zu verstehen und zu testen, wenn möglich auch plattformübergreifend. Die Entwicklung von Systemen zur Fernabfrage für Forschungszwecke (*remote researcher queries*) erfordert zudem einen angemessenen Verwaltungs- und Ethikrahmen sowie Verfahren zur Festlegung von Forschungsprioritäten.

Forscher:innen legen weiterhin nahe, dass ML-Modelle so entwickelt oder überwacht werden können, dass sie keine diskriminierenden Muster, wie beispielsweise Geschlechterstereotypen, hervorbringen. Die Idee bestehe nicht darin, die verwendeten Daten zu begrenzen (d. h. alle geschlechtsbezogenen Daten zu entfernen), sondern zu verhindern, dass Algorithmen daraus geschlechtsspezifische (diskriminierende) Muster ableiten.<sup>144</sup> Die Risikominderung könnte beispielsweise Maßnahmen zur Eindämmung von vorurteilsbehafteten Inhalten (wie traditionalistische und heteronormative Rollenverständnisse) in den Trainingsdatensätzen eines Algorithmus beinhalten.<sup>145</sup> Grundsätzlich sollten Transparenz und Inklusivität durch eine Einbeziehung von intersektionalen feministischen Perspektiven bei der Entwicklung und Verwendung von Algorithmen gesichert werden.

## Fazit

Die Normalisierung misogynen Gewalt, die Belästigung und Einschüchterung von Frauen und LGBTQ+-Personen sowie die Aufrechterhaltung patriarchalischer Geschlechternormen sind in »größere Muster systemischer Gewalt eingebettet, die darauf abzielen, die Autonomie der betroffenen Person zu kontrollieren, zu erniedrigen und erheblich einzuschränken«.<sup>146</sup> Online-Manifestationen von GBV untergraben die Sicherheit, die freie Meinungsäußerung und die Teilnahme am öffentlichen Leben von Frauen und der LGBTQ+-Community.

In diesem Zusammenhang betont dieser Bericht das Kontinuum der Gewalt, innerhalb dessen Misogynie als ideologisches Bindeglied zwischen verschiedenen Formen des gewalttätigen Extremismus dienen kann. Eine Anerkennung struktureller patriarchalischer Normen innerhalb der Gesellschaft und der Gefahr von Misogynie als Ideologie, die als Ausgangspunkt für Radikalisierung dienen kann, sollte sich in strukturellen Ansätzen and Maßnahmen der Plattformen, Regierungen und Zivilgesellschaft niederschlagen.

Neben unmittelbaren Maßnahmen zur Unterstützung der Handlungsfähigkeit von Nutzer:innen, bei der die Opfer und Überlebenden im Mittelpunkt stehen, sollten Plattformen die durch OGBV verursachten individuellen und gesellschaftlichen Risiken bewerten und eindämmen. Dabei sollten die Community-Richtlinien und Risikobewertungen auch die Verbindung zwischen misogynen und gewalttätigen extremistischen Inhalten berücksichtigen. Lücken in der Forschung zu gewalttätigem Extremismus und Misogynie verdeutlichen, dass diese komplexen Phänomene weiter untersucht werden

müssen, unter anderem durch einen verbesserten Datenzugang und eine standardisierte Transparenzberichterstattung. Evidenzbasierte Entscheidungen und Maßnahmen sind von zentraler Bedeutung, um mögliche negative Folgen für die Gleichstellung der Geschlechter und den Schutz der Meinungsfreiheit zu vermeiden.

Plattformen sollten inklusive Community-Richtlinien entwickeln und ausreichend in eine konsequente Durchsetzung investieren. Sie sollten Risiken von OGBV bewerten und eindämmen, die sich aus der Funktionsweise ihrer Systeme ergeben, zu denen auch algorithmische Empfehlungssysteme gehören. Insbesondere sollten die Produktteams prüfen und messen, inwiefern algorithmische Systeme patriarchalische Geschlechternormen widerspiegeln und reproduzieren, die wiederum die Verbreitung von misogynen Inhalten verstärken können.

Ein traumainformierter »Safety and Privacy by Design«-Ansatz sollte zudem Sicherheitstests gewährleisten, bei denen mehrdimensionale Perspektiven und Diskriminierungserfahrungen berücksichtigt werden, bevor neue Dienste und Funktionen eingeführt werden. Ein Zugang zu Plattformdaten über APIs für zugelassene Forscher:innen bleibt nach wie vor von wichtiger Bedeutung, um Daten zu sammeln, Forschungsergebnisse sinnvoll zu interpretieren und Plattformen zur Rechenschaft zu ziehen. Diese Empfehlungen sollten als Ergänzung zu gesamtgesellschaftlichen und gesamtstaatlichen Lösungen verstanden werden. Plattformen können hier eine entscheidende Rolle bei der Unterstützung solcher Bemühungen spielen.

# Endnoten

- 1 Public Service Commission of New Zealand. (n.d.). Diversity and inclusion glossary. Abrufbar unter: <https://www.publicservice.govt.nz/guidance/glossary/diversity-and-inclusion/>
- 2 Office of the United Nations High Commissioner for Human Rights (OHCHR). (n.d.). Gender stereotyping. Abrufbar unter: <https://www.ohchr.org/en/women/gender-stereotyping>
- 3 European Commission. (n.d.). Gender-based violence. Abrufbar unter: [https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence\\_en](https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-based-violence_en)
- 4 UN Women. (2023). Expert Group Meeting report: Technology-facilitated violence against women: Towards a common definition. Abrufbar unter: <https://www.unwomen.org/en/digital-library/publications/2023/03/expert-group-meeting-report-technology-facilitated-violence-against-women>
- 5 Institute for Strategic Dialogue (ISD). (n.d.). The 'Manosphere'. Abrufbar unter: <https://www.isdglobal.org/explainers/the-manosphere-explainer/>
- 6 Southern Poverty Law Center. (n.d.). Male supremacy. Abrufbar unter: <https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy>
- 7 Connell, R. (1987). Gender and Power. Sydney: Allen and Unwin.
- 8 Manne, K. (2019). Down girl: the logic of misogyny. New York: Oxford University Press.
- 9 Gentry, C. E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Abrufbar unter: <https://doi.org/10.1080/17539153.2022.2031131>
- 10 Bailey, M., & Trudy. (2018). On Misogynoir: Citation, Erasure, and Plagiarism. Feminist Media Studies, 18(4), 1–7. Abrufbar unter: <https://www.tandfonline.com/doi/abs/10.1080/14680777.2018.1447395?journalCode=rfms20>
- 11 Gentry, C.E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Abrufbar unter: <https://doi.org/10.1080/17539153.2022.2031131>
- 12 Manne, K. (2019). Down girl: the logic of misogyny. New York: Oxford University Press.
- 13 United Nations Office of the High Commissioner for Human Rights. (2018). The impact of online violence on women human rights defenders and women's organisations. Abrufbar unter: <https://www.ohchr.org/en/statements/2018/06/impact-online-violence-women-human-rights-defenders-and-womens-organisations?LangID=E&NewsID=23238>
- 14 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Abrufbar unter: <https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics>
- 15 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Abrufbar unter: <https://www.icfj.org/our-work/chilling-global-study-online-violence-against-women-journalists>
- 16 Di Meco, L. (2023). Monetizing Misogyny. ShePersisted. Abrufbar unter: [https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted\\_MonetizingMisogyny.pdf](https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted_MonetizingMisogyny.pdf)
- 17 United Nations Digital Library. (2018). Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective. A/HRC/38/47. Abrufbar unter: <https://digitallibrary.un.org/record/1641160>
- 18 Dieser Bericht wählt die Schreibweise »Täter« anstelle von »Täter:innen«, um anzuerkennen, dass es sich in den allermeisten Fällen von geschlechtsspezifischer Gewalt um männliche Täter handelt. Dabei sollte betont werden, dass die Tatmotive komplex sind und geschlechtsspezifische Gewalt und Misogynie ebenso von Frauen ausgehen kann.
- 19 Gentry, C. E. (2022). Misogynistic terrorism: it has always been here. Critical Studies on Terrorism 15: 209–224. Abrufbar unter: <https://doi.org/10.1080/17539153.2022.2031131>
- 20 Georgetown Institute for Women, Peace and Security. (n.d.). Recognizing the violent extremist ideology of incels. Abrufbar unter: <https://giwps.georgetown.edu/resource/recognizing-the-violent-extremist-ideology-of-incels/>
- 21 U.S. Department of State. (2022). 2022 Roadmap for the global partnership for action on gender-based online harassment and abuse. Abrufbar unter: <https://www.state.gov/2022-roadmap-for-the-global-partnership-for-action-on-gender-based-online-harassment-and-abuse/>
- 22 APC Women's Rights Programme. (2015). Briefing paper on VAW. Abrufbar unter: [https://www.apc.org/sites/default/files/HRC%2029%20VAW%20a%20briefing%20paper\\_FINAL\\_June%202015.pdf](https://www.apc.org/sites/default/files/HRC%2029%20VAW%20a%20briefing%20paper_FINAL_June%202015.pdf)
- 23 Gentry, C. E. (2022). Misogynistic Terrorism: It Has Always Been Here. Critical Studies on Terrorism 15: 209–224. Abrufbar unter: <https://www.tandfonline.com/doi/full/10.1080/17539153.2022.2031131>; Johnston, M. F., M. Iqbal, & J. True. (2020). The Lure of (Violent) Extremism: Gender Constructs in Online

- Recruitment and Messaging in Indonesia. *Studies in Conflict & Terrorism* 0731 (May): 1–19. Abrufbar unter: <https://doi.org/10.1080/1057610X.2020.1759267>;
- Phelan, A. (2021). *Terrorismus, Geschlecht und Frauen: Auf dem Weg zu einer integrierten Forschungsagenda*. Abrufbar unter: <https://www.routledge.com/Terrorism-Gender-and-Women-Toward-an-Integrated-Research-Agenda/Phelan/p/book/9780367623104>;
- Meiering, D., A. Dziri, & N. Foroutan. (2020). Connecting Structures: Resistance, Heroic Masculinity and Anti-Feminism as Bridging Narratives within Group Radicalization." *International Journal of Conflict and Violence (IJCV)* 14 (2): 1–19. Abrufbar unter: <https://doi.org/10.4119/IJCV-3805>
- 24 Phelan, A., White, J., Wallner, C., & Paterson, J. (2023). Hostile Beliefs And The Transmission Of Extremism: A Comparison Of The Far-Right In The UK And Australia. Centre for Research and Evidence on Security Threats (CREST). Abrufbar unter: <https://crestresearch.ac.uk/resources/misogyny-hostile-beliefs-and-the-transmission-of-extremism/>
- 25 Institute for Strategic Dialogue (ISD). (n.d.). The Manosphere. Abrufbar unter: <https://www.isdglobal.org/explainers/the-manosphere-explainer/>
- 26 Southern Poverty Law Center. (n.d.). Male supremacy. Abrufbar unter: <https://www.splcenter.org/fighting-hate/extremist-files/ideology/male-supremacy>
- 27 Ibid.
- 28 Institute for Strategic Dialogue (ISD). (n.d.). Incels. Abrufbar unter: <https://www.isdglobal.org/explainers/incels/>
- 29 Gheorghe, R. M. (2023). "Just Be White (JBW)": Incels, Race and the Violence of Whiteness. *Feminist Inquiry in Social Work*, 1-19. Abrufbar unter: <https://doi.org/10.1177/08861099221144275>
- 30 Institute for Strategic Dialogue (ISD). (n.d.). Incels. Abrufbar unter: <https://www.isdglobal.org/explainers/incels/>
- 31 Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. *Studies in Conflict & Terrorism*, 43(7), 565-587. Abrufbar unter: <https://doi.org/10.1080/1057610X.2020.1751459>
- 32 Public Prosecution Service of Canada. (2023). Court rules that murder and attempted murder were terrorist activity in youth case. Abrufbar unter: [https://www.ppsc-sppc.gc.ca/eng/nws-nvs/2023/27\\_07\\_23.html](https://www.ppsc-sppc.gc.ca/eng/nws-nvs/2023/27_07_23.html)
- 33 Davey, D., Comerford, M., Guhl, J., Baldet, W., & Colliver, C. (2021). A taxonomy for the classification of post-organisational violent terrorist content. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/a-taxonomy-for-the-classification-of-post-organisational-violent-terrorist-content/>
- 34 Fielitz, M., Ebner, J., Guhl, J., & Quent, M. (2018). *Loving Hate: Anti-Muslim Extremism, Radical Islamism and the Spiral of Polarization*. Institute for Democracy and Civil Society (IDZ) / Institute for Strategic Dialogue (ISD). Abrufbar unter: [https://www.idz-jena.de/fileadmin/user\\_upload/antimuslimextremism\\_radicalislamism\\_polarization.pdf](https://www.idz-jena.de/fileadmin/user_upload/antimuslimextremism_radicalislamism_polarization.pdf)
- 35 Cook, J. & Roose, J. (2023). *Supreme Men, Subjugated Women: Gender Inequality and Violence in Jihadist, Far-Right, and Male Supremacist Ideologies*. GNET Research. Abrufbar unter: <https://gnet-research.org/2023/01/10/supreme-men-subjected-women-gender-inequality-and-violence-in-jihadist-far-right-and-male-supremacist-ideologies/>
- 36 Ibid.
- 37 Antisemitism Policy Trust (2019). *Misogyny and Antisemitism: Briefing*. Abrufbar unter: <https://antisemitism.org.uk/wp-content/uploads/2019/05/5982-Misogyny-and-Antisemitism-Briefing-April-2019-v1.pdf>
- 38 Crawford, C. (2022). *Sleeping with the Enemy: Sex, Sexuality and Antisemitism in the Extreme Right*. Institute for Community and Social Research. Abrufbar unter: <https://icsr.info/2022/06/22/sleeping-with-the-enemy-sex-sexuality-and-antisemitism-in-the-extreme-right/>
- 39 Anti-Defamation League. (2023). *Antisemitism & Anti-LGBTQ+ Hate Converge in Extremist and Conspiratorial Beliefs*. Center on Extremism. Abrufbar unter: <https://www.adl.org/resources/blog/antisemitism-anti-lgbtq-hate-converge-extremist-and-conspiratorial-beliefs>
- 40 Anti-Defamation League. (2022) *What is "Grooming?" The Truth Behind the Dangerous, Bigoted Lie Targeting the LGBTQ+ Community*. Center on Extremism. Abrufbar unter: <https://www.adl.org/resources/blog/what-grooming-truth-behind-dangerous-bigoted-lie-targeting-lgbtq-community>
- 41 Anti-Defamation League. (2023). *Antisemitism & Anti-LGBTQ+ Hate Converge in Extremist and Conspiratorial Beliefs*. Center on Extremism. Abrufbar unter: <https://www.adl.org/resources/blog/antisemitism-anti-lgbtq-hate-converge-extremist-and-conspiratorial-beliefs>
- 42 Sykes, S. & Dr. Hopner, V. (2023). *Tradwives: The Housewives Commodifying Right-Wing Ideology*. GNET Research. Abrufbar unter: <https://gnet-research.org/2023/07/07/tradwives-the-housewives-commodifying-right-wing-ideology/>
- 43 Veilleux-Lepage, Y. et al. (2023). *Gendered radicalisation and 'everyday practices': An analysis of extreme right and Islamic State women-only forums*. *European Journal of International Security* (2023), 8, 227–242. Abrufbar unter: [https://researchmgt.monash.edu/ws/portalfiles/portal/461281934/420784008\\_oa.pdf](https://researchmgt.monash.edu/ws/portalfiles/portal/461281934/420784008_oa.pdf)

- 44 Chan, E. (2023). Technology-Facilitated Gender-Based Violence, Hate Speech, and Terrorism: A Risk Assessment on the Rise of the Incel Rebellion in Canada. *Violence Against Women*. Vol. 29, Issue 9, 1687-1718. Abrufbar unter: <https://doi.org/10.1177/10778012221125495>
- 45 Martiny, C. & Lawrence, S. (2023). A year of hate: Anti-drag mobilization efforts targeting LGBTQ people in the US. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/a-year-of-hate-anti-drag-mobilization-efforts-targeting-lgbtq-people-in-the-us/>
- 46 Argentino, M.-A., Raja, A. & Gallagher, A. (2022). She Drops: How QAnon Conspiracy Theories Legitimize Coordinated and Targeted Gender Based Violence. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/she-drops-how-qanon-conspiracy-theories-legitimize-coordinated-and-targeted-gender-based-violence/>
- 47 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Forthcoming report]
- 48 Kelly, L. (1988). *Surviving sexual violence*. Cambridge, UK: Polity Press.
- 49 Dunn, S. (2021). Is it Actually Violence? Framing Technology-Facilitated Abuse as Violence. In: Bailey, J., Flynn, A. & Henry, N. (eds) *The Emerald International Handbook of Technology-Facilitated Violence and Abuse* (Bingley, UK: Emerald Publishing, 2021). Abrufbar unter: <https://doi.org/10.1108/978-1-83982-848-520211002>
- 50 The Economist Intelligence Unit (EIU). (2021). Measuring the prevalence of online violence against women. Abrufbar unter: <https://onlineviolencewomen.eiu.com/>
- 51 PEN America. (n.d.). Defining online harassment: A glossary of terms. Abrufbar unter: <https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms/#cross>
- 52 The Global Partnership. (2023). Technology-facilitated gender-based violence: Preliminary landscape analysis. The Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership). Abrufbar unter: <https://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis>
- 53 International Center for Journalists (ICJF). (2022). The Chilling: A global study of online violence against women journalists. Abrufbar unter: <https://www.icjf.org/our-work/chilling-global-study-online-violence-against-women-journalists>
- 54 Jankowicz, N. et al. (2021). Malign creativity: How gender, sex, and lies are weaponized against women online. Wilson Center. Science and Technology Innovation Program. Abrufbar unter: <https://www.wilsoncenter.org/publication/malign-creativity-how-gender-sex-and-lies-are-weaponized-against-women-online>
- 55 Lenhart, A., Ybarra, M., Zickuhr, K. & Price-Feeney, M. (2016). Online Harassment, Digital Abuse, and Cyberstalking in America. Data & Society Research Institute. Abrufbar unter: <https://datasociety.net/library/online-harassment-digital-abuse-cyberstalking/>
- 56 Argentino, M.-A., Raja, A. & Gallagher, A. (2022). She Drops: How QAnon Conspiracy Theories Legitimize Coordinated and Targeted Gender Based Violence. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/she-drops-how-qanon-conspiracy-theories-legitimize-coordinated-and-targeted-gender-based-violence/>
- 57 Dunn, S. (2020). Technology-Facilitated Gender-Based Violence: An Overview. Supporting a Safer Internet Paper No. 1. Centre for International Governance Innovation (CIGI). Abrufbar unter: <https://www.cigionline.org/publications/technology-facilitated-gender-based-violence-overview/>
- 58 The Global Partnership. (2023). Technology-facilitated gender-based violence: Preliminary landscape analysis. The Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership). Abrufbar unter: <https://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis>
- 59 Safety Net Canada (2013). *Assessing Technology in the Context of Violence Against Women & Children: Examining Benefits & Risks*. Vancouver, BC: Safety Net Canada. Abrufbar unter: <https://bcsth.ca/wp-content/uploads/2016/10/Assessing-Technology-in-the-Context-of-Violence-Against-Women-Children.-Examining-Benefits-Risks.pdf>
- 60 The Global Partnership. (2023). Technology-facilitated gender-based violence: Preliminary landscape analysis. The Global Partnership for Action on Gender-Based Online Harassment and Abuse (Global Partnership). Abrufbar unter: [URLhttps://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis](https://www.gov.uk/government/publications/technology-facilitated-gender-based-violence-preliminary-landscape-analysis)
- 61 PEN America. (n.d.). Defining online harassment: A glossary of terms. Abrufbar unter: <https://onlineharassmentfieldmanual.pen.org/defining-online-harassment-a-glossary-of-terms/#cross>

- 62 United Nations Digital Library. (2018). Report of the Special Rapporteur on violence against women, its causes and consequences on online violence against women and girls from a human rights perspective. A/HRC/38/47. Abrufbar unter: <https://digitallibrary.un.org/record/1641160>
- 63 Dunn, S. (2020). Technology-Facilitated Gender-Based Violence: An Overview. Supporting a Safer Internet Paper No. 1. Centre for International Governance Innovation (CIGI). Abrufbar unter: <https://www.cigionline.org/publications/technology-facilitated-gender-based-violence-overview/>
- 64 Henry, N. et al. (2020). Image-based Sexual Abuse. A Study on the Causes and Consequences of Non-consensual Nude or Sexual Imagery. Routledge. Abrufbar unter: <https://doi.org/10.4324/9781351135153>
- 65 Wittes, B., Poplin, C., Jurecic, Q. & Spera, C. (2016). Sextortion: Cybersecurity, teenagers, and remote sexual assault. Brookings Institut. Abrufbar unter: <https://www.brookings.edu/articles/sextortion-cybersecurity-teenagers-and-remote-sexual-assault/>
- 66 Palmer, T. (2018). Rape pornography, cultural harm and criminalization. Northern Ireland Legal Quarterly. Vol. 69 No. 1 (2018): Spring. Abrufbar unter: <https://doi.org/10.53386/nllq.v69i1.77>
- 67 Viola, M., & Voto, C. (2023). Designed to abuse? Deepfakes and the non-consensual diffusion of intimate images. Synthese. 201. Article number: 30. Abrufbar unter: <https://doi.org/10.1007/s11229-022-04012-2>
- 68 Citron, D. (2019). Sexual Privacy. Yale Law Journal 128 (7): 1870–1960.
- 69 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Abrufbar unter: <https://www.icfj.org/our-work/chilling-global-study-online-violenceagainst-women-journalists>
- 70 Veletsianos, G., Houlden, S., Hodson, J. & Gosse, C. (2018). Women scholars' experiences with online harassment and abuse: Self-protection, resistance, acceptance, and self-blame. New Media & Society 20 (12): 4689–708.
- 71 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Forthcoming report]
- 72 United Nations Office of the High Commissioner for Human Rights. (2018). The impact of online violence on women human rights defenders and women's organisations. Abrufbar unter: <https://www.ohchr.org/en/statements/2018/06/impact-online-violencewomen-human-rights-defenders-and-womens-organisations?LangID=E&NewsID=23238>
- 73 International Center for Journalists (ICFJ). (2022). The Chilling: A global study of online violence against women journalists. Abrufbar unter: <https://www.icfj.org/our-work/chilling-global-study-online-violenceagainst-women-journalists>
- 74 Hoffman, B., Ware, J., & Shapiro, E. (2020). Assessing the threat of incel violence. Studies in Conflict & Terrorism, 43(7), 565-587. Abrufbar unter: <https://doi.org/10.1080/1057610X.2020.1751459>
- 75 Simmons, C. & Fourel, Z. (2022). Hate in Plain Sight: Abuse Targeting Women Ahead of the 2022 Midterm Elections on TikTok & Instagram. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/hate-in-plain-sight-abuse-targeting-women-ahead-of-the-2022-midterm-elections-on-tiktok-instagram/>
- 76 Binder, J. F., & Kenyon, J. (2022). Terrorism and the internet: How dangerous is online radicalization?. Frontiers in psychology, 6639. Abrufbar unter: <https://doi.org/10.3389/fpsyg.2022.997390>
- 77 Kavanagh, E. & Brown, L., (2020). "Towards a research agenda for examining online gender based violence against women academics", Journal of Further and Higher Education, vol. 44, No. 10. Abrufbar unter: <http://dx.doi.org/10.1080/0309877X.2019.1688267>
- 78 Martiny, M., Visser, F., & Jones, I. (2022). Evaluating Platform Abortion-Related Speech Policies: Were Platforms Prepared for the Post-Dobbs Environment? Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/evaluating-platform-abortion-related-speech-policies-were-platforms-prepared-for-the-post-dobbs-environment/>
- 79 Jane, E. (2018). Gendered cyberhate as workplace harassment and economic vandalism. Feminist Media Studies, 18(4), 575–591.
- 80 UN Women. (2023). Brief: The state of evidence and data collection on technology-facilitated violence against women. Abrufbar unter: <https://www.unwomen.org/en/digital-library/publications/2023/04/brief-the-state-of-evidence-and-data-collection-on-technology-facilitated-violence-against-women>
- 81 Plan International. (2022). State of the World's Girls Report 2021: Communications Report. Abrufbar unter: <https://plan-international.org/uploads/2022/02/sotwgr2021-commsreport-en.pdf>



- 82 UNFPA. (n.d.). Preventing Technology-Facilitated Gender-Based Violence (TF GBV). Abrufbar unter: [https://www.un.org/techenvoy/sites/www.un.org/techenvoy/files/GDC-Submission\\_UNFPA.pdf](https://www.un.org/techenvoy/sites/www.un.org/techenvoy/files/GDC-Submission_UNFPA.pdf)
- 83 U.S. Secret Service Media Relations. (2023). New Secret Service Research Examines for the First Time Five Years of Mass Violence Data. Abrufbar unter: <https://www.secretservice.gov/newsroom/releases/2023/01/new-secret-service-research-examines-first-time-five-years-mass-violence>
- 84 U.S. Department of State. (2023). Roadmap for the Global Partnership for Action on Gender-Based Online Harassment and Abuse. Abrufbar unter: <https://www.state.gov/2023-roadmap-for-the-global-partnership-for-action-on-gender-based-online-harassment-and-abuse/>
- 85 She Persisted. (2023). Monetizing Misogyny. Abrufbar unter: [https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted\\_MonetizingMisogyny.pdf](https://she-persisted.org/wp-content/uploads/2023/02/ShePersisted_MonetizingMisogyny.pdf)
- 86 UNESCO. (2021). The Chilling: Global trends in online violence against women journalists. Research Discussion Paper. Abrufbar unter: <https://unesdoc.unesco.org/ark:/48223/pf0000377223>
- 87 Global Engagement Center. (2023). Gendered Disinformation: Tactics, Themes, and Trends by Foreign Malign Actors. Abrufbar unter: <https://www.state.gov/gendered-disinformation-tactics-themes-and-trends-by-foreign-malign-actors/>
- 88 ISD-Recherchen untersuchten die Richtlinien von X (ehemals Twitter), Metas Facebook und Instagram, YouTube, TikTok und Telegram.
- 89 Martiny, M., Visser, F., & Jones, I. (2022). Evaluating Platform Abortion-Related Speech Policies: Were Platforms Prepared for the Post-Dobbs Environment? Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/evaluating-platform-abortion-related-speech-policies-were-platforms-prepared-for-the-post-dobbs-environment/>
- 90 Oversight Board. (2023). Oversight Board Overturns Meta's Original Decision in Image of Gender-Based Violence Case. Abrufbar unter: <https://oversightboard.com/news/813577783586004-oversight-board-overturns-meta-s-original-decision-in-image-of-gender-based-violence-case/>
- 91 X Help Center (2023). Hateful Conduct. Abrufbar unter: <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>
- 92 Martiny, M., Simmons, C., Visser, F., Bhatnagar, R., Jones, I., & Castillo Small, A. (2023). A Retrospective Study of Online Gendered Abuse in 2022 in the United States [Preliminary title]. Institute for Strategic Dialogue (ISD) [Unpublished report].
- 93 Ibid.
- 94 O'Connor, C. (2021). Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/hatescape-an-in-depth-analysis-of-extremism-and-hate-speech-on-tiktok/>
- 95 Wired UK. (2022). Telegram Has a Serious Doxing Problem. Abrufbar unter: <https://www.wired.co.uk/article/telegrams-doxing-problem>
- 96 Holnburger, J. (2023). Chronology of a radicalization: How Telegram became the most important platform for conspiracy ideologies and right-wing extremism. CeMAS. Abrufbar unter: <https://cemas.io/en/publications/chronology-of-a-radicalization/>
- 97 Gerster, L., Kuchta, R., Hammer, D. & Schwieter, C. (2022). Telegram as a Buttress: How Far-Right Extremists and Conspiracy Theorists Are Expanding Their Infrastructures via Telegram. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/telegram-as-a-buttress-how-far-right-extremists-and-conspiracy-theorists-are-expanding-their-infrastructures-via-telegram/>
- 98 Council for Media Services in Slovakia (CMS) & Reset (2023). The Bratislava Shooting Report on the role of online platforms. Abrufbar unter: [https://rpms.sk/sites/default/files/2023-03/CMS\\_RESET\\_Report.pdf](https://rpms.sk/sites/default/files/2023-03/CMS_RESET_Report.pdf)
- 99 Deceptive Design. (n.d.). Types of deceptive pattern. Abrufbar unter: <https://www.deceptive.design/types>
- 100 Sinders, C., Shukla, V. & Voegeli, E. (2021). Trust Through Trickery. Common Place. Abrufbar unter: [URLhttps://commonplace.knowledgefutures.org/pub/trust-through-trickery/release/1](https://commonplace.knowledgefutures.org/pub/trust-through-trickery/release/1)
- 101 Ibid.
- 102 Ibid.
- 103 eSafety Commissioner. (n.d.). Safety by Design. Abrufbar unter: <http://www.esafety.gov.au/industry/safety-by-design>
- 104 Ada Lovelace Institute. (2021). Technical methods for regulatory inspection of algorithmic systems. Abrufbar unter: <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/>
- 105 Integrity Institute. (n.d.). Ranking by Engagement. Abrufbar unter: <https://integrityinstitute.org/blog/ranking-by-engagement>
- 106 Thomas, E. & Balint, K. (2022). Algorithms as a Weapon Against Women: How YouTube Lures Boys and Young Men into the Manosphere. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere/>

- 107 Simmons, C. & Fourel, Z. (2022). Hate in Plain Sight: Abuse Targeting Women Ahead of the 2022 Midterm Elections on TikTok & Instagram. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/hate-in-plain-sight-abuse-targeting-women-ahead-of-the-2022-midterm-elections-on-tiktok-instagram/>
- 108 Sykes, S. & Dr. Hopner, V. (2023). Tradwives: The Housewives Commodifying Right-Wing Ideology. GNET Research. Abrufbar unter: <https://gnet-research.org/2023/07/07/tradwives-the-housewives-commodifying-right-wing-ideology/>
- 109 Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. Abrufbar unter: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- 110 Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine Bias. ProPublica. Abrufbar unter: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- 111 Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366, 447–453. Abrufbar unter: <https://pubmed.ncbi.nlm.nih.gov/31649194/>
- 112 Vlasceanu, M. & Amodio, D. M. (2022). Propagation of societal gender inequality by internet search algorithms. *PNAS*. Vol. 119, No. 29. Abrufbar unter: <https://www.pnas.org/doi/full/10.1073/pnas.2204529119#body-ref-r2-2>
- 113 Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Proceedings of the First Conference on Fairness, Accountability and Transparency*, 77–91. PMLR. Abrufbar unter: <https://proceedings.mlr.press/v81/buolamwini18a.html>
- 114 Fosch-Villaronga, E., Poulsen, A., Søråa, R. A., & Custers, B. H. M. (2021). Gendering Algorithms in Social Media. *ACM SIGKDD Explorations Newsletter*, 23(1), 24–31. Abrufbar unter: <https://doi.org/10.1145/3468507.3468512>
- 115 Scott, M. (2023) We Don't Talk About Fixing Social Media. POLITICO. Abrufbar unter: <https://www.politico.eu/newsletter/digital-bridge/we-dont-talk-about-fixing-social-media/>
- 116 Es ist wichtig zu betonen, dass sich die Art der Daten, die als »öffentlich zugänglich« gelten, ändern kann. Dabei sollten mögliche Änderungen auf einer gemeinsamen Taxonomie aufbauen, die festlegt, welche Daten »öffentlich zugänglich« sein sollten.
- 117 »Im öffentlichen Interesse liegende Forschung« bezieht sich im Allgemeinen auf Forschung mit dem Ziel, das kollektive Wissen der Gesellschaft zu erweitern. Regulatorische Präzedenzfälle legen nahe, dass im öffentlichen Interesse liegende Forschung unabhängig von kommerziellen Interessen sein und die Quelle ihrer Finanzierung transparent sein muss. Die Forscher:innen sind nicht notwendigerweise mit akademischen Einrichtungen verbunden und können auch Forscher:innen umfassen, die gemeinnützigen oder Medienorganisationen angehören.
- 118 Bundtzen, S. & Schwieter, C. (2023). Researcher Access to Social Media Data: Lessons Learnt & Recommendations for Strengthening Initiatives in the EU & Beyond. Institute for Strategic Dialogue (ISD). Abrufbar unter: <https://www.isdglobal.org/isd-publications/researcher-access-to-social-media-data-lessons-learnt-recommendations-for-strengthening-initiatives-in-the-eu-beyond/>
- 119 Wilton Park. (2022). Building a Shared Agenda on the Evidence Base for Gender-Based Online Harassment and Abuse. WP3057. Abrufbar unter: <https://www.wiltonpark.org.uk/event/building-a-shared-agenda-on-the-evidence-base-for-gender-based-online-harassment-and-abuse/>
- 120 UN Women. (2023). Technology-Facilitated Violence Against Women: Taking Stock of Evidence and Data Collection. Abrufbar unter: <https://www.unwomen.org/en/digital-library/publications/2023/04/technology-facilitated-violence-against-women-taking-stock-of-evidence-and-data-collection>
- 121 Zu den Methoden gehören Erhebungsdaten, die durch Haushalts-, bevölkerungsbezogene oder experimentelle und spezialisierte Online-Umfragen gewonnen werden; quantitative Verwaltungsdaten, einschließlich dienstleistungsbezogener, programmatischer oder Transparenzberichtsdaten; qualitative Daten, die durch Umfragen, Fokusgruppendifkussionen oder digitale Ethnographien gewonnen werden, sowie Daten, die durch gemischte Methoden gewonnen werden, einschließlich des maschinellen Lernens zur Erfassung und Analyse von Plattformdaten.
- 122 Conway, M. (2016). Determining the Role of the Internet in Violent Extremism and Terrorism: Six Suggestions for Progressing Research. *Studies In Conflict & Terrorism*. 2017, Vol. 40, No. 1, 77–98. Abrufbar unter: <https://www.tandfonline.com/doi/full/10.1080/1057610X.2016.1157408>
- 123 European Commission. (2022). 2022 Strengthened Code of Practice on Disinformation. Abrufbar unter: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>

- 124 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Abrufbar unter: <https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics>
- 125 UN Women. (2023). Technology-Facilitated Violence Against Women: Taking Stock of Evidence and Data Collection. Abrufbar unter: <https://www.unwomen.org/en/digital-library/publications/2023/04/technology-facilitated-violence-against-women-taking-stock-of-evidence-and-data-collection>
- 126 Office of the United Nations High Commissioner for Human Rights (OHCHR). (2023). Report on freedom of expression and the gender dimensions of disinformation. Abrufbar unter: <https://www.ohchr.org/en/calls-for-input/2023/report-freedom-expression-and-gender-dimensions-disinformation>
- 127 Sinders, C., Shukla, V. & Voegeli, E. (2021). Trust Through Trickery. Common Place. Abrufbar unter: <https://commonplace.knowledgefutures.org/pub/trust-through-trickery/release/1>
- 128 PEN America. (n.d.). No Excuse for Abuse. What Social Media Companies Can Do Now to Combat Online Harassment and Empower Users. Abrufbar unter: <https://pen.org/report/no-excuse-for-abuse/>
- 129 Anti-Defamation League (ADL). (2018). When Women Are the Enemy: Intersection Misogyny and White Supremacy. Abrufbar unter: <https://www.adl.org/resources/report/when-women-are-enemy-intersection-misogyny-and-white-supremacy>
- 130 eSafety Commissioner. (n.d.). Basic Online Safety Expectations. Abrufbar unter: <https://www.esafety.gov.au/industry/basic-online-safety-expectations>
- 131 GitHub. (n.d.). Harassment Manager. Abrufbar unter: <https://github.com/conversationai/harassment-manager>
- 132 Global Internet Forum to Counter Terrorism (GIFCT). (n.d.). Content Incident Protocol. Abrufbar unter: <https://gifct.org/content-incident-protocol/>
- 133 Christchurch Call. (2022). Christchurch Call Initiative on Algorithmic Outcomes. Abrufbar unter: <https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/>
- 134 European Commission. (2022). 2022 Strengthened Code of Practice on Disinformation. Abrufbar unter: <https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation>
- 135 National Democratic Institute (NDI). (2022). Interventions to End Online Violence Against Women in Politics. Abrufbar unter: <https://www.ndi.org/publications/interventions-end-online-violence-against-women-politics>
- 136 Perspective (n.d.) Developers. About the API. Abrufbar unter: [https://developers.perspectiveapi.com/s/about-the-api?language=en\\_US](https://developers.perspectiveapi.com/s/about-the-api?language=en_US)
- 137 Dias Oliva, T., Antonialli, D.M. & Gomes, A. (2021). Fighting Hate Speech, Silencing Drag Queens? Artificial Intelligence in Content Moderation and Risks to LGBTQ Voices Online. *Sexuality & Culture* 25, 700–732. Abrufbar unter: <https://doi.org/10.1007/s12119-020-09790-w>
- 138 The End Violence Against Women Coalition, Glitch, Refuge, Carnegie UK, NSPCC, 5Rights, Professorin Clare McGlynn and Professorin Lorna Woods (2022). Violence Against Women and Girls (VAWG) Code of Practice. Abrufbar unter: <https://carnegieuktrust.org.uk/publications/violence-against-women-and-girls-vawg-code-of-practice/>
- 139 Official Journal of the EU (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). Volume 65. Stand: 27. Oktober. Abrufbar unter: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065&from=EN>
- 140 Council of Europe. (2017). Algorithms and Human Rights: Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Abrufbar unter: <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>
- 141 Andrew M. Guess et al. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign?. *Science* 381,398-404. Abrufbar unter: <https://www.science.org/doi/10.1126/science.abp9364>
- 142 Habib, H. et al. (2022). Making a Radical Misogynist: How Online Social Engagement with the Manosphere Influences Traits of Radicalization. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2), Article No. 450, 1–28. Abrufbar unter: <https://dl.acm.org/doi/10.1145/3555551>
- 143 Christchurch Call. (2022). Christchurch Call Initiative on Algorithmic Outcomes. Abrufbar unter: <https://www.christchurchcall.com/media-and-resources/news-and-updates/christchurch-call-initiative-on-algorithmic-outcomes/>
- 144 Fosch-Villaronga, E., Poulsen, A., Søraa, R. A., & Custers, B. H. M. (2021). Gendering Algorithms in Social Media. *ACM SIGKDD Explorations Newsletter*, 23(1), 24–31. Abrufbar unter: <https://doi.org/10.1145/3468507.3468512>

- 145 Gebru, T., et al. (2022). Excerpt from datasheets for datasets. In *Ethics of Data and Analytics*, K. Martin (Ed.), Auerbach Publications, 148–156. Abrufbar unter: <https://www.taylorfrancis.com/chapters/edit/10.1201/9781003278290-23/excerpt-datasheets-datasets-timnit-gebru-jamie-morgenstern-briana-vecchione-jennifer-wortman-vaughan-hanna-wallach-hal-daum%C3%A9-kate-crawford>
- 146 Dunn, S. (2021). Is it Actually Violence? Framing Technology-Facilitated Abuse as Violence. In: Bailey, J., Flynn, A. & Henry, N. (eds) *The Emerald International Handbook of Technology-Facilitated Violence and Abuse* (Bingley, UK: Emerald Publishing, 2021). Abrufbar unter: <https://doi.org/10.1108/978-1-83982-848-520211002>
-

**ISD** | Institute  
for Strategic  
Dialogue

Amman | Berlin | London | Paris | Washington DC

Copyright © Institute for Strategic Dialogue (2023).  
Das Institute for Strategic Dialogue (gGmbH) ist beim  
Amtsgericht Berlin-Charlottenburg registriert (HRB 207 328B).  
Die Geschäftsführerin ist Huberta von Voss. Die Anschrift lautet:  
Postfach 80647, 10006 Berlin. Alle Rechte vorbehalten.

[www.isdgermany.org](http://www.isdgermany.org)

gefördert durch:



Auswärtiges Amt