Österreichisches Institut für
angewandte Telekommunikation

Autor:innen:
Julia Krickl
Natalie Trell
Lena Müller-Naendrup

Jänner 2026 Report

Desinformation auf Social Media Plattformen

TGuard

Einleitung

Ob Desinformationskampagnen in Wahlkampfzeiten, propagandistische Inhalte während Kriegs- und Krisensituationen oder täuschend echte KI-generierte Inhalte zu polarisierenden Themen, Desinformation findet sich auf allen Sozialen Medien. Facebook, Instagram, YouTube, TikTok und Telegram sind zentrale Vermittler von Informationen im Alltag, werden aber zunehmend auch zur Verbreitung von Desinformationskampagnen und -narrativen genutzt. Die verbreitete Desinformation reicht dabei von wahlbezogenen Manipulat-ionsversuchen über pseudo-, irreführende oder direkt gefälschte wissenschaftliche Inhalte bis hin zu gezielter KI-Propaganda.

Mis- und Desinformation verbreiten sich dabei nicht nur über direkte Interaktionen, sondern insbesondere über algorithmische Empfehlungsstrukturen, gezielte Werbung und Gruppenmechanismen. Diese Dynamik stellt Behörden, Zivilgesellschaft und Forschung vor neue Herausforderungen. Der Digital Services Act (Verordnung 2022/2065) verpflichtet Very Large Online Platforms (VLOPs, Plattformen mit mehr als 45 Millionen Nutzer:innen in der EU) dazu, systematisch gegen illegale Inhalte und systemische Risiken wie Desinformation vorzugehen. Zentrale Elemente sind Melde- und Abhilfeverfahren, Transparenzpflichten und die Zusammenarbeit mit Trusted Flaggers - anerkannte, besonders sachkundige Einrichtungen, deren Meldungen zu illegalen Inhalten von Plattformen vorranging bearbeitet werden müssen. Facebook, Instagram, YouTube und TikTok fallen unter den VLOP-Status.

Im Rahmen des **Forschungsprojekts TGuard**[1] wurde das Ausmaß von Desinformation auf den großen Sozialen Medien Plattformen sowie deren Maßnahmen zur Eindämmung von Desinformation untersucht. Im Zentrum stand die Frage, wie wirksam die von Meta (Facebook und Instagram), YouTube, TikTok und Telegram eingesetzten Moderations- und Regulierungsmaßnahmen gegen Desinformation im Kontext des DSA tatsächlich sind und wo strukturelle Lücken zwischen Selbstauskünften der Plattformen und den Beobachtungen unabhängiger Untersuchungen bestehen.

Aufbau und Vorgehen

Die Analyse folgt einem plattformvergleichenden Vorgehen, um Ausmaß und Charakter von Desinformation im digitalen Raum einordnen zu können. Ausgewertet wurden die aktuellsten öffentlich verfügbaren Transparenzberichte der Plattformen, aggregierte Daten der DSA-Transparenzdatenbank für das Jahr 2025 sowie themenspezifische Studien und Fallanalysen unabhängiger Organisationen. Folgende Datenquellen werden berücksichtigt:

TRANSPARENZ-BERICHTE
Halbjahresberichte der Plattformen im Rahmen des Code of Practice on Disinformation[2] (CoPD) - Selbstauskünfte der Plattformen.

DATEN
aus der DSA-Transparenzdatenbank[3] mit den „Statement of Reasons“ (SoRs) zu Moderationsentscheidungen.

THEMENSPEZIFISCHE STUDIEN
Untersuchungen unabhängiger Organisationen (z. B. European Fact-Cechking Standards Network, Maldita, Science Feedback, AI Forensics, Democracy Reporting International)

[1] https://www.tguard.at/
[2] Seit Februar 2025 heißt es Conduct on Misinformation. Analyse: CoPD-Bericht Nr. 6 für Meta, Nr. 5 für TikTok und Google/YouTube, Berichtszeitraum bis Ende 2024 bzw. Mitte 2025. Die Berichte sind hier verfügbar: https://disinfocode.eu/reports
[3] https://transparency.dsa.ec.europa.eu/

Wo Plattformangaben und externe Untersuchungen divergieren, werden beide Perspektiven kenntlich gemacht. Die Auswertung zielt darauf ab, ein systematisch vergleichbares Bild plattformspezifischer Praktiken und ihrer Grenzen zu zeichnen, sowie zentrale Beispielnarrative plattformübergreifend aufzuarbeiten. Aufbauend darauf wurden für Meta (Facebook und Instagram), YouTube und TikTok jeweils drei Ebenen betrachtet:

1. die geltende Plattformrichtlinien zu Desinformation, politische Werbung, KI-Inhalte und problematische Konten;

2. die von den Plattformen eingesetzten Verfahren zur Erkennung und die daraus abgeleiteten Gegenmaßnahmen; sowie

3. die Wirksamkeit dieser Maßnahmen anhand unabhängiger Untersuchungen und dokumentierter Fallbeispiele.

Telegram wird gesondert in einem Exkurs behandelt, da die Plattform nicht als Very Large Online Platform (VLOP) im Sinne des DSA eingestuft ist und sich auch am CoPD nicht beteiligt, womit zentrale Vergleichsdaten fehlen. Der Exkurs folgt daher einer alternativen Herangehensweise und stützt sich stärker auf die Entwicklung der Plattform, Nutzungsdynamiken und Erkenntnissen aus wissenschaftlichen sowie zivilgesellschaftlichen Analysen.

Datenlage

DSA-Transparenzdatenbank: Statements of Reasons 2025

Die DSA-Transparenzdatenbank dokumentiert jede Moderationsmaßnahme der Plattformen mit einer Begründung („Statements of Reasons", SoRs) sowie der Art der Maßnahme. Für Desinformation ist keine eigene Kategorie ausgewiesen; relevante Verstöße können unter verschiedene Kategorien fallen, ins besondere „Negative effects on civic discourse or elections" und „Illegal or harmful speech".

Bezüglich der Verstoßarten zeigt sich: Bei TikTok entfielen fast 50 % aller SoRs auf die für Desinformation relevanten Kategorien. Die Meta-Plattformen verzeichneten dagegen keine SoRs unter „Negative effects on civic discourse or elections"; die größte Kategorie neben „Scope of platform" war „Scams and/or fraud". YouTube wies zwar relevante Fälle aus, anteilsmäßig machten diese aber weniger als 1 % der gesamten SoRs aus.

Bei der Art der Maßnahmen zeigen sich unterschiedliche Muster: TikTok reagierte überwiegend mit Sichtbarkeitsbeschränkungen (990 Mio. Fälle) – in knapp 60 % der Fälle wurden Inhalte entfernt -, Facebook primär mit Kontobeschränkungen

Plattformen	SoRS gesamt 2025	davon: negative effects on civic discourse	davon: illegal or harmful speech
TikTok	1.002.728.224	45.586.463	434.875.107
Facebook	453.069.643	-	15.821.557
Instagram	126.151.906	-	2.812.557
YouTube	104.659.285	178.267	56.478

Tabelle 1: SoRs 2025 nach Plattform (konsolidiert nach für Desinformation relevanten Kategorien). Quelle: DSA-Transparenzdatenbank, eigene Darstellung.

(359 Mio. Fälle, davon 258 Mio. Suspendierungen), YouTube mit Inhaltsentfernungen (95,8 Mio.) und Instagram mit Kontokündigungen (104,6 Mio.). Die große Mehrheit aller Moderationsmaßnahmen ging auf freiwillige Initiativen der Plattformen zurück. Meldungen nach Art. 16 DSA, welcher öffentlich und leicht zugängliche Meldeverfahren für Nutzer:innen vorschreibt, und Hinweise von Trusted Flaggers spielten eine untergeordnete Rolle. Meldungen von Nutzer:innen machten jeweils nur zwischen knapp 150.000 (TikTok) bis maximal unter 500.000 (YouTube) der Maßnahmen aus und Trusted Flagger Meldungen lagen bei unter 500 Fällen pro Plattform.

CoPD-Berichte: Moderationsmaßnahmen HJ 1 / 2025

Der CoPD ist ein freiwilliger, ursprünglich 2018 eingeführter und 2022 überarbeiteter Selbstregulierungsrahmen. Unterzeichner, darunter die hier analysierten Plattformen mit Ausnahme von Telegram, verpflichten sich, halbjährlich detaillierte Transparenzberichte entlang definierter „Commitments“ vorzulegen, die sowohl quantitative Daten wie qualitative Beschreibungen beinhalten. Für dieses Dossier besonders relevant sind die Commitments zu „Scrutiny of Ad Placements“ (Einschränkung der Monetarisierung von Desinformation), „Integrity of Services“ (Bekämpfung koordinierter Manipulation) und „Empowering Users“ (Kontextualisierung irreführender Inhalte). Die CoPD-Halbjahresberichte machen plattform-spezifische Angaben zu Moderationsmaßnahmen entlang eigener Policy-Kategorien. **Tabelle 2** konsolidiert die zentralen Zahlen für das erste Halbjahr 2025 nach entfernten Inhalten oder Konten, je nach Thema.

Die absoluten Zahlen sind aus zwei Gründen nur eingeschränkt vergleichbar: Erstens definieren die Plattformen die Policy-Bereiche unterschiedlich, zweitens erfassen sie Inhalte auf unterschiedlichen Ebenen (einzelne Videos vs. Konten). Aussagekräftig ist eher das Muster: TikTok stellt die umfangreichsten und detailliertesten Daten bereit, YouTube die transparentesten Aufschlüsselungen nach Bereichen, Meta die geringste Granularität. Für die für Desinformation relevanten Kategorien werden Facebook und Instagram teils nicht getrennt ausgewiesen.

Ein weiterer in den CoPD-Berichten thematisierter Aspekt sind Netzwerke an organisierter, verdeckter Einflussnahme bzw. „Coordinated Inauthentic Behavior“ (CIB), die vor allem zur Verbreitung von Desinformationen rund um politische Prozesse und zur Beeinflussung der öffentlichen Meinung eingesetzt werden. Die Dokumentation des Vorgehens gegen solche Netzwerke unterscheidet sich zwischen den Plattformen: Der Fokus liegt teils auf quantifizierten Daten, teils auf ausführlicheren qualitativen Beschreibungen. TikTok entfernte im Berichtszeitraum mindestens 16 CIB-Netzwerke mit insgesamt 31.270 Accounts und knapp 10,5 Millionen Followern. Meta berichtete über die Entfernung mehrerer CIB-Netzwerke, darunter eins mit Fokus auf Rumänien (658 Facebook-Accounts, 14 Seiten, zwei Instagram-Accounts). YouTube wiederum lieferte ausschließlich quantitative Daten ohne inhaltliche Beschreibung: EU-weit wurden 177.272 Fake-Channels gelöscht (838 in Österreich), zudem identifizierte die Plattform 4.284 CIB-Fälle auf EU-Ebene sowie 71 in Österreich.

Plattform	Misinformation		Synthetische Medien		Fake-Accounts / Channels	
	EU	AT	EU	ET	EU	AT
TikTok	207.646	2.709	24.213	585	19.225.031	285.956
Meta	83.000+	1.000+	-	-	~1Mrd (weltweit)	-
YouTube	4.060	62	1.776	29	177.272	838

Tabelle 2: Moderationsmaßnahmen laut CoPD-Transparenzberichten für das erste Halbjahr 2025. Quellen: TikTok 2025, Meta 2025, Google/YouTube 2024/2025, eigene Darstellung.

Beispielnarrative

Unabhängig von der reinen Statistik zeigt sich in Fallstudien, welche Narrative auf den Plattformen kursieren und wie hoch ihre Reichweite trotz Moderation ist:

- → Wahl- und Demokratiebezogene Narrativen: Im Kontext der Europawahlen 2024 waren Social Media Plattformen ein zentrales Ziel für unterschiedliche Desinformationskampagnen und -narrative mit dem Versuch, demokratische Prozesse zu beeinflussen. Algorithmische Empfehlungssysteme spielten ein zentrale Rolle bei der Verbreitung politischer Inhalte (Fundación Maldita 2024).

- → KI-gestützte Kriegsnarrative: In einer ÖIAT-Untersuchung bezogen sich 18 von 21 identifizierten KI-Videos direkt auf die Konflikte in Israel und zwischen Israel und Iran – oft mit vollständig künstlich erzeugten Zivilist:innen oder Soldat:innen. Ein KI-generierter angeblicher Raketenangriff auf Tel Aviv wurde erst nach einer Trusted-Flagger-Meldung entfernt und hatte zu diesem Zeitpunkt bereits über 200.000 Aufrufe erreicht (Strasser und Krickl 2025).

- → Gesundheits- und Klimadesinformation: Ein Facebook-Video, laut dem eine CO_2-Konzentration unter 0,03 % alles Leben auf der Erde vernichten würde, blieb trotz Faktencheck online – bei über 250.000 Aufrufen. Auf YouTube erreichte ein Video mit der Falschbehauptung, das EU-Verbrennerverbot ab 2035 sei aufgehoben, 500.000 Aufrufe ohne Plattformreaktion (Fundación Maldita 2024).

Plattformspezifische Analyse

Meta (Facebook und Instagram)

Richtlinien

Meta untersagt die Verbreitung von Falschinformationen, sofern diese das Potenzial besitzen, physischen Schaden zu verursachen oder demokratische Prozesse zu untergraben. Inhalte, die nachweislich falsch sind, jedoch keine unmittelbare Gefahr darstellen, werden nicht entfernt, sondern durch faktische Gegendarstellungen ergänzt. Seit Oktober 2025 ist politische, wahlbezogene und auf gesellschaftliche Themen bezogene Werbung in der EU auf Meta gänzlich verboten. KI-generierte oder manipulierte Inhalte werden seit Mai 2024 verstärkt gekennzeichnet, insbesondere bei Deepfakes mit gesellschaftlicher Relevanz.

Erkennung und Maßnahmen

Meta setzt auf eine Kombination aus automatisierten Systemen, menschlicher Prüfung und Kooperationen mit externen Fact-Checking-Partnern. In der EU arbeitet Meta (Stand 2024) mit 29 Organisationen zusammen, die Inhalte in 23 Sprachen prüfen. Wird ein Beitrag als falsch eingestuft, erhält er ein sichtbares Label mit Erläuterung sowie eine reduzierte Sichtbarkeit im Algorithmus. Anlässlich der Europawahl 2024 richtete Meta „Voter Information Units“ ein, die laut eigenen Angaben in der EU über 41 Mio. Mal auf Facebook und 58 Mio. Mal auf Instagram aufgerufen wurden (Meta 2024).

Im Januar 2025 kündigte Meta an, das Third-Party-Fact-Checking-Programm in den USA zu beenden und auf ein Community-Notes-Modell umzustellen. In Europa wird das Fact-Checking-Programm aufgrund der DSA-Verpflichtungen vorerst weitergeführt. Die neue Ausrichtung – politisch motiviert und durch das politische Klima in den USA beeinflusst – alarmiert Fact-Checking-Organisationen, insbesondere hinsichtlich eines möglichen weiteren Rückzugs aus CoPD-Verpflichtungen (Schulz-Tomančok 2025).

Wirksamkeit

Trotz umfangreicher Richtlinien bleibt die Wirksamkeit von Metas Maßnahmen umstritten. Eine europaweite Untersuchung der Fakten-Check-Organisation Maldita. es analysierte 1.321 an Plattformen gemeldete Desinformationsbeiträge und zeigte für Meta zwei zentrale Befunde (Fundación Maldita 2024):

- → Hohe Reaktions-, niedrige Löschquote: Auf Facebook wurden in rund 89 % der Fälle Maßnahmen ergriffen (Instagram: 71 %), aber in nur 12 % (Facebook) bzw. 14,6 % (Instagram) der Fälle wurde der Inhalt tatsächlich entfernt. Der Großteil beschränkte sich auf die Anbringung eines Faktencheck-Labels.

- → Thematische Ungleichheit: Bei bestimmten Themen liegt die Reaktionsquote deutlich niedriger – bei klimabezogener Desinformation auf Instagram nur bei 33 %; bei Inhalten zum Nahostkonflikt sowie nationalen und regionalen Themen bei knapp über 50 %.

Recherchen von AI Forensics zeigen erhebliche Schwächen bei der Umsetzung der Werberichtlinien: Rund 60 % der moderierten politischen Anzeigen in der Meta Ad Library verstießen bereits gegen geltende Plattformrichtlinien, während automatisierte Systeme weniger als 5 % der problematischen Anzeigen überhaupt erkannten. Eine dokumentierte pro-russische Propagandakampagne erreichte über 38 Mio. Nutzer:innen in Frankreich, Deutschland, Italien und Polen (Bouchaud u. a. 2024). Science Feedback identifizierte zudem 2024 mindestens 21 Facebook-Konten mit Verbindungen zu sanktionierten russischen oder belarussischen Akteuren; zum Jahresende waren neun dieser Profile trotz Meldung weiterhin aktiv, 13 neue hinzugekommene Accounts blieben unmoderiert (Terroille u. a. 2024).

YouTube

Richtlinien

YouTube verbietet Inhalte, die absichtlich falsche Informationen mit „erheblichem Schadenspotenzial“ verbreiten – insbesondere zu Wahlen, Volkszählungen und gesundheitsbezogenen Themen. Andere falsche oder irreführende Inhalte bleiben online, werden aber in ihrer Reichweite eingeschränkt und mit Kontext versehen. Seit März 2024 sind Creator:innen verpflichtet, realistisch veränderte oder synthetische Inhalte offen zu kennzeichnen; technische Erkennungsmechanismen wie SynthID ergänzen dies durch unsichtbare Wasserzeichen.

Erkennung und Maßnahmen

YouTube kombiniert automatisierte KI-gestützte Klassifizierung, menschliche Moderation und Kooperationen mit zertifizierten Fact-Checking-Organisationen. Früherkennungssysteme achten auf ungewöhnliche Upload- oder Interaktionsmuster, die auf koordinierte Kampagnen hindeuten. Als Gegenmaßnahmen kommen Entfernung, Reichweitenreduktion, Informationspanels und die Hervorhebung „autoritativer Quellen“ zum Einsatz.

Wirksamkeit

YouTube weist im CoPD-Bericht für das erste Halbjahr 2025 vergleichsweise niedrige absolute Zahlen aus: 4.060 Misinformation-Löschungen EU-weit, 62 in Österreich. Bei koordinierter Content-Amplifikation waren es 4.284 Fälle EU-weit. Ob diese Zahlen die tatsächliche Belastung widerspiegeln oder auf systematische Untererfassung hindeuten, ist ohne externen Datenzugang schwer zu beurteilen.

Eine Stichprobenuntersuchung von Science Feedback ergab, dass in einer Auswertung von 940 als problematisch eingestuften YouTube-Videos nur bei rund 37 % überhaupt ein Informationspanel eingeblendet wurde – obwohl YouTube dieses Instrument prominent als Maßnahme gegen Desinformation kommuniziert (Samson 2024). Fallbeispiele wie das 500.000 Mal angesehene Video zum angeblich aufgehobenen EU-Verbrennerverbot zeigen zudem, dass reichweitenstarke Falschinformationen ohne Plattformreaktion online bleiben können. Ein weiteres YouTube-Reel mit der Falschbehauptung, COVID-19-Impfungen würden Autoimmunerkrankungen auslösen, wurde allein in Polen über 300.000 Mal angesehen (Fundación Maldita 2024).

TikTok

Richtlinien

TikToks Community Guidelines untersagen „schädliche Falschinformation“ – darunter Inhalte, die die öffentliche Sicherheit, Gesundheit oder demokratische Prozesse gefährden können. Für politische Werbung gilt ein weitgehendes Verbot: Bezahlte politische Anzeigen sind auf TikTok nicht zulässig. Politiker:innen, Regierungen und politische Parteien verlieren auf verifizierten Konten den Zugang zu Werbefunktionen. Synthetische Medien mit realistischer Darstellung realer Personen oder Szenen müssen als solche gekennzeichnet werden.

Erkennung und Maßnahmen

TikTok setzt stark auf automatisierte Moderation: Ein Großteil der Moderationsentscheidungen erfolgt teil- oder vollautomatisiert. Ergänzend arbeitet die Plattform mit dem Global Fact-Checking Program (GFCP) und zertifizierten Fact-Checkern zusammen. Als Reaktion auf identifizierte Desinformation können Videos entfernt, im For-You-Feed gesperrt oder mit Warnhinweisen und Informationsbannern versehen werden. Bei koordinierten Kampagnen werden ganze Netzwerke gelöscht.

Wirksamkeit

TikTok stellt in den offiziellen Berichten die umfangreichsten Zahlen: EU-weit rund 207.000 entfernte Misinformation-Videos und über 19 Mio. entfernte Fake-Accounts im ersten Halbjahr 2025. Ein Fokus liegt auf strukturellen Interventionen gegen koordinierte Manipulation – 16 CIB-Netzwerke mit rund 10,5 Mio. Follower:innen wurden im selben Zeitraum entfernt.

Die tatsächliche Umsetzung im Alltag zeigt jedoch erhebliche Lücken. Eine Untersuchung von Correctiv und weiteren Organisationen zur deutschen Bundestagswahl 2025 dokumentierte, dass von 201 gemeldeten Desinformationsvideos lediglich 12 tatsächlich aus dem Feed entfernt wurden – rund 6 %. Ähnliche Muster zeigen Untersuchungen von Democracy Reporting International zur rumänischen Präsidentschaftswahl 2025, bei der TikTok 323 als problematisch identifizierte Konten erst spät im Wahlprozess adressierte. Auch beim Verbot politischer Werbung dokumentierten unabhängige Studien wiederholt Umgehungen: Trotz Werbeverbot fanden Forscher:innen 100 politische Anzeigen auf TikTok Deutschland innerhalb einer einzelnen Untersuchung (Hohlfeld u. a. 2024; Kimberly und Scherndl 2025).

12 %
werden entfernt

Nur 12 % der auf **Facebook** gemeldeten Desinformationsinhalte werden tatsächlich entfernt. Reagiert wird auf 89 %, aber meist nur mit einem Label.

37 %
gekennzeichnet

Von 940 als problematisch eingestufter Inhalte erhielten auf **YouTube** 37 % ein Informationspanel.

100
politische Anzeigen

Trotz Werbeverbot fanden Forscher:innen 100 politische Werbeanzeigen innerhalb einer Untersuchung auf **TikTok.**

Exkurs: Telegram

Sonderstellung im Regulierungsrahmen

Telegram nimmt eine Sonderstellung ein: Die Plattform gibt an, monatlich unter 45 Mio. aktive Nutzer:innen in der EU zu verzeichnen und wird daher nicht als VLOP im Sinne des DSA eingestuft.

Damit besteht für Telegram keine gesetzliche Verpflichtung, systemische Risiken wie Desinformation zu identifizieren, zu bewerten oder Maßnahmen zu deren Eindämmung zu dokumentieren. Eine freiwillige Beteiligung am CoPD findet ebenfalls nicht statt. Zudem sind private Messaging-Dienste von zentralen DSA-Pflichten ausgenommen – eine Kategorisierung, die Telegram trotz seiner öffentlichen Kanal-Funktion für sich in Anspruch nimmt.

Die aktuellen Nutzungsbedingungen enthalten keinen expliziten Hinweis darauf, dass Desinformation verboten ist. Moderiert werden laut Telegram Inhalte, die zu Gewalt oder terroristischer Propaganda aufrufen; Inhalte zu sexuellem Missbrauch von Minderjährigen werden entfernt. Die Moderation konzentriert sich damit auf klar ausgewählte Inhaltskategorien und folgt einer primär reaktiven Logik. Die Anwendung der Nutzungsbedingungen ist zudem auf öffentlich gekennzeichnete Gruppen und Kanäle beschränkt – Administrator:innen entscheiden allein, ob ein Kanal als öffentlich gilt, wodurch reichweitenstarke Kommunikationsräume formal als privat geführt werden können.

Entwicklung zu einer Plattform mit Desinformation

Während der Covid-19-Pandemie ab 2020 verdoppelte sich die Nutzung von Telegram; ein Großteil der rechtsextremen und verschwörungsideologischen Szene konzentrierte seine Kommunikation nun auf der Plattform. Bereits 2020 betrieben laut dem Institut für Demokratie und Zivilgesellschaft rund 96 % der einflussreichsten Akteure im Bereich des deutschsprachigen Rechtsextremismus einen Kanal auf Telegram (Donner u. a. 2025). Mit dem russischen Angriffskrieg 2022 kam die Rolle als Distributionskanal geopolitischer Desinformation hinzu: Nach der YouTube-Sperre des russischen Staatssenders RT DE wich dieser auf Telegram aus; die pro-russische Influencerin Alina Lipp erreichte im Dezember 2022 durchschnittlich 75.000 Aufrufe pro Nachricht (Holnburger 2023).

Studien zeigen, dass Desinformation auf Telegram nicht primär durch breite Massenverbreitung wirksam wird, sondern durch hochengagierte Minderheiten. Seriöse Beiträge mit Link zu externen Quellen treten zwar bis zu dreimal häufiger auf als irreführende, aber Kanäle mit irreführenden Inhalten sind bis zu 18 % aktiver. Je stärker der Anteil irreführender Quellen, desto reger der interne Austausch. Damit entstehen virale Effekte, die nicht auf algorithmische Verstärkung angewiesen sind – Viralität entsteht primär durch koordinierte Postings, gegenseitige Verlinkungen und die intensive Aktivität kleiner, aber gut vernetzter Gruppen (Herasimenka u. a. 2023).

Einordnung

Telegram kombiniert öffentliche Reichweiten- und Vernetzungslogiken mit dem regulatorischen Selbstbild eines privaten Messengers. Diese Kombination erlaubt es der Plattform, die infrastrukturellen Vorteile sozialer Medien zu nutzen, ohne deren regulatorische Verantwortung zu übernehmen. Die jüngste Einstufung von WhatsApp als VLOP durch die Europäische Kommission am 26. Januar 2026 – ausdrücklich wegen der öffentlichen Kanal-Funktion – unterstreicht die Problematik. Telegram weist funktional vergleichbare, teils weitergehende öffentliche Verbreitungsmechanismen auf, bleibt jedoch bislang unterhalb der VLOP-Schwelle (Digital Strategy EC 2026).

Fazit & Handlungsempfehlungen

Desinformation ist auf allen großen Onlineplattformen präsent und gewinnt insbesondere in politisch und gesellschaftlich sensiblen Kontexten an Relevanz. Sie verbreitet sich nicht nur über einzelne virale Ereignisse, sondern über längerfristige Dynamiken, die öffentliche Debatten, politische Meinungsbildung und Vertrauen in demokratische Institutionen beeinflussen. Wiederholte und miteinander verknüpfte Desinformationsnarrative tragen zur Erosion gemeinsamer Bezugspunkte bei und vertiefen bestehende Polarisierung.

Die im Dossier analysierten Beispiele zeigen, dass Soziale Medien nicht neutrale Verbreitungsinstrumente sind, sondern durch ihre technischen, ökonomischen und kommunikativen Plattformlogiken maßgeblich prägen, welche Inhalte Sichtbarkeit erhalten. Algorithmische Empfehlungssysteme, gruppenbasierte Kommunikationsräume mit Broadcasting-Funktionen und zunehmend auch KI-gestützte Inhalte verstärken die Präsenz und Verbreitung polarisierender, irreführender oder manipulierter Inhalte.

Was der DSA leistet und wo er an Grenzen stößt

Der DSA, die DSA-Transparenzdatenbank und der CoPD schaffen einen zentralen Rahmen, um Desinformation plattformübergreifend vergleichbar zu erfassen und Plattformen stärker in die Verantwortung zu nehmen. Systemische Risikoanalysen und die Verpflichtung zur Risikominderung tragen dazu bei, koordinierten Kampagnen im politischen Bereich entgegenzuwirken. Allerdings bleiben Maßnahmen wie die CoPD-Berichte auf freiwilliger Basis und sind somit abhängig von der Kooperation der großen Social Media Plattformen.

Am Beispiel Telegram wird jedoch deutlich, dass Moderation dort an Grenzen stößt, wo regulatorische Verpflichtungen nicht greifen. Die Einstufung von WhatsApp als VLOP vom Januar 2026, ausdrücklich wegen der öffentlichen Kanal-Funktion, illustriert eine zentrale Schwäche der aktuellen Plattformregulierung: Wo öffentliche Wirkung nicht mit regulatorischer Einordnung zusammenfällt, können sich Desinformationsdynamiken verbreiten und verfestigen.

Grenzen der Selbstauskunft

Die Wirksamkeit plattforminterner Maßnahmen bleibt ohne externe Kontrolle nur begrenzt beurteilbar, weil wesentliche Daten und Bewertungslogiken weiterhin bei den Plattformen selbst liegen und sich Berichtsformate hauptsächlich auf Selbstauskünfte stützen. Der CoPD funktioniert weitgehend als „à la carte"-Instrument: Signatare wählen selbst, welche Verpflichtungen sie für ihren Dienst umsetzen. Dieser Zugang untergräbt die Effektivität und erhöht den Bedarf an regulatorischer und zivilgesellschaftlicher Prüfung.

Umso problematischer ist, dass Plattformen in den letzten Jahren Maßnahmen teils zurückfahren oder nur selektiv umsetzen. Metas angekündigte Umstellung vom Third-Party-Fact-Checking auf Community Notes in den USA sowie die politische Rahmung dieser Neuausrichtung markieren eine sichtbare strategische Verschiebung, deren Auswirkungen in Europa erst mittelfristig sichtbar werden.

Zentrale Handlungsempfehlungen

→ Regulatorische Einstufung an tatsächlicher öffentlicher Wirkung ausrichten: Plattformen mit öffentlichen Verbreitungsmechanismen sollten nicht durch Selbstpositionierung als „privater Messenger" VLOP-Pflichten entgehen können.

→ Selbstauskunft durch unabhängiger Kontrolle und Datenzugang ergänzen: Fact-Checking-Organisationen, zivilgesellschaftliches Monitoring und unabhängige Forschung sind unverzichtbare Ergänzungen zu den Berichten der Plattformen. Voraussetzung dafür ist jedoch ein verlässlicher Zugang zu Plattformdaten, etwa über die in Art. 40. DSA vorgesehenen Datenzugangsrechte für „vetted researchers".

→ CoPD-Verpflichtungen verbindlicher gestalten: Die freiwillige Basis und Möglichkeit der selektiven

Auswahl von Commitments untergräbt die Wirksamkeit des Instruments und sollte durch verbindliche Standards ersetzt werden.

→ Themenübergreifende Konsistenz einfordern: Die geringere Reaktionsquote bei Klimadesinformation, Migration und geopolitischen Themen zeigt, dass Moderationsmaßnahmen nicht nur bei allgemeinpolitischen Inhalten wirksam sein dürfen.

Nur im Zusammenspiel aus klaren Verpflichtungen, unabhängiger Beobachtung und konsistenter plattformübergreifender Einordnung lassen sich strukturelle Desinformationsdynamiken nachhaltig adressieren.

Hinweis zu Quellen und weiterführenden Informationen

Diese Kurzfassung basiert auf dem ausführlichen Dossier „Desinformation auf Social-Media-Plattformen“ (ÖIAT, Jänner 2026), das im Rahmen des Projekts TGuard (K-Pass / KIRAS) erstellt wurde. Das vollständige Dossier enthält detaillierte Datentabellen, methodische Ausführungen sowie ein vollständiges Literaturverzeichnis mit rund 60 Quellen, darunter Berichte der Europäischen Kommission, CoPD-Transparenzberichte der Plattformen, wissenschaftliche Publikationen sowie Untersuchungen unabhängiger Organisationen (EFCSN, Maldita, Science Feedback, AI Forensics, Democracy Reporting International, CeMAS u. a.).

Das vollständige Dossier ist auf Anfrage über **research.oiat.at** verfügbar.

Quellen

Bouchaud, Paul, Marc Faddoul, und Raziye Buse Çetin. 2024. No Embargo in Sight. Meta Lets pro-Russia Propaganda Ads Flood the EU. AI Forensics. https://aiforensics.org/work/meta-political-ads.

Digital Strategy EC. 2026. „Commission designates WhatsApp as Very Large Online Platform under the Digital Services Act“. Januar 26. https://digital-strategy.ec.europa.eu/en/news/commission-designates-whatsapp-very-large-online-platform-under-digital-services-act.

Donner, Christian, Wyn Brodersen, Maik Fielitz, und Michael Schmidt. 2025. „Drei Jahre MATR-Monitoring auf Telegram: Die Rolle der Plattform im rechtsalternativen Online-Aktivismus“. Machine Against the Rage, Nr. Nr 8. https://doi.org/10.58668/matr/08.1.

Fundación Maldita. 2024. Platform Response to Disinformation during the EU Election 2024. Fundación Maldita. https://maldita.es/nosotros/20240611/platforms-disinformation-response-elections-eu/.

Herasimenka, Aliaksandr, Jonathan Bright, Aleksi Knuutila, und Philip N. Howard. 2023. „Misinformation and professional news on largely unmoderated platforms: the case of telegram“. Journal of Information Technology & Politics 20 (2): 198–212. https://doi.org/10.1080/19331681.2022.2076272.

Hohlfeld, Alexander, Anna Semenova, Martin Degeling, Greta Hess, und Kathy Meßmer. 2024. Tik-Tok, DSA O’Clock? auditing TikTok. Interface. https://tiktok-audit.com/blog/2024/Tik-Tok-oclock/.

Holnburger, Josef. 2023. Chronologie einer Radikalisierung: Wie Telegram zur wichtigsten Plattform für Verschwörungsideologien und Rechtsextremismus wurde. CeMAS - Center für Monitoring, Analyse und Strategie. https://report.cemas.io/telegram/.

Kimberly, Nicolaus, und Gabriele Scherndl. 2025. „Melden zwecklos: Tiktok ignoriert Desinformation zur Bundestagswahl“. correctiv.org, Februar 21. https://correctiv.org/faktencheck/hintergrund/2025/02/21/melden-zwecklos-tiktok-ignoriert-desinformation-zur-bundestagswahl/.

Meta. 2024. Code of Practice on Disinformation -Report of Metafor the Period 01 January to 30 June 2024. Nr. 4. Code of Conduct on Disinformation. Transparency Center. https://disinfocode.eu/reports.

Samson, Charlotte. 2024. „Use of information panels on YouTube videos containing misinformation – a ‘quick check’ study -“. Science Feedback, März 19. https://science.feedback.org/information-panels-youtube-videos-containing-disinformation-quick-check-study/.

Schulz-Tomančok. 2025. „Meta gegen Europa: Zensur oder Pflicht zum Fakten-Check?“ https://www.oeaw.ac.at/news/meta-vs-eu-der-kampf-um-faktenchecks.

Strasser, Iris, und Julia Krickl. 2025. Falsch- und Desinformation auf Social Media. Meldung problematischer Inhalte auf TikTok und Instagram als Trusted Flagger. ÖIAT. https://research.oiat.at/fileadmin/Research/Dokumente/Falsch-Desinfo_SocialMedia_%C3%96IAT_15072025.pdf.

Terroille, Charles, Kateryna Chystoforova, und Saman Nazari. 2024. Sanctioned but Thriving: How Online Platforms Fail To Address the Widespread Presence of Entities Under EU Sanctions. Science Feedback. https://science.feedback.org/sanctioned-but-thriving-how-online-platforms-fail-to-address-the-widespread-presence-of-entities-under-eu-sanctions/.

Impressum

Enstanden im Rahmen des Kooperatives F&E-Projekts TGuard und finanziert im Sicherheitsforschungs-Förderprogramm.

Bundesministerium
Finanzen

FFG
Forschung wirkt.

K-PASS

Durchführung der Recherche

Autor:innen: Julia Krickl MA | Natalie Trell MA | Lena Müller-Naendrup MSc

ÖIAT Research
Österreichisches Institut für angewandte Telekommunikation
www.research.oiat.at
Ungargasse 64-66/3/404
1030 Wien

Stand: Jänner 2026