

Löschung rechtswidriger Hassbeiträge bei Facebook, YouTube und Twitter

Ergebnisse des Monitorings von Beschwerdemechanismen jugendaffiner Dienste

Die Vielzahl fremdenfeindlicher und rassistischer Hasskommentare im Netz führte 2015 zur Bildung der Task Force "Umgang mit rechtswidrigen Hassbotschaften im Internet" des Bundesministeriums der Justiz und für Verbraucherschutz (BMJV). Die beteiligten Unternehmen (Google, Facebook, Twitter) sicherten die unverzügliche Löschung rechtswidriger Hassbeiträge und die anwenderfreundliche Gestaltung von Meldemöglichkeiten zu.

In einem vom Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) und vom BMJV finanzierten Projekt überprüft jugendschutz.net die Reaktionszeiten der Plattformen. Der jüngste Test fand Anfang 2017 statt.

Testaufbau

Für den Test ermittelte jugendschutz.net 540 strafbare Beiträge (§§ 130 und 86a StGB) und meldete sie den Diensten zunächst über einen Standard-User-Account, der jugendschutz.net nicht zugeordnet ist. Inhalte, die binnen einer Woche nicht gelöscht waren, wurden danach über einen akkreditierten Account erneut gemeldet (nur bei YouTube und Twitter möglich). Alle Inhalte, die nach einer weiteren Woche verblieben waren, gab jugendschutz.net abschließend an einen direkten E-Mail-Kontakt weiter.

Die Aufrufbarkeit der gemeldeten Inhalte überprüfte jugendschutz.net jeweils nach 24 Stunden, 48 Stunden und einer Woche. Ein Vortest im April/Mai 2016 diente dazu, das Testscenario zu erproben und für weitere Tests anzupassen. Der erste Haupttest wurde im Juli/August 2016 durchgeführt, der aktuelle Test im Januar/Februar 2017 (jeweils 8 Wochen).

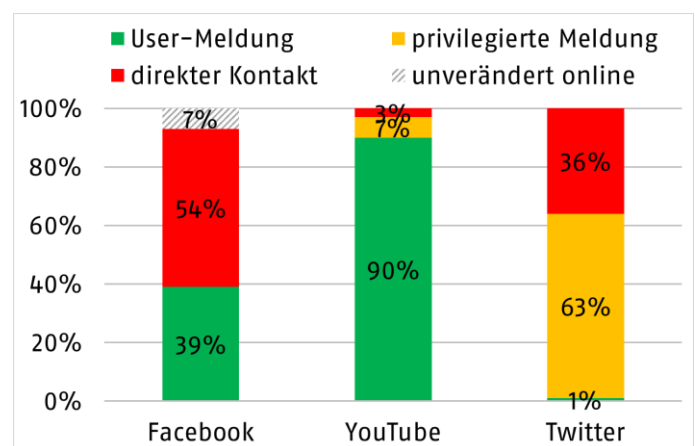
Löschquoten

Von den strafbaren Inhalten, die jugendschutz.net als User meldete, löschten oder sperrten Facebook 39 % (minus 7 % im Vergleich zum letzten Test), YouTube 90 % (plus 80 %) und Twitter weiterhin 1 % (grüne Markierung im Schaubild).

Bei YouTube wurden nach der Meldung als akkreditierter User weitere 7 % gelöscht oder gesperrt, bei Twitter 63 % (gelbe Markierung).

Nach direkten Kontakten per E-Mail löschte Facebook weitere 54 % (Löschquote 88 %), YouTube und Twitter entfernten alle verbliebenen Testfälle (rote Markierung).

Insgesamt löschten somit Facebook 93 % aller gemeldeten strafbaren Inhalte (plus 2 % im Vergleich zum letzten Test), YouTube 100% (plus 2 %) und Twitter 100 % (plus 18 %).



Löschquoten differenziert nach Meldemechanismen. Die Prozentzahlen beziehen sich auf alle gemeldeten Fälle.

Fazit

Die Vereinbarung in der Taskforce, die Mehrzahl der gemeldeten rechtswidrigen Hassbotschaften innerhalb von 24 Stunden zu entfernen, wird derzeit nur von YouTube eingehalten.

Im Vergleich zum letzten Test hat YouTube die Löschrquote bei der Meldung strafbarer Inhalte als User enorm verbessern können (von 10 % auf 90 %), Facebook löschte weniger (Rückgang von 46% auf 39 %). Twitter reagierte anhaltend schlecht auf User-Meldungen (weiterhin 1 %).

Bei Berücksichtigung aller Maßnahmen, die die Plattformen nach User-Meldung, Trusted Flagging und direktem Kontakt ergriffen haben, zeigten sich bei allen dreien Verbesserungen und sehr gute Löschrquoten (93 % bzw. 100 %).

Facebook, YouTube und Twitter bieten generell gute Meldemöglichkeiten für unzulässige Inhalte. Bei YouTube stehen sie ausschließlich angemeldeten Usern zur Verfügung.

Die Content-Richtlinien aller Dienste müssten optimiert werden. Zwar ist das Verbreiten von Hassinhalten überall ausgeschlossen, deutsche Rechtsverstöße sind jedoch nicht vollständig abgebildet.