

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.

Im Rahmen eines vom Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ) und vom BMJV finanzierten Projektes recherchierte jugendschutz.net die Reaktionszeiten der Plattformen bei der Meldung von rechtswidrigen Hassinhalten sowie die Gestaltung der Beschwerdeoptionen.

Testaufbau

Für den Test ermittelte jugendschutz.net 622 strafbare Beiträge (§ 130 StGB Volksverhetzung, Holocaustleugnung; § 86a Verwendung von Kennzeichen verfassungswidriger Organisationen) und meldete sie den Diensten zunächst als normaler User. Strafbare Inhalte, die nach einer Woche nicht gelöscht waren, meldete jugendschutz.net als privilegierter User (nur YouTube und Twitter) und übermittelte sie abschließend direkt an die Support-Abteilungen.

Ausgewertet wurden die Art der ergriffenen Maßnahmen (Löschung oder Sperrung für deutsche User) und die Reaktionszeit der Dienste (weniger als 24 Std.). Ein Vortest im April/Mai 2016 diente dazu, das Testszenario zu erproben und mit den Betreibern abzustimmen. Der Haupttest fand mit einer Dauer von acht Wochen im Juli/August 2016 statt.

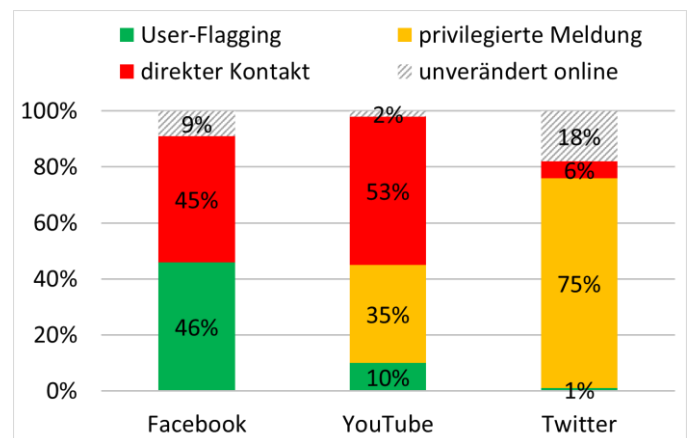
Löschquoten

Bei den strafbaren Inhalten, die jugendschutz.net als User gemeldet hat, erreichte Facebook eine Löschquote von 46 %. YouTube entfernte oder sperrte 10 % der als User gemeldeten strafbaren Inhalte, bei Twitter war es 1 %.

Meldeooptionen für privilegierte User stehen nur bei YouTube und Twitter zur Verfügung. Bei YouTube konnten über die Meldung als Trusted Flagger 39 % der verbliebenen Fälle gelöscht/gesperrt werden. Dies entspricht, bezogen auf die Gesamtzahl der gemeldeten strafbaren Inhalte, einer Löschquote von weiteren 35 %. Bei Twitter konnten über die Meldung als akkreditierter User 76 % der verbliebenen Fälle gelöscht/gesperrt werden und damit weitere 75 % aller gemeldeten strafbaren Inhalte.

Nach direkten Kontakten per E-Mail löschte Facebook 84 % der dann noch verbliebenen strafbaren Inhalte, YouTube 96 % und Twitter 24 %. Dies entspricht, bezogen auf die

Gesamtzahl der gemeldeten strafbaren Inhalte, bei Facebook einer Löschquote von weiteren 45 %, bei YouTube von weiteren 53 % und bei Twitter von weiteren 6 %.



Löschquoten differenziert nach Meldemechanismen. Die %-Angaben beziehen sich jeweils auf alle gemeldeten Fällen.

Schlussfolgerungen

Im Vergleich zum Vortest haben alle Dienste ihr Beschwerdemanagement verbessert. Beim User-Flagging gibt es aber weiterhin Nachholbedarf. Die Zusicherung in der Task Force, die Mehrzahl der gemeldeten rechtswidrigen Hassbotschaften innerhalb von 24 Stunden zu entfernen, wird noch von keinem Dienst eingelöst.

Bei Berücksichtigung aller Maßnahmen, die die Plattformen nach User-Meldungen, Trusted Flagging und direkten Kontakten ergriffen haben, ergeben sich gute Löschquoten (Facebook 91 %, YouTube 98 %, Twitter 82%; jeweils bezogen auf alle Fälle). Bei der Gesamtzahl der Löschungen und der Reaktionszeit waren Steigerungen im Vergleich zum Vortest zu verzeichnen. Nur Twitter liegt bei der Reaktionszeit deutlich zurück.

Facebook, YouTube und Twitter bieten generell gute Meldemöglichkeiten für unzulässige Inhalte. Bei YouTube stehen sie ausschließlich angemeldeten Usern zur Verfügung. Die Meldung von strafbaren Inhalten muss aber noch einfacher gestaltet werden und für jeden Nutzer möglich sein.

In den Content-Richtlinien aller Dienste ist das Verbreiten von Hassinhalten ausgeschlossen, deutsche Rechtsverstöße (Kennzeichen/Holocaustleugnung) werden jedoch nicht explizit erwähnt. Hier sollten die Dienste ihre Richtlinien entsprechend ergänzen.