

WIEN / 09. September 2021

Stellungnahme

**Verordnung der
Kommunikationsbehörde
Austria über die
Ausgestaltung der Berichte
und zum Umfang der
Berichtspflicht von
Diensteanbietern
(Transparenzberichte-VO)**

Für epicenter.works
Thomas Lohninger

 **EPICENTER
WORKS**
for digital rights



GENERELLE BEMERKUNGEN

Wir bedanken uns für die Möglichkeit im Konsultationsverfahren¹ Stellung zu nehmen. Bereits im Zuge des Gesetzgebungsprozess des Kommunikationsplattformen-Gesetz (KoPIG) durften wir uns mit einer Stellungnahme zu Wort melden und haben bereits damals die Transparenzverpflichtungen begrüßt und die Vorgaben zur vergleichbaren Ausgestaltung dieser Berichte durch die KommAustria befürwortet.²

Generell begrüßen wir den Entwurf und finden ihn sehr gelungen. Unser größter Kritikpunkt ist der blinde Fleck der automatisierten Inhaltsmoderation. Weitere Verbesserungsvorschläge betreffen aussagekräftigere Kennziffern für die Dauer von Überprüfungsverfahren und künftige Standardisierungen der grafischen Ausgestaltung von Verfahrensabläufen. Am wichtigsten finden wir die bereits getroffene Klarstellung, dass Moderationsentscheidungen auf Basis von Gemeinschaftsstandards nicht als Umgehungsmöglichkeit der Transparenzverpflichtungen genutzt werden dürfen.

Wir freuen uns, dass mit dem gegenständlichen Konsultationsverfahren auch ein inklusiver Prozess für deren Ausgestaltung stattfindet und hoffen dieser steigert den gesellschaftlichen Nutzen der gewonnen Erkenntnisse über die Moderationsvorgänge auf Kommunikationsplattformen.

Andere Positionen unserer Organisation im Bezug auf das KoPIG sind nicht Gegenstand dieser Stellungnahme, wobei wir anmerken wollen, dass viele der von uns hier im Bezug auf besonders große Kommunikationsplattformen als positiv bewerte Regelungen, für kleinere Anbieter existenzbedrohend sein könnten.

Wir würden einen ähnlichen inklusiven Konsultationsprozess zur Ausgestaltung der KommAustria Verordnung bezüglich Mindeststandards für Meldeformulare gemäß § 3 Abs 7 KoPIG anregen und hierbei insbesondere auf die Expertise der im Bereich „Hass im Netz“ tätigen Organisationen verweisen.

Inhaltsverzeichnis

| | |
|---|---|
| Generelle Bemerkungen..... | 2 |
| Spezifische Anmerkungen..... | 3 |
| Vollständigkeit im Sinne der auf Gemeinschaftsstandards basierenden Moderationsverfahren..... | 3 |
| Graphische Ausgestaltung des Moderationsprozesses..... | 3 |
| (teil)automatisierte Moderationsentscheidungen..... | 3 |
| Lösungsvorschläge..... | 4 |
| Schutz der Nutzerdaten..... | 4 |
| Overblocking..... | 4 |
| Aussagekräftigere Kennziffern zur Dauer von Überprüfungsverfahren..... | 4 |
| Organisatorische, personelle und technische Ausstattung..... | 4 |

1 <https://www.rtr.at/KonsultationVerordnungTransparenzberichteKopl-G>

2 <https://epicenter.works/document/2942>

SPEZIFISCHE ANMERKUNGEN

Vollständigkeit im Sinne der auf Gemeinschaftsstandards basierenden Moderationsverfahren

Wir begrüßen die Klarstellung im Verordnungsentwurf, dass auch Moderationsentscheidungen auf Basis von Gemeinschaftsstandard i.S.d §2 Z 3 unter die Transparenzverpflichtungen des KoPIG fallen, sofern diese durch die Meldefunktionalität gemäß § 3 KoPIG initiiert wurden. Da die allermeisten Gemeinschaftsstands illegale Inhalte bereits verbieten, wird nur durch die von der KommAustria vorgelegte Regelung in § 8 Abs 1 Z 1 verhindert, dass Diensteanbieter die Regelungen des KoPIG großflächig umgehen.

Eine Bearbeitung der Meldung auf Basis der Gemeinschaftsstandards ist immer billiger für den Diensteanbieter und bietet weniger prozedurale Garantien für die Nutzer*innen als das KoPIG mittels des dort normierten Beschwerde- und Überprüfungsverfahrens vorsieht. Deshalb ist es wichtig alle eingelangten Meldungen in die Transparenzberichtspflichten des Diensteanbieters aufzunehmen, da sich ansonsten ein Schlupfloch ergeben würde.

Graphische Ausgestaltung des Moderationsprozesses

Wir begrüßen die graphische Ausgestaltung der Moderationsprozesse gemäß § 6 Z 4 und Überprüfungsverfahren gemäß § 7 Z 2. Dieser Zugang öffnet ein wichtiges Thema einer breiteren Öffentlichkeit, das viele Betroffene oft in kafkaesker Ratlosigkeit zurück lässt. Es ist essentiell in diesbezüglichen Darstellungen auch auf Gemeinschaftsstandards basierende Entscheidungsprozesse zu inkludieren. Für eine künftige Revision der Transparenzverordnung wird weitere Konkretisierung auf etablierte Standards für Programmablaufdiagramme (Flussdiagramme) angeregt.³

(teil)automatisierte Moderationsentscheidungen

Unser größter Kritikpunkt im gegenständlichen Entwurf ist sein großer blinder Fleck im Bezug auf teil- oder vollautomatisierte Inhaltsmoderation. Seit mehreren Jahren gibt es einen sehr starken Trend weg vom „human in the loop“ und viele CEOs von amerikanischen Technologiekonzernen antworten auf Kritik an ihrer Inhaltsmoderation mit dem Versprechen einer Besserung durch weitere Automatisierung der Vorgänge (AI, machine learning, o.ä.). Derartige Automatisierungen führen jedoch eben nicht zu einer Verbesserung der Qualität oder Objektivität von Moderationsentscheidungen, sondern meist nur vermehrt zu strukturellem Overblocking.⁴ Insbesondere marginalisierte Gruppen sind oft bereits in den zugrundeliegenden Trainingsdaten von Diskriminierungen betroffen und werden durch den vermehrten Einsatz dieser Systeme strukturell in ihrem Recht auf Meinungsäußerungsfreiheit beschränkt.⁵ Wir bewerten diesen Trend als äußerst gefährlich und gesellschaftspolitisch diskussionswürdig.

Der gegenständliche Entwurf auferlegt Diensteanbietern in § 5 Z 3 lediglich eine makroskopische Betrachtung des Verhältnisses teil- oder vollautomatisierter Inhaltsmoderation im Vergleich zu menschlicher Inhaltsmoderation. Die Beschreibung des konkreten Einsatzes im Rahmen eines Meldeverfahrens gemäß § 6 Z 2 stellt lediglich auf den Zweck der verwendeten Software ab und nicht

3 Siehe DIN 66001

4 <https://blog.witness.org/2020/03/as-content-moderators-go-home-content-could-go-down/>

5 <https://techcrunch.com/2019/08/14/racial-bias-observed-in-hate-speech-detection-algorithm-from-google/?guccounter=1>

deren Hersteller oder vollen Funktionsumfang. Es wäre für die internationale Forschungsgemeinschaft unheimlich wertvoll, würden im Datenüberblick auch Statistiken über die Verwendung von teil- oder vollautomatisierten Systemen für Melde- und Überprüfungsverfahren ausgewiesen.

Lösungsvorschläge

§ 6 Z 2 sollte wie folgt ergänzt werden:

*Einsatz von **teil- oder voll**automatisierten Hilfsmittel bei der Bearbeitung von Meldungen mit Angabe der genauen Zwecke, **Softwarehersteller und des Funktionsumfangs**;*

§ 8 Abs 1 Z 1 und 2 sollte jeweils ein lit c ergänzt werden:

c. danach, ob die getroffene Entscheidung jeweils teil- oder vollautomatisiert getroffen wurde

Schutz der Nutzerdaten

Wir begrüßen die Anforderung im Sinne des Datenschutzes gemäß § 4 Abs 1 Z 2.

Overblocking

Wir begrüßen die Transparenzverpflichtung bezüglich Anstrengungen von Diensteanbietern zur Vermeidung von Overblocking in § 5 Z 4.

Aussagekräftigere Kennziffern zur Dauer von Überprüfungsverfahren

In § 7 Z 4 wird lediglich die Angabe der durchschnittlichen Dauer von Überprüfungsverfahren vorgegeben. Da mit besonders hohen Fallzahlen zu rechnen ist und einige Verfahren erwartbar sehr schnell (automatisiert) gelöst werden, wären weitere statistische Angaben wie Mittelwert und Streuung sinnvoll.

Organisatorische, personelle und technische Ausstattung

Wir begrüßen die detaillierten Angaben bezüglich der personellen Ausgestaltung der Moderationsteams und insbesondere die Transparenzverpflichtung bezüglich Maßnahmen zur Bewältigung gesundheitlicher Belastungen.