

Community Notes auf dem Prüfstand

VB verfassungsblog.de/community-notes-auf-dem-pruefstand/



Marc Bovermann

08 January 2024

Die größten Social Media Plattformen haben ein Problem mit Desinformation. Insbesondere auf X, vormals Twitter, war nach dem Terroranschlag der Hamas am 07.10.2023 und dem Beginn des Krieges in der Ukraine eine Flut an Falschinformationen feststellbar. Daher hat die EU-Kommission vor Kurzem mitgeteilt, dass sie ein förmliches Verfahren nach Art. 66 Abs. 1 Digital Services Act (DSA) gegen X eingeleitet hat. Gegenstand der Untersuchung ist unter anderem, ob die Plattform hinreichend gegen dieses Problem vorgeht. Tut sie das nicht, verstößt sie gegen ihre Risikominderungspflicht aus Artt. 35 Abs. 1, 34 Abs. 1 UAbs. 2 lit. c DSA mit der Konsequenz, dass die Kommission die Plattform verpflichten muss, einen Aktionsplan zur Risikominderung nach Art. 75 Abs. 2 DSA vorzulegen. Ebenso droht ihr eine Geldbuße nach Art. 74 Abs. 1 lit. a DSA.

X setzt dabei alles auf eine Karte: Wie aus dem X Transparency Report vom 03.11.2023 geschlossen werden kann, unterliegen Desinformationen nicht der sog. Content Moderation, sondern ihnen soll allein durch den Einsatz eines neuen Tools entgegengewirkt werden. Das heißt, dass die Nutzerinhalte auf X von Seiten des Betreiberunternehmens weder durch Algorithmen noch durch dazu beauftragte Personen auf Falschinformationen kontrolliert werden. Dieses Vorgehen spart nicht nur Kosten, sondern entspricht auch vollkommen dem nach der Übernahme durch Musk („free speech absolutist“) neu gefundenen Selbstverständnis von X. Und das Tool, in dem sich dieses spiegelt, sind die Community Notes (Deutsch: Kollektive Anmerkungen).

Die Community Notes als Allheilmittel

Community Notes sind Anmerkungen, die andere X Nutzer zu einem Tweet schreiben, der potenziell irreführend oder falsch ist. Die Anmerkungen werden neben dem jeweiligen Tweet angezeigt und sollen diesen, durch das Bereitstellen weiterer Informationen, kontextualisieren. Dafür gibt es eine eigene Backend Plattform, getrennt von der Hauptplattform. Hier können die Beitragenden sich anonym austauschen und sog. Notes verfassen. Hilfreiche Notes können dabei positiv, nicht hilfreiche Notes negativ bewertet werden. Auch die Antwort „teilweise hilfreich“ ist möglich. Haben genug Beitragende eine Note als hilfreich bewertet, so wird sie auch auf der Hauptplattform angezeigt. Die so ausgewählte Note kann im Nachhinein aber auch überstimmt werden.

Das Tool unterscheidet sich auch durch weitere Gestaltungsmerkmale deutlich von der Hauptplattform: So soll die Veröffentlichung des sog. Source Codes auf GitHub für eine gesteigerte Transparenz der Funktionsweise des Dienstes sorgen. Weiter basiert die Abstimmung darüber, ob eine Note öffentlich angezeigt werden soll, nicht auf dem Mehrheitsprinzip, sondern auf einem sog. Bridging-based Ranking System. Dieses ordnet die Beitragenden anhand deren früheren Abstimmungsverhaltens einer bestimmten Perspektive zu. Nur, wenn eine Note genügend positive Abstimmungen von Personen verschiedener Perspektiven erhält, wird sie auf der Plattform angezeigt. Personen, die sich zuvor uneinig waren, müssen sich über eine Note einig sein. So soll erreicht werden, dass nur im Ton neutrale und im Inhalt richtige Notes auf der Hauptplattform angezeigt werden.

Zuletzt hat X noch einige Änderungen vorgenommen, um die Anzeige von Notes auf der Hauptplattform zu beschleunigen. Dies soll dafür sorgen, dass mehr Nutzer die klarstellenden Notes sehen. Auch wurde eingeführt, dass für Tweets, die mit einer Note versehen sind, keine Werbung mehr geschaltet wird. Damit soll verhindert werden, dass Nutzer mit reißerischen oder irreführenden Aussagen Geld verdienen können.

Eingeführt wurde das Feature im Januar 2022, noch unter dem Namen des Vorgängerprojekts „Birdwatch“. Schon damals konnte Twitter Erfolge im Kampf gegen Desinformation erzielen. Nach der Übernahme durch Musk im November 2022 wurde das Feature schnell ausgebaut. Ende November dieses Jahres wurde dann die Möglichkeit, irreführende Tweets nach den X AGB zu melden, gänzlich abgeschafft. Die Community Notes nehmen somit die Rolle eines Allheilmittels im Kampf gegen Desinformation auf X ein.

Die Behauptung es handle sich um ein Allheilmittel, wäre für Musk wohl sogar noch eine Untertreibung. Nach seinen Vorstellungen liefert seine Plattform die akkuratesten Informationen im ganzen Internet. Auf dem Prüfstand befindet sich damit nicht nur die Frage, ob das Tool wirklich ausreicht, um das Systemrisiko der Plattform hinreichend zu mindern, sondern auch das neue Selbstverständnis der Plattform selbst.

Die Pflicht zur Bekämpfung von Desinformation

Die wichtige Vorfrage für die Kommission ist daher, ob X überhaupt eine Pflicht zu einem solchen Vorgehen trifft und wenn ja, wie dieses auszusehen hat.

Wegen Art. 33 Abs. 1 DSA kommt eine Pflicht zur Bekämpfung von Desinformation nur für sehr große Online-Plattformen (VLOPs) in Betracht. Zu diesem ausgewählten Kreis gehört auch X. Für die Frage, wie der DSA diese Pflicht aber en détail konstatiert, lohnt sich ein genauerer Blick in das Regelwerk.

Dem Normtext selbst lässt sich diese Pflicht nicht entnehmen. Stattdessen wird auf Artt. 35 Abs. 1, 34 Abs. 1 UAbs. 2 lit. c DSA abgestellt, die eine Pflicht zur Minderung von Systemrisiken für die gesellschaftliche Debatte normieren. Ein Blick in die Erwägungsgründe

bestätigt schnell, dass Desinformationen ein Systemrisiko darstellen (Erwgr. Nr. 83, 84, 88, 95, 104, 106). Explizit formuliert aber nur Erwgr. Nr. 104, der eigentlich der Erläuterung des Konzepts der Ko-Regulierung dient, dass Desinformationen auch ein systemisches Risiko für die Demokratie darstellen. Legt man Artt. 34 und 35 DSA also im Lichte dieses Erwägungsgrundes aus, lässt sich so eine Pflicht zur Bekämpfung von Desinformationen begründen – Problem gelöst, könnte man meinen.

Die Frage danach, wie die Risikominderung genau auszusehen hat, ist damit nicht gelöst. Der DSA bezieht hierzu keine explizite Stellung. Nach Erwgr. Nr. 2 bildet das Vorgehen gegen Online-Desinformation aber ein Hauptziel des DSA. Desinformationen seien auch ein Paradebeispiel für die gesellschaftlichen Risiken, die von VLOPs ausgehen. Vor diesem Hintergrund scheint es daher verwunderlich, dass der DSA das Problem im Rahmen der Erwägungsgründe nicht einmal klar verortet. In Erwgr. Nr. 82, der die systemischen Risiken für die gesellschaftliche Debatte näher erläutern soll, verliert der Verordnungsgeber jedenfalls kein Wort zu Desinformationen.

Die Kommission verweist zur Beantwortung der Frage nur auf den „Code of Practice on Disinformation“ (CoP). Diesen hat die Kommission im Jahre 2022 auch unter der Mitwirkung von Twitter beschlossen. Es finden sich dort viele, teils sehr spezifische Pflichten, die sich die teilnehmenden Unternehmen selbst auferlegt haben, um Desinformationen auf ihrer Plattform effektiv zu bekämpfen.

Verhaltenskodizes (wie dieser) sind nach der Konzeption des DSA in Art. 45 Abs. 1 weiterhin freiwillig. Dies stimmt für VLOPs aber nur in der Theorie, wie schon Erwgr. Nr. 104 andeutet: Ob eine VLOP hinreichende Risikominderungsmaßnahmen ergriffen hat, bestimmt sich auch danach, ob sie sich an einem Verhaltenskodex beteiligt. Sonst kann sie nur schwer nachweisen, dass sie ihrer Systemrisikominderungspflicht nachgekommen ist. Diese Wechselwirkung hat der Verordnungsgeber sogar kodifiziert: Die Kommission berücksichtigt eine Selbstverpflichtung bei dem Verfassen von Aktionsplänen nach Art. 75 Abs. 2 DSA. Letztlich erscheint es auch nur schwer möglich, dass eine VLOP ihr Systemrisiko hinreichend mindert, ohne sich gleichzeitig an einen entsprechenden Verhaltenskodex zu halten.

Identifizieren VLOPs daher also ein Systemrisiko, für das es einen Verhaltenskodex gibt, so ist die Einhaltung desselben faktisch verpflichtend. Auch für X gilt damit der CoP – sei es freiwillig oder unfreiwillig. Für Musk, der zuvor seine Beteiligung an eben diesem Verhaltenskodex beendet hat, heißt es daher: Welcome back!

Das Problem mit den Community Notes

Die Community Notes sollen also dazu dienen, den sich so ergebenden Verpflichtungen nachzukommen. Doch sind sie das versprochene Allheilmittel zu X' Desinformationsproblem?

Über die Halbwertszeit eines Tweets lässt sich nur wenig herausfinden. Klar ist nur: Sie ist sehr kurz. Das heißt, dass die meisten Nutzer einen Tweet in nur kurzer Zeit nach dessen Veröffentlichung sehen. Eine Diskussion und Abstimmung über korrigierende Notes können in dieser Zeit nicht stattfinden. Ein Großteil der Nutzer nimmt daher eine Falschinformation auf, ohne jemals eine möglicherweise klarstellende Community Note gesehen zu haben. Es handelt sich um ein strukturelles Grundproblem, das die akzessorische Stellung des Features mit sich bringt.

Als Ausgleich hat X ein System aufgesetzt, mit dem Nutzer benachrichtigt werden, die mit einem Tweet interagiert haben, der im Nachhinein mit einer Community Note versehen wurde. Dabei werden nur die Nutzer, die mit einem Tweet interagiert (Like, Retweet, Kommentar) haben, benachrichtigt. Nutzer, die den Tweet nur gelesen haben, werden nicht benachrichtigt. Beachtlich ist, dass die Anzahl der Nutzer, die einen Tweet nur gelesen, nicht aber mit ihm interagiert haben, meist wesentlich größer ist als die Anzahl der interagierenden Nutzer.

Ob die Beschränkung auf interagierende Nutzer ausreichend ist, scheint daher fraglich. Eine Antwort gibt Art. 33 Abs. 1 DSA. Dieser regelt die Vergabe des Status „VLOP“. Er wird einer Online-Plattform aufgrund der hohen Zahl der „aktiven Nutzer“ verliehen. Bei der Frage, welche Nutzer „aktiv“ sind, wendet sich Erwgr. Nr. 77 explizit dagegen nur Nutzer zu erfassen, die mit einer Information auf der Plattform interagieren. Dies entspricht ebenso dem CoP, der nur auf die „visibility“ von Desinformationen abstellt. Auch Nutzer, die sich die Inhalte auf der Plattform nur ansehen, sind damit „aktiv“. Auch sie sind den gesellschaftlichen Risiken der VLOPs ausgesetzt. Eine Risikominderung gegenüber dieser Nutzergruppe wird daher nicht zuverlässig erreicht.

Begrüßenswert ist zwar auch der Einsatz eines Bridging-based Ranking. Beachtlich ist aber, dass das System eigentlich ein Gegenmodell zum Engagement-based Ranking bilden soll. Letzteres basiert auf einem Kernprinzip der Plattformökonomie, das jedem Empfehlungssystem von VLOPs zugrunde liegt – auch dem von X: Es geht um die größtmögliche Interaktion mit einem Tweet und die Maximierung der auf der Plattform verbrachten Zeit. Dass nun ein neues Ranking System für die Community Notes eingeführt wird, welches die Defizite des eigentlich eingesetzten Ranking Systems ausgleichen soll, wirkt so, als würde man ein gebrochenes Bein mit einem Pflaster behandeln wollen.

Weiterhin sind die Community Notes auch der Manipulation durch die Nutzer ausgesetzt. Die Plattform unternimmt wenig dagegen, dass die Nutzer durch eine abgesprochene Koordinierung des Abstimmungsverhaltens das Ranking System manipulieren. Oftmals besitzen diese Nutzer auch mehr als nur einen Account. Hier könnte die Plattform durch striktere Kriterien bei der Nutzerauswahl und eine Moderation der Community Notes entgegenwirken. Das Gleiche gilt im Übrigen auch für die bereits bestehenden Accounts; hier könnte X noch strengere Maßstäbe anlegen.

Auch die Auswahl der Posts, die mit einer Community Note versehen werden sollen, obliegt den Nutzern. Es wurde dabei ein Fokus auf politische Inhalte festgestellt, obwohl diese nur einen geringen Anteil der Informationen auf der Plattform ausmachen. Auch bestimmte Accounts (beispielsweise von Elon Musk) sind besonders von Community Notes betroffen. Kombiniert mit der Möglichkeit der gezielten Manipulation des Dienstes durch kollusive Absprachen der Beitragenden, läuft das System so Gefahr, den politischen Interessen Dritter ausgeliefert zu sein. Die Plattform könnte dem entgegenwirken, indem sie die Auswahl der Notes, die ein Beitragender bewerten kann, so kuratiert, dass die Gefahr der Manipulation oder der Belästigung einer bestimmten Person minimiert wird.

Letztlich bleibt zu kritisieren, dass die Plattform im Übrigen jegliche Moderation von Falschinformationen eingestellt hat. Die Community Notes sind zwar eine begrüßenswerte Maßnahme zur Bekämpfung von Desinformation. Ein funktionierendes Moderationssystem können sie aber nicht ersetzen. Auch der CoP sieht die Einbindung der Nutzer nur als Teil eines größeren Gesamtkonzepts. Durch die fehlende Beachtung von Desinformation bei der Content Moderation fehlt so auch in der Unternehmensstruktur eine Verankerung des Ziels der Bekämpfung von Desinformation. X sollte beachten, dass systemische Risiken systemische Antworten erfordern.

Der DSA auf dem Prüfstand

Dies ist das erste förmliche Verfahren der Kommission unter dem DSA. Es befinden sich damit nicht nur die Community Notes auf dem Prüfstand, sondern auch das Regelwerk selbst. Das Verfahren greift dabei Kernanliegen des DSA auf. Mit Blick auf die erforderliche Rechtsfortbildung ist daher auf einen Nichteinhaltungsbeschluss nach Art. 73 Abs. 1 (iVm Art. 75 Abs. 2) DSA zu hoffen, in dem die Kommission die Möglichkeit hat, die Verpflichtungen des DSA zu konturieren.

Dies ist insbesondere für die besonders unklaren Systemrisikobewertungs- und –minderungspflichten zu hoffen. Besonders interessant wird dabei ein Zwiespalt, der sich schon bei Abschluss des CoP aufgetan hat: VLOPs sind nach Artt. 35, 34 DSA verpflichtet auch legale Inhalte, die Falschinformationen enthalten, zu bekämpfen. Gleichzeitig sind sie nach Art. 14 Abs. 4 DSA auch grundrechtsgebunden; eine Meinungsäußerung kann daher nicht einfach so von der Plattform entfernt werden. Dieser Widerstreit fiel auf X bisher unter der Kontrolle Musks zugunsten der „Free Speech“ aus – eine Option, die in der EU mangels First Amendments eigentlich nicht gewählt werden kann. Hier könnte die Kommission besonders unter Beweis stellen, dass der DSA scharfe Zähne hat und der digitale Raum nicht einfach den Vorstellungen Musks überlassen ist.

Ein Fazit in 280 Zeichen


Die Community Notes sind kein Allheilmittel. X ist seiner Risikominderungspflicht daher nicht hinreichend nachgekommen ist. Es liegt nun an der Kommission die (digitale) Öffentlichkeit auf X zurückzuerobern.

LICENSED UNDER CC BY SA

EXPORT METADATA

Marc21 XMLMODSDublin CoreOAI PMH 2.0

SUGGESTED CITATION Bovermann, Marc: *Community Notes auf dem Prüfstand*, *VerfBlog*, 2024/1/08, <https://verfassungsblog.de/community-notes-auf-dem-prufstand/>, DOI:

[10.59704/d39353743945e3b8](https://doi.org/10.59704/d39353743945e3b8) .

Explore posts related to this:

Other posts about this region:

Europa

LICENSED UNDER CC BY SA