

Facebook hat sich selbst eine unabhängige Kontrollinstanz gegeben, die als eine Art oberster Gerichtshof fungiert: Das Oversight Board hat im Herbst 2020 seine Arbeit aufgenommen und bislang sieben Fälle entschieden. PD Dr. Matthias C. Kettemann, Forschungsprogrammleiter „Regelungsstrukturen und Regelbildung in digitalen Kommunikationsräumen“ am Leibniz-Institut für Medienforschung | Hans-Bredow-Institut, forschte dazu, wie Facebook seine Gemeinschaftsstandards entwickelt und legitimiert. *tv diskurs* sprach mit ihm über den Einfluss, den das Board auf die Plattform haben könnte.



„Ein neues Zeitalter der Plattformmoderation!“

Was sind die Aufgaben des Oversight Boards, wie arbeitet es?

Das Board hat die Aufgabe, die Löschung einzelner Inhalte letztinstanzlich auf ihre Rechtmäßigkeit zu untersuchen. Ein Beispiel ist Donald Trump. Dessen Profil wurde gelöscht. Dem Board liegt aktuell die Frage vor, ob Facebook das durfte oder das Profil wieder ins Netz stellen muss.

Im Moment geht es also nur um Inhalte, die gelöscht wurden?

Ganz genau. Mit der Zeit soll das Board dann auch über Inhalte entscheiden, die Facebook nicht gelöscht hat.

Nach welchen Kriterien wird eine Vorauswahl getroffen?

Das Board soll für Einzelfälle Entscheidungen treffen, die aber eine grundsätzliche, über den jeweiligen Fall hinausweisende Bedeutung haben müssen. Wichtiger als die Einzelfallentscheidungen sind die Empfehlungen, die das Board geben kann. So hat das Board z. B. Facebook empfohlen, interne Regeln transparenter zu machen, klarer Entscheidungen zu kommunizieren, schneller zu reagieren, seine Algorithmen besser unter Kontrolle zu bekommen. Dadurch könnte sich tatsächlich mehr ändern als durch die Entscheidung, ob nun ein einzelner Inhalt wieder online geht oder nicht.

Das Board versucht, über die Empfehlungen Einfluss zu nehmen?

Ja. Sonst würde das Experiment gleich scheitern. Die Expertinnen und Experten müssen sich natürlich von Facebook freispielen. Das sind ja sehr renommierte Vertreterinnen und Vertreter der Zivilgesellschaft – eine Nobelpreisträgerin, Rechtswissenschaftlerinnen, Menschenrechtsaktivistinnen, der Ex-Chefredakteur des „Guardian“ – die durchaus einen Ruf zu verlieren haben. Sie versuchen natürlich, sich als eigene Instanz zu etablieren. Facebook ist einerseits stolz auf das, was man geschaffen hat.

Andererseits möchten man sich natürlich nicht zu sehr vom Board beeinflussen lassen. Das ist ein richtig spannendes Experiment! Wir hatten noch nie ein globales Inhalte-Privatgericht.

Das ist eine deutliche Kehrtwende hinsichtlich der Politik, die Facebook in Sachen Inhalte-Moderation bisher vertreten hat: Diese wurde bislang überwiegend an Drittanbieter in Billiglohnländer ausgelagert, es drang kaum etwas nach außen darüber, nach welchen Regeln Inhalte gelöscht oder beibehalten wurden, wie die Moderation abläuft. Facebook hat sich lange Zeit auf das Plattformargument zurückgezogen – wir sind nicht verantwortlich für die Inhalte, die kursieren. Jetzt machen sie genau das Gegenteil: öffentlichkeitswirksame Übernahme von Verantwortung.

Da muss man differenzieren. Das Geschäftsmodell hat sich nicht geändert, und Content-Moderation findet immer noch auf die von Ihnen beschriebene Weise mit unterbezahlten, schlecht behandelten Menschen statt, die im Hinblick auf die Kontrolle der Inhalte den Großteil der Schmutzarbeit machen. Aber die Schaffung des Boards stellt dennoch einen wichtigen ersten Schritt in Richtung von mehr Transparenz und Rechenschaftspflicht dar. Es wird gesellschaftlich nicht mehr akzeptiert, dass Unternehmen große und wichtige Moderationsentscheidungen ganz allein treffen. Die Bedeutung dieser Plattformen wächst immer weiter. Das haben wir inzwischen erkannt – und die Plattformen auch. Mit Blick auf die Pandemie und den letzten amerikanischen Präsidentschaftswahlkampf haben die Plattformen gesehen, dass sie stärker moderierend eingreifen müssen. Wir erleben eine Zeitenwende. So viel moderiert wie im letzten halben Jahr haben Facebook, Twitter und die anderen noch nie. Sie haben Tweets von Donald Trump mit Warnhinweisen versehen, Desinformationen zu Corona gelöscht, ganze Netzwerke von Rassistinnen und Rassisten offline genommen. Das ist ein neues Zeitalter der Plattformmoderation. Auch wenn sich nicht gleich alles ändern wird.

Gibt es überhaupt eine Rückbindung zwischen dem Board und der Content-Moderation?

Das Board ist unabhängig konstruiert. Sie haben eigenes juristisches Fachpersonal und sind nicht verbunden mit Facebooks Moderationsarmee. Inhalte, die ohne guten Grund gelöscht wurden, sollen wiederhergestellt werden. Hinsichtlich der Empfehlungen des Boards muss Facebook in den Regeln seine Moderationsrichtlinien ändern. Das dauert schon. Es geht also immer über das Unternehmen selbst. Es gibt keinen direkten Zugriff, keine Durchgriffsrechte vom Board auf die Moderation.

Auch nicht umgekehrt? Nutzer, die gesperrt wurden, können sich beim Board beschweren. Aber Moderatorinnen und Moderatoren können nicht direkt Zweifelsfälle an das Board weitergeben?

Da haben Sie recht. Das machen sie nur intern. Da bestehen interne Hierarchien. Akute Fälle, etwa konkrete Drohungen, werden „eskaliert“, schnell zu zuständigem Facebook-Personal geleitet. Bei nicht so dringenden Zweifelsfällen dauert das länger. Moderatorinnen und Moderatoren, die allerdings zu oft vermerken, dass sie unsicher sind, bekommen interne Strafpunkte. Strukturell scheint es so zu sein, dass eher Geschwindigkeit honoriert wird als tiefere Reflexion über die Feinheiten der Inhalte. Was bei so einem Knochenjob aber auch schwer möglich ist.

Man hat sich mit dem Board also eine Parallelstruktur geschaffen ...

Ja, die Moderation ist das normale Businessmodell, während das Board ein Luxusgremium ist. Ein öffentlichkeitswirksames Beratungsorgan für entscheidende Fragen. Die tagtägliche Moderation beeinflusst das Board nicht, zumindest nicht direkt.

Man darf aber nicht vergessen, dass bis zu 95 % aller gelöschten Inhalte ohnehin automatisiert erkannt werden. Auch die Arbeit der Moderatorinnen und Mode-

ratoren macht nur einen relativ kleinen Teil der gelöschten Inhalte aus.

Wie verhält sich das Board zu diesen versteckten Steuerungsmechanismen?

In einem der ersten Fälle ging es um die automatisierte Löschung eines Bildes, das eine weibliche Brustwarze im Rahmen einer Brustkrebskampagne zeigte. Facebook selbst hat die Löschung rückgängig gemacht, gesagt, das sei ein Versehen gewesen, und das Board gebeten, diesen Fall nicht anzunehmen. Das Board hat den Fall mit Blick auf seine grundsätzliche Bedeutung trotzdem verhandelt. Eine Kernaussage aus diesem ersten Urteil des Boards ist, dass Facebook viel mehr tun muss, um seine Algorithmen unter Kontrolle zu bekommen.

Hat Facebook gesagt, was man tun wird, um die Algorithmen stärker zu kontrollieren?

Facebook hat den überwältigenden Anteil der Empfehlungen angenommen und gesagt, man werde diese implementieren. Wie genau die Umsetzung aussieht, steht noch nicht bis ins letzte Detail fest. Dass Facebook in Richtung transparentere und menschenrechtssensiblere Moderation nun mehr macht, ist gut und richtig. Ob das jetzt nur wegen des Boards ist, sei dahingestellt. Manche der angekündigten Maßnahmen, so scheint es, waren auch ohne Entscheidung des Boards im Köcher des Unternehmens. Es ist aber jedenfalls gut, dass wir jetzt mehr darüber reden, wer bestimmt, was wir online sagen dürfen. Das ist schon ein Wert an sich. Generell ist es ein Problem, dass die breite Öffentlichkeit zu sehr auf die Entscheidungen des Boards schaut und weniger darauf, welche daraus folgende Implementierung Facebook umsetzt. Die Öffentlichkeit blickt auch zu sehr auf plakative Einzelfälle, wie z. B. die Löschung des Accounts von Donald Trump, und weniger auf die strukturellen Bedingungen, wie die Empfehlungsalgorithmen, die ein Kommunikationsphänomen wie Trump erst möglich gemacht haben.

Auf welcher Grundlage trifft das Board seine Entscheidungen? Sind es Facebooks Community Standards?

Dem Board wurde in die internen Regeln geschrieben, dass sie primär auf Grundlage der Gemeinschaftsstandards und Werte von Facebook entscheiden, darüber hinaus aber internationale Menschenrechte miteinbeziehen können. Das Board hat jedoch sehr schnell eine spannende Kehrtwende vollzogen. Es hat nämlich in einer der ersten Entscheidungen gesagt, dass man seine Entscheidungen zwar immer auch an den Gemeinschaftsstandards misst. Wobei es den Mitgliedern dabei vor allem darum geht, zu schauen, ob die Gemeinschaftsregeln konsequent umgesetzt werden. Aber: Facebook habe sich auch zu den internationalen Menschenrechten bekannt und die würden maßgeblich – in den meisten Entscheidungen weit überwiegend – herangezogen werden. Das konnte man so schreiben, weil Facebook sich zu den UNO-Leitprinzipien für Wirtschaft und Menschenrechte bekannt hat. Die helfen dem Board, wichtige Fragen zu stellen: Werden die internationalen Menschenrechte eingehalten? Was ist verhältnismäßig? Wo hört Ironie auf, wann beginnt Hassrede? Wie muss man die Rechte Einzelner und die Rechte der Gesamtheit gegeneinander aufwiegen? Wir sehen hier sehr schön, wie globale Menschenrechte importiert werden in diese privaten Ordnungen. Das ist ein großer Vorteil des Boards.

Sie haben auch dazu geforscht, wie Facebook seine Gemeinschaftsstandards entwickelt hat?

Ganz am Anfang war die Regel so, dass das, was den Moderatorinnen und Moderatoren komisch vorkam, gelöscht wurde. Das war für die ersten paar Monate noch okay. Dabei haben sich dann Regeln herausgebildet, die zuerst intern entwickelt und in inoffiziellen Dokumenten gesammelt wurden. Erst mit der Zeit haben sich diese Standards immer mehr ausdifferenziert. Wir konnten 2019 in der Unternehmenszentrale

von Facebook plattformethnografische Forschung betreiben und haben uns angeschaut, wie Facebook – noch zwei Jahre, bevor das Board ins Leben gerufen wurde – versucht hat, die Gemeinschaftsstandards legitimer zu machen.

Und wie?

Durch eine Erhöhung der Inputlegitimität, also durch eine Erhöhung der Qualität und Repräsentativität des Inputs von außen. Deswegen gibt es bei Facebook eine Abteilung, die sich dem Stakeholder-Engagement widmet, die also eine Rückbindung an andere Stakeholder, gerade die Zivilgesellschaft, sucht. Mitarbeitende telefonieren dann, wenn eine neue Regel in Arbeit ist, mit wichtigen gesellschaftlichen Akteuren, NGOs, Wissenschaftlerinnen und Wissenschaftlern, und testen ab, wie die geplante Regel wirken würde – und wie sie besser gemacht werden könnte. Das Problem ist, dass es neben diesem ernsthaften und guten Versuch, sinnvolle Regeln zu setzen, immer eine interne Parallelstruktur gab. Einzelne mächtige Manager oder der Chef, Mark Zuckerberg selbst, haben – das wissen wir von Leaks – in mehreren Fällen Entscheidungen darüber, was online bleibt und wie Algorithmen eingesetzt werden, getroffen. Eine strikte Interpretation der Gemeinschaftsstandards im Lichte der Minimierung potenzieller gesellschaftlicher Gefährdungen wurde nach politischer Opportunität ausgehebelt. Ein Beispiel dafür ist der Bereich „Holocaustleugnung“. Im amerikanischen Facebook war Holocaustleugnung bis zum Sommer 2020 erlaubt. Noch 2019 sagte Zuckerberg in eine Rede, er finde die Meinung abscheulich, müsse sie aber als Teil der „free speech“ schützen. Und dann, von einem Tag auf den anderen, beschloss die Unternehmensführung, dass das nun doch problematisch und Holocaustleugnung nicht mehr erlaubt sei. Inhaltlich ist das natürlich eine gute Entscheidung! Aber wie die Entscheidung getroffen und durchgesetzt wurde, fordert das ganze System heraus,

das eigentlich rational abgesichert sein sollte. In diesem Prozess hilft das Oversight Board. So eine Frage könnte man vorlegen und diskutieren lassen, wie sie das auch gerade mit der Löschung von Trumps Account machen.

Wie bewerten Sie ganz allgemein die Balance zwischen dem Recht auf freie Meinungsäußerung und der Durchsetzung von Regeln, die den Einzelnen oder bestimmte Werte auf der Plattform schützen sollen? Findet da eher ein Overblocking statt, auch durch Inhalte-Moderation, oder müsste Facebook noch viel stärker eingreifen, um z. B. etwas gegen Hassrede zu unternehmen?

Aus der Forschung der letzten Jahre weiß man, dass von ungefähr hundert Inhalten fünf gelöscht werden. Von diesen fünf gelöschten wurden in der Regel 90 % wegen Gemeinschaftsstandards gelöscht und 10 % wegen Rechtswidrigkeit. Das bedeutet, schon jetzt löschen die Plattformen sehr viel mehr, als sie rechtlich müssten. Das kann zu einem großen Problem für die Meinungsäußerung werden, wenn die internen Regeln, nach denen gelöscht wird, nicht klar und legitim sind – oder wenn Staaten durch die Hintertür versuchen, nicht verbotene, aber unerwünschte Inhalte löschen zu lassen. Gegen Hassrede sollte Facebook stärker vorgehen, vor allem gegen die Empfehlungsalgorithmen, die Gruppen zusammengeführt haben, die im echten Leben niemals hätten zusammengeführt werden sollen.

Das ist primär das Werk von Algorithmen, diese Gruppenzusammenführungen?

Es ist nachgewiesen, dass Empfehlungsalgorithmen in der letzten Zeit den Effekt gehabt haben, separate Gruppen zusammenzuführen, die nur marginale Anknüpfungspunkte hatten, beispielsweise QAnon und die Impfgegnerbewegung oder in Deutschland die Impfgegner und die Reichsbürger. Die beiden haben einen einzigen

Überschneidungspunkt: eine gewisse Distanz zu Autoritäten, ein Misstrauen gegenüber dem Staat. Die Empfehlungsalgorithmen sind inhaltsblind. Diese Gruppenzusammenführung war fatal. Man darf die Algorithmen aber auch nicht überschätzen. Bei einer Studie zu den Algorithmen von YouTube konnte nicht nachgewiesen werden, dass bewusst Desinformation Inhalte gepusht wurden. Die Algorithmen haben keine nur negative Rolle, können aber, wenn sie nicht kontrolliert werden, auch Negatives zusammenführen und verstärken.

Was tut Facebook bislang, um diese negativen Effekte der Algorithmen einzugrenzen?

Vor der US-Wahl hat man gesagt, man empfehle keine politischen Gruppen mehr, man lasse auch keine Werbung für bestimmte politische Inhalte mehr zu. Man hat versucht, die gesamte QAnon-Bewegung und die Proud Boys aus Facebook zu löschen. Im Bereich der Impfgegnerschaft werden coronabezogene Impfinformationen entweder gelöscht oder deamplifiziert und demonetarisieren, zumindest werden sie nicht verstärkt. Es gibt keine entsprechenden Vorschläge mehr – diese künstlichen Verbindungen werden reduziert. Das Board hatte im Übrigen festgestellt, dass Facebook etwas zu offensiv coronabezogene Desinformation löscht, und mit Blick auf die freie Meinungsäußerung empfohlen, mehr zu tolerieren. Aber Facebook hat sich gegen diese Empfehlung entschieden und ist da sehr rigoros. Ich finde, in dieser Frage ist das Board etwas zu akademisch vorgegangen. Wichtig ist hier aber etwas anderes: dass Facebook erklärt, wen und was es warum löscht – wir haben ein Recht auf Rechtfertigung, wann immer unsere Rechte von privaten Akteuren berührt werden.