

عصر
فضای
مجازی

عصر
فضای
مجازی

گزارش شماره ۶۸

خرداد ۱۴۰۰



مرکز ملی فضای مجازی
پژوهشگاه فضای مجازی

واقعیات و اینکه کجامه توان آنها را یافت: پژوهش تجربه بر روی سگوهای اینترنتی و تعدیل محتوا (دَفنه کِلروپدی لیرسن)

محتوای انتشار یافته در این اثر
الزاماً بیانگر دیدگاه مرکز ملی فضای مجازی نیست

تهیه شده در پژوهشگاه فضای مجازی
(گروه مطالعات بنیادین)

مترجم: محمد امین شفیع خانی

ناظر علمی: دکتر علیرضا کاظمی - دکتر حسین
مطلبی کریمندی

حقوق مادی و معنوی این اثر متعلق به مرکز ملی فضای
مجازی است و استفاده از آن با ذکر منبع مجاز می باشد.

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نیش
خیابان ۱۶ غربی، پلاک ۲۰
تلفن: ۰۲۱-۸۶۱۵۱۰۶۱
کد پستی: ۱۵۱۵۶۷۴۳۱۱

فهرست

۵ سخن نخست

۹ چکیده

۱۳ مقدمه

بخش اول

حذف و قوانین قابلیت مسئولیت واسطه‌ای — ۲۱

بخش دوم

منابع اطلاعات — ۲۳

بخش سوم

پژوهش مستقل — ۵۵

بخش چهارم

نتایج حذف محتوای سگوها — ۶۵

بخش پنجم

پرسشهای تجربی درباره‌ی حذفیات محتوای سگو — ۷۱

جمع‌بندی — ۷۷

منابع — ۸۱

سخن نخست



فضای مجازی با شتاب شگرف و رو به تزایدی که در حال بسط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را درنوردیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می‌دهد. لذا به نظر می‌رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالاخص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزاری کنار سایر ابزارهای بشری تصویر می‌کرد که تنها طریقت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره‌کننده فضای مجازی و سایه گسترتری آن در حوزه‌ها و شئون بشر در یک دهه اخیر آن را چون سکویی می‌داند که بسیار فراتر از شأن ابزاری حیات انسان‌ها را سامان جدیدی داده و ادعای تمدن نوینی را دارد. رویکردی که از قضا از چشمان بصیر رهبر انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته‌اند.

در همین راستا گزارش‌های عصر فضای مجازی تلاش می‌کند تا فهم سازمان‌ها و دستگاه‌های مرتبط با حوزه فضای مجازی را ارتقاء بخشیده و آن‌ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه مهیا سازد.

سید ابوالحسن فیروزآبادی

دیسرورای عالی‌ترین مرکز ملی فضای مجازی

چکیده



مقاله «واقعیات و اینکه کجا می‌توان آنها را یافت: پژوهش تجربی بر روی سگّوهای اینترنتی و تعدیل محتوا» نوشته دَفنی کِلِر، مدیر برنامه تنظیم‌گری سکو در مرکز سیاست‌گذاری سایبری استنفورد و پدی لیرِسِن از دانشگاه آمستردام است که در سال ۲۰۲۰ در کتاب «رسانه‌های اجتماعی و مردم‌سالاری» چاپ انتشارات دانشگاه کمبریج به چاپ رسیده است. تمرکز این مقاله روی ادبیات موجود مربوط به حذف محتوای اینترنتی توسط دولت و سگّوها است. نویسندگان به داده‌های موجودی اشاره می‌کنند که توسط دولت‌ها، دانشگاهیان، سازمان‌های مردم‌نهاد (سمن‌ها) و خود سگّوها منتشر شده است که نشان می‌دهد حذف محتوا چه‌قدر و به چه دلایلی رخ می‌دهد. حذف محتوا مبتنی بر قواعد داخلی سگّوها و مقررات حکومت‌ها تبدیل به یک رویه پیچیده و رایج شده است که علاوه بر آثار مثبت (در حذف محتواهای مشکل‌آفرین)، آثار منفی فراوانی (همچون خطا در حذف محتوا و سوگیری و تعصب‌های نژادی، سیاسی و مذهبی) داشته است. بررسی میزان موفقیت رویه‌های حذف محتوا، آسیب‌شناسی مشکلات آن‌ها و همچنین ارائه پیشنهاد مؤثر برای اصلاح این رویه‌ها نیازمند پژوهش‌های تجربی گسترده روی فعالیت‌های

حذف محتوای سکوها و قوانین و قواعد موجود است که روش‌های متنوع تجربی را می‌طلبد. این مقاله که در آن بالغ بر ۱۲۰ اثر به‌روز بررسی شده است، به خوبی آخرین نتایج و تحلیل‌های تجربی از قوانین، قواعد و رویه‌های حذف محتوای سکوها را ارائه می‌کند و مهم‌ترین مسائل باقی‌مانده برای پژوهش‌های آینده و همچنین پیشنهادهایی برای بهبود عملکرد سیاست‌های حذف محتوا را مطرح می‌کند. از این نظر، این مقاله می‌تواند برای سیاست‌گذاران فضای مجازی کشور برای مؤثرتر ساختن سیاست‌های حذف محتوا و کاستن از نقایص آن‌ها بسیار مفید باشد.

واژگان کلیدی: تعدیل محتوا، حذف محتوا، خطوط راهنمای اجتماع

سکوها، فیلترینگ

مقدمه



«واقعیت‌ها اینک که می‌توان آنها را یافت: پژوهش تجربی بر روی سک‌وهای اینترنتی و تعدیل محتوا»

ما در دورانی زندگی می‌کنیم که در آن نگرانی‌های فزاینده‌ای وجود دارد که سک‌وهای اینترنتی مانند فیس‌بوک^۲ یا توئیتر^۳، در گفتگوهای برخط ما وساطت می‌کنند، همچنین نفرت می‌پراکنند، اطلاعات غلط نشر می‌دهند و دستاوردهای سیاسی را منحرف می‌کنند. به عنوان نمونه، انتخابات ریاست جمهوری ۲۰۱۶ ایالات متحده، موجی از دغدغه‌ها درباره‌ی آسیب‌های ناشی از سک‌وها را برانگیخت. سیاست‌گذاران سراسر جهان قوانینی را خواستار شده‌اند که برای سک‌وها به منظور مقابله‌ی بیشتر با محتوای غیر قانونی و حتی محتوای به کلی مخرب، مورد نیاز است.

از منظر خود سک‌وها، این پیشنهادها وجوه اشتراک زیادی دارند. جدای از فرامین مجزا - برای مثال پرداختن به محتوایی که گمراه‌کننده، نفرت‌افکن یا خشن است - تبعیت از همه این فرمان‌ها مستلزم فرایندهای عملیاتی مشابه هستند. سک‌وها هم‌اکنون این فرایندها را برای اعمال قوانین موجود و خطوط راهنمای اختیاری اجتماع^۴ شان داشته‌اند. هر تلاش جدیدی برای تنظیم برخط محتوا احتمالاً بر پایه‌ی سامانه‌ها، کارکنان و ابزارهای موجود ایجاد خواهد شد - و طبعاً هم نقاط قوت و هم نقاط ضعفشان را

۱. دفنی کالر (Daphne Keller) برنامه‌ی تنظیم‌گری سک‌وها را در مرکز سیاست سایبری استنفورد اداره می‌کند و سابقاً در شورای عمومی گوگل (Google) عضو بوده است. پدی لیرسین (Paddy Leerssen) دانشجوی دکتری مؤسسه‌ی حقوق اطلاعات (IVIR) در دانشگاه آمستردام است، و همکار غیر مستقر مرکز اینترنت و جامعه در استنفورد است.

2. Facebook

3. Twitter

4. Discretionary Community Guidelines

اطلاعات قابل اعتماد درباره‌ی سامانه‌های حذف‌کننده‌ی محتوای سکوها، سال‌ها بود که به دشواری به دست می‌آمد؛ اما به طور پیوسته داده‌ها و پرده‌برداری‌هایی در نتیجه‌ی تمرکز پژوهشگران بر روی موضوع و تلاش‌های سکوها برای افزایش شفافیت خود، بروز می‌یابند. این فصل، منابع جاری و احتمالاً آینده‌ی اطلاعات را مرور می‌کند. برخی از حذفیات محتوایی^۱ نیازمند قانون هستند، درحالی‌که دیگر موارد داوطلبانه اجرا می‌شوند. حذفیات حقوقی توسط قوانین مسئولیت واسطه‌ای،^۲ که به سکوها می‌گوید که چه مسئولیتی در قبال محتوای غیر قانونی‌ای که توسط کاربرانشان منتشر می‌شوند داشته باشند، سامان می‌یابند. عموم سکوها تحت چارچوب‌هایی قانونی مانند «اقدام حق نشر هزاره‌ی دیجیتال ایالات متحده»^۳ (DMCA) یا «رهنمود تجارت الکترونیک اتحادیه‌ی اروپا»^۴ الزامات حقوقی‌شان را در به‌کارگیری سامانه‌های «تذکر و حذف»^۵ لحاظ می‌دارند. سکوهای بزرگ‌تر بر روی این عملیات سرمایه‌گذاری سنگینی می‌کنند یا حتی بعضاً این سرمایه‌گذاری‌ها را با تلاش‌هایی فعالانه به قصد تعریف و حذف مواد غیر قانونی تکمیل می‌کنند. هر چند شواهدی که ما بررسی کردیم نشان می‌دهند که سکوها در پایبندی مستمر به قانون، در کل ضعیف عمل کرده‌اند.

اما حذف داوطلبانه‌ی محتوای سکوها بر مجموعه قواعد خصوصی‌ای بنا شده که عبارت‌اند از: خطوط راهنمای اجتماع.^۶ این استانداردهای خصوصی غالباً حاشیه گسترده‌ای از گفتار قانونی را معین می‌کنند که اقدام فزاینده‌تر از آن، تخطی از آن خواهد بود. خطوط راهنمای اجتماع ممکن است بر اساس باورهای اخلاقی خود اپراتور سکو و یا بر

1. Takedown

2. Intermediary Liability Laws

3. The US Digital Millennium Copyright Act: DMCA

4. The EU's eCommerce Directive

5. Notice-and-takedown

6. Community Guidelines

اساس هنجارهای اجتماعی ترسیم شود. این راهنماها حتی می‌توانند به دنبال شکل‌دهی به تجربیات کاربر با اهداف کسب‌وکاری باشند. مثلاً یک پایگاه آگهی املاک و مستغلات، بایستی از انتشار تصاویری غیر از تصاویر ساختمان‌ها و املاک، ممانعت به عمل بیاورد. عمده‌ی پایگاه‌های اینترنتی از انتشار هرزنامه، هرزه‌نگاری و نتایج مشابه آزاردهنده جلوگیری می‌کنند. سگوها همچنین می‌توانند از خطوط راهنمای اجتماع به منظور تسهیل دشواری‌های ناشی از الزام حقوقی و خودداری از درگیری با دولت‌ها استفاده کنند، از این طریق که گفتگوی فراتر از حدّ قانون را ممنوع کنند. دولت‌ها، به جهت نقشی که دارند، ممکن است به خاطر قانون اساسی یا قانون حقوق بشر از چالش اجتناب کنند، اگر حذف محتوای قانونی به فعل خصوصی، و نه دولتی، نسبت داده شده باشد.^۱ بنابراین، از آن روی که سیاستگذاران و مردم به طور فزاینده‌ای مطالبه می‌کنند که سگوها محتوای زیان-آور یا توهین‌آمیز را حذف کنند، هرچند آن محتوا لزوماً غیر قانونی نباشد، این قواعد اختیاری مهم‌تر از همیشه شده‌اند.

تلاش‌های جدید سگوها برای کنار گذاشتن محتوای ممنوع، ناگزیر اشتراکات زیادی با این سامانه‌های موجود خواهند داشت. این مسئله از یک وجه خبری خوب است، چون سیاستگذاران روی سابقه‌ی سفید و خالی حساب باز نمی‌کنند. حقوقدانان، پژوهشگران و کارمندان سگوها دو دهه تجربه دربارهِ راه‌هایی که سامانه‌های حذف محتوا در عمل اتخاذ می‌کنند، اندوخته‌اند. البته همین مسئله، از وجه دیگر خبر بدی است. شواهد نشان می‌دهند که سگوها کاری جدّی برای اعمال قواعد گفتگو انجام نمی‌دهند. حتی

هنگامی که خطوط راهنمای اجتماع را که خودشان تعریف کرده‌اند هم به کار می‌بندند، باز هم نتایج اغلب آشفته به نظر می‌رسند. برای پژوهشگران مستقل، ارزیابی کمی صحت عملکرد سکوها در به‌کارگیری استانداردهای خطوط راهنمای اجتماع دشوار است، چرا که خود قواعد، به خوبی فهم نشده‌اند و اطلاعات موثق ناچیزی درباره‌ی متون، تصاویر، ویدیوها یا دیگر محتویات محذوف در دست است. عملکرد سکوها تحت قوانینی همچون «قانون حق نشر ایالات متحده»^۱ یا «قانون گفتگوی نفرت آلمان»^۲ برای پژوهش، آسان‌تر در دسترس‌اند چون قواعد به جای آنکه از خطوط راهنمای اختیاری سکوها دربیایند، از قانون عمومی درمی‌آیند. همچنین در این باره که سکوها چگونه قواعد حقوقی را بکار ببندند و اینکه کدام محتوای خاص را حذف می‌کنند، اطلاعات بیشتری در دسترس است.^۳ این امر به متخصصان اجازه‌ی ارزیابی اعمال مربوط به حذف در سکوها - و نیز اجازه‌ی مستندسازی درباره مشکلات مهم را داده است. سکوهایی که تذکراتی دریافت می‌کنند که ادعای محتوای غیرقانونی دارند و در نتیجه آن‌ها را در معرض ریسک حقوقی قرار می‌دهند، معمولاً جانب احتیاط را می‌گیرند و حتی اطلاعات قانونی را هم حذف می‌کنند. این «فراحذف‌کنندگی»^۴ محصول فرعی دائمی سامانه‌های تذکر و حذف است. سکوها اطلاعات متنوعی از سنخ مطبوعاتی و ویدئوهای مستند از خشونت‌های پلیس در اکوادور^۵ گرفته تا پوشش رسانه‌ای تحقیق و بررسی کلاهبرداری‌ها در ایالات متحده،^۶ نقد سازمان‌های مذهبی^۷ و گزارش‌های علمی^۸ را حذف کرده‌اند.

1. US Copyright Law

2. German Hate Speech Law

۳. این فصل به دنبال فهرست نمودن پژوهش‌ها درباره‌ی حذف محتوا در کشورهای محدودکننده‌تر گفتگو مانند چین نیست. رک: Human Rights Watch (2006).

4. Over-removal

5. Vivanco 2014

6. Cushing 2017

7. Galperin 2008

8. Oransky 2013; Timmer 2013

سکّوها در طول تاریخ خود از مشوّق‌های کمی برای برای در میان گذاشتن اطلاعات جزئی حذف محتوا با عموم بهره برده‌اند. گردآوری سوابقی که شامل فرایندهای حذف محتوا می‌شود، که می‌تواند از ابزارها و استانداردهایی مختلف استفاده برده باشد یا اینکه توسط گروه‌های درونی متفاوتی در طول زمان مدیریت شده باشد، کار دشواری است؛ و هر افشایی، به ویژه آن مواردی که بر خطای سکّو صحّه می‌گذارد، می‌تواند علیه سکّوها در دادگاه یا خبرگزاری‌ها استفاده شود. با این وجود منافع بلندمدت‌تر شفافیت بیشتر، هم برای جامعه و هم برای خود سکّوها، بیش از پیش موضوعیت یافته است. بدون این رویکرد، مباحثات عمومی درباره‌ی مسئولیت سکّوها می‌تواند شبیه بیافریند. قوانین صادره‌ای که ریشه‌ای عملیاتی در دنیای واقعی عملیات‌ها و قابلیت‌های سکّوها نداشته باشند، می‌توانند برای شرکت‌ها و کاربرانشان مشقّت‌بار باشند و هنوز در رسیدن به اهداف مشروع قانون‌گذاران ناکام باشند.

خواه متأثر از تشخیص این مسئله و خواه در نتیجه‌ی افزایش فشارها از سوی جامعه مدنی، دانشگاهی یا دیگر بخش‌ها، برخی از سکّوها شفافیت عمومی بیشتری در سال‌های اخیر تأمین نموده‌اند. این فصل منابع اصلی اطلاعاتی که سکّوها ارائه کرده‌اند، و همچنین پژوهش‌های مستقل در مورد عملیات‌های حذف محتوا را بررسی خواهد نمود. بخش «حذف و قوانین مسئولیت واسطه‌ای» را با بررسی خیلی مختصر قانون مسئولیت واسطه‌ای، که نقشی محوری در ساختاردهی به عملیات‌های حذف محتوای سکّوها دارد، آغاز خواهیم کرد. یک مطالعه‌ی قوی دانشگاهی به ترسیم شیوه عمل متداول

حذف در سگوها و موضوعات پژوهشی یاری خواهد رساند. بخش «منابع اطلاعات»، که عمده‌ی فصل را به خود اختصاص داده است، تحلیل عمیق‌تری درباره‌ی ادبیات تجربی جاری و احتمالاً منابع اطلاعات آینده به دست می‌دهد. نخست، درباره‌ی افشاگری‌های سگوها و دیگر همکارانشان، نظیر کاربران و دولت‌ها، در مورد تعدیل محتوا بحث می‌کنیم. دوم، درباره‌ی پژوهش مستقل از طرف اشخاص ثالثی همچون دانشگاهیان و مطبوعاتی‌ها، که شامل تحلیل داده‌ها، مصاحبه‌ها و بررسی‌هاست بحث می‌کنیم. در آخر، پیش از نتیجه‌گیری فصل، پرسش‌ها و حوزه‌هایی ویژه برای پژوهش‌های تجربی آینده فهرست خواهیم نمود.

مباحث پیرامون قوانین پیشنهادی جدید که تنوع آن از تنظیم‌گری محتوای تروریستی اتحادیه‌ی اروپا تا قانون «اخبار جعلی» سنگاپور را در برمی‌گیرد، بایستی توسط ارزیابی‌های تجربی صورت‌گرفته از ظرفیت‌های سگوها برای تبعیت و پیامدهای ناخواسته‌ی بالقوه‌ی تلاش‌های ایجابی آنها، مطلع / آگاه شوند. بدون اطلاعات بهتر درباره‌ی قوت‌ها و ضعف‌های واقعی سگوها به عنوان تنظیم‌گران گفتگو، نمی‌توان انتظار مشاهده‌ی قوانینی به خوبی طراحی‌شده را داشت.

بخش اول

حذف قوانین قابلیت مسئولیت واسطه‌ای



حذف قوانین قابلیت مسئولیت واسطه‌ای

این بخش درباره‌ی قوانین مسئولیت واسطه‌ای صحبت می‌کند، قوانینی که زمینه‌ی حقوقی مباحث تعدیل محتوا را شکل می‌دهند. همان‌طور که توضیح دادیم، این قوانین به این می‌پردازند که سگوها چه هنگام و چگونه به لحاظ قانونی نیاز به حذف محتوا دارند. به دنبال یک تحلیل قانونی مختصر، نشان می‌دهیم که چنین قوانینی چگونه در عمل، به کمک یک آموزه‌ی بسیار ریزبینانه اجرا می‌شوند: «تذکر و حذف در شیوه عمل روزمره»^۱.

قوانین مسئولیت واسطه‌ای

قوانین مسئولیت واسطه‌ای به واسطه‌های اینترنتی‌ای مانند ISPها،^۲ موتورهای جستجوگر یا شرکت‌های رسانه‌های اجتماعی می‌گوید که هرکدام چه مسئولیت‌های قانونی‌ای در قبال گفتگوی کاربران‌شان دارند. به عنوان قانون روشن و پذیرفته شده،^۳ قانون مسئولیت واسطه‌ای عموماً از دکترین‌های حقوقی اساسی زیربنایی‌ای که مواردی همچون هتک حرمت یا گفتگوی نفرت‌آمیز را تعریف می‌کنند، جدا هستند. ولی قوانین مسئولیت واسطه‌ای با تجویز اینکه سگوها چه

1. Urban, Karaganis, and Schofield 2016

۲. مخفف عبارت Internet Service Provider که به شرکت‌های ارائه‌دهنده‌ی خدمات اینترنت اطلاق می‌شود -م.

3. black-letter law

هنگام و چگونه بایستی اقدام کنند، به شدت بر این که چه گفتاری در واقع حذف می‌شود اثر می‌گذارند.

در مرتبه‌ی بالا، قوانین مسئولیت واسطه‌ای بایستی بین سه هدف غالباً در رقابت با هم، توازن برقرار کنند. جزئیات حقوقی واقع در قانون ملی، نوعاً بازتاب قضاوت قانون‌گذاران در این باره است که چگونه بهترین توازن را بین آن اهداف برقرار کنند. یک هدف، «جلوگیری از آسیب» است. عموماً هر چه یک قانون، سگوها را بهتر به حذف محتوای غیرقانونی یا دیگر محتویات آسیب‌زا تشویق کند، این هدف بهتر تأمین می‌شود. هدف رقیب دیگر، «محافظت از گفتگوها و اطلاعات برخط قانونی» است. قانونی که محتاج اعمال غلیظ و شدید خط‌مشی‌ها توسط سگوهاست، ممکن است در اجرای این هدف با گرفتاری مواجه شود و به منظور جلوگیری از ریسک حقوقی، سگوها را به سمت حذف گفتگوهای با ارزش و قانونی بکشاند. هدف سوم «شاعه نوآوری» است. قوانین مسئولیت واسطه‌ای ابتدایی به عنوان ابزاری برای حفاظت از صنایع نوپا تلقی می‌شدند. امروزه، قوانین مسئولیت واسطه‌ای می‌توانند عمیقاً روی رقابت بین سگوهای مستقر و شرکت‌های نوپا^۱ اثر بگذارند.

توازن اولویت‌ها بین این سه هدف، یکی از موضوعات ارزش‌های ملی و گزینش‌های خط‌مشی است؛ اما این پرسش که در عمل، کدام قواعد مشخص حقوقی در تحقق هر کدام از این اهداف سهم دارند، سؤالی تا حدی تجربی است که به عملیات واقعی سگوهایی که به مطالبات و مشوق‌های قانون پاسخ می‌دهند، گره می‌خورد.^۲ در فضای بین‌الملل، عمده‌ی قوانین مسئولیت واسطه‌ای، دو عنصر پایه‌ای

1. Start-Ups

۲. خیلی از قوانین برای دعاوی مختلف، قواعد مختلفی را تأمین می‌کنند (مانند حق مؤلف در برابر تروریسم) یا برای واسطه‌های مختلف نیز قواعد مختلف دارند (مثلاً هاست‌ها (Hosts) در برابر SPها). برای بازخوانی تنوع نگاه‌ها، نک: Keller, 2019a.

مشترک دارند. نخست، اینکه سگوها در مقابل دعای حقوقی‌ای ناشی از گفتگوهای غیر قانونی کاربرانشان مصونیت دارند، مادامی که در گسترش چنین گفتگوهایی هم دخالت بیش از حد نداشته باشند. قوانین ملی بین اینکه چگونه بایستی مصونیت سگوهای «خنثی» ارزیابی شود و میزان تعدیل محتوایی که این سگوها می‌توانند بدون قرار گرفتن در معرض مسئولیت انجام دهند، تفکیک می‌کنند. یک مورد نسبتاً تافته‌ی جدا بافته در این موضوع، «اقدام نزاکت ارتباطات» در ایالات متحده^۱ است که به سگوها مصونیت‌های وسیع غیرعادی‌ای اعطا می‌کند، حتی هنگامی که از محتوای غیر قانونی آگاه باشند، و این مصونیت را به قصد صریح تشویق سگوها به تعدیل و کنار گذاشتن محتوای «مورد نقد» اعطا می‌کند.^۲

دوم، اینکه بیشتر قوانین مسئولیت واسطه‌ای، سگوها را وقتی مُلزم می‌سازند که به محتوای غیر قانونی «دانش» داشته باشند. در اکثر کشورهای جهان، سگوهایی که از موضوعاتی نظیر هتک حرمت یا پروپاگانداى تروریست‌ها مطلع می‌شوند، موظف به حذف محتوای مربوطه یا مواجه شدن با تبعات قانونی می‌شوند. هر چند قوانین به طور اساسی درباره‌ی اینکه مصداق «دانستن» چیست، اختلاف دارند. ذیل برخی از قواعد ملی، سگوها تنها وقتی به لحاظ قانونی بایستی گفتگوی کاربران را قطع کنند که دادگاهی رأی به غیر قانونی بودن آن بدهد.^۳ در جاهای دیگر، قانون دست سگوها را در تصمیم‌گیری در این باره که کدام گفتگوها به حریم قانون تعدی می‌کنند، باز می‌گذارد.

با این چهارچوب، ریشه‌ی مهم تمایزات در رویه‌هایی یافت می‌شود

1. The US Communications Decency Act: CDA

2. 47 U.S.C. § 230.

3. Marco Civil da Internet, Federal Law no. 2014) 12.965) (Brazil); Copyright Act, Law No. 20.435 (Chile); Corte Suprema de Justicia de la Nacion, 2014/10/8, Rodriguez, Maria Belen c. Google Inc. / da os y perjuicios (Argentina).

که قانون برای حذف محتوای سگوها به کار می‌گیرد. اقدام حق نشر هزاره‌ی دیجیتال ایالات متحده (DMCA) یکی از قوانین مسئولیت واسطه‌ای است که بیشترین جزئیات رویه‌ای را داراست.^۱ این اقدام، پیش‌نیازهای رسمی برای تذکرات رسمی وارده از صاحبان حق نشر، مراحل «مقابله با تذکر»^۲ از سمت کاربرانِ متهم شده و دیگر جزئیات را که شامل جریمه برای تذکرها‌ی دارای سوء نیت علیه گفتگوی قانونی است، را به دقت بیان نموده است. اصول مانیلا^۳ که مجموعه‌ای از قواعد مسئولیت واسطه‌ای الگو است که توسط گروه‌های جامعه‌ی مدنی در سراسر دنیا تأیید شده و با ادبیات حقوق بشری پشتیبانی شده است، فهرستی از محافظت‌های رویه‌ای بیشتر را تهیه نموده است، که شامل لوازم شفافیت عمومی برای حذف خطاها، سوگیری یا سوءاستفاده در سامانه‌های تذکر و حذف است.^۴

بحث به سرعت در حال توسعه‌ی خط‌مشی مسئولیت واسطه‌ای، به این می‌پردازد که الزامات بالقوه‌ی سگوها برای پایش یا اعمال خط‌مشی بر گفتگوی کاربران چیست. تا دوران اخیر، عمده‌ی قوانین کشورها بر پایه‌ی این فرضیه بنا می‌شد که سگوها، اصلاً نباید گفتگوی کاربر را به طور پیوسته پایش کنند و بر آن مبنای، به صراحت برچسب غیرقانونی بودن را بزنند. قوانین مهمی مانند «حق نشر هزاره‌ی دیجیتال» آمریکا و دستورالعمل تجارت الکترونیک اروپا، صراحتاً الزامات پایشگری را نفی نموده‌اند؛^۵ الزاماتی که سگوها را فقط درباره‌ی محتوای غیر قانونی‌ای که از آن، معمولاً از طریق تذکر

1. 17 U.S.C. § 512.

2. Counter-notice

3. Manila

4. Manila Principles on Intermediary Liability, www.manilaprinciples.org;

دیوید کی، گزارشگر ویژه‌ی سازمان ملل در حوزه‌ی آزادی بیان، گزارش نموده است که اصول مانیلا «محافظت از خط اساسی واسطه‌ها با عنایت به استانداردهای آزادی بیان را پایه‌گذاری کرده است» (Kaye 2017).

5. 17 USC 512(m); Council Directive 2000/31/2000 O.J. (L 1 (178 (EC) ("eCommerce Directive"), Article 15.

اشخاص ثالث آگاه می‌شوند، مسئول می‌سازد. ابزارهای فیلترینگ به طور داوطلبانه توسعه‌یافته‌ی سکوها، تا کنون انتظارات سیاستگذاران را دستخوش تغییر نموده است، هرچند اجرای دقیق این ابزارها، به خوبی فهمیده نشده است. یک قانون اصلی جدید، یعنی دستورالعمل رسمی حق نشر اروپا، به طور مؤثری نیازمند فیلترینگ است.^۱ دیگر موارد پیشنهادی در حوزه‌هایی نظیر محتوای تروریستی در دست تصمیم‌گیری هستند.^۲ طرح انتقادات که از فناوریان تا سه گزارشگر سازمان ملل را در بر می‌گیرد، نگرانی‌هایی جدی درباره‌ی فیلترها را مطرح کرده‌اند.^۳ همان‌طور که در ۲۰۱۹ در نامه‌ای از طرف سازمان‌های جامعه‌ی مدنی، شامل حقوق دیجیتال اروپائیان،^۴ سازمان بند ۱۹ و اتحادیه‌ی آزادی‌های مدنی آمریکایی^۵ بیان نموده است، فیلترها همچنان «فناوری‌هایی ناآزموده و کمتر درک‌شده برای محدود ساختن بیانات برخط» هستند، که بسیار مستعدّ خاموش کردن بیان محافظت‌شده، از تقلیدهای هنری طنزآمیز گرفته تا گزارش‌های حقوق بشری است، که در نتیجه به «ارزش‌های مردم‌سالارانه و حقوق بشر فردی» لطمه می‌زند.^۶

مسئله‌ی پدیدآمده‌ی دیگر، وابستگی سکوها و دولت‌ها به خطوط راهنمای اجتماع، به جای اتکایشان به قانون، به عنوان مبنای حذف محتوای برخط است. قواعد اختیاری سکوها، غالباً مطالب مطابق قانون را هم ممنوع می‌کنند، و تا همین اخیر هم عموماً فرض بر

1. Directive 2019/790/2019 O.J. (L92 (130 (EC) ("Copyright Directive"), Article 17.

۲. پارلمان اروپا (European Parliament)، رأی تقنینی در ۱۷ آوریل ۲۰۱۹ در قالب طرح پیشنهادی برای تنظیم‌گری پارلمان و شورای اروپا در مورد جلوگیری از اشاعه‌ی محتوای تروریستی برخط (ویرایش موقت)، TA-PA-0421(2019)PROV .0421(2019)EN.pdf_0421-2019-8-www.europarl.europa.eu/doceo/document/TA

3. Cannataci et al. 2018; O'Brian and Malcolm 2018

4. European Digital Rights: EDRI

5. The American Civil Liberties Union: ACLU

۶. مرکز مردم‌سالاری و فناوری. جامعه‌ی مدنی در فوریه ۲۰۱۹ نامه‌ای به پارلمان اروپا درباره‌ی تنظیم‌گری‌های پیشنهادی برای جلوگیری از اشاعه‌ی محتوای تروریستی برخط نوشته است (ویرایش موقت). . <https://cdt.org/Civil-Society-Letter-to-European-Parliament-on-Terrorism-Database.pdf/02/2019/files>

این بود که سگّوها در این مسئله حدود اختیارات شدیداً گسترده‌ای دارند.^۱ مؤسسات ملّی و قوانین حقوق بشری کاربران اینترنت را از دخالت حکومت در اعمال قانونی مطابق حقوق گفتگوبشان محافظت می‌کند، اما سگّوها عمدتاً در ممنوع کردن هر گفتگویی که بخواهند، آزادند؛ و از آنجا که خطوط راهنمای اجتماع به طور خصوصی تعریف و تحمیل شده است، تصمیمات سگّوها عموماً در معرض بررسی دادگاه‌ها قرار نمی‌گیرند.

اما در سال‌های اخیر، دولت‌ها به ویژه در اروپا، به طور فزاینده‌ای متوجه خطوط راهنمای اجتماع سگّوها به عنوان سازوکار تحمیل شده هستند. به طور مثال، کد هدایت گفتگوی نفرت‌افکن کمیسیون اروپا و هم کد شیوه عمل در مورد اطلاعات کاذب آن کمیسیون، به سگّوها اعلام کرده‌اند که داوطلبانه محتواهای مشخصی را حذف کنند، که اغلب متکی بر خطوط راهنمای اجتماع هستند.^۲ بازوهای اعمال قانون شامل یوروپل،^۳ به خاطر نقشی که دارند، غالباً از خطوط راهنمای اجتماع یا شرایط استفاده از خدمات به جای قانون، به عنوان مبنای درخواست از سگّوها به منظور حذف محتوا استفاده می‌کنند.^۴ سازمان‌های آزادی مدنی به این چالش خُرده گرفته‌اند، و می‌گویند که آنها قانون‌گذاری مردم‌سالارانه و دادگاه‌ها را با سامانه‌هایی خصوصی شده و غیر پاسخگو تعویض نموده‌اند.^۵

۱. دادگاه‌های آمریکا، پیوسته حق سگّوها برای قطع گفتگوی قانونی کاربران را سلب می‌کنند. اما در عرصه‌ی بین‌المللی، تصویری متفاوت می‌تواند تجلی کند. در سال ۲۰۱۸، نخستین دادگاه‌ها به عنوان نمونه در آلمان و برزیل، دعای کاربران را در چنین وضعیتی سلب کردند. رک: Keller (2019b). نگرانی خبرنگاران در این باره که ممکن است حکومت‌ها، به الزامات حقوق بشر در خطوط راهنمای خصوصی اعضا به قصد جلوگیری از مطالب و اطلاعات بر خط تعذی کنند، رو به افزایش است. رک: Angelopolous et al. (2016). pp. 50-51; Kuczerauw (2017).

2. European Commission (2016) (agreement with Microsoft, YouTube, Twitter, and Facebook); در حالیکه کد گفتار نفرت‌افکن (2016 European Commission) درخواست حذف محتوایی که توسط کاربران منتشر شده است دارد، کد اطلاعات کاذب (2018b European Commission) به محتویات تبلیغاتی می‌پردازد و علیه جلوگیری از انتشار اطلاعات «کاذب» توسط کاربران عادی مشورت می‌دهد.

3. Europol

4. Europol 2016; Chang 2018

5. European Digital Rights 2016; Chang 2018. همچنین رک

یک مطالعه‌ی موردی: تذکر و حذف در عملکرد روزمره

کامل‌ترین تحلیل درباره‌ی عملیات تبعیت از مسئولیت واسطه‌ای، با اختلاف «تذکر و حذف در عملکرد روزمره» است.^۱ برای رسیدن به این نتیجه، پژوهشگران تذکرات حذفی تأثیر گذاشته روی حدود ۴ هزار URL^۲ صفحه وب در نتایج جستجوی وب و جستجوی تصویر گوگل، کاربران مصاحبه شدهٔ سگوها، صاحبان حق و دیگر اعضای زیست‌بوم «تذکر و حذف» را واریسی کردند. هرچند این مطالعه روی حق نشر متمرکز است (حوزه‌ای که در آن، غنی‌ترین مجموعه مطالب عمومی در دسترس است)، تحلیل دقیق آن روندها و مسائل مشابه برای حذف محتوا تحت قوانین دیگر را نیز مستند می‌کند. بسیاری از درس‌های آن، به ویژه آنهایی که به خودکارسازی^۳ و عملیات در مقیاس بزرگ مرتبطاند، به خوبی به حذف بر اساس خطوط راهنمای اجتماع هم ربط دارند. یکی از مهم‌ترین یافته‌های این مطالعه، واگرایی میان عملیات و ظرفیت‌های کلان سگوهایی مانند گوگل و دیگر واسطه‌های اینترنتی کوچک‌تر است.^۴ شرکت‌های کوچک‌تر و سنتی‌تر عموماً تیم‌هایی سه نفره یا با نفرات کمتر برای این امر به کار گرفته‌اند و به بررسی فردی مجزای هر تذکر می‌پردازند.^۵ بسیاری از افراد توصیف می‌کنند که «حذف محتوا را حتی آن هنگام که درباره‌ی قوت ادعای ضمنی یقین ندارند، می‌پذیرند» تا از قرار گرفتن در معرض مسئولیت قانونی اجتناب کنند.^۶ آنها همچنین «بازی تعمدی» تذکردهندگان با فرایند حذف را هم گزارش کرده‌اند، که «شامل آزار رقبا، برنده شدن در منازعات شخصی، ساکت کردن صدای منتقد یا تهدید [سگو]» می‌شود.^۷

1. Urban et al. 2016

۲. مخفف Uniform Resource Locator است که شامل یک نشانی اینترنتی است که محل یک پوشه‌ی اطلاعات یا منبعی در فضای اینترنت را مشخص می‌کند -م.

3. automation

5. Urban et al. 2016, pp. 36-29

4. Urban et al. 2016, pp. 74-73, 29-28

6. Urban et al. 2016, p. 41

7. Urban et al. 2016, p. 40

اما در مورد بازیگران بزرگ‌تر، قضیه بسیار متفاوت است. تفاوت از مقیاس آغاز می‌شود: در مقابل صدها و هزاران تذکری که شرکت‌های کوچک‌تر دریافت می‌کنند، گوگل بیش از ۱۰۸ میلیون تذکر حذف در مورد جستجوهای وب در بازه‌ی شش ماهه مطالعه دریافت کرده است.^۱ تذکردهندگان بزرگ‌مقیاس و همچنین سگوها آتکای زیادی به خودکارسازی کرده‌اند. به طور مثال صاحبان حق نشر و فروشندگان برون‌سپاری شده‌ی ایشان، فرایند تذکر را با استفاده از جستجوهای به منظور شناسایی فهرستی از URLها، خودکار ساخته‌اند.^۲ نتیجه‌ی «ربات‌های تذکر»^۳ بعضی اوقات همراه با خطاهایی است، مانند درخواست از طرف یک پیش‌نواز موسیقی برای حذف نسخه‌ی سینمایی «سقوط خانه‌ی پیش‌نواز» اثر ادگار آلن پو.^۴

سگوها گزارش دادند که به طور خودکار بسیاری از این درخواست‌های خودکار را، علی‌الخصوص از منابع «مورد اعتماد»، می‌پذیرند. در نتیجه، برخی هیچ بازبینی انسانی برای عمده تذکرات خودکاری که دریافت می‌کنند ندارند.^۵ برخی همچنین محتوا را به صورت پیش‌کنش‌گر،^۶ با استفاده از «اقداماتی چون سیستم‌های فیلترینگ مبتنی بر پیش‌بینی،^۷ سامانه‌های پایین نگه‌دارنده مبتنی بر تطبیق هاش^۸ و امتیازات حذف مستقیم طرف سرور^۹ برای صاحبان حق مورد اعتماد»^{۱۰} خط‌مشی‌گذاری می‌کنند.^{۱۱}

ارین^{۱۲} و همکارانش در تحلیل کمی خود، خطاهای قابل توجهی در

1. Urban et al. 2016, pp. 77, 29

2. Urban et al. 2016, p. 92

3. Robonotices

4. Edgar Allan Poe

5. Urban et al. 2016, pp. 91-90

6. Urban et al. 2016, p. 29

7. Proactive

8. Ex ante filtering systems

9. Hash-Matching based 'staydown' systems

10. Direct back-end takedown privileges

11. Urban et al. 2016, p. 29

12. Urban

عملیات‌های DMCA را مستند نموده‌اند. به طور مثال، آنها از میان تذکره‌هایی که به گوگل ارائه شده، ۲۸ درصد را ادعای قانونی سؤال‌برانگیز یافته‌اند.^۱ حدود ۴/۲ درصد از آنها - که طبق برون‌یابی، تا ۴/۵ میلیون درخواست طی یک دوره ۶ ماهه از مجموعه داده‌ها برآورد شده است - خطاهایی ساده به نظر می‌رسیدند، که شامل درخواست حذف موادی بوده است که به ادعای قانونی تذکردهندگان مرتبط نبوده است.^۲ از میان تذکرات ارائه‌شده، آنهایی که درباره‌ی نتیجه‌ی جستجوی تصاویر گوگل بوده‌اند، این عدد ۳۸ درصد بوده است.^۳ اگر محاسبات شامل رگبار یک فرد برای درخواست‌های حذف نامناسب شود، ۷۰ درصد از تذکرات جستجوی تصویر مورد تردید جدی قرار می‌گیرند.^۴

همچنین سگوهایی که در مطالعه‌ی اِربن مورد مصاحبه قرار گرفتند، میزان ضد تذکرات DMCA را که از سمت کاربرانی آمده که حذف‌های اشتباه را به چالش می‌کشیدند، پایین گزارش کردند. بسیاری از سگوها اصلاً هیچ ضد تذکری را دریافت نکرده‌اند.^۵ این یافته با نمودارهایی که توسط انجمن تصویر متحرک آمریکا^۶ در ۲۰۱۳ ترسیم شده، سازگار است، که نشان می‌دهند میزان ضد تذکرات برای درخواست حذف در DMCA توسط اعضای شرکت‌ها ۰.۰۰۰۳۲ درصد بوده است؛ و تنها ۸ ضد تذکر برای ۲۵،۲۳۵،۱۵۱

1. Urban et al. 2016, p. 88

2. Urban et al. 2016, p. 88

3. Urban et al. 2016, pp. 99-98

4. Urban et al. 2016, pp.

۹۸-۹۹ هرگونه ارزیابی از مطالعه‌ی میزان خطا، به ویژه آن هنگام که برای سیاستگذاری ضروری باشد، محل چالش است - زیرا نیازمند آنست که پژوهشگران هم، مانند سگوها، نظر خود را در حوزه‌های غیر شفاف قانونی اظهار کنند. برخی از انتقادات وارده به روش‌شناسی مطالعه‌ی اِربن، کرگینس (Karaganis) و اسکوفیلد (Schofield) (2016) درباره‌ی این حوزه‌ها و دیگر زمینه‌ها بوده است. برای نمونه، رک: Urban, Karaganis, and (2017) Ford. Schofield (2017) و دیگر خبرگان که -شامل نویسندگی یکی از فصول این کتاب و پروفیسور حقوق، آنبری بردی (Annemarie Bridy) هم می‌شود و- از استانداردهای مطالعه‌ی طبقه‌بندی تذکرات، دفاع کردند. رک: Urban et al. (2017) Bridy and Keller ; (2017) al.

5. Urban et al. 2016, p. 44

6. Motion Picture Association of America

URL تذکر داده شده شناسایی شد.^۱ به طور مشابه گزارش‌های شفافیت سکو، معمولاً میزان ضد تذکرات برای ادعاهای مربوط به حق نشر را کاملاً زیر ۱ درصد گزارش نموده‌اند.^۲ با توجه به مستندسازی گسترده حذف بیش از اندازه، این نمودارها اذعان می‌کنند که حذفیات نادرست مورد چالش قرار نمی‌گیرند.

1. Boyden 2013
2. Bridy and Keller 2016

بخش دوم

منابع اطلاعات



این بخش منابع متعدد اطلاعات تجربی را هم از جانب سگّوها و دیگر عوامل همکار در حذف محتوا و هم از جانب پژوهشگران مستقل، بررسی می‌نماید. برخی موارد، که شامل گزارش‌های شفافیت اصلی سگّوها می‌شود، در زمان‌بندی‌های منظم منتشر می‌شوند و می‌توانند منابعی مفید برای اطلاعات آینده باشند.

افشا توسط سگّوها و دیگر مشارکت‌کنندگان در حذف محتوا

سگّوها از چند راه، اطلاعات درباره‌ی تعدیل محتوا را ابراز می‌کنند. این راه‌ها شامل (۱) گزارش‌های دوره‌ای شفافیت؛ (۲) منابع اطلاعاتی اصلی به اشتراک گذاشته شده با آرشیوهای پژوهش دانشگاهی نظیر پایگاه داده‌ی لومین^۱ در هاروارد؛ (۳) اعلان به افراد متأثر شده درباره‌ی تصمیمات مربوط به حذف؛ و (۴) اظهارات عمومی اتفاقی و دیگر افشاگری‌ها درباره‌ی مسائل خاص محتوایی است. به طور روزافزونی، (۵) دولت‌ها نیز سگّوها را ملزم به اجرای بایگانی عمومی درباره‌ی تعدیل محتوا می‌کنند، و حتی برخی از دولت‌ها داده‌های مربوط به دخالتشان در رویه‌ی حذف را نشر می‌دهند. (۶) بازرسان شخص ثالث

نیز اطلاعات درباره‌ی فعالیت‌های حذف را آشکار می‌کنند، همان‌طور که (۷) اطلاعات درز یافته از سگوها چنین می‌کنند.

گزارش‌های شفافیت

بسیاری از سگوها گزارش‌های دوره‌ای شفافیت منتشر می‌کنند، گزارش‌هایی که معمولاً داده‌های گردآوری‌شده درباره‌ی درخواست‌های حذف محتوا را افشا می‌کنند. نمایه‌ای از گزارش‌های شفافیت که توسط سازمان جامعه‌ی مدنی اکسس ناو^۱ ایجاد شده، گزارش‌های از بیش از ۷۰ شرکت^۲ را که شامل گوگل،^۳ فیسبوک،^۴ توئیتر،^۵ آمازون،^۶ تومبلیر،^۷ مدیوم،^۸ ردیت،^۹ گیت‌هاب^{۱۰} و ورد پرس^{۱۱} می‌شود را فهرست می‌کند.^{۱۲} این می‌تواند امکان یک فرانگیری کمی مهم به تصویر کلان - یا دست‌کم بخشی از آن تصویر - را فراهم نماید. آن گزارش‌ها معمولاً داده‌های مربوط به درخواست‌های حذف را گردآوری می‌کنند، امری که وابسته به میزان تمکین سگوهاست. آنها همچنین می‌توانند میزان تناوبی که کاربران متهم به عملکرد نادرست درخواست تجدید نظر می‌دهند یا تصمیم سگورا به چالش می‌کشند، برملا کنند. گزارش‌های شفافیت به‌طور سنتی بر روی درخواست‌های قانونی حذف متمرکز هستند. البته در ۲۰۱۸، فیسبوک،^{۱۳} توئیتر^{۱۴} و یوتیوب،^{۱۵} نخستین گزارش‌های اعمال خطوط راهنمای اجتماعشان را منتشر کردند.

گزارش‌های شفافیت محدودیت‌هایی جدی دارد. داده‌های گردآوری‌شده در گزارش‌های شفافیت تنها ارزیابی‌های خود سگورا را نشان

1. Acces Now
2. Access Now (2016)
3. Google (n.d.).
4. Facebook (2018a).
5. Twitter (2017)
6. Amazon (2015)

7. Tumblr (n.d.).
8. Medium (2015).
9. Reddit, Inc. (2015).
10. GitHub (2015).
11. WordPress
12. Automattic (n.d.).

می‌دهند، و مزیت‌های موجود در موارد زیربنایی را نشان نمی‌دهند. این، به آن معناست که پژوهشگران نمی‌توانند صحت تصمیمات به حذف را ارزیابی کنند یا متوجه روندهایی از اعمال ناسازگار شوند. همچنین، بیشتر گزارش‌های شفافیت تنها دسته‌ای جزئی از حذفیات را پوشش می‌دهند - غالباً تنها آن دسته‌هایی که توسط دولت‌ها یا صاحبان حق نشر مطرح گردیده‌اند. این سبب می‌شود تا این گزارش‌ها سوالات درباره‌ی پاسخ سگوها به مدعیات حقوقی مطرح شده توسط افراد، مثلاً تحت قانون فرانسوی هتک حرمت یا قانون برزیلی حریم شخصی را بی‌پاسخ باقی گذارد.

گزارش‌های شفافیت همچنین در روش دسته‌بندی داده‌ها، تنوع گسترده‌ای می‌یابند که مقایسه‌ی نظیر به نظیر شرکت‌ها را دشوار می‌سازد. به طور نمونه، گزارش‌هایی که در پی آن‌اند که یک شرکت چه مقدار تذکر دریافت می‌کند، نمی‌تواند برای مقایسه با گزارش‌هایی که تعداد درخواست حذف آیتم‌های محتوا را رصد می‌کند به کار آیند، زیرا یک تذکر ممکن است همه‌ی موارد را فهرست نموده باشد.

گزارش‌های شفافیت در جزئیاتشان هم تنوع فراوانی دارند. به طور مثال، گزارش‌های خطوط راهنمای اجتماع را که ذکر شد، در نظر بگیرید. گزارش یوتیوب تعداد کانال‌ها و ویدیوهایی را که به دلیل یازده نوع مختلف از تخلفات استاندارد (از جمله هرزنامه، برهنگی، ترویج خشونت و افراط‌گرایی) حذف شده‌اند، مستند ساخته است.^۱ این گزارش همچنین مشخص می‌کند که این ویدیوها چگونه حذف شده‌اند، از طریق نشان‌گذاری خودکار، نشان‌گذاران فردی قابل

اطمینان، کاربران، سَمَن‌ها^۱ یا عوامل دولتی. گزارش فیسبوک حتی از این هم جزئی‌تر است؛ این گزارش ثبت می‌کند که چند وقت یکبار کاربران علیه تصمیمات حذفی درخواست تجدید نظر داده‌اند و چه میزان محتوا بعد از حذف، بازیابی شده‌اند (خواه توسط فیسبوک و خواه در نتیجه‌ی درخواست کاربران^۲). افزون بر این گزارش عددی، فیسبوک جزئیاتی درباره‌ی عملیات‌هایش را هم بازگو می‌کند. بیش از ۲۰ هزار نفر، تعداد کارکنان مورد انتظاری است که در حوزه‌ی تعدیل محتوا فعالیت می‌کنند، که شامل گویشوران مادرزاد بیش از ۵۰ زبان و گروه‌هایی است که شبانه‌روزی کار می‌کنند. جدا از آن، فیسبوک یک نسخه‌ی عمومی تفصیلی از استانداردهای جامعه‌اش^۳ و راهنمای فهم نمودارهای گزارش را هم منتشر نموده است.^۴ در سوی دیگر، گزارش توئیتر قرار دارد که به طور مشخص، جزئیات اندکی دارد. این گزارش تنها تعداد حساب‌های کاربری منفردی را که گزارش شده‌اند و مطابق شش دسته‌ی مختلف تخلفات با آنها برخورد شده است، مستند ساخته است، بی‌آنکه میزان پذیرش‌ها یا بازگردانی‌ها یا سازوکار گزارش‌دهی را مگر در مواردی که از نهادهای دولتی شناخته‌شده دریافت کرده، مشخص سازد.^۵

یک ارزیابی بیرونی مهم از قوت و ضعف گزارش‌های شفافیت شرکتها را می‌توان در گزارش دوره‌ای «چه کسی پشت تو در می‌آید؟»^۶ مربوط به «بنیاد آخرین یافته‌های الکترونیک»^۷ یافت^۸ مورد دیگر، در «نمایه‌ی مسئولیت‌پذیری شرکت سهامی رده‌بندی حقوق دیجیتال»^۹ یافتنی است، نمایه‌ای که شرکت‌های فناوری را با معیارهای متعدد مرتبط با شفافیت و محافظت از آزادی بیان کاربران و

۱. مخفف سازمان مردم‌نهاد (Non Governmental Organization: NGO) - م.

2. Facebook 2018b

3. Facebook 2018b

4. Facebook 2018

4. Twitter 2018

6. Who Has Your Back

7. Electronic Frontier Foundation

8. Gebhart 2018

9. Ranking Digital Rights Corporate Accountability Index

حقوق حریم شخصی رتبه‌بندی می‌کند.^۱ «جعبه ابزار گزارش شفافیت مؤسسه‌ی فناوری باز»^۲ نیز یک مقایسه‌ی ارزشمند میان گزارش‌های موجود و همچنین توصیه‌هایی برای بهبود عملکرد به دست می‌دهد.^۳ این جعبه ابزار، از اصول سانتا کلارا،^۴ که «اعداد، تذکر و تجدید نظر» را به عنوان عناصر حیاتی در عملیات‌های حذف محتوای سگوها اعلان می‌دارد،^۵ بهره می‌برد.

اطلاعات منبع اولیه اشتراک‌گذاری شده توسط سگوها

پژوهشگران برای نتیجه‌گیری درباره‌ی عملیات حذف محتوای سگوها، نیازمند دانستن این هستند که کدام محتواها در عمل حذف می‌شوند. تا امروز، بهترین منبع برای چنین اطلاعاتی، پایگاه داده‌ی لومِن است، آرشیوی که مرکز برکمن کلاین^۶ در هاروارد میزبان آن است.

لومِن تذکرات حذف قانونی در هر سگو - یا فرستنده - ای را که تصمیم به اشتراک‌گذاری گرفته است، آرشیو می‌کند. اطلاعات شخصی فرستنده، و بسته به مورد - همان‌گونه که در مورد محتوای کودک‌آزاری رخ می‌دهد - URL محلّ محتوا مخفی می‌ماند، اما در غیر این موارد، پژوهشگران می‌توانند همه‌ی مکالمه را بررسی کنند. در آخرین واری، پایگاه داده حدود ۹.۳ میلیون تذکر را ذخیره کرده، که تقریباً ۳.۳۵ میلیارد URL را هدف می‌گیرند.^۷ عمده‌ی آنها از گوگل می‌آیند؛ و در بقیه‌ی آن توییتر، وُردپرس، آرشیو اینترنت،^۸ کیک‌استارتر^۹،

1. The Ranking Digital Rights 2018 Corporate Accountability Index. <https://rankingdigitalrights.org/index2018/>

2. The Open Technology Institute's Transparency Reporting Toolkit

3. Open Technology Institute 2018

4. Santa Clara Principles

5. The Santa Clara Principles. <https://santaclaraprinciples.org/>

6. Berkman Klein

۷. مکاتبه‌ی ایمیلی با آدم هلند (Adam Holland)، مدیر پروژه‌ی لومِن.

9. Kickstarter

8. Internet Archive

ردیت و ویمئو^۱ سهم دارند. از آنجا که تذکرات، مطالبی را شناسایی می‌کنند که غالباً هنوز در URLهای فهرست شده در دسترس هستند، پژوهشگران می‌توانند محتوایی را که غیر قانونی انگاشته شده ببینند و ارزیابی کنند که آیا واقعاً سکو تصمیم درستی گرفته است؟ پایگاه داده‌ی لومِن امکان پژوهش‌های گسترده‌ی دانشگاهی را فراهم نموده است.^۲

تقاضا برای دیگر انواع «منابع اولیه»ی شفافیت رو به افزایش است. به طور مثال، یک پیشنهاد مهم اخیر از طرف دولت فرانسه درخواست شفافیت مکفی برای بازرسان به منظور بررسی موارد خاص تصمیم حذف محتوا دارد.^۳ همین‌طور اصلاحیه‌های پیشنهادی به قانون الزام شبکه‌ی آلمان^۴ که با جزئیات بیشتر در بخش «بایگانی‌های عمومی و دیگر افشاگری‌های دولت» مورد بحث قرار گرفته است - همچنین درخواست بازرسی فعالیت‌های تعدیل محتوا و ایجاد یک «مؤسسه پایاپای»^۵ عمومی برای داوری شکایت کاربران در مورد حذفیات غلط دارد.^۶ چنین سازوکارهای نظارتی مستقلی اگر به لحاظ اجرایی تقویت شوند و به لحاظ اقتصادی امکان‌پذیر شوند، چه بسا که موجبات پژوهش‌های جزئی‌تر شخص ثالثی درباره‌ی ذات تصمیمات تعدیل محتوا را فراهم نماید.

1. Vimeo

۲. این شامل پژوهش اِربِن، کِرگنِس و اسکوفیلد (۲۰۱۷) و تعدادی از دیگر پژوهش‌های دانشگاهی است که شامل آن مواردی است که در بخش III.B در ادامه آمده است. همچنین بنگرید به خلاصه‌ی *Amici Curiae Chilling Inc. v. Google, 10 Perfect, 16-8* Effects Clearinghouse Leaders in Support of Appellee at www.eff.org/document/amicus-5813411 WL 2010, (2011. 9th Cir) 976 3d.F 653 „Inc brief-chilling-effects (که به دیگر کارها ارجاع می‌دهد). شفافیت معادل برای حذف محتوای میزبانی‌شده، هرچند بسیار با ارزش است، دشوارتر است. این، نیازمند آنست که میزبان یا لومِن محتواهای قانونی انگاشته‌شده را هنگام دریافت یک درخواست حذف، فعالانه نگاه‌دارد. البته در بسیاری از موارد، همان‌طور که در پست‌های فیسبوک به طور خصوصی به اشتراک گذاشته شده است، افشای محتوا همچنین می‌تواند با الزامات حفظ حریم خصوصی در تعارض باشد.*

۳. *French Secretary of State for Digital Affairs (2019) French Secretary of State for Digital Affairs*, p. 20. که فراخوان یک «سازوکار مستقل و فوق قضائی برای بررسی تصمیم سکو» را داده است.

4. *Netzwerkdurchsetzungsgesetz: NetzDG*

5. *Clearing House*

6. *German Parliament 2018*

اعلان به افراد متأثر شده

سکوها همچنین اطلاعات بالقوه مفیدی برای کاربران را فراهم می‌کنند که تحت تأثیر درخواست‌های حذف قرار گرفته‌اند. به طور خاص، آنها ممکن است (۱) به فردی که درخواست حذف را داده است، پاسخ دهند که بفهمد آیا درخواست او شایسته‌ی پذیرش است؛ (۲) کاربری که محتوایش حذف شده است را مطلع کنند؛ یا (۳) «سنگ قبر»^۱ی برای محتوای گمشده قرار دهند، و به کاربرانی که به دنبال یک صفحه وب یا اطلاعاتی از دست رفته هستند، فقدان آن را گوشزد کنند. عبارت «این ویدئو در دسترس نیست» یوتیوب شاید آشنا ترین مثال دیداری برای بسیاری از کاربران اینترنت باشد. اطلاعاتی که از موارد تذکرات مربوط به افراد به دست می‌آید، اغلب چرخه‌ای از اخبار را درباره‌ی تصمیمات به طور خاص بحث‌برانگیز به راه می‌اندازند. این اطلاعات، نمی‌تواند تصویری کلان به دست پژوهشگران دهد، اما می‌تواند با قراردادن اطلاعات در دست مردمی که به احتمال زیاد برایشان مهم است و اقدامی انجام می‌دهند، نقش مهمی در بازکردن خطاها ایفا کند.

افشاجری‌های سکوها در مسائل خاص

سکوها در مسائل تعدیل محتویات مربوط به موضوعات حساس برای جامعه، به گونه‌ای فزاینده اطلاعاتی‌های عمومی دقیق و جامع را برای تبیین سیاست‌هایشان مطرح می‌کنند. برخی از سکوها اطلاعات جزئی از ارزیابی موارد فردی توسط سکوها را عرضه می‌کنند، چیزی که گزارش‌های شفافیت جمع‌شده معمولاً فاقد آن هستند.

یک مثال مهم، حمله‌ی تروریستی به کلیسای کریست^۱ در ۱۵ مارس ۲۰۱۹ است. تمام برنامه‌ی حمله از طریق فیسبوک پخش زنده شد و در بسیاری از وب-سایت‌های دیگر نیز به سرعت انتشار یافت. فیسبوک فوراً دو اعلامیه‌ی عمومی منتشر کرد که تلاش‌هایش برای حذف این پخش تصویری را توضیح می‌داد. آنها جزئیات لحظه به لحظه‌ی رویدادها را که با نخستین پخش زنده شروع شده بود، پوشش دادند، همچنین داده‌ها درباره‌ی این که این ویدئو در ادامه چه قدر دیده شده، به اشتراک‌گذاری شده، بارگذاری مجدد شده و در نهایت حذف گردیده را فراهم کردند.^۲ در واکنش به حادثه‌ی کلیسای کریست، فیسبوک حتی در یک اقدام بی‌سابقه، کیت کلونیک^۳ را که یک حقوق‌دان دانشگاهی است، به همراهی با تیم واکنش به این رخداد دعوت کرد. کلونیک بعداً مشاهداتش را منتشر کرد.^۴

فیسبوک همچنین اعلامیه‌های مختلفی در دفاع از تلاش‌های خود در حذف «رفتارهای هماهنگ‌شده‌ی غیرمعتبر»^۵ به منظور محافظت از انتخابات کشورهایمانند اکراین،^۶ اسرائیل،^۷ هند و پاکستان^۸ و هدف‌گیری حساب‌های کاربری جعلی ایجاد شده از روسیه^۹ و ایران^{۱۰} صادر کرده است.^{۱۱} این اعلامیه‌های پست‌شده شیوه‌شناسایی رفتارهای غیرمعتبر توسط سکو، الگوهایی کشف‌شده و تصمیمات اتخاذشده شامل مثال‌ها و همچنین داده‌های تجمیع‌شده را توضیح می‌دهد.

1. Christchurch

۲. «نخستین گزارش از ویدئوی اصلی، ۲۹ دقیقه بعد از شروع ویدئو، و ۱۲ دقیقه بعد از اینکه پخش زنده تمام شد دریافت شد. در ۲۴ ساعت نخست، ما بیش از ۱/۲ میلیون ویدئوی در حال بارگذاری از حمله را حذف کردیم... تقریباً ۳۰۰ هزار کمی افزودنی پس از آن که پست شدند، حذف شدند.» (Facebook ۲۰۱۹a)

3. Kate Klönick

4. Klönick 2019

5. Coordinated inauthentic behavior

6. Facebook (2019c).

7. Facebook (2019d).

8. Facebook (2019e).

9. Facebook (2019f).

10. Facebook (2019g).

۱۱. روشن است است که نمی‌توان منکر سیاسی‌کاری‌های صورت گرفته حتی در این قبیل نگاه‌هاست. انتخاب این قبیل متون جهت ترجمه با توجه به اصل محتوای مورد بحث در این مقالات بوده است و در ترجمه مقالات رعایت اصل امانت صورت گرفته است و گرنه با توجه به روشن شدن حمایت‌های شبکه‌های اجتماعی مانند توئیتر و فیسبوک از سیاست‌های دولت آمریکا- مانند آنچه در ماجرای ترور شهید سلیمانی و حذف محتوای اعتراضی در این خصوص رخ داد- بدیهی است که گزارش‌هایی که این شبکه‌ها بر علیه کشورمان و یا سایر رقبای بین‌المللی خود منتشر می‌کنند چندان قابل اعتنا نیست-م.

کلود فلیر^۱ که یک شرکت زیرساخت وب است، یک پُستِ بلاگ تأثیرگذار درباره‌ی تعدیل محتوای مربوط بعد از شورش‌های «شارلوتس‌ویل»^۲ در ۱۱ آگوست ۲۰۱۸ منتشر کرد. این شرکت توضیح داد که چرا ارائه‌ی خدمات به «دیلی استورمر»^۳، یک وبسایت سفیدبرترانگار^۴ را متوقف کرده است. نویسنده که مدیر عامل شرکت، متیو پرینس^۵ بود به روشنی منتقد خود بود و «ریسک‌های ورود شرکتی همچون کلود فلیر به خط‌مشی‌گذاری در محتوا» را مورد تأکید قرار داد.^۶ شاید فراگیرترین مثال از گزارش‌دهی مسائل مشخص، گزارش گوگل از «سه سال از حق فراموش‌شدن»^۷ باشد.^۸ این سند از نظر میزان جزئیاتی که درباره‌ی فرایندهای درونی شرکت در ارزیابی درخواست‌های فردی حذف به دست می‌دهد، منحصر به فرد است. این گزارش نمونه‌هایی بدون نام از مواردی فردی، همچون درخواست یک نفر برای حذف نتایج جستجوی «مصاحبه‌ای هدایت‌شده پس از جان به در بردن از یک حمله‌ی تروریستی» یا درخواستی مشابه برای «یک مقاله‌ی خبری درباره‌ی صدور حکم براءت برای خشونت‌های خانگی در مواردی که هیچ گزارش پزشکی‌ای برای تأیید جراحات قربانی توسط قاضی در آنها ارائه نشده» را آورده است.^۹ این گزارش شاخصه‌ها و طبقه‌بندی‌های گوناگون ویژه‌ای را که گوگل برای پاسخ‌دهی به درخواست‌ها استفاده می‌کند، فهرست نموده است. به طور مثال یک شاخصه، اصالت «هویت درخواست‌کننده» است. گوگل هویت درخواست‌کننده را به تناسب هر محتوای برخط مورد بحث، در شش دسته در جدول ۱۰.۱ طبقه‌بندی نموده است.^{۱۰} گوگل

1. Cloudflare

2. Charlottesville

3. Daily Stormer

4. White supremacist

5. Matthew Prince

6. Cloudflare 2017

7. Three Years of the Right to Be Forgotten.

8. Google 2018b

9. Google 2018b, p.

۱۰؛ گوگل هر دو را حذف کرد

10. Google 2018b, p. 5

حق فراموش‌شدن این حق است که اطلاعات شخصی فرد از جستجوهای اینترنتی و دیگر مکان‌های آنلاین تحت شرایط خاص حذف شود. م.

بر پایه‌ی این تقسیم‌بندی ریز، ارقام کلان و تحلیل‌های آماری تولید می‌کند.

جدول ۱۰۱ خرابی همه‌ی URL‌های درخواست‌شده پس از ژانویه ۲۰۱۶ توسط دسته‌های مختلف هویت در خواست‌کننده

نکته: افراد خصوصی بیشترین حجم درخواست‌ها را داشته‌اند.

میزان حذف از فهرست	خرابی	URL‌های درخواست‌شده	هویت درخواست‌کننده
۴۴٫۷٪	۸۴٫۵٪	۸۵۸،۸۵۲	افراد خصوصی
۷۸٪	۵٫۴٪	۵۵،۱۴۰	فوقی
۳۵٫۵٪	۴٫۱٪	۴۱،۲۱۳	اشخاص عمومی غیر دولتی
۱۱٫۷٪	۳٫۳٪	۳۳،۹۳۷	مقام رسمی حکومتی یا سیاستمدار
۰٫۰٪	۲٫۳٪	۲۲،۷۳۹	نهاد شرکتی ^۱
۲۷٫۲٪	۰٫۴٪	۴،۴۰۲	شخص متوقفا

این گزارش از این بابت نیز ارزشمند است که به گونه‌ای محسوس، ترسیم می‌کند که چگونه یک سکو بایستی دعاوی پیچیده را به عناصر استانداردشده برای پردازش سریع در مقیاس بالا بشکند و تفکیک کند. پژوهشگران مستقل می‌توانند این عناصر را بازبینی کنند تا مثلاً به ارزیابی این امر پردازند که این عناصر چه قدر به عنوان جایگزینی برای بررسی قضایی حقوق آزادی بیان و حریم خصوصی متضاد طرفین، کفایت دارند. آنها همچنین می‌توانند شاخصه‌ها و عناصر گزارش‌شده را به عنوان یک نقطه‌ی شروع محسوس و قابل بحث در گفتگو بر سر اینکه کدام اطلاعات را

کارمندان سگوها بایستی در مورد هر تصمیم حذفی، به طور معقول رصد کنند و گزارش دهند، مورد استفاده قرار دهند.

بایگانی‌های عمومی و دیگر افشاگری‌های دولت

گاهی اطلاعات ارزشمندی درباره‌ی عملیات سگوها از طریق دادگاه‌ها یا دیگر بایگانی‌های عمومی به سطح عمومی تسری می‌یابد.^۱ اسنادی که مثلاً در مورد درگیری حقوقی وایکام^۲ با یوتیوب عمومی شد، سرخط‌های خبری را به دلیل افشاگری‌ها در مورد طرفین دعوا به خود اختصاص داد.^۳ اطلاعات افشا شده در پی مشاوره‌های دولت و ارکان فراملی در مطبوعات منتشر شده است که شامل گزارش کارکنان کمیسیون اروپا در ۲۰۱۲^۴ و گزارش‌هایی از دفتر گزارشگر آزادی بیان سازمان ملل^۵ بود.

پس از آن، سگوهای بزرگی نظیر فیسبوک، یوتیوب و توئیتر، گزارش‌هایی مهم به عنوان بخشی از هم‌نوایی‌شان با قانون NetzDG آلمان منتشر کردند. این قانون به خاطر قوت کم‌نظیر قواعد حذف محتوا بهترین قانون شناخته‌شده است، اما اقتضائات گزارش‌دهی عمومی بی‌سابقه‌ای را تحمیل می‌کند. گزارش‌های دوسالانه‌ی سگوها شامل اطلاعاتی نظیر ارقام مربوط به کارکنان، منابع سلامتی در دسترس کارکنان، فرایندهای عملیاتی، تعداد مشاوره‌های صورت‌گرفته با مشاوران حقوقی بیرونی و مدت زمان صرف‌شده برای پاسخ‌دهی به تذکرات است که طبق تخطی‌های حقوقی مشخص ادعا شده، و

۱. به عنوان نمونه، رک:

European Commission (2018a); US Copyright Office (2015); US Patent and Trademark Office (2015); Torrent Freak (2018) (citing testimony of Google legal director disclosing use of hash matching on Google Drive).

2. Viacom

3. Anderson 2010; YouTube 2010

4. European Commission 2012

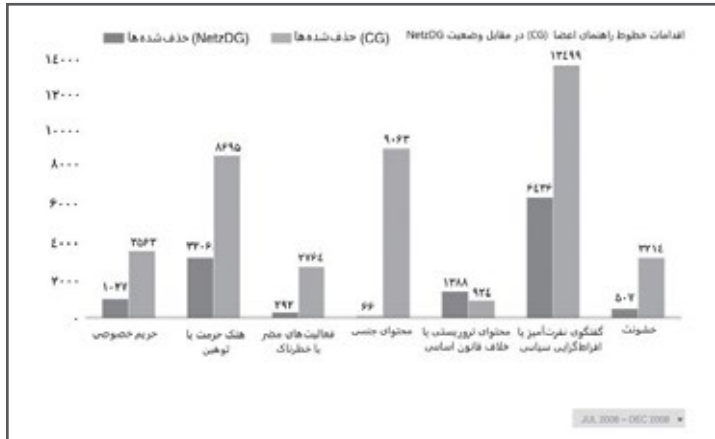
5. Kaye 2017

تفکیک شده است. در حالی که پژوهشگران هیچ ابزار مستقلی برای دسترسی به صحت آمار حقوقی سکوها ندارند، این گزارش‌ها سرشار از دیگر آمارها و جزئیات اجرایی هستند.

مثلاً در نیمه‌ی دوم ۲۰۱۸، یوتیوب تذکرات NetzDG را که ۲۵۰ هزار آیتیم در آن مشخص شده بود، دریافت کرد. بیشترین مقدار گزارش‌ها از دسته‌ی گفتار نفرت‌افکن یا افراط‌گرایی سیاسی بود (بیش از ۸۳ هزار شکایت) و پس از آن هتک حرمت یا توهین (بیش از ۵۱ هزار شکایت) و محتوای جنسی (بیش از ۳۶ هزار مورد) بود. یوتیوب در پاسخ، ۵۴۶۴۴ آیتیم را با میزان‌های مختلفی از حذف محتوا در هر دسته، حذف نمود.^۱ این سند همچنین نشان می‌دهد که گزارش‌ها توسط کاربران ارسال شده‌است یا عوامل دولتی آلمان.

همان‌گونه که نمودار ۱۰.۱ از گزارش نشان می‌دهد، یوتیوب همچنین اتکای سنگینی به خطوط راهنمای اجتماع داشته است. با این وجود، یوتیوب برای رفع وضعیت حقوقی بیش از ۱۰،۰۰۰ آیتیم به این قانون آلمانی نگاه کرده است.

نمودار ۱۰۰۱ اعمال خطوط راهنمای اجتماع در مقابل اساس نامه‌های NetZDG^۱



اما گزارش‌های NetZDG درباره‌ی فیسبوک، تصویر به شدت متفاوتی دارد. در دوره‌ی زمانی مشابه با یوتیوب، فیسبوک تنها ۵۰۰ شکایت NetZDG را از طریق گزارش دریافت کرد که شامل ۱۰,۰۴۸ آیتم محتوا می‌شد - که تنها کسری از آن چیزی است که از یوتیوب و دیگر سکوهای اصلی دریافت کردند.^۲ این احتمالاً بخاطر اینست که فرم شکایت NetZDG در وبسایت فیسبوک، در مقایسه با یوتیوب و توئیتر به خوبی در دید نیست. اکنون حذفیات فیسبوک بر مبنای خطوط راهنمای اجتماع، در دوره‌ی زمانی مشابه به رقم میلیونی رسیده است، و این آمار در گزارش NetZDG گنجانده نشده است. فقط در ۲ جولای ۲۰۱۹، اداره‌ی عدالت فدرال آلمان، فیسبوک را به خاطر گزارش‌دهی ناقص، ۲ میلیون یورو جریمه کرد، با این استدلال که چون گزارش فیسبوک شفاف مشخص نمی‌کند که کدام شکایات

که بر پایه‌ی خطوط راهنمای اجتماع هستند، در واقع موارد غیر قانونی هستند: «رقم شکایات دریافتی درباره‌ی محتوای غیر قانونی ناقص است» و بنابراین، گزارش‌های فیسبوک یک «تصویر تحریف شده» ایجاد می‌کنند.^۱ این مسئله، پرسش‌هایی را در این باره طرح می‌کند که آیا NetzDG سگوها را ملزم به ارزیابی و گزارش‌دهی روی قانونی بودن تصمیمات مربوط به محتوایی می‌کند که به آن‌ها در چهارچوب NetzDG اشاره‌ای نشده‌است.

همچنین کمیسیون اروپا توییت‌ر، گوگل و فیسبوک را متقاعد نموده است تا ماهانه گزارش‌های مربوط به شکایات را در آغاز انتخابات می ۲۰۱۹ اتحادیه‌ی اروپا، به عنوان بخشی از کد شیوه عمل در مورد اطلاعات کاذب منتشر کنند.^۲ این گزارش‌ها طیفی از فعالیت‌های مرتبط با اطلاعات کاذب را تعریف می‌کند، که شامل سواد رسانه و تلاش‌های صحت‌سنجی^۳ و همچنین صورت‌های معینی از تعدیل محتوا است. مهم این‌ها که این‌ها از معدود گزارش‌هایی هستند که اعمال استانداردهای تبلیغاتی، شامل اطلاعات فرایندی از سازوکارهای تأیید، بازبینی و شفافیت تبلیغاتی سیاسی و همچنین داده‌های کمی روی فعالیت تبلیغاتی و تصمیمات حذف را به بحث می‌گذارند. به عنوان نمونه، در می ۲۰۱۹، گوگل ۱۶,۶۹۰ حساب کاربری مستقر در اروپا را به عنوان متخلف از خط‌مشی‌های ارائه‌ی اطلاعات نادرست^۴ شناسایی کرد، و توییت‌ر ۱,۴۱۸ تبلیغ خلاف خط‌مشی فعالیت کسب‌وکار غیرقابل قبول^۵ خود را حذف کرد (که برای مثال محتوای گمراه‌کننده را ممنوع می‌کند).^۶ دست آخر، کمیته‌ی اطلاعات مجلس نمایندگان ایالات متحده نیز مجموعه داده‌هایی مهم را که از فیسبوک

1. Bundesamt für Justiz 2019
2. European Commission 2018b
3. Fact-checking

4. Misrepresentation policies
5. Unacceptable Business Practice Policy
6. Google 2019; Twitter 2019

به دست آمده و درباره‌ی تبلیغات سیاسی روس‌ها حین انتخابات ۲۰۱۶ ریاست جمهوری ایالات متّحده بود، منتشر نمود.^۱

سکّوها همچنین اطلاعات حذفشان را به کمیسیون اروپا به منظور گنجانده شدن در گزارش کمیسیون روی کد رفتار گفتار نفرت افکن^۲ ارائه می‌کنند.^۳ تحت این کد رفتار، سازمان‌های متخصص متذکر نقش سکّوها درباره‌ی محتویاتی که سازمانشان به عنوان گفتگوی نفرت افکن غیر قانونی می‌داند، می‌شوند. در نخستین بررسی کمیسیون، سکّوها از ۲۸ درصد این تذکرات تبعیت کردند و در دومین بررسی این عدد به ۵۹ درصد رسید - که کمیسیون آن را «پیشرفتی مهم» نامید. اما هنوز این ارقام تنها بیانگر میزان توافق تذکره‌دهندگان و سکّوها درباره‌ی اینکه چه محتوایی می‌بایست حذف شود هستند، و اگرچه کد رفتار به گفتگوی نفرت افکن «غیر قانونی» اشاره می‌کند، سکّوها (و چه بسا تذکره‌دهندگان) تذکرات را ذیل خطوط راهنمای اجتماع خودشان ارزیابی می‌کنند. بدون دسترسی مستقل به تذکرات و محتویات متأثر شده، ما نمی‌توانیم بدانیم هر یک از طرفین چه استانداردی را اعمال می‌کند، این استانداردها با چه دقت و سطح سازگاری‌ای اعمال می‌شوند، یا اینکه چگونه به قوانین کشورها ربط داده می‌شوند.^۴

برخی از افشاگری‌های مرتبط اما کم‌اهمیت‌تر توسط خود دولت‌ها، گاهی به همراه سکّوها، ارائه می‌شوند. به طور مثال، یک گزارش از یوروپل در ۲۰۱۶ درباره‌ی محتوای تروریستی‌ای که توسط واحد مراجعات اینترنتی^۵ آن به سکّوها برای حذف ارجاع داده شده، بحث می‌کند.^۶

1. US House of Representatives 2019

2. Hate Speech Code of Conduct

3. European Commission 2017, 2016

۴. این کد «گفتار نفرت افکن غیر قانونی» را پوشش می‌دهد، و کمیسیون می‌گوید گزارش‌های آن، میزان تبعیت سکّوها از «هشدارهایی در مورد گفتگوی نفرت افکن غیر قانونی» را نشان می‌دهد. در عمل، سکّوها حتی هشدارهای محتوایی را محترم می‌شمارند که از قانون تخطی نمی‌کند ولی از خطوط راهنمای اجتماع سکّو تخطی می‌کند - و کمیسیون اروپا این موارد را هم «موقفیت‌آمیز» در نظر می‌گیرد. افزایش از تبعیت ۲۸ درصدی به ۵۹ درصدی می‌تواند به این معنا باشد که هشداردهندگان در پیش‌بینی قواعد داخلی سکّوها بهتر شده‌اند.

5. Internet Referral Unit (IRU)

6. Europol 2016

از آن زمان، واحد مراجعات اینترنتی ۹،۷۸۷ آیتم را به ۷۰ سگوی مختلف ارجاع داده است. میزان موفقیت این واحد بالای ۹۰ درصد است. هر چند از آنجا واحدهای مراجعات اینترنتی به دنبال حذف محتوا تحت خطوط راهنمای اجتماع سکوها هستند، این ارقام صرفاً بازتاب موفقیت در پیش‌بینی کاربری‌های این قواعد توسط سکوها - یا متقاعد کردن سکوها برای پذیرش تفسیر آژانس‌های اعمال قوانین است. پژوهشگران مستقل هیچ ابزاری برای فهمیدن اینکه چه مقدار از محتواهای ارجاعی یا حذف‌شده، ناقض هر کدام از قوانین هستند، ندارند. برخی از دولت‌های اروپایی نیز یک واحد مراجعات اینترنتی برای خودشان در سطح ملی ایجاد کرده‌اند. این اقدامات آنها به دلیل فقدان شفافیت مورد نقد قرار گرفته، اما گاهی افشاگری‌هایی هم درباره‌ی آنها رخ داده است. واحد مراجعات اطلاعات ضد تروریستی انگلستان^۱ در دسامبر ۲۰۱۶ داده‌هایی را در وبسایت خود منتشر نمود با این مدعا که این واحد در حذف بیش از ۲،۰۰۰ قطعه محتوا در هفته نقش داشته است و در مجموع، در جریان حذف بیش از ۲۵۰ هزار قطعه محتوا تا پایان سال بوده است.^۲ ضمناً این واحد اطلاعات جلسه‌ی استماع پارلمان^۳ و پاسخگویی به درخواست‌های آزادی اطلاعات را که از طرف استاد حقوق دانشگاه کالج دوبلین^۴ تی. جی. مکینتایر^۵ مطرح شده بود نیز منتشر کرده است. گزارش‌دهی بیشتر دولت‌ها درباره‌ی دخالتشان در تعدیل محتوا، ممکن است یک روز توسط قانون الزامی شود. مثلاً پیش‌نویس اتحادیه‌ی اروپا درباره‌ی تنظیم‌گری محتوای تروریستی شامل استلزامات دقیق برای شفافیت حول ارجاعات اعمال قانون به

1. The UK's Counter-Terrorist Information Referral Unit (CTIRU)
 2. UK Metropolitan Police 2016
 3. Hansard. 2017. HL Deb 787 Col. 1261. <http://bit.ly/2kctmPL>
 4. University College Dublin
 5. McIntyre (2018)

سگوهاست.^۱

بازرسی‌ها

گزارش‌های منتشرشده از سوی بازرسان مستقل، نمایانگر نوعی کوچک ولی احتمالاً روبه‌رشد از افشاگری هستند. سازمان ابتکار شبکه‌ی جهانی^۲ گزارش‌هایی از حریم خصوصی و عملیات حذف محتوا منتشر کرده است که به عملکرد شرکت‌هایی نظیر فیسبوک، گوگل، لینکدین^۳، مایکروسافت و یاهو در سال ۲۰۱۳ اشاره دارد. این گزارش‌ها که از اطلاعات درونی اما عادی به اشتراک گذاشته شده توسط شرکت‌ها استفاده می‌کند، شامل ارزیابی‌های عمومی و موردکاوی‌ها می‌شود.^۴ دیگر بازرسی‌های غیر دوره‌ای و موردی به یک واکنش مرسوم رسوایی‌های صنایع فناوری تبدیل شده‌اند و ممکن است اطلاعات مرتبطی را در نتیجه‌ی ادامه‌ی بازرسی ایجاد کنند.^۵ یکی از گزارش‌های «بازرسی حقوق بشری از بنیاد پاسبانی اینترنت»^۶ بینش‌هایی درباره‌ی عوامل مهم غیر سگویی دخیل در حذف محتوا ارائه داده است.^۷ بنیاد پاسبانی اینترنت انگلستان^۸ سازمانی خصوصی و

1. European Parliament. Legislative resolution of 17 April 2019 on the proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online (provisional edition), P8_TA-PROV(0421(2019. www.europarl.europa.eu/doceo/document/TA2019-8-

2. (The Global Network Initiative)

یک سازمان غیر دولتی است که به دنبال مخالفت با سانسور اینترنت توسط دولت‌های اقتدارگرا با انگیزه‌های مردم‌سالارانه و حمایت از حقوق حریم خصوصی افراد است -م.

3. LinkedIn

4. Global Network Initiative, Company Assessments. <https://globalnetworkinitiative.org/company-assessments>

۵. به عنوان نمونه، فیسبوک اخیراً به اتهامات درباره‌ی سوگیری در مواجهه با اقدامات جدید بازرسی پاسخ داده است، که این پاسخ عبارتست از: «یکی از مشاوران یک بازرسی از اثر فیسبوک روی اقلیت‌ها و رنگین‌پوستان انجام می‌دهد، در حالی که مشاور دیگر به شرکت در مورد سوگیری بالقوه علیه چشم‌اندازهای محافظه‌کارانه مشورت می‌دهد» (Ong ۲۰۱۸).

6. Human Rights Audit of the Internet Watch Foundation ("IWF Audit")

7. MacDonald

۲۰۱۴، یک منبع ارزشمند دیگر که سازوکارهای حذف محتوا توسط بنیاد پاسبانی اینترنت و دیگر نهادهای ذریبط انگلستان را مستند ساخته است، اثر مکینتایر (McIntyre ۲۰۱۸) است.

8. The UK's Internet Watch Foundation (IWF)

غیر انتفاعی است که با پلیس، شرکت‌ها و عموم مردم برای تشخیص موارد برخط آزار جنسی کودکان همکاری می‌کند. این بنیاد سپس URLهایی را که بایستی مسدود شود برای نهادهای واسط فهرست می‌کند. سند بازرسی از این بنیاد که به درخواست یک متخصص حقوق بشر خارج از آن آماده شد، عملیات‌های داخلی این گروه را به طور جزئی بررسی می‌کند و فرایندهایی بهبودیافته را برای مثلاً تجدیدنظرها و درخواست‌های قضاوت حقوقی دشوار پیشنهاد می‌کند. این گزارش در میان پژوهش‌های تجربی، به دلیل تمرکزش بر اثرگذاری دوسویه‌ی کنش حکومت‌ها و بخش خصوصی، منحصر به فرد - و برای توسعه به اموری فراتر از حفاظت از کودکان مهم - است. مثلاً این گزارش درباره‌ی نقشی که بنیاد پاسبانی اینترنت به عنوان یک سازمان خصوصی در سرعت بخشیدن به درخواست‌های حذف مطرح‌شده توسط پلیس دارد بحث می‌کند - درخواست‌هایی که در حالت عادی و بدون فعالیت این بنیاد، نیازمند فرایندهای قضایی اضافی خواهد بود.¹

اطلاعات درز یافته

سگّوها افزون بر خطوط راهنمای اجتماعی که در معرض عموم گذاشته‌اند، قواعد و دستورالعمل‌های جزئی‌تر برای کارمندان بخش تعدیل محتوایشان را هم صادر می‌کنند. این مستندات محرمانه‌اند، اما در موقعیت‌های متعددی به خبرگزاری‌ها درز می‌یابند. این اطلاعات درز یافته بر روی اینکه اصول عمومی سگّوها چگونه در عمل اجرا می‌شوند، پرتو می‌افکند؛ سگّوها به منظور تسهیل فعالیت

1. MacDonald 2014, p. 5

وسیع نیروهای تعدیل‌گر خود، غالباً مجبورند گفتگوهای پیچیده را در حد قواعد سرانگشتی فروبکاهند.

به عنوان نمونه، راهنماهای دستورالعمل فیسبوک به گاردین^۱ درز پیدا کرد، که در آن به تعدیل‌گران محتوا گفته شده بود که عبارت «یکی باید ترامپ را بکشد» یک مصداق قطعی از رفتار خشونت‌آمیز است، اما «بیایید کودکان چاق را کتک بزنیم» مصداق آن نیست.^۲ در اسنادی که به Süddeutsche Zeitung درز کرد، فیسبوک به تعدیل‌گران دستور داده بود که با مردم مادامی که «در گزارش‌های خبری دو سال اخیر، حداقل پنج‌بار از آنها نام‌برده شده باشد»،^۳ به عنوان شخصیت‌های عمومی^۴ و با حریم خصوصی کمتر حفاظت شده رفتار کنند. درز اطلاعات ۲۰۱۲ گوکر،^۵ مدعی بود که به پیمانکاران فیسبوک دستور داده شده بود که به کارمندان فیسبوک در مورد هرگونه محتوای مربوط به «نقشه‌ی کردستان»، «آتش زدن پرچم ترکیه» و «همه‌ی جسارت‌ها به آتاتورک (متنی و تصویری)» سخت بگیرند - که گویای این بود که فیسبوک تسلیم فشارهای عمومی از طرف ترکیه شده بود.^۶ پس از آن، نیویورک تایمز هم اسنادی مشابه را منتشر نمود.^۷

1. The Guardian
2. Hopkins 2017
3. Krause and Grasseger 2016
4. Public figures
5. Gawker
6. Chen 2012
7. Fisher 2018

بخش سوم

پژوهش مستقل



بخش مهم و در حال رشد اطلاعات درباره‌ی عملیات حذف در سگّوها از پژوهش‌های بیرونی حاصل می‌شود. برخی از این پژوهش‌ها (۱) داده‌هایی را که توسط خود سگّوها بروز داده شده است، تحلیل می‌کنند، درحالی‌که برخی دیگر (۲) نظرسنجی‌ها و مصاحبه‌هایی با کارکنان سگّوها و دیگر مشارکت‌کنندگان در زیست‌بوم حذف محتوا را انجام می‌دهند.^۱ (۳) و دیگر موارد نیز آزمون‌ها و تجربیات خودشان را بر روی سرویس‌های سگّوها و نرم‌افزار وابسته اجرا می‌کنند.

تحلیل داده‌های افشاشده توسط سگّوها

همان‌گونه که در بخش «منابع اولیه‌ی اطلاعات اشتراک‌گذاری شده توسط سگّوها» ذکر شد، پژوهشگرانی که از پایگاه داده‌ی لومین استفاده می‌کنند می‌توانند کارهای مهمی را انجام دهند که سایرین نمی‌توانند: بررسی محتویاتی که سگّوها حذف نموده‌اند. چند مورد از گزارش‌های بر پایه‌ی لومین، نظیر «تذکر و حذف»، همین رویکرد را دارند.^۲

۱. سگّوها اخیراً قدم‌هایی در راستای افشاکاری‌های بزرگتر به پژوهشگران مستقل داشته‌اند. مثلاً فیسبوک در پروژه‌ی اشتراک‌گذاری علوم اجتماعی واحد (Social Science One) که در متن اصلی بحث شد، مشارکت نموده و توییت را آزمایش ابزارهای جدید قواعد به‌سطح‌آوری (surfacing) بر روی کاربران در یک پروژه‌ی مشترک با پژوهشگران موافقت نمود (Benesch and Matias ۲۰۱۸). این موارد پیشرفت‌های مهمی هستند، اما روشن نیست که آیا به اطلاعات عمومی جدیدی به ویژه در موضوع حذف محتوا منجر خواهد شد یا خیر.

2. Urban et al. 2016

اکثر این گزارش‌ها به حق نشر می‌پردازند چرا که عمده داده‌هایی که لومین نگه می‌دارد به اقدام حق نشر هزاره دیجیتال آمریکا مربوط است.^۱

یک مورد استثناء وجود دارد که از طرف استاد حقوق، یوجین ولوخ^۲ مطرح شده، کسی که کشف کرده است که تعداد زیادی از شاکیان، احکام جعلی دادگاهی داشتند و از آن‌ها برای متقاعد کردن گوگل برای حذف محتوا از نتایج جستجویش استفاده کردند.^۳ کار کارآگاهی ولوخ، که شامل اجیر کردن یک کارآگاه واقعی بود، سرخ‌هایی را به دست داد که به تاریخ و ارقام مهرهای اداری‌ای که بر روی اسناد منتسب به دادگاه قرار داشت، مربوط می‌شد - جزئیاتی که تنها به دلیل اینکه لومین اجازه‌ی کپی دقیق گرفتن از آنها را به پژوهشگران داده بود، در دسترس بود.

گزارش‌دهی و مصاحبه‌ها با همکاران در فرایندهای حذف

دیگر پژوهشگران کار مشقت‌بار دنبال کردن توسعه‌های جهانی و جستجو و مصاحبه با افراد دخیل را دنبال کرده‌اند. ربکا مک‌کینون^۴ کار اساسی مهمی را در کتاب ۲۰۱۲ خود با عنوان «اجازه امر شبکه‌ای شده»^۵ بنا نهاد.^۶ گزارش تذکر و حذف، که در بخشی از این کتاب با

۱. هاینز و بکلس (2005 Heins and Beckles) یافته‌اند که 47 درصد از تذکرات با دلایل ضعیف اقامه شده‌اند یا شامل گفتگوهایی شده‌اند که دفاعیات حقوقی محکمی دارند؛ اربن و کیلتر (2006 Urban and Quilter, pp. 621, 641) گفته‌اند که ۵۵ درصد تذکرات شامل مشاجرات بین رقبای کاری می‌شود و ۳۱ درصد از آنها پرسش‌های حقوقی قابل توجهی دارند؛ بار-زیو و الکین کورن (2017 Bar-Ziv and Elkin-Koren) یافته‌اند که تعداد اندکی از تذکره‌دهندگان مسئول بیش از ۶۶ درصد از درخواست‌های سوءاستفاده آمیز، روی دامنه‌ی اسرائیل (il) است، اما مابقی درخواست‌ها تا کمتر از ۵ درصدشان نامعتبر یا سوال برانگیز است؛ سینگ (2015 Seng, 1 p.) یافته است که ۸۳ درصد از تذکرات، خطای «صوری» دارند و ۱۳ درصد از آنها خطای محتوایی دارند. یک نظرسنجی از اعضا توسط اتحادیه‌ی حق نشر که یک سازمان وکالتی حامی حق نشر است، یک آمار کمی نادر از تجربیات هشداردهندگان فراهم آورده‌اند. از میان ۲۱۹ پاسخ‌دهندگان به پیمایش، ۶۲ درصدشان احساس می‌کنند که تذکرات دریافتی از DMCA هرگز توسط سگوها پاسخ داده نمی‌شود. مثال‌های این گزارش گویای این بودند که بسیاری از آنها ممکن است شرکت‌هایی کوچک‌تر باشند. (رک: 2016 Copyright Alliance).

2. Eugene Volokh
3. Volokh and Levy 2016
4. Rebecca McKinnon
5. Consent of the Networked
6. MacKinnon 2012

عنوان «یک مطالعه‌ی موردی: تذکر و حذف در عملیات روزمره» بحث شده است، مبتنی بر مصاحبه‌ها و مستندات گسترده‌ای از اقدامات خود گزارشی صاحبان حق نشر، سگوها و دیگر بازیگران در زیست‌بوم تذکر و حذف است.^۱

مساهمت‌های علمی شایان توجه مربوط به اعمال خطوط راهنمای اجتماع شامل مقاله‌ی «حکمرانان جدید: مردم، قواعد و فرایندهای حاکم بر گفتگوهای برخط» نوشته کیت کلونیک^۲ است، که نویسنده‌اش برای آن گزارش‌هایی که تاکنون روی این موضوع بوده را مرور کرده و با تعدیل‌کنندگان محتوای سگوهای اولیه مصاحبه کرده است تا رشد قواعدشان را بفهمد. به طور مثال فیسبوک با قواعدی مبهم کار خود را آغاز کرد، تا حدی که یکی از کارمندان این وضعیت را با این عنوان توصیف می‌کرد: «احساس می‌کنی بد است؟ حذفش کن».^۳ تارلتن گیلسپی^۴ درباره‌ی تعدیل محتوا در کتاب ۲۰۱۸ خود کاوش نموده است.^۵ دیگر پژوهشگران درباره‌ی قواعد حذف مستند شده‌ی سگوها که در دسترس عموم هستند، گزارش نموده‌اند.^۶

گزارشگرانی همچون جولیا آنگوین^۷ درباره‌ی میزان تحمل محتوا در سگوها آزمایش انجام داده‌اند که شامل مستندسازی مواردی نظیر عباراتی با هدف تبلیغ یهودستیزی در فیسبوک می‌شود.^۸ دانشگاهیان و مشاوران جامعه‌ی مدنی‌ای که به سایت Onlinecensorship.org منتسب هستند از جمع‌سپاری برای گردآوری و احصاء گزارش تجربیات کاربران بهره برده‌اند^۹ که شامل اثر نامتجانس ظاهری بر روی گروه‌های آسیب‌پذیر و اقلیت‌ها

1. Urban et al. 2016, pp. 117-116/13-10

3. Klonick 2018

5. Gillespie 2018

7. Julia Angwin

9. Anderson, Carlson et al. 2016; Anderson, Stender et al. 2016

2. Kate Klonick

4. Tarleton Gillespie

6. Venturini et al. 2016; York et al. 2018

8. Angwin, Varner, and Tobin 2017

می‌شود.^۱ مستندسازان حداقل دو فیلم درباره‌ی تجربه‌ی عملی تعدیل‌گران فردی واقع در خط مقدم، که در کشورهایی همچون هند یا فیلیپین کار می‌کنند و دارای قرارداد با سگوهای متعلق به ایالات متحده هستند، ساخته‌اند.^۲ دنیل کرایس^۳ و شین مک‌گور^۴ مصاحبه‌هایی جامع با کارکنان تبلیغات فیسبوک و گوگل انجام داده‌اند تا بررسی کنند که چگونه استانداردهای شرکت‌ها حول تبلیغات سیاسی توسعه یافته و اعمال شده است؛^۵ و دانشگاهیانی چون سارا رابرتز^۶ نقش - و آسیب‌پذیری - این نیروی کاری جهانی مربوطه را تحلیلی دقیق نموده‌اند.^۷

دانشگاهیان همچنین کارمندان سگوها را در کنفرانس‌ها و دیگر رویدادهای مباحثه‌ای نظیر سلسله جلسات «تعدیل محتوا در مقیاس وسیع»^۸ که توسط پروفیسور اریک گلدمن از دانشگاه سانتا کلارا راه‌اندازی شده است، به بحث می‌کشاند.^۹

آزمون‌ها و آزمایش‌های مستقل

به این خاطر که تصمیمات سگوها بر حذف محتوا پشت درهای بسته گرفته می‌شود، برخی از پژوهشگران خلاقانه به دنبال آن‌اند

1. York and Gullo 2018
2. The Cleaners (2018), directed by Hans Block and Moritz Riesewieck (Berlin: Gebrueder Beetz Filmproduktion); Field of Vision - The Moderators (2017), directed by Ciaran Cassidy and Adrian Chen (New York: First Look Media), YouTube video, April 14. www.youtube.com/watch?v=k9m0axUDpro
3. Daniel Kreiss
4. Shannon McGregor
5. Kreiss and McGregor 2019
6. Sarah Roberts
7. Roberts 2019
8. the Content Moderation at Scale (or "COMO")
9. anta Clara University School of Law, 2018. Content Moderation & Removal at Scale conference, Santa Clara, CA, February 2. <https://law.scu.edu/event/content-moderation-removal-at-scale>.

در جلسه‌ای که بازتاب خبری وسیعی داشت، امنا لانسو (Emma Llansó) از مرکز مردم‌سالاری و فناوری و مایک منسینک (Mike Masnick) از بلاگ تک‌دیرت (The blog Techdirt) مخاطبان خود را به «ایجاد فراخوان» درباره‌ی یک‌سری از تصمیمات آزاردهنده‌ی حذف دعوت کردند - و یافتند که حتی با وجود همگنی نسبی فرهنگی میان مخاطبان واکنش‌گفتن دی سی، هیچ توافقی در این باره صورت نگرفت (2018 Goldman).

که به اطلاعاتی مفید در این باره پی ببرند. مثلاً پژوهشگران اروپایی در اوایل دهه ۲۰۰۰، از آثار مشهور ادبی خارج از حق نشر - مثلاً در یک مورد، کتاب «درباره‌ی آزادی»^۱ جان استوارت میل^۲ و مقاله‌های از طنزپرداز هلندی قرن نوزدهم، مولتاتولی^۳ در یک مورد دیگر^۴ - پست می‌گذاشتند و پس از آن، تذکرات نقض حق نشر درباره‌ی آنها ارائه می‌دادند تا ببینند آیا واسطه‌ها آن پست‌ها را حذف خواهند کرد یا خیر. عمده‌ی واسطه‌ها این تذکرات را پذیرفتند. کمی بعدتر، پژوهشگری به نام ریشاب دارا^۵ مجموعه‌ای از درخواست‌های حذف محتوا را به واسطه‌هایی در هند فرستاد و پاسخ‌هایشان را با جزئیات پیگیری نمود.^۶ سگوها عموماً در این مورد هم در توجه به درخواست‌ها به خطا افتادند. پژوهش دارا از نظر تمرکز بر ادعاهای غیر از حقوق نشر، پژوهشی نسبتاً منحصر به فرد است. مثلاً او با استناد به یک قانون علیه حمایت از قماربازی، سبب شد که یک سایت خبری نظرات کاربران در مورد یک تغییر پیشنهادی در قانون قماربازی هند را حذف کند.^۷ میان پرل^۸ و نیوا ال‌کین کورن^۹، پژوهشگران دانشگاه حيفا^{۱۰} پژوهشی مشابه در اسرائیل درباره‌ی ارزیابی استفاده از الگوریتم‌ها در فرایندهای حذف بر اساس حق نشر - روشی که آن را «وررفتن با جعبه سیاه»^{۱۱} نامیده‌اند - انجام داده‌اند.^{۱۲} پژوهش مستقل ممکن است به طور خاص برای مطالعه سامانه‌های فیلترینگ الگوریتمی محتوا مرتبط باشد، چیزی که به شدت در حال تبدیل شدن به محور مباحثات قانونی و فعالیت‌های تعدیل سگوهای

1. On Liberty

3. Multatuli

5. Rishabh Dara

7. Dara 2011, p.

8. Mayaan Perel

10. Haif

12. Perel and Elkin-Koren 2017

2. John Stuart Mill

4. Leyden 2004

6. Dara 2011

9. Niva Elkin-Koren

11. Black Box Tinkering

۱۵ این پژوهش ممکن است به خاطر این که اخلاقیات در مورد هدف قرارداد قانونی کاربران اینترنت جهان واقعی برای حذف غیرموجهانه دغدغه‌مند است، قابل تکرار نباشد.

بزرگ است. اتکا به این فناوری‌ها فعالان حقوق مدنی را نگران کرده است، از آن روی که در مورد تصمیماتی که نیازمند ارزیابی‌های زمینه‌ای با ریزه‌کاری هستند، این فناوری‌ها ضعیف عمل می‌کنند.^۱ به عنوان نمونه تمایز قائل شدن میان پروپاگاندای تروریستی و یادداشت‌های مطبوعاتی در تفسیر تروریسم، یا تمیز بین سرقت ادبی و تقلید ادبی یا دیگر ابزارهای مجاز ادبی، برای یک سامانه‌ی خودکار سخت است.

مرکز مردم‌سالاری و فناوری در گزارش ۲۰۱۸ خود، فیلترهای بر پایه‌ی متن و به لحاظ تجاری در دسترس را مورد بررسی قرار داده‌اند و یافته که میزان دقت آن، در حدود ۷۰-۸۰ درصد است.^۲ فیلترها به طور خاص، در ارزیابی لطیفه‌ها و طعنه‌ها یا درباره‌ی زبان‌هایی که توسط توسعه‌دهندگان سکو صحبت نمی‌شدند، ضعیف عمل کرده بودند.^۳ گزارش ۲۰۱۷ نیک فمستر^۴، استاد علوم رایانه‌ی پرینستون^۵ به همراه اوان انگستروم^۶ از گروه حمایت از کسب‌وکارهای نوپای «انجین»^۷ با تحلیل اکوپرینت^۸ جزئیات فنی بیشتری را فراهم آورد؛ اکوپرینت یکی از معدود ابزارهای فیلترینگ متن‌باز است (که بنابراین به طور عمومی قابل بررسی است).^۹ نویسندگان میزان خطای ۱ تا ۲ درصدی را در انجام تطبیق ساده دوتایی یافتند، که شامل مثبت کاذب و منفی‌های کاذب می‌شد.

تحت یک نگرش وسیع‌تر به تعدیل محتوا، سگوها همچنین به گفتمان از

1. Article 2018 .1. Joint Letter on European Commission regulation on online terrorist content. www.article19.org/resources/joint-letter-on-european-commission-regulation-on-online-terrorist-content/; Reda (2017).

2. Center for Democracy and Technology 2017

3. Center for Democracy and Technology 2017, pp. 19, 14

4. Nick Feamster

5. Princeton

6. Evan Engstrom

7. Engine

8. Echoprint

9. Engstrom and Feamster 2017

طریق طراحی الگوریتم‌های رتبه‌بندی و پیشنهاددهنده‌شان، مانند پیش‌خوان خبری فیسبوک^۱ و ویدئوهای پیشنهادشده یوتیوب، شکل می‌دهند.^۲ بخش در حال رشدی از ادبیات علوم رایانه و علوم ارتباطات به تعیین نحوه‌ی عملکرد و آثار این سامانه‌های پیچیده می‌پردازد.^۳ طراحی این الگوریتم‌ها معمولاً تنظیم‌نشده است، اما تعدادی از دولت‌ها اخیراً پیشنهاد تنظیم‌گری آن‌ها را داده‌اند.^۴ اکثر این اقدامات همچنین به صراحت تقاضای شفافیت بیشتر در پیشنهادهای الگوریتمی را دارند.^۵

1. Facebook's News Feed
2. Keller 2019b

۳. برای نمونه، رک: Hargreaves et al. (2018). که الگوهای پیشنهاد محتوا در فیسبوک برای رسانه‌های خبری ایتالیایی را بر پایه‌ی داده‌های مشاهده‌ی حساب‌های کاربری ساختگی مستند نموده است؛ رک: Cornia et al. (2018) که آثار تغییرات پیش‌خوان خبری فیسبوک در سازمان‌های خبری مختلف را آزمایش نموده است.

۴. Helberger, Leerssen, and van Drunen (2019).

همچنین رک: French Secretary of State for Digital Affairs (2019), p. ۳، که یک «تعهد به شفافیت کاربرد محتوای سفارشی» و «وظیفه‌ی احتیاط درباره‌ی کاربری اسگوها» را پیشنهاد می‌دهد؛ European Commission (2018b) در کد عملیات خود درباره‌ی ضد اطلاعات، سگوها را نیازمند «تسهیل امکان مشاهده‌ی ضد اطلاعات از طریق تقویت مشاهده‌ی محتوای قابل اطمینان» توصیف می‌کند.

5. European Commission (2018b). Regulation 1150/2019 (EU). همچنین رک.

بخش چهارم

نتایج حذف محتوای سئوها



نتایج حذف محتوای سگوها

بیشتر پژوهش‌های تجربی بر روی حذفیات محتوای سگوها صرفاً بر روی خود تصمیمات حذف متمرکز است. یافتن پژوهش بر روی پرسش‌های پیچیده‌تر درباره‌ی اینکه حذف‌ها چگونه بر افراد کاربر و جامعه در سطح کلان اثر می‌گذارد، دشوار است. یک استثنا احتمالاً حوزه روبه‌رشد پژوهش روی اثرگذاری برخط و تحریف فرایندهای سیاسی مردم‌سالارانه است. حوزه‌ی پژوهش تجربی مربوطه شامل «اخبار جعلی»، دخالت انتخاباتی روس‌ها، تقویت پیام‌های مبتنی بر بات‌ها و سوگیری سیاسی دخیل در خط‌مشی‌های تعدیل محتوای سگوها می‌شود. کارهای در حال انجام و احتمالاً کارهای آینده در این حوزه از نظر تطبیقی قوی هستند و در بخش‌های مختلف این کتاب هم مورد بحث قرار گرفته‌اند. یک منبع نویدبخش برای پژوهش‌های آینده، پروژه‌ی Social Science One فیسبوک در همکاری با شورای پژوهش علوم اجتماعی است، پروژه‌ای که دسترسی‌هایی به داده‌های بی‌نام کاربران برای پژوهشگران مستقل درباره‌ی «آثار رسانه‌های اجتماعی بر روی مردم‌سالاری و انتخابات‌ها» فراهم خواهد کرد.¹

1. Social Science One. <https://socialscience.one>

دیگر موضوع مربوط، اتهامی است که به طور فزاینده‌ای در ایالات متحده، آلمان و دیگر کشورها در حال مطرح شدن است که سگّوهای اصلی مستقر در کالیفرنیا علیه محافظه‌کاران سیاسی سوگیری دارند. تصمیمات فردی حذف غالباً موجب پوشش خبری یا نگرانی در شبکه‌های اجتماعی درباره‌ی این امکان می‌شود. البته برای بررسی این ادعا، پژوهشگران نیازمند اطلاعاتی به مراتب بیش از این درباره‌ی الگوهای کلی حذف هستند. حتی با این داده‌ها، پژوهشگران می‌توانند هنوز در مورد آنچه که یک گفتگوی سیاسی مشروع لحاظ می‌شود و اینکه کدام کاربران در دسته‌ی «محافظه‌کاران» جای می‌گیرند، اختلاف نظر داشته باشند. به طور مثال، مفسران در مورد طبقه‌بندی مناسب حزب نازی آمریکایی اختلاف نظر دارند.^۱

فراتر از موضوعات مرتبط با انتخابات، پژوهش تجربی بر روی آثار وسیع‌تر تصمیمات حذف سگّوها نادر هستند. یک پرسش به طور خاص مهم در مورد رابطه‌ی بین گفتگوی برخط و خشونت‌ورزی برون‌خط است. شاهدانی از سراسر دنیا متوجه اثر رسانه‌های اجتماعی به عنوان یک عامل علی در خشونت‌ورزی در مناطق مختلف، از میمانمار تا لیبی بوده‌اند.^۲ یک مطالعه در ۲۰۱۸ از آلمان که مدعی بود اثر فیسبوک روی حملات فیزیکی علیه مهاجران را احصاء کرده است، سرخط عناوین خبری را به خود اختصاص داد و به خاطر روش‌شناسی‌اش مورد نقد جدی قرار گرفت.^۳

پژوهش بر روی تروریسم، افراط‌گرایی و جذب نیرو در مقام مقایسه با دیگر موضوعات، پیشرفته محسوب می‌شود اما متخصصان در مورد نقش واقعی محتوای برخط اختلاف نظر دارند. مثلاً یک بررسی از

1. Hanania 2019; Graves 2019
2. Walsh and Zway 2018; McLaughlin 2018
3. Taub and Fisher 2018; Masnick 2018

ادبیات به روز موضوع در ۲۰۱۷ نظراتی واگرا را ذکر نموده که البته گویای حرکتی به سمت «اجماع بر اینکه اینترنت به تنهایی علت افراط‌گرایی نیست، اما می‌تواند به عنوان یک تسهیلگر و کاتالیزور در سیر افراد به سمت کنش‌های خشن سیاسی عمل کند»، است.^۱ دیگر پژوهشگران داده‌هایی را آورده‌اند که مطرح می‌کند سگوهای بازی که رؤیت‌پذیری عمومی و گفتار مقابل را مجاز می‌دانند، در مقایسه با محیط‌های گفتگوی بسته‌ی اینترنتی کمتر ممکن است به خشونت در جهان واقعی منجر شوند.^۲ یک پرسش تجربی مرتبط، درباره‌ی گفتگوی برخاسته و مشارکت همگانی صورت-گرفته توسط اعضای گروه‌های آسیب‌پذیر یا اقلیت‌ها است. به طور مثال، بسیاری از اندیشمندان این نگرانی‌شان را بیان می‌کنند که تحمل گفتگوی قانونی اما توهین‌آمیز یا تهدیدآمیز بر روی سگوهایی همچون توییتر، به طرز مؤثری حضور اقلیت‌های قومی، زنان و دیگر گروه‌های تحت حمله در میان عموم را کاهش می‌دهد.^۳ همچنین سازمان‌های حقوق مدنی سگوها را به ساکت نمودن ناشایسته‌ی اعضای اقلیت‌ها متهم نموده‌اند.^۴ در غیاب مجموعه داده‌های نمایانگر که تصمیمات حذف محتوای فردی را آشکار کند، ارائه پاسخ صحیح به پرسش‌ها درباره‌ی اثر نابرابر یا سوگیری در اجرای حذفیات غیرممکن است. دیگر نتایج اجرائیات حذف در سگوها ممکن است بر هر کاربری اثر بگذارد.

1. Meleagrou-Hitchens and Kaderbhai 2017, pp. 39, 19; Keller 2018

2. Benesch 2014; Munger

۲۰۱۷. برای نمونه، پژوهش سوزان بنش (Susan Benesch) نشان داده است گفتگوی که آن را به خشونت‌های حین انتخابات کنیا مرتبط می‌دانستند، در مقایسه با تبادلات عمومی مباحث در توییتر، بیش از حد به بحث‌های محدود‌شده‌ی فیسبوک ربط داده شده‌اند. (Benesch 2014).

3. West 2017

۴. برای مثال، در سال ۲۰۱۷ هفتاد سازمان حقوق مدنی و عدالت اجتماعی به فیسبوک نامه‌ای برای شکایت از غرض‌ورزی‌ها در تصمیمات حذف محتوا نوشتند (Levin 2017). در سال 2018، یوتیوب با ابراز اعتراض از سوی کاربرانی نا دگر جنس‌گرا (LGBTQ) مواجه شد که می‌گفتند ویدئوهایشان به طرز نادرستی جریمه می‌شوند (The Guardian 2017)؛ همچنین رک: Duguay, Burgess, and Suzor (2018). یکی از نویسندگان این فصل استدلال نموده است که تلاش‌های سگوها ناشی از گرایش شدیدشان به مقابله با افراط‌گرایی اسلامی می‌تواند منجر به آسیب‌های ناشایسته‌ای به کاربرانی شود که عرب‌زبان هستند و یا راجع به اسلام صحبت می‌کنند (Keller, 2018, pp. 20-26).

مثلاً افرادی که توسط سگوه‌های اصلی‌ای همچون فیسبوک و گوگل از حساب کاربری‌شان بیرون نگه داشته شده‌اند، ممکن است در دسترسی به دیگر خدمات برخطی هم که وابسته به اطلاعات ورود به حساب کاربری‌شان باشد محروم شوند. آن‌هایی که برای آثار هنری یا نوشتاری‌شان متکی بر خدمات میزبانی هستند ممکن با حذف تنها کپی‌شان از آن اثر مواجه شوند؛^۱ و مطالعات متعددی مطرح می‌کنند آن دسته از کاربران اینترنت که معتقدند گفتگوی‌شان پایش می‌شود، نوشتار و پژوهش‌شان را محدود می‌کنند.^۲

1. Macdonald 2016

2. Marthews and Tucker 2017, Pen America 2013, Penney 2016

بخش پنجم

پرسش‌های تجربه‌ی درباره‌ی حذفیات
محتوای سگو



پرسش‌های تجربه‌درباره‌ی حذفیات محتوای سگّو

پژوهش‌های تجربی‌ای که در این فصل مرور شد، به برخی از پرسش‌ها درباره‌ی حذفیات محتوای سگّو پاسخ می‌دهد و برخی دیگر را روشن می‌سازد. ملاحظاتی کلیدی که به تصمیمات خط‌مشی‌گذاری آگاهی می‌دهند، در اینجا فهرست شده است. پژوهش‌های جاری و آینده‌ای که به این پرسش‌ها می‌پردازند، هم فهم ما و هم تصمیم‌گیری عمومی ما را درباره‌ی پرسش‌هایی که شامل سگّوها و گفتگوهای برخط می‌شود، تقویت خواهند کرد.

• میزان دقت‌ها در شناسایی موارد ممنوعه:

- در تذکرات دریافتی از اشخاص ثالث به طور کلی

- در تذکرات دریافتی از اشخاص ثالث متخصص یا «مورد اطمینان»

- در موارد نشانه‌گذاری شده توسط ابزارهای خودکار

- در تصمیم‌گیری‌های سگّوها

• حوزه‌های دقت بالاتر یا پایین‌تر

- برای ادعاهای مختلف (نظیر هتک حرمت یا حق نشر)

- برای انواع مختلف محتوا (نظیر تصاویر در مقابل متن؛ زبان

انگلیسی در مقابل زبان هندی؛ مقالات خبری در مقابل اشعار)

- برای انواع مختلف تذکره‌دهندگان (نظیر «خبرگان مورد اطمینان»)
- میزان موفقیت سازوکارهای طراحی شده برای جلوگیری از حذف‌های بیش از اندازه
- تعهدات قانونی یا جریمه‌هایی برای تذکره‌دهندگان
- تعهدات قانونی یا جریمه‌هایی برای سگوها
- ضد تذکر توسط کاربرانی که به انتشار پست‌های غیر قانونی متهم شده‌اند
- بازرسی‌ها توسط سگوها
- بازرسی‌ها توسط اشخاص ثالث
- شفافیت عمومی
- هزینه‌ها
- هزینه‌های اقتصادی یا دیگر هزینه‌های حمل شده بر سگوها
- هزینه‌های اقتصادی یا دیگر هزینه‌های حمل شده بر اشخاص ثالث، وقتی که سگوها کمتر از حد لازم حذف انجام می‌دهند (محتوای ممنوعه در سگو باقی می‌ماند)
- هزینه‌های اقتصادی یا دیگر هزینه‌های حمل شده بر اشخاص ثالث، وقتی که سگوها بیش از حد حذف انجام می‌دهند (هنگامی که سگو، محتوای قانونی یا مجاز را حذف می‌کند)
- فیلترها
- دقت در شناسایی کپی برداری‌ها
- دقت در طبقه‌بندی محتوای پیش از این دیده نشده
- توانایی تشخیص یا ارزیابی در مواقعی که یک آیتم محتوای واحد در یک بافت جدید پدیدار می‌شود (نظیر گزارش خبری)

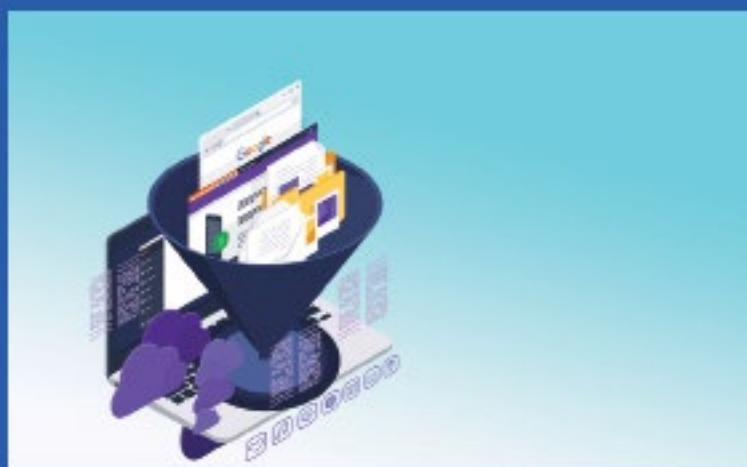
- دقت نسبی درباره‌ی انواع مختلف محتوای ممنوعه (نظیر
برهنگی در برابر حمایت از تروریسم)
- دقت نسبی درباره‌ی انواع مختلف فایل‌ها و رسانه‌ها (نظیر متن
در برابر MP3)
- اثربخشی بررسی انسانی توسط کارمندان سگو برای اصلاح
خطاهای فیلترینگ
- هزینه‌های پیاده‌سازی و نگهداری بارشده روی سکوهایی که از
فناوری فیلترینگ شخص ثالث استفاده می‌کنند
- تأثیر روی توسعه فنی متعاقب (مانند محدود ماندن به
طراحی‌های فنی مشخص)
- خطوط راهنمای اجتماع
 - قواعد اعمال شده
 - فرایندها، شامل درخواست تجدیدنظر
 - دقت و هزینه اعمال
 - نقش دولت‌ها در تعیین خطوط راهنمای اجتماع
 - نقش دولت‌ها در تصمیمات خاص حذف محتوا
 - عواقب حذف، حذف بیش از اندازه و حذف کمتر از اندازه
 - اطلاعات عمومی و افشا، شامل اعتماد به رسانه‌ها
 - خروجی‌های انتخاباتی
 - خشونت
 - علائق تجاری تذکره‌دهندگان
 - علائق تجاری کسب‌وکارهایی که از حذفیات اثر پذیرفته‌اند
 - تأثیر نابرابر بر اساس تفاوت در نژاد، جنسیت و غیره.

جمع بندی



فهم عمومی از اجرائیات حذف محتوا در سگّوها، حتی در میان پژوهشگران متخصص هم دیر زمانی است که محدود بوده است. این خلاً اطلاعات موجب ضعیف ماندن سیاست گذاران در مقام پاسخگویی به دغدغه‌ها در مورد سگّوها، گفتگوی برخط و مردم‌سالاری می‌شود. البته بخش در حال رشدی از پژوهش‌های مستقل و افشاگری‌های شرکت‌ها، در حال علاج این وضعیت است. از طریق شفافیت عمومی تقویت‌شده توسط سگّوها، پرسشگری‌های اندیشمندانه و ارزشیابی توسط خبرگان مستقل، ما می‌توانیم به سمت بینش‌های جدید و تصمیمات سیاستی سالم‌تر حرکت کنیم.

منابع



- Access Now. (2016). Transparency Reporting Index. Access Now report. www.accessnow.org/transparency-reporting-index/
- Ahlert, C., Marsden, C., & Yung, C. (2004). How “liberty” disappeared from cyberspace: The mystery shopper tests internet content self-regulation. Programme in Comparative Media Law & Policy at the Oxford Centre for SocioLegal Studies research paper.
- Amazon. (2015). Amazon Information Request Report. Amazon report. http://d0.awsstatic.com/certifications/Transparency_Report.pdf
- Anderson, J., Carlson, K., Stender, M., West, S. M., & York, J. C. (2016). Censorship in Context: Insights from Crowdsourced Data on Social Media Censorship. Onlinecensorship.org report. <https://onlinecensorship.org/news-and-analysis/onlinecensorship-org-launches-second-report-censorship-in-context-pdf>
- Anderson, J., Stender, M., West, S. M., & York, J. C. (2016). Unfriending Censorship: Insights from Four Months of Crowdsourced Data on Social Media Censorship. Onlinecensorship.org report. <https://onlinecensorship.org/news-and-analysis/onlinecensorship-org-launches-first-report-download>
- Anderson, N. (2010). Smoking guns, dark secrets aplenty in YouTube-Viacom filings. Ars Technica, March 18. <https://arstechnica.com/tech-policy/03/2010/smokingguns-dark-secrets-spilled-in-youtube-viacom-filings/>
- Angelopolous, C., Brody, A., Hins, A. W. et al. (2016). Study of Fundamental Rights Limitations for Online Enforcement Through Self-

Regulation. Report. <https://perma.cc/8QAW79-QT>

- Angwin, J., Varner, M., & Tobin, A., (2017). Facebook enabled advertisers to reach “Jew haters.” ProPublica, September 14. www.propublica.org/article/facebook-enabled-advertisers-to-reach-jew-haters
- Automattic. (n.d.). Transparency Report. Automattic report. <https://transparency.automattic.com>
- Bar-Ziv, S., & Elkin-Koren, N. (2017). Behind the scenes of online copyright enforcement: Empirical evidence on notice & takedown. Connecticut Law Review, 46-1 ,(2)50.
- Benesch, S. (2014). Countering dangerous speech to prevent mass violence during Kenya’s 2013 elections. Dangerous Speech Project website, February 9. <https://dangerousspeech.org/countering-dangerous-speech-kenya2013->
- Benesch, S., & Matias, J. N. (2018). Launching today: New collaborative study to diminish abuse on Twitter. Medium, April 6. <https://medium.com/@susanbenesch/launching-today-new-collaborative-study-to-diminish-abuse-ontwitter2-b91837668cc>
- Boyden, B. (2013). The Failure of the DMCA Notice and Takedown System. George Mason University Center for the Protection of Intellectual Property. <https://sls.gmu.edu/cpip/wp-content/uploads/sites/08/2013/31/Bruce-Boyden-The-Failure-of-theDMCA-Notice-and-Takedown-System1.pdf>
- Bridy, A., & Keller, D. (2016). U.S. Copyright Office Section 512 Study: Comments in Response to Notice of Inquiry, March 31.

- (2017). U.S. Copyright Office Section 512 study: Comments in response to second notice of inquiry. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2920871
- Bundesamt fuer Justiz. (2019). Bundesamt für Justiz erlässt Bußgeldbescheid gegen Facebook. Bundesjustizamt.de. www.bundesjustizamt.de/DE/Presse/Archiv/20190702/2019.html
- Cannataci, J., Kaye, D., & Ní Aoláin, F. (2018). Mandates of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression; the right to privacy; and promotion and protection of human rights and fundamental freedoms while countering terrorism. United Nations Special Rapporteur communication OTH 2018/71, December 7. <https://spcommreports.ohchr.org/TMResultsBase/DownloadPublicCommunicationFile?gId=24234>
- Center for Democracy and Technology. (2017). Mixed Messages? The Limits of Automated Social Media Content Analysis. Center for Democracy and Technology report. <https://cdt.org/insight/mixed-messages-the-limits-of-automated-social-mediacontent-analysis/>
- Chang, B. (2018). From Internet Referral Units to international agreements: Censorship of the Internet by the UK and EU. Columbia Human Rights Law Review, 212–114,(2)49.
- Chen, A. (2012). Inside Facebook’s outsourced anti-porn and gore brigade, where “camel toes” are more offensive than “crushed heads.” Gawker.com, February 16. <http://gawker.com/5885714/inside-facebooks-outsourced-anti-porn-and-gorebrigade-where-camel-toes->

are-more-offensive-than-crushed-heads

- Cloudflare. (2017). Why we terminated Daily Stormer. Cloudflare.com (blog), August 16. <https://blog.cloudflare.com/why-we-terminated-daily-stormer/>
- Copyright Alliance. (2016). Comments of the Copyright Alliance Before the U.S. Copyright Office, Docket No. 7-2015. <https://copyrightalliance.org/wp-content/uploads/11/2016/Copyright-Alliance-Section-512-Comments1.pdf>
- Cornia, A., Sehl, A., Levy, D., & Nielsen, R. K. (2018). Private Sector News, Social Media Distribution, and Algorithm Change. Reuters Institute Digital News Project report. <https://reutersinstitute.politics.ox.ac.uk/our-research/private-sector-news-social-media-distribution-and-algorithm-change>
- Cushing, T. (2017). Someone under Federal Indictment Impersonates a Journalist to File Bogus DMCA Notice. Techdirt, May 23. www.techdirt.com/articles/09500537404/20170518/someone-under-federal-indictment-impersonates-journalist-to-file-bogus-dmca-notice.shtml
- Dara, R. (2011). Intermediary Liability in India: Chilling Effects on Free Expression on the Internet. <https://cis-india.org/internet-governance/intermediary-liability-in-india.pdf>
- Duguay, S., Burgess, J., & Suzor, N. (2018). Queer women's experiences of patchwork platform governance on Tinder, Instagram, and Vine. *Convergence*, 252–237 ,(2)26. <https://doi.org/1354856518781530/10.1177>

- Engstrom, E., & Feamster, N. (2017). The Limits of Filtering: A Look at the Functionality & Shortcomings of Content Detection Tools. Engine report. www.engine.is/the-limits-of-filtering
- European Commission. (2016). Code of Conduct on Countering Illegal Hate Speech Online: First Results on Implementation. European Commission report. http://ec.europa.eu/information_society/newsroom/image/document/50-2016/factsheetcode-conduct40573_8-.pdf
- (2017). Code of Conduct on Countering Online Hate Speech: Results of Evaluation Show Important Progress. European Commission report. http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=71674
- (2018a). Public consultation on measures to further improve the effectiveness of the fight against illegal content online. Europa.eu. <https://ec.europa.eu/digital-singlemarket/en/news/public-consultation-measures-further-improve-effectiveness-fightagainst-illegal-content-online>
- (2018b). EU Code of Practice on Disinformation. Europa.eu. <https://ec.europa.eu/digital-single-market/en/news/code-practice-disinformation>
- European Digital Rights. (2016). Europol: Non-transparent cooperation with IT companies. EDRI.org. <https://edri.org/europol-non-transparent-cooperation-withit-companies/>
- Europol. (2016). EU Internet Referral Unit: Year One Report. Europol report. www.europol.europa.eu/sites/default/files/documents/eu_

iru_1_year_report_highlights.pdf

- Facebook. (2018a). Facebook Government Requests. Facebook report. <https://govtrequests.facebook.com>
- (2018b). NetzDG Transparency Report (January–June 2018). Facebook report. https://fbnewsroomus.files.wordpress.com/07/2018/facebook_netzdg_july_2018_english1-.pdf
- (2019a). A further update on the New Zealand terrorist attack. Facebook Newsroom, March 20. <https://newsroom.fb.com/news/03/2019/technicalupdate-on-new-zealand/>
- (2019b). Update on New Zealand. Facebook Newsroom, March 18. <https://newsroom.fb.com/news/03/2019/update-on-new-zealand/>
- (2019c). More coordinated inauthentic behavior from Russia. Facebook Newsroom, May 6. <https://newsroom.fb.com/news/05/2019/more-cib-from-russia/>
- (2019d). Removing coordinated inauthentic behavior from Israel. Facebook Newsroom, May 16. <https://newsroom.fb.com/news/05/2019/removing-coordinated-inauthenticbehavior-from-israel/>
- (2019e). Removing coordinated inauthentic behavior and spam from India and Pakistan. Facebook Newsroom, April 1. <https://newsroom.fb.com/news/04/2019/cib-and-spam-from-india-pakistan/>
- (2019f). More coordinated inauthentic behavior from Russia. Facebook Newsroom, May 6. <https://newsroom.fb.com/news/05/2019/more-cib-from-russia/>
- (2019g). Removing more coordinated inauthentic behavior from

Iran. Facebook Newsroom, May 28. <https://newsroom.fb.com/news/05/2019/removing-more-cibfrom-iran/>

- Fisher, M. (2018). Inside Facebook's secret rulebook for global political speech. *New York Times*, December 27. www.nytimes.com/27/12/2018/world/facebook-moderators.html
- Ford, G. S. (2017). Notice and takedown in everyday practice: A review. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2963230
- French Secretary of State for Digital Affairs. (2019). Interim Mission Report: Creating a French Framework to Make Social Media Platforms More Accountable. Mission report. www.numerique.gouv.fr/uploads/Regulation-of-social-networks_Missionreport_ENG.pdf
- Galperin, E. (2008). Massive takedown of anti-Scientology videos on YouTube. Electronic Frontier Foundation, September 5. www.eff.org/deeplinks/09/2008/massive-takedown-anti-scientology-videos-youtube
- Gebhart, G. (2018). Who has your back? Censorship edition 2018. Electronic Frontier Foundation, September 10. www.eff.org/who-has-your-back2018-
- German Parliament. (2018). Proposal of Representative Kunast and Others to Further Develop the Network Enforcement Law. Document 5950/19. <http://dip21.bundestag.de/dip21/btd/1905950/059/19.pdf>
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. New Haven, CT: Yale University Press.

- GitHub. (2015). GitHub's 2014 Transparency Report. Github report. <https://github.com/blog/-1987github-s-2014-transparency-report>
- Goldman, E. (2018). COMO: Content moderation at scale conference recap. Eric Goldman (blog). <https://blog.ericgoldman.org/archives/07/2018/como-content-moderation-at-scale-conference-recap.htm>
- Google. (n.d.). Google Transparency Report. Google report. <https://transparencyreport.google.com>
- (2018a). YouTube Community Guidelines Enforcement Report. Google report. <https://transparencyreport.google.com/youtube-policy/removals?hl=en>
- (2018b). Three Years of the Right to be Forgotten. Google report. https://g.co/research/rtbf_report
- (2018c). Removals under the Network Enforcement Law. Google report. <https://transparencyreport.google.com/netzdg/youtube>
- (2019). EU Code of Practice: May Report. Google report. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60042
- Graves, Z. (2019). Does Twitter have an anti-conservative bias, or just a Nazi bias? Techdirt, July 18. www.techdirt.com/articles/16154641652/20190221/does-twitter-have-anti-conservative-bias-just-anti-nazi-bias.shtml
- The Guardian. (2017). YouTube changes restrictions on gay-themed content following outcry. The Guardian, March 21. www.theguardian.com/music/2017/mar/21/youtube-changes-restrictions-gay-lgbtq-themed-content-tegan-sarah

- Hanania, R. (2019). It isn't your imagination: Twitter treats conservatives more harshly than liberals. Quillette, July 18. <https://quillette.com/12/02/2019/it-isnt-yourimagination-twitter-treats-conservatives-more-harshly-than-liberals/>
- Hargreaves, E., Agosti, C., Menasche, D., Neglia, G., Reiffers-Mason, A., & Altman, E. (2018). Biases in the Facebook News Feed: A case study on the Italian elections. arXiv.org. <https://arxiv.org/abs/1807.08346>
- Heins, M., & Beckles, T. (2005). Will Fair Use Survive? Free Expression in the Age of Copyright Control. Brennan Center for Justice report. www.brennancenter.org/publication/will-fair-use-survive-free-expression-age-copyright-control
- Helberger, N., Leerssen, P., & van Drunen, M. (2019). Germany proposes Europe's first diversity rules for social media platforms. LSE Media Policy Project (blog), May 29. <https://blogs.lse.ac.uk/mediapolicyproject/29/05/2019/germany-proposes-europesfirst-diversity-rules-for-social-media-platforms/>
- Hopkins, N. (2017). Revealed: Facebook's internal rulebook on sex, terrorism and violence. The Guardian, May 21. www.theguardian.com/news/2017/may/21/revealed-facebook-internal-rulebook-sex-terrorism-violence
- Human Rights Watch. (2006). "Race to the Bottom": Corporate Complicity in Chinese
- Internet Censorship. Human Rights Watch report. www.hrw.org/reports/2006/china0806/china0806web.pdf

- Kaye, D. (2017). Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, A/HRC/22/35, United Nations, August 18. <https://documents-dds-ny.un.org/doc/UNDOC/GEN/G46/077/17/PDF/G1707746.pdf?OpenElement>
- Keller, D. (2018). Internet platforms: Observations on speech, danger and money. Hoover Institution Aegis Paper Series No. 1807. www.hoover.org/research/internet-platforms-observations-speech-danger-and-money
- (2019a). Build your own Intermediary Liability law: A kit for policy wonks of all ages. Balkinization (blog), June 11. <https://balkin.blogspot.com/06/2019/build-your-own-intermediary-liability.html?m=1>
- (2019b). Who do you sue? State and platform hybrid power over speech. Hoover Institution Aegis Paper Series No. 1902. www.hoover.org/sites/default/files/research/docs/who-do-you-sue-state-and-platform-hybrid-power-over-online-speech_0.pdf
- Klonick, K. (2018). The new governors: The people, rules, and processes governing online speech. *Harvard Law Review*, 1669–1599, (6) 131.
- (2019). Inside the team at Facebook that dealt with the Christchurch shooting. *The New Yorker*, July 18. www.newyorker.com/news/news-desk/inside-the-team-atfacebook-that-dealt-with-the-christchurch-shooting
- Krause, T., & Grasegger, H. (2016). Facebook's secret rules of deletion. *Süddeutsche Zeitung* Internet Archive. <https://web.archive.org/web/20170103040043/http://international.sueddeutsche.de/post/154543271930/facebooks-secret-rules-of-deletion>

- Kreiss, D., & McGregor, S. (2019). The “arbiters of what our voters see”: Facebook and Google’s struggle with policy, process, and enforcement around political advertising. *Political Communication*, 2(36). www.tandfonline.com/eprint/9IUFBkGmZ4YHtNXM7sQ2/full?target=10584609.2019.1619639/10.1080
- Kuczerawy, A. (2017). The power of positive thinking: Intermediary liability and the effective enjoyment of the right to freedom of expression. *Journal of Intellectual Property, Information Technology and E-Commerce Law*, 237–226 ,(3)8.
- Levin, S. (2017). Civil rights groups urge Facebook to fix “racially biased” moderation system. *The Guardian*, January 18. www.theguardian.com/technology/2017/jan/18/facebook-moderation-racial-bias-black-lives-matter
- Leyden, J. (2004). How to kill a website with one email: Exploiting the European Ecommerce Directive. *The Register*, September 10. www.theregister.co.uk/14/10/2004/isp_takedown_study
- Macdonald, F. (2016). Google’s deleted an artist’s blog, along with 14 years of his work. *Science Alert*, July 18. www.sciencealert.com/google-has-deleted-an-artist-s-blogwith-14-years-of-his-work
- MacDonald, K. (2014). A Human Rights Audit of the Internet Watch Foundation. Lord MacDonald Report. www.iwf.org.uk/sites/default/files/inline-files/Human_Rights_Audit_web.pdf
- MacKinnon, R. (2012). *Consent of the Networked: The Worldwide Struggle For Internet Freedom*. New York: Hachette Book Group.
- Marthews, A., & Tucker, C. E. (2017). *Government surveillance and*

internet search behavior. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2412564

- Masnick, M. (2018). Dubious studies and easy headlines: No, a new report does not clearly show Facebook leads to hate crimes. Techdirt, August 3. www.techdirt.com/articles/00122840491/20180823/dubious-studies-easy-headlines-no-newreport-does-not-clearly-show-facebook-leads-to-hate-crimes.shtml?threaded=true
- McIntyre, T. J. (2018). Internet censorship in the United Kingdom: National schemes and European norms. In L. Edwards (Ed.), *Law, Policy and the Internet* (pp. 330–291). Oxford: Hart Publishing.
- McLaughlin, T. (2018). How Facebook's rise fueled chaos and confusion in Myanmar. *Wired*, September 10. www.wired.com/story/how-facebooks-rise-fueled-chaosand-confusion-in-myanmar
- Medium. (2015). Medium's 2015 Transparency Report. Medium report. <https://blog.medium.com/medium-s-2015-transparency-report5-c6205c48afe>
- Meleagrou-Hitchens, A., & Kaderbhai, N. (2017). Research Perspectives on Online Radicalisation: A Literature Review, 2016–2006. VOX-Pol report. www.voxpol.eu/new-vox-pol-report-research-perspectives-online-radicalisation
- Munger, K. (2017). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 649–629 ,(3)39. <https://link.springer.com/article/10.1007/s5-9373-016-11109>
- O'Brian, D., & Malcolm, J. (+70).(2018 internet luminaries ring the alarm

on EU copyright filtering proposal. Electronic Frontier Foundation, June 12. www.eff.org/deeplinks/06/2018/internet-luminaries-ring-alarm-eu-copyright-filtering-proposal

- Ong, T. (2018). Facebook is recruiting external advisers to tackle claims of bias on its platform. The Verge, May 2. www.theverge.com/17310894/2/5/2018/facebookrecruiting-external-advisers-bias-minority-conservative
- Open Technology Institute. (2018). The Transparency Reporting Toolkit: Content Takedown Reporting. Open Technology Institute report. www.newamerica.org/oti/reports/transparency-reporting-toolkit-content-takedown-reporting/
- Oransky, I. (2013). WordPress removes Anil Potti posts from Retraction Watch in error after false DMCA copyright claim. Retraction Watch, February 5. <https://retractionwatch.com/05/02/2013/wordpress-removes-anil-potti-posts-fromretraction-watch-in-error-after-false-dmca-copyright-claim>
- PenAmerica (2013). Chilling Effects: NSA Surveillance Drives U.S. Writers to SelfCensor. Pen America report. <https://pen.org/chilling-effects>
- Penney, J. (2016). Chilling effects: Online surveillance and Wikipedia use. Berkeley Technology Law Journal, 172 ,(1)31. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2769645
- Perel, M., & Elkin-Koren, N. (2017). Black box tinkering: Beyond disclosure in algorithmic enforcement. Florida Law Review, 181(69). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2741513

- Regulation 2019 ,1150/2019 O.J. (L 57 (186 (EU)
- Reda, J. (2017).When filters fail: These cases show we can't trust algorithms to clean up the internet. Julia Reda website. <https://juliareda.eu/09/2017/when-filters-fail/>
- Reddit, Inc. (2015). Reddit, Inc. Transparency Report, 2015. www.reddit.com/wiki/transparency/2015
- Roberts, S. T. (2016). Commercial content moderation: Digital laborers' dirty work. Media Studies Publications, Paper No. 12. <https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=1012&context=commpub>
- (2019). Behind the Screen: Content Moderation in the Shadows of Social Media. New Haven, CT: Yale University Press.
- Seng, D., (2015). "Who watches the watchmen?" An empirical analysis of errors in DMCA takedown notices. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563202
- Taub, A., & Fisher, M. (2018). Facebook fueled anti-refugee attacks in Germany, new research suggests. New York Times, August 21. www.nytimes.com/21/08/2018/world/europe/facebook-refugee-attacks-germany.html
- Timmer, J. (2013). Site plagiarizes blog posts, then files DMCA takedown on originals. Ars Technica, February 5. <https://arstechnica.com/science/02/2013/site-plagiarizesblog-posts-then-files-dmca-takedown-on-originals>
- Torrent Freak. (2018). Google Drive uses hash matching to detect pirated content. TorrentFreak.com. <https://torrentfreak.com/google->

drive-uses-hash-matching-detectpirated-content

- Tumblr. (n.d.). Transparency Report. Tumblr report. www.tumblr.com/transparency
- Twitter. (2017). Twitter Transparency Report Copyright Notices January – June 2017. Twitter Report. <https://transparency.twitter.com/en/copyright-notices.html#copyrightnotices-jan-jun2017->
- (2018a). Twitter Rules Enforcement Report. Twitter report. <https://transparency.twitter.com/en/twitter-rules-enforcement.html>
- (2018b). Twitter Netzwerkdurchsetzungsgesetzbericht: Januar – Juni 2018. Twitter report. <https://cdn.cms-twdigitalassets.com/content/dam/transparency-twitter/data/download-netzdg-report/netzdg-jan-jun2018-.pdf>
- (2019). EU Code of Practice: May Report. Twitter report. https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60043
- UK Metropolitan Police. (250,000). (2016th piece of online extremist/terrorist material to be removed. Archive.org. <https://web.archive.org/web/20180306202125/http://news.met.police.uk/news/250000th-piece-of-online-extremist-slash-terrorist-material-to-be-removed208698->
- Urban, J. M., Karaganis, J., & Schofield, B. (2016). Notice and takedown in everyday practice. UC Berkeley Public Law Research, Paper No. 2755628. <http://ssrn.com/abstract=2755628>
- (2017). Response to “Notice and takedown in everyday practice: A review.” The Takedown Project. <http://takedownproject.org/2-770>

- Urban, J. M., & Quilter, L. (2006). Efficient process or “chilling effects”: Takedown notices under Section 512 of the Digital Millennium Copyright Act. Santa Clara High Tech Law Journal, 693–621 :(4)22.
- US Copyright Office. (2015). Section 512 Study. US Copyright Office report. www.copyright.gov/policy/section512/
- US House of Representatives. (2019). Social Media Advertisements. Intelligence.House.Gov. <https://intelligence.house.gov/social-media-content/social-media-advertisements.htm>
- US Patent and Trademark Office. (2015). Multistakeholder Forum on the DMCA Notice and Takedown System. USPTO.Gov. www.uspto.gov/learning-and-resources/ippolicy/copyright/multistakeholder-forum-dmca-notice-and-takedown-system
- Venturini, J., Louzada, L., Maciel, M., Zingales, N., Stylianou, K., & Belli, L. (2016). Terms of Service and Human Rights: An Analysis of Online Platform Contracts. Report. <http://bibliotecadigital.fgv.br/dspace/handle/18231/10438>
- Vivanco, J. M. (2014). Censorship in Ecuador has made it to the Internet. Human Rights Watch news release, December 15. www.hrw.org/news/15/12/2014/censorshipecuador-has-made-it-internet
- Volokh, E., & Levy, P. A. (2016). Dozens of suspicious court cases, with missing defendants, aim at getting web pages taken down or deindexed. Washington Post, October 10. www.washingtonpost.com/news/volokh-conspiracy/wp/10/10/2016/dozens-of-suspicious-court-cases-with-missing-defendants-aim-at-getting-web-pages-taken-down-or-

deindexed/?utm_term=.b61e8a5967b2

- Walsh, D., & Zway, S. A. (2018). A Facebook war: Libyans battle on the streets and on screens. New York Times, September 4. www.nytimes.com/04/09/2018/world/middleeast/libya-facebook.html
- West, L. (2017). I've left Twitter. It is unusable for anyone but trolls, robots and dictators. The Guardian, January 3. www.theguardian.com/commentisfree/2017/jan/03/ive-left-twitter-unusable-anyone-but-trolls-robots-dictators-lindy-west
- York, J. C., Faris, R., Deibert, R., & Heacock, R. (2018). Policing content in the quasipublic sphere. OpenNet Initiative bulletin, September. <https://opennet.net/policingcontent-quasi-public-sphere>
- York, J. C., & Gullo, K. (2018). Offline/Online project highlights how the oppression marginalized communities face in the real world follows them online. Electronic Frontier Foundation, March 6. www.eff.org/deeplinks/03/2018/offlineonlineproject-highlights-how-oppression-marginalized-communities-face-real
- YouTube. (2010). Broadcast yourself. YouTube (official blog), March 18. <https://youtube.googleblog.com/03/2010/broadcast-yourself.html>



مرکز ملی فضای مجازی
پروژه نگاه فضای مجازی

csri.majazi.ir