

تحلیل سازوکارهای پلتفرم‌های رسانه‌های اجتماعی برای مقابله با اطلاعات کذب و اخبار جعلی؛ فیس‌بوک و توئیتر

 10.52547/ami.2022.1651.1165

حسین حسینی / استادیار گروه مطالعات فضای مجازی، پژوهشگاه فرهنگ، هنر و ارتباطات، تهران، ایران.*
hassani@ricac.ac.ir

دریافت: ۱۴۰۱/۴/۱۲ - پذیرش: ۱۴۰۱/۵/۱۰

چکیده

پلتفرم‌ها با انواع محتواهای زیان‌آور از جمله اخبار جعلی در قالب فرایندی به‌نام بازبینی محتوا مقابله می‌کنند. این فرایند پیچیده و دشوار است و با کمک عوامل انسانی و ماشینی انجام می‌شود. در این مقاله، مراحل و روش‌های بازبینی محتوا، در پلتفرم‌های فیس‌بوک و توئیتر به‌روش اسنادی تحلیل شده است تا سازوکارهای آن‌ها جهت مقابله با محتواهای خبری جعلی شناسایی شود. یافته‌ها نشان می‌دهد هر کدام از این دو پلتفرم با آن‌که در یک زمینه سیاسی و اقتصادی یکسان و تحت سیاست‌گذاری رسانه‌ای واحد عمل می‌کنند، سازوکارهای متفاوتی برای مقابله با اطلاعات کذب دارند که از رویکرد کلی آن‌ها نشأت می‌گیرد. رویکرد کلی فیس‌بوک دفاع از شفافیت و دادن صدا به مردم است و بنابراین به‌جای حذف، کاهش رؤیت‌پذیری این نوع محتواها را درپیش گرفته است، اما توئیتر با هدف تضمین سلامت مکالمه، با سهولت بیش‌تری اقدام به حذف محتواهای جعلی می‌کند. در نهایت، باید گفت شرکتی شدن مقابله با اطلاعات نادرست و اخبار جعلی، دریافت اخبار واقعی توسط شهروندان را دشوارتر و حوزه عمومی و دموکراسی را با مخاطرات فزاینده پیچیده‌تری مواجه می‌کند. از طرف دیگر، سلطه شرکت‌ها بر فرایند نظارت بر محتواهای خبری، دولت‌هایی همانند دولت ایران را با چالش‌های جدیدی از منظر حکمرانی اینترنت روبه‌رو می‌کند.

کلیدواژه‌ها: رسانه‌های اجتماعی، اخبار جعلی، اخبار کذب، بازبینی محتوا.

A Biannual Scientific Research Journal

Art & Media Studies

Vol.4, No.7, Spring and Summer 2022

pp.155-184

Analysis of Social Media Platform Mechanisms to Moderate Disinformation and Fake News; Facebook and Twitter

 10.52547/ami.2022.1651.1165

Hossein Hassani / Assistant Professor Department of Virtual Space Studies , Research Center For Culture, Art And Communications, Tehran, Iran.*

hassani@ricac.ac.ir

Received: 2022/7/3 - **Accepted:** 2022/8/1

Abstract

This article aims to analyze the policy and mechanisms of global social media platforms to deal with fake news and disinformation. This process is complex and challenging and is done with the help of human and machine agents. In this article, the steps and methods of content moderation on Facebook and Twitter platforms have been analyzed in a documentary way to identify their mechanisms to deal with fake news content. The findings show that these two platforms, although they operate in the same political and economic context and under the same media policy, have different mechanisms to deal with false information, which originates from their general approach. Facebook's general approach is to defend transparency and give voice to people. Therefore, instead of deleting it, it has taken the reduction of the visibility of this type of content. Still, Twitter, to ensure the health of the conversation, removes fake content more easily. Finally, it must be said that the corporatization of dealing with false information and fake news makes it more difficult for citizens to receive real news, and the public sphere and democracy face increasingly complex risks. On the other hand, the dominance of companies over monitoring news content makes governments like the Iranian government face new challenges from the point of view of internet governance.

Keywords: social media, fake news, disinformation, content moderation.

مقدمه

تولید و انتشار اخبار، بیش از پیش فرایند پلتفرمی شدن را سپری می‌کنند (Poell, et al., 2021) و سازمان‌های خبری بیش از پیش به پلتفرم‌های آنلاین برای تولید و انتشار اخبار وابسته می‌شوند، اما در عین حال، پلتفرم‌ها از جمله رسانه‌های اجتماعی و پیام‌رسان‌ها به مجرای اصلی تولید، انتشار و هم‌رسانی اخبار جعلی نیز تبدیل شده‌اند. در نتیجه، یک انقلاب پلتفرمی (فن‌دایک، ۱۳۹۶)، برای میلیاردها نفر در سرتاسر جهان این امکان را فراهم کرده است که انبوه محتوای کاربرساخته خود را بدون نظارت پیشینی یا پیش‌بازبینی^۱ در حساب‌های کاربری خود به سرعت به اشتراک بگذارند. به‌طور طبیعی، بخشی از این محتوای کاربرساخته، محتوای ناهنجار یا به‌تعبیر شرکت فیس‌بوک زیان‌آور^۲ هستند که توسط برخی از کاربران با انگیزه‌ها و اهداف گوناگون به اشتراک گذاشته می‌شود.

رسانه‌های اجتماعی یا پیام‌رسان‌ها در اصل به‌منزله^۳ یک سکوی فعالیت مجازی واسطه‌ای عمل می‌کنند که طیفی از محتوای متناسب یا نامتناسب از جمله اخبار جعلی و اطلاعات کذب^۴ به‌واسطه آن‌ها منتشر می‌شوند. اجتماع کثیری از کاربران عادی و غیرحرفه‌ای در تولید، اشتراک‌گذاری و مصرف محتوای فرهنگی و رسانه‌ای در این پلتفرم‌ها نقش دارند، اما به‌طور طبیعی، بسیاری از آن‌ها فاقد مهارت، اخلاق حرفه‌ای و یا تعهد اجتماعی به هنجارها و ارزش‌های فرهنگی و نیز اصول اخلاقی تولید محتوا از جمله محتوای خبری هستند. بنابراین حجم گسترده‌ای از به اصطلاح محتوای زیان‌آور همانند اطلاعات نابهنجار را - بنا به تعریف شرایط خدمت پلتفرم‌های بزرگ - در این پلتفرم‌ها، به اشتراک می‌گذارند. این نوع محتواها بر اساس استانداردهای اجتماع^۴ - به‌طور مثال، در فیس‌بوک - یا شرایط ارائه خدمت اجازه انتشار در پلتفرم‌ها را ندارند و به‌شیوه‌های مختلفی مشمول تنظیم‌گری می‌شوند.

اخبار جعلی به‌دلیل آسیب‌های زیان‌باری که دارند، از جمله آسیب به دموکراسی، مغشوش کردن واقعیت و پوشاندن واقعیت، تضعیف امنیت ملی، خدشه به اعتماد عمومی، تضعیف انسجام اجتماعی و عمل، به‌منزله^۵ ابزار پروپاگاندا بیش از پیش

به‌عاملی نگران‌کننده و چالش‌برانگیز از جمله برای دولت‌ها تبدیل شده‌اند. به‌دلیل فراگیری فزاینده این نوع آسیب‌ها و مخاطرات، اخیراً مقامات اتحادیه اروپا در قالب قانون خدمات دیجیتال^۵، از شرکت‌های بزرگ پلتفرمی خواسته‌اند تا ضمن احترام به آزادی بیان، اقدامات سخت‌گیرانه‌تری را برای مقابله با اطلاعات کذب به اجرا بگذارند (Chan and Casert, 2022).

به‌عبارت دیگر، انتشار اخبار جعلی در فضای آنلاین و پیامدهای آن برای زندگی واقعی، سبب شده است تا پلتفرم‌های اشتراک‌گذاری محتوا بیش از پیش با فشار برای کاستن از سرعت انتشار اخبار جعلی و نیز اطلاعات نادرست^۶ مواجه شوند. با وجود پیچیدگی و دشواری مقابله با اخبار جعلی، پلتفرم‌های آنلاین، سازوکارهایی را برای مقابله با این نوع محتواهای زیان‌آور از جمله اخبار جعلی تدبیر کرده‌اند. عقب‌افتادگی تنظیم‌گرایانه نهادهای دولتی و نهادهای قانون‌گذاری در شناسایی، تعریف و تصویب قوانین به‌روز و متناسب در مورد اخبار جعلی، حتی در کشورهایی که زمینه رشد پلتفرم‌های رسانه‌های اجتماعی هستند، یک عامل اساسی برای واگذاری و نیز درخواست جدی از پلتفرم‌ها برای مقابله با انتشار این نوع محتواهای کذب در پلتفرم‌ها است. از آن‌جا که فرایند خودحکمرانی در این پلتفرم‌ها، امری شرکتی و تابع خط‌مشی‌های کلان و فلسفه مدیران آن‌ها است، به‌طور طبیعی، نوع تعاریف، سازوکارهای تعبیه شده و نیز نگرش مدیران آن‌ها به‌حدود، شیوه‌ها، میزان و نحوه مقابله با اخبار جعلی متفاوت است. به‌عبارت دیگر، در رسانه‌های اجتماعی که اساساً تولید محتوا در آن‌ها توسط کاربران انجام می‌شود، تولید اخبار جعلی، اطلاعات نادرست، اطلاعات کذب و کژاطلاعات به‌یکی از دغدغه‌های اساسی راستی‌آزمایی محتوا و صحت‌سنجی آن، به‌یک چالش و دغدغه اساسی برای مدیران این پلتفرم‌ها تبدیل شده است.

سازوکاری که تحت عنوان بازبینی محتوا^۷ نامیده می‌شود، عمده‌ترین روشی است که رسانه‌های اجتماعی برای مقابله و یا جلوگیری از انتشار بیش‌تر محتواهای نامناسب به‌کار می‌برند. بازبینی محتوا را می‌توان به‌مثابه بخش عمده خودحکمرانی یا حکمرانی در پلتفرم‌ها قلمداد کرد. خودحکمرانی طیفی از تدوین اسناد، مقررات و دستورالعمل‌ها

برای تعریف و تعیین انواع محتواهای زیان‌آور و تدبیر سازوکارهای عملی برای تطبیق رفتار و گفتار در پلتفرم‌ها با این دستورالعمل‌ها است. بازبینی محتوا به معنای نظارت بر محتواهای کاربرساخته^۸، بررسی تطابق آن‌ها با این قواعد از پیش تعیین‌شده، تصمیم‌گیری در مورد بقاء یا حذف محتواهای چالش‌برانگیز و درنهایت، اقدام عملی در مورد آن‌ها است.

بنابراین، هدف این مقاله، مطالعه رویکردهای کلی مدیران دو پلتفرم فیس‌بوک و توئیتر به اخبار جعلی و فهم روش‌هایی است که آن‌ها در قالب فرایند بازبینی محتوا با انتشار این نوع محتواها در رسانه اجتماعی تحت هدایت و مدیریت خود مقابله می‌کنند. هدف اساسی دیگر، فهم پیامدهای وابستگی دولت‌هایی همانند ایران به پلتفرم‌های بزرگ جهانی برای مقابله با انتشار اخبار جعلی است. بنابراین مسئله اساسی این مقاله این است که این پلتفرم‌ها چه سازوکارهایی را برای مقابله با انتشار اخبار جعلی و اطلاعات کذب در پیش گرفته‌اند؟ در قالب این مسئله محوری، تلاش می‌شود به پرسش‌های فرعی زیر پاسخ داده شود:

۱. پلتفرم‌های بزرگ جهانی چه نگرشی به انتشار اخبار جعلی در پلتفرم تحت حاکمیت خود دارند؟

۲. این نگرش چگونه در سازوکارهای عملی بازبینی محتوا نمود پیدا کرده است؟

۳. پیامدهای این نگرش‌ها و سازوکارهای عملی برای اکوسیستم‌های پلتفرمی دیگر و نیز دولت‌هایی همانند ایران چیست؟

ادبیات نظری این مقاله به‌طور کلی، شامل مفهوم‌سازی اخبار جعلی و مفاهیم مرتبط با آن است. سپس رویکرد بازبینی محتوا را که سازوکار اصلی مقابله با اخبار جعلی در پلتفرم‌ها است، به‌شکل گذرا مرور می‌کنیم.

۱.۱. اخبار جعلی، اطلاعات کذب و اطلاعات نابهنجار

یکی از نگرانی‌های فزاینده در سرتاسر جهان، کاربرد پلتفرم‌های دیجیتال برای انتشار اطلاعات نادرست، اطلاعات کذب و نیز اخبار جعلی است. انتشار این نوع

اطلاعات و محتواهای خلاف واقع عموماً از انتخابات ریاست جمهوری امریکا در سال ۲۰۱۶ میلادی شدت گرفت. هرچند، همان طور که مسینتایر^۹ (۲۰۱۸) نوشته است، اخبار جعلی نه با انتخابات ریاست جمهوری امریکا در سال ۲۰۱۶ و نه پس از ابداع رسانه‌های اجتماعی آغاز شدند، تولید اخبار جعلی هم‌زمان با مفهوم «اخبار» پدیدار شده است و حتی می‌توان مدعی شد انتشار اخبار جعلی پس از ابداع دستگاه چاپ در سال ۱۴۳۹ میلادی شروع به گسترش کرد، زیرا در آن زمان، صحت‌سنجی اخبار «واقعی» دشوار بود و هیچ تصویری نسبت به اخلاق یا عینیت روزنامه‌نگاری وجود نداشته است.

عموماً اصطلاح رایج برای توصیف محتواهای خبری که برخلاف واقعیت و حقیقت هستند، اصطلاح اخبار جعلی است، اما به هر حال، اصطلاحات و تقسیم‌بندی‌های دیگر نیز برای توصیف خبرها و اطلاعات کذب و غیرواقعی نیز به کار رفته است که در ادامه آن‌ها را مرور می‌کنیم.

گیوستی و پیراس^{۱۰} (۲۰۲۱: ۲) کوشیده‌اند تا تفاوت‌های میان این مفاهیم را توضیح دهند. به‌نوشته آن‌ها،

«تا این اواخر اطلاعات کذب عموماً در گفتمان سیاسی برای مشخص کردن مسایل مرتبط با دست‌کاری اطلاعات به کار می‌رفته است که هدف آن، تولید اطلاعات کاذب با هدف فریب افکار عمومی بوده است. این اصطلاح در مقایسه با پروپاگاندا که از قرن شانزدهم رایج شده و به‌معنای استفاده‌گزینشی از اطلاعات با هدف تولید برخی از اثرات سیاسی مطلوب بوده است، جدیدتر است. گرچه اطلاعات کذب، محصول برساخت یک ناحقیقت هدفمند است، اما اطلاعات نادرست نشان‌گر یک بخش ناکامل، مبهم، کذب یا ابهام‌آمیز از اطلاعات است. اطلاعات نادرست غالباً متشکل از واقعیت واقعی است که از زمینه خود جدا شده و یا با مطالب کذب آمیخته شده تا از پیام مورد نظر حمایت کند و عمدتاً بخشی از یک طرح یا دستورکار کلان‌تر است.»

به‌نظر می‌رسد مهم‌ترین تفاوت اطلاعات نادرست از اطلاعات گمراه‌کننده، تعمدی

بودن انتشار این نوع اطلاعات است. بنابراین بخش زیادی از محتوایی که توسط کاربران به اشتراک گذاشته می‌شوند، باید در دسته اطلاعات نادرست قرار گیرند، زیرا هر چند این نوع محتواها خلاف واقع هستند، اما تنها قصد اشتراک گذاری از آن‌ها آگاه کردن دیگران از یک خبر است؛ هر چند امکان دارد آن محتوا یا اطلاعات نادرست در ابتدا به شکل عمدی منتشر شده باشد که در دسته اطلاعات کذب قرار می‌گیرد. رویه‌های رایج در پیام‌رسان‌ها، به شکل «بازفرست شده» و «بازفرست شده به تعداد بسیار» نقشی مهمی در چرخه تغییر ماهیت انواع اطلاعات کذب دارند.

در همان راستای تعریف اخبار نادرست، اخبار جعلی را می‌توان به منزله نسخه پیشرفته و فن‌آورانه اطلاعات نادرست تلقی کرد که در اینترنت و به طور خاص، در رسانه‌های اجتماعی منتشر می‌شوند. پیش از هر چیز، انتشار اخبار جعلی با زندگی سیاسی در برهه‌های حساس آن‌گره خورده است که از جمله آن‌ها می‌توان به انتخابات و بحران‌های بین‌المللی اشاره کرد، اما به سرعت به تمام وجوه سیاست و زندگی نفوذ کرده است. ترکیب اخبار جعلی و رسانه‌های اجتماعی سبب تولید اثرات ناخودآگاه می‌شود (Giusti and Piras, 2021).

آزادی و نیک‌گو (۱۳۹۸: ۱۲۱) اخبار جعلی را به این شکل تعریف کرده‌اند:

«خبر جعلی، اطلاعات مخربی است که به صورت حرفه‌ای و سازمانی در پوشش خبرهای واقعی و در راستای ایجاد جنگ روانی، بیش‌تر اوقات بر بستر رسانه‌های غیررسمی تولید و منتشر می‌شود».

گرچه در این تعریف، عناصر عمدی بودن و حتی نهادی بودن تعیین‌کننده اخبار جعلی است، اما ویژگی‌هایی که گیوستی و پیراس (۲۰۲۱) برای اخبار جعلی برشمرده‌اند، از جمله این‌که اخبار جعلی نسخه پیشرفته و فن‌آورانه اطلاعات نادرست است که به طور خاص، در رسانه‌های اجتماعی منتشر می‌شود، فاقد عنصر عمدی بودن است.

این مفاهیم گاهی به جای یکدیگر به کار می‌روند و با هم هم‌پوشانی دارند، اما آن‌ها اطلاعات کذب را زیرمجموعه اطلاعات نادرست قلمداد می‌کنند. از نظر آن‌ها، اطلاعات

کذب، اطلاعات نادرستی هستند که در تولید و انتشار آن‌ها عنصر عمدی بودن وجود دارد و با هدف گمراه کردن منتشر می‌شوند، اما انتشار اطلاعات نادرست ممکن است به شکل سهوی یا عمدی صورت گیرد. اما وقتی این اقدام به‌شکلی سازمانی و نهادی صورت گیرد، با اطلاعات کذب مواجه هستیم. اخبار جعلی نیز به‌منزله نوعی از اطلاعات کذب هستند. اخبار جعلی عموماً و به‌شکل کذب به‌شکلی طراحی می‌شوند که شبیه مقالات و گزارش‌های خبری سازمان‌های خبری معتبر باشند.

براساس این تعاریف، می‌توان نتیجه گرفت که وقتی محتواهای کذب به‌شکل عمدی تولید و منتشر شوند، می‌توانیم آن‌ها را اطلاعات کذب بدانیم که می‌تواند به‌شکل نهادی باشد، اما اطلاعات نادرست عموماً فاقد عنصر نیت‌مندی هستند و بخش عمده اطلاعات نادرستی که در رسانه‌های اجتماعی و پیام‌رسان توسط کاربران عادی به اشتراک گذاشته می‌شوند، در این دسته قرار می‌گیرند. اخبار جعلی به‌منزله شکلی از اطلاعات کذب عمدتاً عامدانه و تقلیدی از قالب‌های خبری رسانه‌های معتبر هستند. در قسمت پیشین، تلاش شد تا به اختصار تفاوت‌های انواع محتوای کذب بیان شود. اما بیان یک رویکرد نظری دیگر به این نوع محتواهای اطلاعاتی نیز سودمند به‌نظر می‌رسد. کالر واردل و حسین درخشان به‌جای این مفاهیم از مفهوم ناهنجاری اطلاعاتی^{۱۱} استفاده کردند. واردل و درخشان^{۱۲} (۲۰۱۷) اعتقاد دارند، اصطلاح اخبار جعلی نمی‌تواند به‌اندازه کافی پدیده پیچیده آلودگی اطلاعاتی را توصیف کند. همچنین سیاست‌مداران از این اصطلاح برای پرچسب‌زنی به‌دسته‌ای از سازمان‌های خبری استفاده کرده‌اند که از نحوه پوشش اخبار توسط آن‌ها رضایت ندارند. بنابراین آن دو ترجیح داده‌اند مفهوم ناهنجاری اطلاعاتی را به‌کار ببرند. در قالب این چارچوب مفهومی، سه نوع مختلف اطلاعات نادرست، اطلاعات کذب و کژاطلاعات شناسایی شده است. اطلاعات زمانی نادرست تلقی می‌شود که اطلاعات کذب به اشتراک گذاشته می‌شود، اما منظور از آن، آسیب رساندن یا وارد کردن زیان نیست؛ زمانی که اطلاعات کذب آگاهانه و با هدف ایجاد زیان یا آسیب به اشتراک گذاشته می‌شود، ما با اطلاعات کذب مواجه هستیم. زمانی که اطلاعات معتبر که باید به‌شکل خصوصی

نگهداری شوند، در حوزه عمومی منتشر می‌شوند تا سبب آسیب زدن به یک فرد یا نهاد شوند، ما با کژاطلاعات مواجه هستیم.

۲. بازبینی محتوا

محتوای زبان‌آور یا نامتناسب بر اساس یک فرایند حکمرانی درونی، خودتنظیم‌گری و یا حکمرانی توسط پلتفرم‌ها^{۱۳} (Gillespie, 2018)، به اصطلاح مورد بازبینی محتوا قرار می‌گیرند که می‌تواند به شکل بازبینی پیشین یا بازبینی پسین انجام شود. بخش قابل توجهی از محتواهایی که در این دوران تولید و منتشر شده‌اند، از جمله اطلاعات نادرست، اطلاعات کذب و اخبار جعلی که وارد دل و درخشان (۲۰۱۷) آن‌ها را در قالب اصطلاح فراگیر ناهنجاری اطلاعاتی طبقه‌بندی کرده‌اند، به منزله نوعی محتوای زبان‌آور تلقی می‌شوند.

بازبینی محتوا که فرایندی است که طبق آن تعیین می‌شود چه چیزی در فضای آنلاین باقی بماند و چه چیزی حذف شود، یکی از جنبه‌های اجتناب‌ناپذیر نظارت بر محتوا در رسانه‌های اجتماعی است. همان‌طور که برت^{۱۴} آشکارا گفته است، بدون این عمل رسانه‌های اجتماعی در هر زمانه‌ها، ارباب اشخاص، نوشته‌های ملی‌گرایان افراطی، تهدید به سر بریدن توسط تروریست‌ها و سوءاستفاده جنسی از کودکان غرق می‌شود. به گفته رابرتس^{۱۵} (۲۰۱۷) بازبینی محتوا، کرداری سازمان‌یافته است که طی آن بر محتوای کاربر ساخته که در رسانه‌های اجتماعی و پلتفرم‌های دیگر منتشر می‌شوند، نظارت می‌شود. هدف از این اقدام، تعیین تناسب محتوا با یک مکان، موقعیت و حوزه قضایی است. در هر زمینه اجتماعی، برخی هنجارهای فرهنگی، اصول اخلاقی و بندهای قانونی و حقوقی وجود دارد که ضرورت دارد محتواهای به اشتراک گذشته شده با مقررات حاکم بر آن منطبق باشد.

بازبینی محتوا از پلتفرمی به پلتفرم دیگر متفاوت است و در برخی از آن‌ها، این عمل به شکل تصادفی، غیرسازمان‌یافته یا غیرمنسجم روی دهد؛ یعنی بر اساس گزارش‌های موردی کاربران یا دستگاه‌های ناظر یا مواجهه مدیران و کاربران یک پلتفرم خاص،

اما در پلتفرم‌های دیگر این اقدام سازمان‌یافته، روال‌مند است و فرایندهای مشخص و معینی باید طی شود تا در مورد حذف محتوای غیرمجاز تصمیم‌گیری شود. بازبینی محتوا یا توسط افراد داوطلب صورت می‌گیرد - نظارت عموم کاربران بر محتوا و امکان گزارش محتواهای مسئله‌دار در این دسته قرار می‌گیرد - و یا توسط شرکت‌های ثالثی انجام می‌شود که برخی بازبین‌ها را استخدام و برای بررسی محتواهای پلتفرم‌ها به آن‌ها دستمزد پرداخت می‌کنند.

زوارت در مقاله‌ای درباره نحوه بازبینی محتوا توسط رسانه‌های اجتماعی می‌نویسد که حکم‌رانی پلتفرم‌های آنلاین یا حکم‌رانی در پلتفرم‌ها، شامل اصول و قواعد گوناگون اخلاقی، حقوقی و فرهنگی و نیز ابزارهای فنی است که توسط خود پلتفرم‌ها با هدف قاعده‌مندسازی محتوای در حال انتشار در بستر آن‌ها اجرا می‌شود که از ارسال هشدار و اخطار تا حذف محتواهای زیان‌آور و مسدودسازی کامل حساب کاربری کاربران متخلف را شامل می‌شود.

«شرایط استفاده از خدمات»^{۱۶} معمولاً مهم‌ترین سند مورد اتکای پلتفرم‌ها برای تعیین مصادیق محتواهای غیرمجاز و زیان‌آور است. البته طیفی از دستورالعمل‌ها، راهنماها و خط‌مشی‌های دیگر همانند خط‌مشی داده و حریم خصوصی نیز وجود دارد که مبنای عمل متولیان نظارت بر محتوا در پلتفرم‌های آنلاین و رسانه‌های اجتماعی است. به دلیل این که محتواهای غیرمجاز و قالب‌های تولید محتوای زیان‌آور دائماً به‌روزشونده است، معمولاً پلتفرم‌های بزرگ به‌طور منظم تلاش می‌کنند تا سندهای راهنمای خود را به‌روز کنند تا هرچه بیش‌تر و تا حد ممکن از انتشار محتواهای نامناسب در بستر خود جلوگیری کنند. بازبینی، صرفاً کرداری انسانی نیست و به دلیل حجم گسترده محتواهایی که هر روز در رسانه‌های اجتماعی به اشتراک گذاشته می‌شوند، فن‌آوری هوش مصنوعی و یادگیری ماشین در فرایند نظارت به خدمت گرفته شده است و با آموزش مداوم به عوامل مصنوعی تلاش می‌شود تا نقش آن‌ها در فرایند بازبینی توسعه پیدا کند. بازبینی محتوا گاهی کاملاً انسانی، گاهی کاملاً ماشینی و گاهی ترکیبی است. پلتفرم در تلاش هستند تا نوع بازبینی ماشینی را بیش‌ازپیش

توسعه دهند که دلایل آن برای نمونه، کاهش هزینه نیروی انسانی، حجم گسترده محتواها و ضرورت واکنش سریع به محتواهای مسئله‌دار برای تأمین ایمنی و امنیت کاربران است.

مقابله با اخبار جعلی به‌منزله نوعی محتوای زیان‌آور یا مسئله‌دار در پلتفرم‌های بزرگ رسانه‌های اجتماعی در قالب فرایند بازبینی محتوا صورت می‌گیرد که هم می‌تواند به‌وسیله عامل انسانی انجام شود و هم عامل ماشینی.

۳. روش پژوهش

روش‌شناسی این مقاله از نوع توصیفی - تحلیلی است. روش گردآوری در این مقاله، از نوع اسنادی است.

صادقی فسایی و عرفان‌منش (۱۳۹۴: ۶۹)، روش اسنادی را نوعی روشی کیفی محسوب کرده‌اند که طی آن، پژوهش‌گر می‌کوشد با کاربرد منظم داده‌های حاصل از اسناد به کشف، استخراج، طبقه‌بندی و ارزیابی مطالب مرتبط با پژوهش خود اقدام کند. آن‌ها مراحل اجرای روش اسنادی را به این شکل پیشنهاد کرده‌اند:

۱. انتخاب موضوع، تعیین اهداف و سؤالات.
 ۲. بررسی‌های اکتشافی و پیشینه پژوهش.
 ۳. انتخاب رویکرد نظری.
 ۴. جمع‌آوری منابع و تکنیک‌های بررسی منابع (تعیین واژگان کلیدی، مرور نظام‌مند، طبقه‌بندی مفاهیم و تقلیل داده‌ها).
 ۵. پردازش، نگارش و گزارش که شامل جمع‌بندی و ارائه نقطه‌نظر است.
- به‌عبارت دیگر، برای تحلیل و مطالعه اقدامات عملی و نیز استخراج رویکردهای کلان پلتفرم‌های فیس‌بوک و یوتیوب مطالعه اسناد و مدارک و سپس تحلیل و استخراج خط‌مشی‌ها به‌عنوان شیوه اصلی، به‌کار گرفته شده است.
- اسناد مورد مطالعه در این مقاله، طیف گسترده‌ای را دربرمی‌گیرند. برای انجام این مطالعه، انواع اسناد و دستورالعمل‌های خود این پلتفرم‌ها، مقررات مرتبط با حکم‌رانی

پلتفرم‌ها، شیوه‌نامه‌های بازبینی محتوا، مصاحبه‌ها شامل مصاحبه‌های مدیران پلتفرم‌ها گرفته تا مصاحبه‌های دولت‌مردان، منتقدان، کارشناسان و قانون‌گذاران، گزارش‌های فصلی، مقالات روزنامه‌ها، مقالات علمی، کتاب‌ها، اطلاعات و آمار و تمام داده‌های دیگر مرتبط با این موضوع گردآوری و تحلیل شده‌اند. علاوه بر این، هر کدام از این پلتفرم‌ها نیز به‌طور موردی مطالعه شده‌اند.

یک موضوع مهم که در مرحله تحلیل یافته‌ها در تحلیل اسنادی اهمیت دارد، تقلیل داده‌ها است که شامل کاربرد کدگذاری است که در نهایت، به رایۀ مقوله‌بندی‌های کلان منجر می‌شود (فلیک، ۱۳۸۸). با توجه به هدف و مسئله این مقاله، تحلیل یافته‌ها به‌شکل رایۀ مقوله‌های کلان و تحلیل و تفسیر آن‌ها انجام می‌شود. گردآوری مدارک برای نگارش این مقاله در نیمه دوم سال ۱۴۰۰ انجام شده است. با توجه به این‌که به‌روزرسانی یک موضوع محوری در تدوین دستورالعمل‌ها (همانند شرایط رایۀ خدمت) و روش‌های عملی کاربرد آن‌ها توسط پلتفرم‌ها است، احتمال دارد تاکنون تغییراتی در تعاریف، شیوه مقابله با محتواهای زیان‌آور و تغییر آن‌ها با ملاحظه قوانین ملی یا بین‌المللی مصوب جدید رخ داده باشد. این موضوع در عرصه پلتفرم‌ها امری رایج است.

۴. تحلیل یافته‌ها

همان‌طور که گفته شد، هدف این مقاله، تحلیل رویکرد و نیز سازوکارهای دو پلتفرم بزرگ رسانه اجتماعی جهانی در زمینه اخبار جعلی است. در بخش تحلیل یافته‌ها، ابتدا پلتفرم فیس‌بوک، سپس توئیتر ارزیابی می‌شوند. تحلیل یافته‌ها بر اساس تحلیل محتوای کیفی کلیه مدارک مطالعه شده مرتبط با موضوع بازبینی محتوا انجام شده است. این مقوله‌های محوری عبارتند از: ساختار اقتصادی و سیاسی امریکا به‌منزله عاملی تعیین‌کننده؛ شفافیت و اتصال به‌منزله اصول حاکم بر بازبینی محتوا در فیس‌بوک؛ بازبینی محتوای اخبار جعلی در فیس‌بوک و حراست از آزادی بیان؛ نحوه انجام راستی‌آزمایی محتواهای کذب در فیس‌بوک؛ سلامت گفت‌وگو به‌منزله اصل

حاکم بر بازبینی محتوا در توئیتر و مقابله توئیتر با اخبار جعلی و اطلاعات نادرست است. در ادامه، مقوله‌های مختلف و توضیحات مرتبط با آن‌ها در ارتباط با هر کدام از پلتفرم‌ها ارائه می‌شوند.

۴-۱. ساختار اقتصادی و سیاسی امریکا به منزله عاملی تعیین کننده

ساختار اقتصادی و سیاسی امریکا نقشی تعیین کننده در خط‌مشی کلی، رقابت پلتفرم‌ها و به‌طور خاص، شیوه نظارت بر محتوای کاربرساخته دارد. اصول اقتصادی سرمایه‌داری و بازار آزاد، تعیین کننده نحوه فعالیت شرکت‌ها است. بنابراین، اصول رقابت آزاد و جلوگیری از انحصار از موضوعات اساسی قانون‌گذاری و تنظیم‌گری پلتفرم‌ها در امریکا محسوب می‌شود.

همان‌طور که باچر (۲۰۲۱) به‌درستی اشاره کرده است، تنظیم مقررات برای پلتفرم‌ها در زمینه امریکا باید بر مبنای لیبرالیسم بازار درک شود که طی تصور می‌شود شرکت‌ها می‌توانند بدون مداخله دولت به خودتنظیم‌گری بپردازند که این خودبخشی از روند کلان شرکتی‌سازی همه‌امور در این کشور است.

از سوی دیگر، باید اشاره شد که حراست و حفاظت از آزادی بیان، یک اصل اساسی در مورد انتشار محتواهای رسانه‌ای در امریکا محسوب می‌شود. به‌طوری‌که بسیاری از تلاش‌ها برای تنظیم‌گری پلتفرم‌ها و حتی مقابله با محتواهای زیان‌آور با ملاحظه عدم خدشه در آزادی بیان شهروندان مورد توجه قرار می‌گیرد. اصل آزادی بیان به‌طور خاص در متمم اول قانون اساسی امریکا تضمین شده است که تاریخ ۱۵ دسامبر ۱۷۹۱ میلادی تصویب شد. این متمم یکی از ده متممی است که منشور حقوق ایالات متحده امریکا^{۱۲} را تشکیل می‌دهند. مؤسسان ایالات متحده امریکا در آن زمان با هدف حفاظت از آزادی‌های فردی و محدود کردن قدرت حکومت خودکامه، این متمم‌ها را به قانون اساسی این کشور تازه تأسیس اضافه کردند. متن متمم اول به این شرح است:

«کنگره در خصوص ایجاد مذهب، یا منع پیروی آزادانه از آن یا محدود

ساختن آزادی بیان یا مطبوعات یا حق مردم برای برپایی اجتماعات آرام و دادخواهی از حکومت برای جبران خسارت هیچ قانونی وضع نمی‌کند» (Congress.gov).

گرافاناکی (۲۰۱۸) اعتقاد دارد که متمم اول قانون اساسی آمریکا، عمده‌ترین مانع در برابر سانسور دولتی و نقض آزاد بیان کاربران در پلتفرم‌های آنلاین محسوب می‌شود. از سوی دیگر، بخش ۲۳۰ قانون نزاکت ارتباطات^{۱۸} سال ۱۹۹۶ میلادی را باید یکی از بهترین ابزارهای حفاظت از آزادی بیان و نوآوری در اینترنت دانست. طبق آن «هیچ کدام از تأمین‌کنندگان یا کاربران یک خدمت رایانه‌ای تعامل نباید به‌منزله ناشر یا گوینده هر نوع اطلاعاتی که توسط تأمین‌کننده دیگر محتوای ارتباطی ارائه شده است قلمداد شوند» (EFF, 2022).

بنابراین در زمینه اجتماعی آمریکا که بزرگ‌ترین پلتفرم‌های جهان در آن بنیان نهاده شده و توانسته‌اند اکوسیستمی جهانی شکل دهند، ملزم به حفاظت از آزادی بیان کاربران هستند. به‌همین دلیل، هر نوع مداخله حتی برای حذف محتواهای غیراخلاقی، با مناقشه‌های گسترده‌ای در سطح افکار عمومی این کشور همراه است. به‌همین دلیل، شرکت‌های پلتفرمی همواره روی لبه تیغ نقض آزادی بیان کاربران و نیز حفاظت از سلامت محیط تعامل کاربران حرکت کنند.

۴-۲. شفافیت و اتصال به‌منزله اصول حاکم بر بازبینی محتوا در فیس‌بوک
فیس‌بوک در حال حاضر بزرگ‌ترین رسانه اجتماعی در جهان است. فیس‌بوک (متا) خود را یک شرکت فن‌آوری (Bucher, 2021) معرفی می‌کند و نه یک شرکت رسانه‌ای؛ تا از این طریق، از مسئولیت‌های حقوقی مرتبط با آن شانه خالی کند. گشوده‌تر و متصل کردن بیش از پیش جهان، مأموریت و آرمان فیس‌بوک است که گرچه ظاهراً با هدف گسترش فهم و درک مشترک میان انسان‌ها پیگیری می‌شود (فن‌دایک، ۱۳۹۶).

در بخش «درباره» فیس‌بوک، مأموریت فیس‌بوک، «قدرت‌مند کردن مردم برای ساختن اجتماع و نزدیک‌تر کردن جهان با هم ذکر شده است». در فیس‌بوک، هر

روز بیش از صد میلیارد پیام به اشتراک گذاشته می‌شود که به افراد کمک می‌کند تا حتی وقتی از هم دور هستند نزدیک به هم بمانند. هر روز بیش از یک میلیارد استوری (روایت) با دیگران به اشتراک گذاشته می‌شود که به افراد کمک می‌کند تا خود را بیان و با دیگران پیوند برقرار کنند. همچنین در فیس‌بوک بر حل مسئله به شکل جمعی و همکاری برای متصل کردن افراد در سرتاسر جهان تأکید شده است (About Face- book, 2021).

گرچه دلالت‌های ضمنی این رویکرد به این معنا است که معماری فن‌آورانه و اقتصادی فیس‌بوک سبب برساخت پیکربندی خاصی از روابط اجتماعی و هنجارهای اجتماعی می‌شود (Bucher, 2021). برای نمونه، گرایش کاربران به افشای بیش از پیش داده‌های شخصی و بازنمایی زندگی خصوصی که توسط فیس‌بوک شکل داده می‌شود، پیش از هر چیز در خدمت تولید انبوه کلان داده برای این شرکت است. بخش عمده‌ای از سود این شرکت حاصل تحلیل این کلان داده‌ها است.

به عبارت دیگر، فیس‌بوک در پس آرمان شفافیت و بسط ارتباط، کسب بیش‌ترین سود مادی از امکان نقض حریم خصوصی کاربران و گردآوری آن‌ها به شکل داده و آنالیتیکس آن‌ها با هدف پیش‌برد مدل تجاری این شرکت است. جذب آگهی و آگهی‌دهندگان فلسفه محوری فیس‌بوک است و بنابراین تا حد امکان تلاش می‌کند تا این پلتفرم به منزله فضای بازی برای جریان سریع و بی‌پایان محتواها عمل کند و هر نوع مانع تحمیلی بیرون با هر هدفی همچون حفاظت از اجتماع و داده‌های کاربران را به شیوه‌های مختلف و به‌ظاهر قانونی دور می‌زند.

داده‌های کاربران و انواع گوناگون آن‌ها، دارایی‌های ارزشمندی برای فیس‌بوک به‌شمار می‌رود که اکنون متا نامیده می‌شود. همان‌طور که لیور و همکاران (۱۴۰۰) در بحث از اینستاگرام مطرح کرده‌اند، فیس‌بوک پس از خرید اینستاگرام بلافاصله تمام داده‌های کاربران را در اختیار خود گرفت که هم شامل داده‌ها-اطلاعات مرتبط با آن عکس و نیز فراداده‌ها است، یعنی اطلاعات درباره آن عکس که توسط دستگاه‌های گوشی تلفن همراه ثبت می‌شوند. در نتیجه، این اطلاعات می‌تواند مورد استفاده مفید

و همین‌طور غیرمفید قرار گیرد که نمونه آن، جنجال کمبریج آنالیتیکا بود.

۴-۳. بازبینی محتوای اخبار جعلی در فیس‌بوک و حراست از آزادی بیان

در فیس‌بوک نیز همانند سایر پلتفرم‌ها، بازبینی محتوا به کمک عوامل انسانی و ماشینی صورت می‌گیرد و فیس‌بوک تلاش می‌کند تا با کاربرد هم‌زمان این دو دسته از عوامل فیس‌بوک را به‌مکانی ایمن و به‌دور از محتوای زیان‌آور تبدیل کند. به‌گفتهٔ سینگ^{۱۹} (۲۰۱۹) بزرگ‌ترین اقدامات مرتبط با بازبینی محتوا در فیس‌بوک انجام می‌شود و علاوه بر این، بیش‌ترین میزان تحقیق و بررسی در مورد کردارها و تصمیم‌گیری‌ها در مورد بازبینی محتوا مرتبط با فیس‌بوک است. یکی از دلایل آن، این است که این پلتفرم بزرگ‌ترین رسانهٔ اجتماعی در جهان است. فیس‌بوک از ابزارهای پایش و رصد محتوا در مرحلهٔ پیش از انتشار بهره می‌برد؛ به‌طوری‌که همهٔ محتوای کاربر ساخته به‌محض اشتراک‌گذاری یا پست شدن در این شبکه، در طی یک فرایند خودکار بازبینی می‌شوند. این پلتفرم از روش بازبینی محتوا به‌روش پس از رخداد نیز استفاده می‌کند و از مجموعه‌ای از الگوریتم‌های متفاوت برای شناسایی و نظارت بر محتوای قابل اعتراض نیز استفاده می‌کند.

پل و همکاران (۲۰۲۱) حکم‌رانی توسط پلتفرم‌ها را به‌سه دسته اصلی تقسیم کرده‌اند:

۱. تنظیم‌گری^{۲۰} که شامل تعیین استانداردها، اصول راهنما یا دستورالعمل‌ها و خط‌مشی‌ها است.

۲. ساماندهی^{۲۱} که شامل مقوله‌بندی و نظم‌بخشی به محتوا و خدمات است.

۳. بازبینی که شامل اعمال حکم‌رانی توسط پلتفرم‌ها است. در فیس‌بوک، آن‌چه که از نظر این پژوهش‌گران مراحل تنظیم‌گری نامیده می‌شود، شرایط آرایه خدمت، خط‌مشی داده و استانداردهای^{۲۲} اجتماع است.

مهم‌ترین سند یا دستورالعملی که بازبینی محتوا در فیس‌بوک را هدایت می‌کند، استانداردهای اجتماع است. به‌طور کلی، هدف استانداردهای اجتماعی مشخص کردن

محتواهای مجاز و غیرمجاز در فیس‌بوک است. فیس‌بوک بنا بر ادعای خود به دنبال ایجاد فضایی برای خودابرازی و دادن صدا به مردم است. این موضوع مبتنی بر ارزش‌هایی همانند سندیت، ایمنی، منزلت و حریم خصوصی است که در تدوین استانداردها یا معیارهایی برای اجتماع کاربران فیس‌بوک به‌منزلهٔ تابلو راهنما عمل می‌کند. بنابراین، مرحلهٔ بازبینی شامل کاربرد اسناد و مجموعه دستورالعمل‌های فوق برای مواجهه و مقابله با انتشار اخبار جعلی و اطلاعات نادرست است. استانداردهای اجتماعی فیس‌بوک در واقع مجموعه‌ای از باید‌ها و نباید‌ها است. در مجموع این استانداردها شامل شش دستهٔ کلی خشونت و رفتار جنایی، ایمنی، محتوای ناخوشایند،^{۲۳} درست‌کاری و سندیت،^{۲۴} احترام به مالکیت معنوی و درخواست‌ها و تصمیمات مرتبط با محتوا است. فیس‌بوک اخبار جعلی را در دستهٔ هم‌بستگی و صحت، طبقه‌بندی کرده است.

پیش از بحث دربارهٔ این که، فیس‌بوک از چه رویه‌ها و چه ابزارهایی جهت انتشار اخبار جعلی و انواع اطلاعات نابهنجار در پلتفرم و زیرمجموعه‌های خود استفاده می‌کند، توجه به رویکرد کلی فیس‌بوک به مقولهٔ اخبار جعلی سبب می‌شود نحوهٔ نگرش کلی مالکان آن به این مقوله خطیر آشکار شود.

فیس‌بوک در بخش مرکز شفافیت خود و بخش استانداردهای اجتماعی، سیاست یا خط‌مشی خود دربارهٔ اخبار کذب^{۲۵} را توضیح داده است. فیس‌بوک «کاهش گسترش اخبار کذب» در فیس‌بوک را یک «مسئولیت» برای خود بیان می‌کند. این پلتفرم کمک به مطلع بودن مردم، بدون ایجاد اختلال در گفت‌وگو عمومی و نیز وجود مرزی باریک بین مطالب هجوآمیز و عقاید افکار را عمده‌ترین سیاست فیس‌بوک دربارهٔ اخبار کاذب بیان می‌کند که همان کاستن میزان این نوع محتواها است. روش فیس‌بوک جلوگیری از انتشار گسترده آن‌ها در بخش خوراک خبری آن است (Facebook, 2021).

بنابراین رویکرد فیس‌بوک که خود را متعهد به ایجاد فضای گفت‌وگو و اطلاع‌رسانی شفاف می‌داند، محدودسازی میزان انتشار و کاهش میزان دیده شدن این نوع محتواهاست و نه حذف آن‌ها. این پلتفرم خود را متعهد می‌داند تا از آزادی بیان افراد

حتی اگر بیان به شکل تولید و اشتراک گذاری محتوای کذب باشد، پاسداری کند. بنابراین، به وسیله استناد به آزادی بیان و یا به عبارت دیگر، دستاویز قرار دادن آن، مسئولیت اجتماعی فیس بوک به جای حذف و مسدودسازی به محدودسازی تقلیل پیدا می کند. بخشی از این اقدامات محدودسازانه شامل موارد زیر است:

- مختل کردن انگیزه های اقتصادی برای افراد، صفحه ها و دامنه هایی که اطلاعات نادرست را ترویج می کنند.

- استفاده از علائم مختلف شامل بازخوردهای دریافتی از اجتماع کاربران فیس بوک برای پیش برد یک مدل یادگیری ماشینی که پیش بینی می کند کدام گزارش ها نادرست هستند.

- کاهش انتشار محتواهایی که شرکت های ثالث مستقل راستی آزمایی تشخیص داده اند که آن محتواها کذب هستند.

- توان مندسازی مردم برای تصمیم گیری برای خودشان برای این که چه چیزی را بخوانند، به چه چیزی اعتماد کنند و هم رسانی کنند که این کار از طریق آگاهی بخشی به آن ها با ارایه زمینه بیش تر درباره خبر و ترویج سواد خبری صورت می گیرد.

- همکاری با دانشگاهیان و سایر سازمان ها برای کمک به حل این مسئله چالش برانگیز (Ibid).

به عبارت دیگر، همان طور که لیونز^{۲۶} مدیر محصولات فیس بوک، سه راهبرد فیس بوک را خلاصه وار بیان کرده است: حذف حساب های کاربری و محتواهایی که از استانداردهای اجتماعی یا سیاست های آگهی فیس بوک تخطی می کنند؛ کاهش انتشار اخبار جعلی و محتوای نامستند؛ و آگاهی بخشی به مردم با ارایه زمینه بیش تر به آن ها درباره پست هایی که آن ها مشاهده می کنند.

آدام ماسری^{۲۷} (۲۰۱۷) یکی از مدیران وقت متا و رئیس کنونی اینستاگرام (زیرمجموعه متا)، می نویسد: فیس بوک در سه حوزه کلیدی با گسترش اخبار جعلی مقابله می کند که شامل ایجاد اختلال در انگیزه های اقتصادی برای انتشار اخبار کذب، تولید محصولات جدید و کمک به مردم برای تصمیم گیری آگاهانه تر هنگام مواجه با

اخبار جعلی است. انتظار می‌رود برخی اقدامات همانند همکاری با اجتماع کاربران و سازمان‌های طرف سوم که راستی‌آزمایی مطالب را انجام می‌دهند، سرعت انتشار اطلاعات نادرست کاهش یابد و در عوض، سبب کاهش صرفه اقتصادی این نوع محتواها از جمله ناتوانی در جذب آگهی شود. بهبود رتبه‌بندی مطالب منتشر شده بر اساس مشاهده میزان تمایل عمومی به اشتراک آن، آسان‌تر کردن گزارش محتوای کذب توسط اعضای اجتماع کاربران فیس‌بوک و همکاری با شرکت‌های راستی‌آزمایی اقدامات جدیدی است که فیس‌بوک ارائه کرده است.

همچنین برخی برنامه‌ها همانند پروژه روزنامه‌نگاری فیس‌بوک و همکاری با پروژه سواد خبر^{۲۸} از جمله اقداماتی است که فیس‌بوک به واسطه آن‌ها می‌خواهد سواد خبری کاربران را ارتقاء دهد.

برنامه راستی‌آزمایی فیس‌بوک از سال ۲۰۱۶ برای کارافزار فیس‌بوک راه‌اندازی شد و در سال ۲۰۱۹ به اینستاگرام توسعه یافت. هدف آن، شناسایی و نیز برخورد با اطلاعات نادرستی بود که به شکل ویروسی منتشر می‌شدند. تیم‌ها و شرکت‌های همکار فیس‌بوک عمدتاً بر ادعاهای کذب و به‌ویژه مواردی که به موضوعات داغ روز و مهم تبدیل می‌شوند، تمرکز دارند.

بر اساس اولویت‌های راستی‌آزمایی که در ۲۹ ژوئیه ۲۰۲۱ به روزرسانی شده است،^{۲۹} محتواهای کاربرساخته همانند آگهی‌ها، نوشته‌ها، عکس‌ها، ویدئوها و محتواهایی را که صرفاً متنی هستند بازبینی و رده‌بندی می‌کنند. محتواهای گول‌زننده^{۳۰} مبنای واقعیت داشتن آن‌ها مشخص نیست، محتوایی که به‌منزله عقیده ارائه می‌شود، اما بر اساس اطلاعات نادرست است، ادعایی که کذب بودن آن‌ها قابل اثبات است و به‌ویژه موارد پرتوجه روز در اولویت راستی‌آزمایی توسط فیس‌بوک قرار دارند. اما در مقابل محتواهایی که شامل ادعای قابل اثبات نیستند، محتواهایی که در زمان نوشتن درست نیست، عقاید و سخنرانی‌های سیاست‌مداران و نیز محتوایی که شامل عقاید عمومی باشد تا زمانی که سبب انتشار اطلاعات کاذب نباشند، توسط فیس‌بوک در معرض راستی‌آزمایی قرار نمی‌گیرند.

۴-۴. نحوه انجام راستی‌آزمایی^{۳۱} محتواهای کذب در فیس‌بوک

بر اساس نوشته فیس‌بوک در مورد شیوع عملکرد واری و واقعیت یا راستی‌آزمایی که آخرین بار در ۲۹ ژوئیه ۲۰۲۱ به‌روزرسانی شده است، افرادی که عملیات راستی‌آزمایی را انجام می‌دهند، مستقل از فیس‌بوک هستند و شبکه بین‌المللی راستی‌آزمایی بین‌المللی^{۳۲} آن‌ها را تأیید کرده‌اند.^{۳۳} راستی‌آزمایی در فیس‌بوک با همکاری اعضای این شبکه صورت می‌گیرد. وظیفه اعضای این شبکه، تمرکز بر مشروعیت، صحت و دقت اطلاعات و وظیفه فیس‌بوک آگاه کردن مردم در مورد محتواهایی است که به‌عنوان کذب رده‌بندی شده‌اند.

راستی‌آزمایی در فیس‌بوک، دارای مراحل مختلفی است. اولین مرحله، شناسایی اطلاعات است. فن‌آوری فیس‌بوک به‌روش‌های گوناگون اقدام به شناسایی محتواهای کاربرساخته جعلی یا کذب می‌کند. نحوه واکنش افراد و سرعت انتشار یک مطلب یکی از این نشانه‌ها است. نشان‌دار کردن^{۳۴} یا گزارش دادن بخشی از محتوا، سبب می‌شود تا افراد متولی راستی‌آزمایی با دقت بیشتری به آن توجه کنند. علاوه بر این، کامنت‌های زیر پست که نشان‌دهنده عدم باور آن مطلب باشند، یادگیری ماشینی و نیز شناسایی محتوا توسط خود راستی‌آزمایان روش‌های دیگر شناسایی است. سپس مرحله بازبینی محتوا روی می‌دهد که طی آن افراد راستی‌آزمای محتوا را بازبینی و میزان دقت آن را رده‌بندی می‌کنند. در این مرحله که به‌شکل مستقل از فیس‌بوک صورت می‌گیرد، اقدامات گوناگونی از جمله برقراری تماس با منبع، مشاوره گرفتن از داده‌های عمومی و تعیین موثق بودن عکس‌ها و ویدئوها انجام می‌شود. مرحله بعد شامل برچسب زدن به‌محتوای کذب و مطلع کردن افراد از آن است.

به‌طور کلی، سیاست فیس‌بوک در برابر محتواهای کذب، اطمینان از این است که افراد کم‌تری آن اطلاعات را مشاهده می‌کنند. پس از آن که محتوایی به‌منزله کذب برچسب خورد، در خوراک خبری فیس‌بوک در رده‌های پایین‌تری ظاهر می‌شوند. در اینستاگرام که زیرمجموعه شرکت متا قرار دارد، محتوای مورد نظر در بخش کاوش^{۳۵} فیلتر می‌شود و از میزان برجستگی آن در قسمت‌های فید و استوری به‌شکل قابل

ملاحظه‌ای کاسته می‌شود. این اقدام سبب می‌شود که تعداد افرادی که آن مطلب را ملاحظه می‌کنند، به میزان زیادی کاهش یابد. در صورت تکرار انتشار مطالب کذب توسط صفحه‌ها، گروه‌ها، وبسایت‌ها و نیز حساب‌های کاربری در اینستاگرام برخی مدت‌ها برای مدت زمان معین علیه آن‌ها انجام می‌شود. از جمله، حذف از قسمت پیشنهادها، کاهش میزان توزیع آن‌ها، جلوگیری از توانایی درآمدزایی و آگهی‌دهی و نیز حذف احتمال ثبت آن‌ها به منزلهٔ یک صفحهٔ خبری^{۳۶} است.

۴-۵. سلامت گفت‌وگو به منزلهٔ اصل حاکم بر بازبینی محتوا در توئیتر

توئیتر، امروزه یکی از مهم‌ترین رسانه‌های اجتماعی به‌شمار می‌رود. توئیتر در آغاز به‌منزلهٔ نوع وبلاگ‌نویسی خرد^{۳۷} تلقی می‌شد، اما اکنون خود به یک پلتفرم بزرگ تبدیل شده است. صفحهٔ سیاست‌مداران در کشورهای گوناگون در توئیتر، به یک عرصهٔ بااهمیت در عرصهٔ دیپلماسی و بیان سیاست‌ها و موضوع‌گیری‌های دولت‌ها تبدیل شده است. این پلتفرم بیش‌تر ماهیتی خبری و اطلاع‌رسانی دارد. ازدیاد انتشار اخبار جعلی و اطلاعات نادرست در عصر پسا حقیقت، نحوهٔ بازبینی محتوا در این پلتفرم را حایز اهمیت کرده است.

توئیتر به‌عنوان یکی از بزرگ‌ترین و محبوب‌ترین وبسایت‌های رسانه‌های اجتماعی به‌کاربران امکان می‌دهد تا یک «مکالمه» عمومی وب‌مبنای هم‌زمان را از طریق ارسال پیام که دارای محدودیت ۱۴۰ کاراکتر است (البته از نوامبر ۲۰۱۸ به ۲۸۰ افزایش یافت) انجام دهند. امکان ارسال پیام از طریق خود وبسایت یوتیوب و نیز تلفن‌های همراه و ابزارهای موبایل از جمله تبلت و ساعت نیز وجود دارد. هدف توئیتر این است که کاربران در قالب تعداد مشخصی از کاراکتر به این پرسش که «چه چیزی در حال وقوع است؟» پاسخ بدهند (Murthy, 2018).

توئیتر به‌جای آن‌که خود را واسطه یا میزبانی برای تبادل محتوا و یا اطلاع‌رسانی بداند، اصل و اساس فعالیت خود را مبتنی بر فراهم کردن امکان مکالمه معرفی می‌کند. مکالمه یا گفت‌وگو نشان‌گر تأکید این پلتفرم بر عمومی بودن این فضا است. به

این معنا توئیتر می‌خواهد فضایی را فراهم کند که همه شهروندان در سطوح گوناگون در آن بتوانند با هم صحبت کنند. به عبارت دیگر، توئیتر بستری را فراهم کند که افراد گفت‌وگوهای روزمره خود را از فضای واقعی به فضای توئیتر منتقل کنند و بنابراین تلاش می‌کند تا فضایی سالم را برای شکل‌گیری گفت‌وگو و شنود فراهم کند.

در قسمت درباره وبسایت توئیتر چنین نوشته شده است: «توئیتر چیزی است که هم‌اکنون در حال وقوع است و چیزی است که مردم هم‌اکنون درباره آن صحبت می‌کنند». توئیتر معتقد است که «در خدمت مکالمه همگانی» است و به همین دلیل، برخورداری مردم از «فضایی از آزاد و ایمن» برای این پلتفرم اهمیت دارد و خود را متعهد می‌داند تا از «سلامت مکالمه همگانی» که آن را موضوعی جدی می‌داند حفاظت کند (Twitter, 2020).

بنابراین برای این که افراد بتوانند گفت‌وگوهای زمان واقعی خود را به توئیتر منتقل کنند، آزادی و سلامت مکالمه، یک اصل اساسی است. آزادی به این معنا است که هیچ قدرتی نتواند مانع از آزادی بیان شود و منظور از سلامت نیز عدم اشاعه مطالبی است که در روال جاری گفت‌وگوها خدشه وارد کند. سه سند یا مقررات‌نامه کلی که نحوه انتشار محتوا و رفتار در این پلتفرم را تعیین می‌کنند عبارت‌اند از: شرایط رایانه خدمت، خط‌مشی خرید خصوصی و نیز قواعد و خط‌مشی‌های توئیتر. گرچه در مورد اخبار جعلی برخی قواعد و دستورالعمل‌های مرتبط دیگر وجود دارد که در ادامه، آن‌ها را مرور خواهیم کرد.

۴-۶. مقابله توئیتر با اخبار جعلی و اطلاعات نادرست

همان‌طور که اشاره شد، پلتفرم توئیتر خود را متعهد به «سلامت مکالمه عمومی» می‌داند که هدف آن، ایجاد فرصت گفت‌وگو است. در توئیتر به جای عبارت اطلاعات نادرست یا اخبار جعلی از عبارت «اطلاعات گمراه‌کننده»^{۳۸} استفاده می‌شود و تنها در صورتی این پلتفرم در روند مکالمه مداخله می‌کند که محتواهای کاربرساخته از این نوع قواعد توئیتر را نقض کنند (Twitter Help Center, 2020a).

برخی از خط‌مشی‌های مجزای توئیتر دربارهٔ اطلاعات گمراه‌کننده عبارت‌اند از: خط‌مشی اطلاعات نادرست در شرایط بحران،^{۳۹} خط‌مشی اطلاعات گمراه‌کننده مرتبط با کووید ۱۹، خط‌مشی رسانه‌های دست‌کاری‌شده و ساختگی^{۴۰} و خط‌مشی یکپارچگی مدنی.^{۴۱}

در توئیتر محتوای گمراه‌کننده یا «اطلاعات نادرست» شامل ادعاهایی است که کارشناسان متخصص خارج از این پلتفرم کذب بودن آن را تأیید کنند و یا این به‌روش‌های فریبنده و یا گیج‌کننده منتشر شوند (Twitter Help Center, 2022b). توئیتر طیفی از اقدامات مختلف را در سطح پلتفرم برای مقابله با محتواها و نیز رفتارهای ناقض مکالمهٔ سالم در توئیتر در سطح مختلف انجام می‌دهد. سطح نخست در سطح توئییت است. برخی اقدامات در سطح توئییت عبارت‌اند از: برچسب زدن به توئییتی که احتمالاً حاوی اطلاعات مورد مناقشه یا گمراه‌کننده است. در صورتی که توئیتر به این نتیجه برسد که توئییتی دارای بر اساس خط‌مشی آن پلتفرم، دارای اطلاعات گمراه‌کننده یا مناقشه‌برانگیز است، برچسبی به آن محتوا افزوده می‌شود (Twitter Help Center, 2022c).

مناقشه و مجادلهٔ میان دونالد ترامپ و توئیتر پس از اعلام نتایج انتخابات ریاست جمهوری سال ۲۰۲۰ در امریکا، یکی از موارد عمده‌ای بود که طی آن توئیتر از این شیوه علیه ترامپ استفاده کرد. این اقدام چالش‌های زیادی را در سطح امریکا از منظر سیاسی و نیز احترام یا تهدید آزادی بیان در این کشور و همین‌طور مجادلهٔ میان دولت فدرال امریکا و شرکت‌های پلتفرمی ایجاد کرد. بنابر گزارشی که نیویورک تایمز در نوامبر ۲۰۲۰ منتشر شده است، توئیتر طی مدت زمان کوتاه به ۳۸ درصد از ۲۹ مورد از توئییت‌های ترامپ برچسب زده بود. ترامپ متهم شده بود که در مورد فرایند انتخابات ریاست جمهوری امریکا اطلاعات گمراه‌کننده را منتشر کرده است (Conger, 2020).

علاوه بر روش برچسب زدن به توئییت، این پلتفرم روش‌های دیگری را در سطح توئییت انجام می‌دهد که شامل محدود کردن دیده شدن یک توئییت، الزام به حذف

توثیت، پنهان کردن توییت نقض کننده در حال انتظار برای حذف آن و نیز تذکر^{۴۲} است که طی موارد استثنایی و در راستای منفعت عمومی، توییت این امکان را ایجاد می کند تا برخی توییت ها با وجود نقض خطمشی های این رسانه اجتماعی همچنان در دسترس باشند.

البته در توییت مرحله تجدیدنظر نیز پیش بینی شده است. یک مرحله مهم پس از تمامی این اقدامات، امکان درخواست تجدیدنظر درباره یک توییت حذف شده است. توییت این امکان را فراهم کرده است تا نسبت به حذف و تعلیق توییت حذف خود اعتراض کنند تا مورد بازبینی قرار گیرد. کاربر باید توضیح دهد که چرا از نظر او آن توییت حذف شده، ناقض قواعد توییت نیست (Twitter, 2022).

هریک از خطمشی های فوق که مبنای توییت برای مقابله با اطلاعات گمراه کننده است، جزئیات دقیقی دارد که راهنمای عمل فوق العاده دشوار حکم رانی و بازبینی محتوای اخبار کذب است. این دستورالعمل ها راهنمای بازبینان انسانی و ماشینی در عصر پساحقیقت است.

نتیجه

در عصر پساحقیقت، گسترش اخبار جعلی و اطلاعات نادرست و کذب یکی از دغدغه های اساسی هم برای شهروندان و هم برای دولت مردان محسوب می شود. گسترش انتشار اخبار جعلی که کاربرد برنامه های نرم افزاری به نام بات در رسانه های اجتماعی سبب پیچیدگی انتشار و دشواری مقابله با آن شده است، می تواند دست یابی به اطلاعات و اخبار واقعی را به مخاطره بیاندازد. امروزه رسانه های اجتماعی زمینه اصلی تولید و اشتراک گذاری اخبار جعلی محسوب می شوند. وابستگی کاربران جهانی به برخی پلتفرم های برتر همانند فیس بوک و توییت و انبوه کاربران که برخی از آن ها نقشی محوری در انتشار اخبار جعلی دارند، به یک معضل جهانی تبدیل شده است. موقعیت جهانی و برون سرزمینی رسانه های اجتماعی بزرگ سبب شده است تا دولت ها در اعمال حاکمیت و نیز کاربرد خطمشی های سنتی انتشار اخبار حقیقی ناتوان شوند.

در نتیجه، دولت‌ها به‌ناگزیر به‌خط‌مشی‌ها، سازوکارها و رویه‌های تدبیرشده توسط صاحبان پلتفرم‌های بزرگ امریکایی تن داده‌اند. از سوی دیگر، پلتفرم‌های امریکایی نیز در یک زمینه اقتصادی سیاسی خاص رشد و توسعه پیدا کرده‌اند که رشد بی‌وقفه و رقابت برای کسب سود به‌منزله بنیان اصلی نظام سرمایه‌داری، راهبر اساسی آن محسوب می‌شود. از طرف دیگر، حفاظت و حراست از آزادی بیان سبب شده تا مداخله دولت فدرال در عملکرد پلتفرم‌ها به‌حد اقل برسد. اصل آزادی بیان حتی در فرایند مقابله با محتوای زیان‌آور نیز یک چالش عمده محسوب می‌شود. زیرا شرکت‌های امریکایی را در معرض اتهام نقض آزادی بیان قرار می‌دهد. همان‌طور که مشخص است، مقابله با اخبار جعلی که اطلاعات کذب که موجب خدشه‌دار شدن اطلاع‌رسانی واقعی و فرایندهای دموکراتیک می‌شوند، تابع خط‌مشی‌های کلان شرکت‌های امریکایی شده است که آن‌ها خود را موظف می‌دانند تابع اصول فلسفه سیاسی امریکایی باشند. شرکتی شدن حکمرانی، یکی از پیامدهای ناگزیر توسعه اکوسیستم پلتفرمی جهانی است و بنابراین دولت‌ها، به‌ویژه دولت‌هایی همانند دولت ایران با بازیگران جدیدی در عرصه روابط بین‌الملل مواجه می‌شوند که اصولاً شرکت‌ها هستند و نه دولت‌ها. برخی از دولت‌های محدود، همانند چین، روسیه و آلمان تلاش کرده‌اند تا به‌روش‌های گوناگون همانند ایجاد اکوسیستم پلتفرمی جایگزین، همانند چین، یا تصویب قوانین الزام‌آور برای پلتفرم‌های جهانی، همانند قانون اجرای شبکه آلمان، با انتشار اخبار جعلی مقابله کنند. هر کدام از این رویکردها، تابع سیاست کلان فضای مجازی این کشورها، ظرفیت‌های فنی و اقتصادی آن‌ها، نوع رویکرد برای مواجهه با سلطه پلتفرم‌های بزرگ جهانی و تفاوت ارزشی با ارزش‌های حاکم بر اکوسیستم پلتفرمی مسلط کنونی امریکایی است. اتخاذ هر کدام از این رویکردها که شامل برخی پیش‌نیازها و الزامات از منظر توان فنی و جایگاه بین‌المللی است، بر نحوه مقابله با انتشار اخبار جعلی همانند سایر محتوای زیان‌آور و ناسالم تأثیر محوری دارد. گرچه پلتفرم‌های محلی ایرانی می‌توانند سازوکارهای پلتفرم‌های بزرگ را درون معماری خود تعبیه کنند و از آن‌ها الگوبرداری کنند، اما با توجه تعداد اندک کاربران پلتفرم‌ها و پیام‌رسان‌های

ایرانی و حضور فراگیر کاربران ایرانی در پلتفرم‌های بزرگی همانند اینستاگرام، مقابله با مخاطرات فضای مجازی از جمله اخبار جعلی مستلزم تعیین سیاست کلی فضای مجازی توسط نهادهای متولی بر اساس واقعیت‌ها است.

پی‌نوشت‌ها

1. pre-moderation
2. harmful
3. disinformation
4. community standards
5. Digital Services Act
6. misinformation
7. content moderation
8. user generated content
9. McIntyre
10. Giusti and Piras
11. information disorder
12. Wardle and Derakhshan
13. governance by platforms
14. Barret
15. Roberts
16. TOS
17. United States Bill of Rights
18. Section 230 of the Communications Decency Act
19. Singh
20. regulation
21. Curation
22. community standards
23. objectionable content
24. integrity and authenticity
25. false news
26. Lyons
27. Mossery
28. News Literacy Project



پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی

29. <https://transparency.fb.com/features/content-fact-checkers-prioritize/>

30. haaxes

31. <https://transparency.fb.com/features/how-fact-checking-works/>

32. International Fact-Checking Network (IFCN)

۳۳. شبکه بین‌المللی راستی‌آزمایی وابسته به مؤسسه پوینتر است که هدف آن، گردآوردن متولیان راستی‌آزمایی در سطح بین‌المللی است. مؤسسه مطالعات رسانه پوینتر یک مدرسه روزنامه‌نگاری و یک سازمان پژوهشی واقع در سن پترزبورگ فلوریدای امریکا است.

34. flagging

35. Explore

36. news page

37. microblogging

38. misleading information

39. crisis misinformation policy

40. synthetic and manipulated media policy

41. civic integrity policy

42. notiice

منابع

آزادی، محمدحسین و نیک‌گو، سعید. (۱۳۹۸). «حباب فیلتر و شخصی‌سازی اطلاعات در رسانه‌های اجتماعی». *رسانه*. ۳۰(۴). ۱۱۰-۱۲۸.

فلیک، اوه. (۱۳۸۸). *درآمدی بر تحقیق کیفی*. ترجمه هادی جلیلی. تهران: نی.

فن‌دایک، یوزه. (۱۳۹۶). *فرهنگ اتصال: تاریخ انتقادی رسانه‌های اجتماعی*. ترجمه

حسین حسنی. تهران: سوره مهر.

صادقی فسایی، سهیلا و عرفان‌منش، ایمان. (۱۳۹۴). «مبانی روش‌شناختی پژوهش

اسنادی در علوم اجتماعی؛ مورد مطالعه: تأثیرات مدرن شدن بر خانواده ایرانی».

راهبرد فرهنگ. ۸(۲۹). ۶۱-۹۱.

لیور، تاما و همکاران. (۱۴۰۰). *اینستاگرام: فرهنگ‌های رسانه‌های اجتماعی دیداری*.

ترجمه محمد مهدی وحیدی. تهران: سوره مهر.

About Twitter. (2021). "Our company". Twitter website. [https://about.](https://about.twitter.com/en/who-we-are/our-company)

[twitter.com/en/who-we-are/our-company](https://about.twitter.com/en/who-we-are/our-company).

- Bucher, Taina. (2021). *Facebook*. Cambridge : Polity.
- Chan, K. and Casert, R. (2020). "EU law targets Big Tech over hate speech, disinformation". *ABC NEWS*. <https://abcnews.go.com/Business/wireStory/eu-law-targets-big-tech-hate-speech-disinformation-84261130>.
- Conger, K. (2020). "Twitter has labeled 38% of Trump's tweets since tuesday". *Nytimes.com*. <https://www.nytimes.com/2020/11/05/technology/donald-trump-twitter.html>.
- US Congress*. (2022). Constitution of the United States. [constitution.congress.gov](https://constitution.congress.gov/constitution/amendment-1/).
- EFF. (2022). *CDA 230: The Most Important Law Protecting Internet Speech*. EFF.org. <https://www.eff.org/issues/cda230#:~:text=Section%20230%20says%20that%20%22No,%C2%A7%20230>.
- Facebook*. (2021). "Facebook Community Standards: False News Policy Rationale". Facebook Transparency Center. <https://transparency.fb.com/en-gb/policies/community-standards/false-news/>.
- Gillespie, T. (2018). *Custodians of The Internet: platforms, content moderation, and the hidden decisions that shape social media*. New Haven and London : Yale University Press.
- Giusti, S. and Piras, E. (2021). *Introduction : In search of paradigms : Disinformation, fake news, and post-truth politics; Democracy and Fake news: Information Manipulation and Post-Truth Politics*. Edited by Serena Giusti and Elisa Piras; edited by Serena Giusti and Elisa Piras, London and Paris : Routledge.

- Grafanaki, S. (2018). "Platforms, the First Amendment and Online Speech : Regulating the Filters". *Pace Law Review*. 39(1).
- Mcintyre, L. (2018). *Post-truth*. Cambridge and London : MIT Press.
- Mossery, A. (2017). "Working to Stop Misinformation and False News", *Meta Website*. <https://www.facebook.com/formedia/blog/working-to-stop-misinformation-and-false-news>.
- Murthy, D. (2018). *Twitter: Social Communication in the Twitter Age*. Cambridge and Medford : Polity.
- Poell, T. et al. (2021). *Platforms and Cultural Production*. Cambridge and Medford : Polity.
- Singh, S. (2019). "Everything in Moderation : An Analysis of How Internet Platforms Are Using Artificial Intelligence to Moderate User Generated Content". *Open Technology Institute*. <https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/the-limitations-of-automated-tools-in-content-moderation>.
- Twitter blog*. (2020). "Protecting the public conversation". https://blog.twitter.com/en_us/topics/company/2020/covid-19#protecting.
- Twitter Help Center*. (2022a). "Appeal an Account Suspension or Locked Account". Twitter.com. <https://help.twitter.com/forms/general?sub-topic=suspended>.
- Twitter Help Center*. (2022b). "How We Address Misinformation on Twitter". Twitter.com. <https://help.twitter.com/en/resources/addressing-misleading-info>.
- Twitter Help Center*. (2022c). "Our Range of Enforcement Options". Twit-

ter.com. <https://help.twitter.com/en/rules-and-policies/enforcement-options>.

- Wardle, C. and Derakhshan, H. (2017). Information Disorder: Toward an interdisciplinary framework for research and policymaking. *Council of Europe*. file:///C:/Users/h.hasani/Downloads/162317GBR_Report%20desinformation%20(1).pdf.
- Roberts, S. T. (2017). Content Moderation; in Encyclopedia of Big Data. edited by Laurie A. Schintler, Connie L. McNeely, Switzerland: Springer International Publishing.

