



Developing a Content Moderation Framework for Fake News on Social Media (A Case Study of X and Instagram)

Somayeh Labafi *

*Corresponding Author, Associate Prof., Department of Media Management, Iranian Research Institute for Information Science and Technology (IranDoc), Tehran, Iran. E-mail: labafi@irandoc.ac.ir

Faeze Zamani

MSc., Department of Media Management, School of Culture & Communication, Sooreh University, Tehran, Iran. E-mail: misszamani75@gmail.com

Abstract

Objective

In today's digital landscape, the rapid spread of unverified and misleading information, particularly on social media, poses significant challenges to modern societies. Fake news has emerged as a critical issue, threatening public trust, fueling social polarization, and damaging the legitimacy of institutions. However, beyond the issue of fake news itself lies a more systemic and strategic problem: the necessity of effective content moderation. As social media platforms play an increasingly central role in shaping public opinion, distributing information, and facilitating discourse, the ability to moderate fake news within these platforms has emerged as a critical dimension of media governance. Content moderation in this context is not limited to deleting or labeling posts; rather, it encompasses a multi-layered process involving policymaking, technological tools, user engagement, and cultural adaptation. This study aims to develop a comprehensive, context-sensitive framework for moderating fake news on domestic platforms by examining the strategies employed by two global leaders in the field—X (formerly Twitter) and Instagram. Through comparative analysis and in-depth engagement with domain experts, the research attempts to localize and adapt these strategies to the specific socio-cultural and regulatory environment of Iran.

Citation: Labafi, Somayeh & Zamani, Faeze (2025). Developing a Content Moderation Framework for Fake News on Social Media (A Case Study of X and Instagram). *Journal of Business Management*, 17(3), 250-278. <https://doi.org/10.22059/JIBM.2024.378826.4812> (in Persian)



Methodology

This research adopts a qualitative, comparative approach. Initially, the study reviewed policy documents, moderation guidelines, and institutional reports from both platforms to identify their respective content moderation mechanisms. In the second phase, data were gathered through semi-structured interviews with experts in digital policy, media regulation, and information governance. Thematic analysis was used to code and synthesize findings in a structured and iterative manner, facilitating the development of a localized and actionable moderation framework.

Findings

The results indicate that effective fake news moderation requires a dual-layered approach involving both off-platform and on-platform interventions. Off-platform strategies include legislative development, promotion of reliable news sources, enhancement of media literacy, institutional empowerment, public education campaigns, fostering critical discourse, development of detection algorithms, and encouraging civic engagement. On-platform strategies encompass internal content moderation guidelines, user verification systems, crowdsourcing fact-checking tasks, collaborations with official news organizations, algorithmic filtering, labeling and ranking content, and enabling transparent user feedback mechanisms. Notably, Iranian experts emphasized the importance of aligning these measures with local legal and cultural norms, ensuring algorithmic transparency, and offering users clear paths for appeal and redress.

Conclusion

Fake news, amplified by viral digital dynamics, can erode social cohesion and destabilize public discourse. However, the deployment of structured and context-aware content moderation strategies can both limit the spread of false information and enhance user trust in digital environments. The framework proposed in this study synthesizes international practices with localized insights to deliver a hybrid model suitable for domestic social media platforms. This model not only serves as a policy guideline for platform developers and regulators but also contributes to the broader agenda of rebuilding public trust and ensuring the credibility of online information ecosystems in Iran and similar contexts.

Keywords: Fake news, Content moderation, Platform regulation, Social media platforms.

ارائه چارچوب تعدیلگری برای اخبار جعلی در پلتفرم‌ها (مورد مطالعه: تعدیلگری

اخبار جعلی در پلتفرم‌های شبکه اجتماعی ایکس و اینستاگرام)

سمیه لبافی *

دانشیار، گروه پژوهشی مطالعات اجتماعی اطلاعات، پژوهشگاه علوم و فناوری اطلاعات ایران (ایراندک)، تهران، ایران. رایانامه:

labafi@irandoc.ac.ir

فائزه زمانی

کارشناس ارشد، گروه مدیریت رسانه، دانشکده فرهنگ و ارتباطات، دانشگاه سوره، تهران، ایران. رایانامه:

misszamani75@gmail.com

چکیده

هدف: در عصر شبکه‌های اجتماعی، گسترش بی‌ضابطه اطلاعات در کنار تضعیف مرجعیت رسانه‌های رسمی، زمینه‌ساز ظهور پدیده‌هایی چون اخبار جعلی شده است که نه تنها خود مسئله نگران‌کننده‌ای است، بلکه چالش اساسی‌تری به نام «تعدیلگری محتوا» را ایجاد کرده است. با توجه به نقش برجسته‌ای که پلتفرم‌های شبکه اجتماعی در توزیع اخبار و شکل‌دهی به افکار عمومی یافته‌اند، تعدیلگری اخبار جعلی در این فضا، به یکی از اصلی‌ترین محورهای حکمرانی رسانه‌ای بدل شده است؛ به طوری که توانایی تعدیلگری محتوا در این بسترها، هم به کاهش انتشار اخبار نادرست کمک می‌کند و هم می‌تواند حافظ اعتماد عمومی، کیفیت تعاملات اجتماعی و حتی، امنیت اطلاعاتی در جامعه باشد. پژوهش حاضر از طریق شناسایی راه کارهای اتخاذ شده توسط دو پلتفرم شبکه اجتماعی ایکس و اینستاگرام و همچنین، نظرها و تجربه‌های خبرگان این حوزه، چارچوب جامعی به منظور تعدیلگری اخبار جعلی برای پلتفرم‌های شبکه‌های اجتماعی داخلی ارائه داده است.

روش: این مطالعه از نوع کیفی و روش اجرای آن، مطالعه تطبیقی بوده است. برای تحلیل داده‌ها، از ابزار تحلیل مضمون استفاده شده است تا راه کارهای تعدیلگری موجود در دو پلتفرم ایکس و اینستاگرام بررسی شود. با هدف بومی‌سازی راه کارها جهت اعمال و کاربست آن‌ها در پلتفرم‌های داخلی، از ابزار جمع‌آوری داده از طریق مصاحبه استفاده شده و با خبرگان مصاحبه شده است.

یافته‌ها: بر اساس نتایج به دست آمده، راه کارهای موجود را می‌توان در دو دسته راه کار «برون پلتفرمی» و «درون پلتفرمی» طبقه‌بندی کرد؛ به طوری که در دسته اول، راه کارهایی همچون تنظیم قوانین و مقررات، تشویق به ترویج منابع اعتمادساز، توسعه آموزش و آگاهی و روشنگری عمومی، تعزیز انضباط رسانه‌ای، ارتقای سواد رسانه‌ای، توانمندسازی نهادهای مربوطه، ترویج فرهنگ انتقاد سازنده، تشویق به توسعه الگوریتم‌های تشخیص اخبار جعلی و تشویق به مشارکت اجتماعی شهروندان قرار گرفته است. در دسته دوم نیز راه کارهایی چون ارائه قوانین و دستورالعمل‌ها، جمع‌سپاری، اعزاز هویت کاربران، همکاری با منابع رسمی خبری،

استناد: لبافی، سیمیه و زمانی، فائزه (۱۴۰۴). ارائه چارچوب تعدیلگری برای اخبار جعلی در پلتفرم‌ها (مورد مطالعه: تعدیلگری اخبار جعلی در پلتفرم‌های شبکه اجتماعی ایکس و اینستاگرام). مدیریت بازرگانی، ۱۷(۳)، ۲۵۰-۲۷۸.

تاریخ دریافت: ۱۴۰۳/۰۲/۱۲

تاریخ ویرایش: ۱۴۰۳/۰۵/۱۸

تاریخ پذیرش: ۱۴۰۳/۰۶/۰۴

تاریخ انتشار: ۱۴۰۴/۰۶/۰۹

doi: <https://doi.org/10.22059/JIBM.2024.378826.4812>

مدیریت بازرگانی، ۱۴۰۴، دوره ۱۷، شماره ۳، صص. ۲۷۸-۲۵۰

ناشر: دانشکده مدیریت دانشگاه تهران

نوع مقاله: علمی پژوهشی

© نویسندگان

صحت‌سنجی، رتبه‌بندی و برجسب‌گذاری محتوا ارائه شده است. در سطح داخلی نیز، خبرگان بر ضرورت شفاف‌سازی الگوریتم‌ها، رعایت ملاحظات فرهنگی و حقوقی و توسعه نظام آموزش عمومی، در کنار اعمال ابزارهای نرم و سخت برای تعدیل محتوا تأکید داشتند.

نتیجه‌گیری: در جهانی که پلتفرم‌های شبکه اجتماعی اهمیت زیادی دارند، اخبار جعلی، به‌عنوان ابزاری مؤثر برای ترویج اطلاعات نادرست و تضعیف اعتماد عمومی، می‌تواند پایه‌های حقیقت و صداقت را در جامعه به خطر بیندازد؛ تأثیر این اخبار و محتوا بر جامعه، از شکاف اجتماعی تا فروپاشی اعتماد، درخور توجه است. شیوه‌ها و ابزارهای تأثیرگذار تعدیلگری می‌تواند از یک سو فعالیت کاربران در تولید محتوای کاربرساخته و از سوی دیگر، اعتماد آن‌ها به محتوای تولید شده در این پلتفرم‌ها را افزایش دهد. چارچوب ارائه‌شده در این پژوهش با ترکیب تجارب بین‌المللی و دیدگاه‌های بومی‌شده، مدلی ترکیبی و عملیاتی برای تعدیلگری محتوای جعلی در پلتفرم‌های اجتماعی داخلی ارائه می‌دهد. این چارچوب می‌تواند به‌عنوان مبنایی برای سیاست‌گذاری ملی، طراحی پلتفرم‌های بومی و نیز، بازسازی اعتماد عمومی به فضای مجازی استفاده شود.

کلیدواژه‌ها: اخبار جعلی، تعدیلگری، تنظیمگری پلتفرم، پلتفرم شبکه اجتماعی.

مقدمه

خبر، رابطه انسان را با بسیاری از مقوله‌های موجود در زمینه‌های مختلف سیاسی، فرهنگی، اجتماعی، اقتصادی، امنیتی، دینی و غیره بیان می‌کند. خبر علاوه بر اطلاع‌رسانی درباره گذشته، راهی است که به آینده جهت می‌دهد. امروزه پلتفرم‌های رسانه‌های اجتماعی بستر مناسبی را برای ایجاد رسانه‌های مختلف و تسهیل انتشار خبر فراهم آورده‌اند؛^۱ رسانه‌هایی که بر پایه محتوای کاربرساخته بنا شده‌اند و امکان ارسال و تبادل اطلاعات را مهیا کرده‌اند و همچنین بستری برای اشتراک اطلاعات و انتشار اخبار فراهم آورده‌اند (کالزادا و گیل،^۲ ۲۰۲۰).

تکثر رسانه‌ای موجب از بین رفتن انحصار تولید و نشر اخبار توسط رسانه‌های رسمی و ملی شده و رسانه‌های اجتماعی، رفته‌رفته به اصلی‌ترین منبع اخبار در جوامع مختلف تبدیل شده‌اند و ضمن به وجود آوردن بستری به منظور تسهیل فرایند گردش اطلاعات و ارتقای آگاهی عمومی، موجب افزایش امکان ایجاد و تکثیر معضلی به نام «اخبار جعلی»^۳ شده‌اند. اخباری که در ماهیت خود دچار تحریف و از بستر و زمینه اصلی‌شان جدا شده و نامعتبرند (نچوشتای و لوئیس،^۴ ۲۰۱۹؛ ایوانز، جکسون و مورفی،^۵ ۲۰۲۳). این پدیده برای نخستین بار در طول انتخابات ریاست جمهوری ۲۰۱۶ آمریکا و به دنبال آن، در همان سال، در جریان برگزیت^۶ به منصه ظهور رسید و پس از آن رواج یافت. نگرانی‌ها در خصوص اخبار جعلی به حدی رسید که در ۲۷ سپتامبر ۲۰۱۹، در هفتاد و چهارمین مجمع عمومی سازمان ملل متحد، بیست کشور از جمله فرانسه، بریتانیا و هند توافق‌نامه‌ای را برای مقابله با اخبار جعلی امضا کردند (ساعی و آزادی، ۱۴۰۰). از آنجایی که مقابله با اخبار جعلی متشکل از کنش‌های دولتی و غیردولتی (پلتفرمی) است، کمیسیون اتحادیه اروپا نیز در سال ۲۰۲۰، با هدف مقابله با گسترش اخبار جعلی در عین حفاظت از حق آزادی بیان، پلتفرم‌های رسانه اجتماعی پرمخاطب همچون متا (فیسبوک) و ایکس (توییتر) را ملزم کرد که بیش از پیش اقدامات لازم را برای مقابله با اخبار جعلی و پالایش محتوای خود انجام دهد (آیین‌نامه عملکرد در خصوص اطلاعات نادرست، کمیسیون اتحادیه اروپا،^۷ ۲۰۲۲).

با توجه اهمیت مقابله با انتشار گسترده اخبار جعلی توسط پلتفرم‌های رسانه‌های اجتماعی، تعدیل محتوا امری ضروری است؛ به طوری که تارلتون ژیلیسی^۸ (۲۰۱۸)، تعدیل محتوا را یکی از مشخصه‌های تعریف شده برای هر پلتفرم و امری «ضروری، تبیین‌کننده و سازنده» می‌داند و تأکید می‌کند که در صورت نبود آن، پلتفرم‌ها دوام نخواهند داشت. در این میان، پلتفرم‌های شبکه‌های اجتماعی ایکس و اینستاگرام، مدت‌هاست که فرایند تعدیل محتوا و مقابله با اخبار جعلی را آغاز کرده‌اند و توانسته‌اند تجربه‌ها و راه کارهای کارآمد و به روز را در این زمینه ارائه دهند (ایوانز و همکاران، ۲۰۲۳؛

1. <https://hawzah.net/fa/Article/View/117398>

2. Calzada & Gil

3. Fake News

4. Nechushtai & Lewis

5. Evans, Jackson & Murphy

۶. خروج انگلیس از اتحادیه اروپا

7. Disinformation weakening democracy on <https://news.stanford.edu/2022/04/21/disinformation-weakening-democracy-barack-obama-said>

8. Tarleton Gillespie

هارامبام، هوبوکن و هلیبرگر^۱، ۲۰۱۸). بنابراین، مطالعه این راه‌کارها، آغاز مسیر در پژوهش‌های این حوزه به حساب می‌آید. از طرفی دیگر، با توجه به اهمیت موضوع و نبود قوانین و مقررات مدون و کارآمد برای فضای مجازی در کشور^۲ و نیز موقعیت حساس و همچنین امکان تأثیر اخبار جعلی بر افکار عمومی، در کنار ایجاد چالش‌های سیاسی و امنیتی در کشور توسط رسانه‌های خارجی، اتخاذ مناسب‌ترین کاربست‌ها برای کاهش خطرهای احتمالی، امری ضروری تلقی می‌شود. به همین منظور، شناسایی و مقایسه راه‌کارهای اتخاذ شده توسط پلتفرم‌های مزبور و نهادهای فعال در این زمینه و استخراج مناسب‌ترین و به‌روزترین راه‌کارهای درون و برون پلتفرمی مورد استفاده، نخستین گام این پژوهش برای ترسیم چارچوب ابتدایی است و از آن جهت که هدف غایی پژوهش حاضر، بومی‌سازی این راه‌کارها و ارائه چارچوب تعدیلگری اخبار جعلی در پلتفرم‌های شبکه‌های اجتماعی داخلی است، تلاش شده است که نظرهای کارشناسان داخلی و خبرگان این حوزه نیز گردآوری و در چارچوب نهایی و توصیه‌های عملی اعمال شود. نتایج به‌دست‌آمده از این پژوهش، می‌تواند نقطه آغاز مناسبی برای سیاست‌گذاران و فعالان این حوزه باشد و همچنین سبب به‌وجودآمدن نقشه‌راهی روشن برای کاربران شود. بنابر آنچه گفته شد، این پژوهش در تلاش است تا به هدف اصلی خود، یعنی استخراج راه‌کارهای کارآمد جهت تعدیلگری اخبار جعلی در پلتفرم‌های شبکه‌های اجتماعی داخلی، با کمک شناسایی و مقایسه راه‌کارهای اتخاذ شده توسط پلتفرم‌های اینستاگرام و ایکس دست یابد.

پیشینه نظری پژوهش

اخبار جعلی

به‌گفته بارسلی^۳ (۲۰۱۸)، «خبر جعلی» جدیدترین عنوان برای هنر باستانی دروغ گفتن است. خبر جعلی، عبارت است از هر خبری که در ماهیت خود دچار تحریف شده، از بستر و زمینه اصلی خود جدا شده و بنابراین معتبر نیست. انگیزه‌های مالی و سیاسی، دو انگیزه عمده‌ای هستند که در منابع مختلف برای تولید اخبار جعلی هدفمند ذکر شده‌اند (مجدی‌زاده و موسوی حق‌شناس، ۱۳۹۹). از طرفی دیگر، ریشه‌های فراگیری اخبار جعلی را می‌توان در موضوعاتی همچون عدم اطمینان، نظام سیاسی کشور و توهم دانش دانست (رحمانیان، ۱۴۰۰). آنچه در این پژوهش حائز اهمیت است، واژه‌های به‌کاربرده شده توسط اینستاگرام، توئیتر (به‌عنوان پلتفرم شبکه‌های اجتماعی انتخاب شده برای انجام مطالعه موردی) و اتحادیه اروپا به‌عنوان یک نهاد نظارتی است. جالب است بدانیم که هر یک از شرکت‌های یادشده، از واژه‌ای متفاوت برای ارجاع کاربران خود به مسئله اخبار جعلی استفاده کرده‌اند. به‌طور مثال، اینستاگرام در تمامی اسناد و دستورالعمل‌های خود، از واژه «غلط»^۴ استفاده می‌کند و در بخش رتبه‌بندی محتوا، آن را این چنین تعریف می‌کند: «محتوایی که هیچ مبنای واقعی ندارد شامل: نقل قول‌های جعلی، ادعاهایی که غیرممکن هستند یا نمی‌توان آن‌ها را تفسیری از چیزی

1. Harambam, Hoboken & Helberger

2. www.farsnews.ir/news/14010113000276/ایران-جزو-معدود-کشورهای-بدون-قانون-فضای-مجازی-است

3. Barsley

4. False

دانست که واقعا اتفاق افتاده یا گفته شده‌اند؛ تئوری‌های توطئه‌ای که رویدادها را به‌عنوان کار مخفی افراد یا گروه‌ها توضیح می‌دهند که ممکن است به اطلاعات واقعی یا تأیید نشده استناد کنند، اما نتیجه‌گیری نامقبولی ارائه کنند؛ محتوای ساختگی از وبسایت‌هایی که خودشان را به‌عنوان رسانه‌های خبری واقعی معرفی می‌کنند؛ محتوای تصویری یا ویدئویی معتبر که در اصل به‌منظور دیگری ارائه شده است».

شبکه اجتماعی توئیتر، اما واژه دیگری را برای ارجاع به این حوزه مد نظر قرار داده و آن «اطلاعات نادرست یا اطلاعات همراه کننده»^۱ است و در توضیح آن این گونه می‌نویسد: «ادعاهایی که توسط کارشناسان خارجی نادرست بودن آن‌ها تأیید شده یا حاوی اطلاعاتی هستند که به‌طور فریبنده و گیج‌کننده به اشتراک گذاشته شده‌اند. این محتوای همراه کننده از طریق ترکیبی از فناوری و مرور انسانی و همچنین، از طریق مشارکت با کارشناسان جهانی شخص ثالث شناسایی می‌شود»^۲.

واژه انتخابی اتحادیه اروپا، اطلاعات دروغین است که آن را «محتوای غلط یا همراه کننده» می‌داند که بدون قصد سوء به اشتراک گذاشته شده است، اگرچه اثرهای آن، همچنان می‌تواند آسیب‌رسان باشد. این واژه همچنین در برنامه اقدام دموکراسی اروپا، به‌عنوان «محتوای غلط یا همراه کننده که با هدف فریب یا تأمین منافع اقتصادی یا سیاسی منتشر می‌شود و ممکن است باعث آسیب عمومی شود» تعریف شده است.

و اما تعریف بومی این واژه که در مصوبه شورای عالی فضای مجازی تحت عنوان «اطلاعات، اخبار و محتوای خبری خلاف واقع» بیان شده، این گونه است: «هرگونه اطلاعات، اخبار و محتوای خبری منتشر شده در فضای مجازی شامل انواع صور متنی، صوتی، تصویری یا چندرسانه‌ای که مابه‌ازایی در واقعیت نارد یا شکل تحریف‌شده‌ای از یک واقعیت است که موجب آسیب و اخلال در نظم عمومی، تشویش اذهان عمومی، کاهش ثبات، انسجام، اعتماد، سلامت و امنیت فرد و جامعه در ابعاد اقتصادی، اجتماعی و سیاسی یا کسب منافع نامشروع می‌شود».

رویکرد دولت‌ها در جهت مقابله با اخبار جعلی

ژیلپی (۲۰۱۸) در کتاب خود، تعدیل محتوا در پلتفرم‌های دیجیتال را فرایندی پیچیده می‌داند که به‌صورت خودکار یا براساس قوانین صریح عمل نمی‌کند؛ زیرا تعدیل محتوا توسط تصمیم‌های مخفی و مبهم انسانی انجام می‌شود و به عواملی نظیر سیاست‌های داخلی شرکت، تفسیر قوانین و معیارها، تأثیر فشارها و نیازهای خارجی و داخلی وابسته است. او بیان می‌کند که این مسئولیت، به جهت گستردگی و خطیر بودن آن، هم برعهده پلتفرم‌ها و هم، بر دوش دولت‌ها و نهادهای حاکمیتی است. نقش دولت‌ها در تعدیل محتوای پلتفرم‌ها متنوع است و به شرایط قانونی و سیاسی هر کشور بستگی دارد (مولر^۳، ۲۰۲۲؛ کواک، لی و لی^۴، ۲۰۲۱). در برخی کشورها، دولت قوانین و مقرراتی را تدوین کرده است که تعدیل محتوا را تحت نظارت قرار می‌دهد و پلتفرم‌ها را ملزم می‌کند تا محتوای مشکوک، غیرقانونی یا مضر را حذف

1. Misleading information

2. <https://help.twitter.com/en/resources/addressing-misleading-info>

3. Møller

4. Kwak, Lee & Lee

کنند. در این حالت، دولت نقش ناظر و نظارتگر را برعهده دارد و مسئولیت تعدیل محتوا به صورت اصولی به پلتفرم‌ها سپرده می‌شود. دولت‌ها می‌توانند همچنین با ایجاد قوانین و مقرراتی در زمینه حریم خصوصی، امنیت داده‌ها، آزادی بیان و سایر موضوعات مرتبط، تأثیری بر تعدیل محتوا در پلتفرم‌ها داشته باشند. این قوانین می‌توانند شرایطی را برای پلتفرم‌ها ایجاد کنند که آن‌ها را به اتخاذ تدابیر تعدیل محتوا مجاب می‌کنند، مانند گزارش‌دهی دربارهٔ محتوای مخل اخلاقی یا غیرقانونی، حفظ اطلاعات کاربران یا همکاری با مقامات قضایی در صورت لزوم. علاوه‌براین، دولت‌ها می‌توانند با ایجاد پس‌زمینه قانونی و سیاسی، پلتفرم‌ها را تشویق کنند تا در تعدیل محتوا در قالب توافق‌نامه‌ها یا کدهای اخلاقی مشارکت کنند. به‌طور مثال، برخی کشورها در زمینهٔ مبارزه با تروریسم و ترویج محتوای تروریستی با پلتفرم‌ها همکاری می‌کنند تا محتوای مشکوک را شناسایی و حذف کنند. در این پژوهش، به‌منظور شناسایی رویکرد دولت‌ها در جهت مقابله با اخبار جعلی، سعی می‌شود تا اقدامات اتحادیهٔ اروپا و جمهوری اسلامی ایران شناسایی شود. بررسی قوانین اتحادیهٔ اروپا، با توجه به طراحی سیاست‌های هماهنگ و اقدامات مشترک کشورهای عضو این اتحادیه، می‌تواند الگوی مناسبی جهت تشخیص قوانین تمامی کشورهای اروپایی عضو این اتحادیه باشد.

در ایران، شورای عالی فضای مجازی نقش مهمی در تعیین سیاست‌ها، ضوابط، مقررات و آیین‌نامه‌ها به‌منظور مقابله با انتشار اخبار جعلی و محتوای مضر و غیرقانونی در فضای مجازی دارد. این شورا در مصوبه‌ای^۱ با هدف صیانت از سلامت و امنیت روانی جامعه و کاهش آسیب‌های موصوف از طریق تعریف و تعیین اقدامات پیشگیرانه و واکنشی و ارتقاء سطح مسئولیت‌پذیری و پاسخ‌گویی هماهنگ و به‌هنگام برای مقابله با ناهنجاری‌ها و پیامدهای تولید و نشر اطلاعات، اخبار و محتوای خلاف واقع سعی در تدوین سند در این زمینه نمود. در قسمت تعاریف این مصوبه، «اطلاعات، اخبار و محتوای خبری خلاف‌واقع» این گونه تعریف می‌شود: «هرگونه اطلاعات، اخبار و محتوای خبری منتشرشده در فضای مجازی شامل انواع صور متنی، صوتی، تصویری یا چندرسانه‌ای که مابه‌ازایی در واقعیت نداشته یا شکل تحریف‌شده‌ای از یک واقعیت است که موجب آسیب و اخلال در نظم عمومی، تشویش اذهان عمومی، کاهش ثبات، انسجام، اعتماد، سلامت و امنیت فرد و جامعه در ابعاد اقتصادی، اجتماعی و سیاسی یا کسب منافع نامشروع می‌شود». در این مصوبه، قوه قضائیه، وزارت فرهنگ و ارشاد اسلامی و وزارت ارتباطات و فناوری اطلاعات، می‌بایست با همکاری مرکز ملی فضای مجازی اقداماتی از جمله تنقیح و متناسب‌سازی مجازات‌ها، تعیین و تفکیک مسئولیت صاحبان درگاه‌ها، تبیین مصادیق جرم، شناسایی و احراز هویت درگاه‌های نشر، پیاده‌سازی سازوکارهای جامع رصد و راستی‌آزمایی، ایجاد پنجرهٔ همکاری درگاه‌ها، نظام اعتبارسنجی و بسیاری اقدامات دیگر انجام دهند. در این میان، درگاه‌های نشر وظایفی همچون ایجاد سازوکار لازم تشخیص اطلاعات، اخبار و محتوای خبری خلاف‌واقع دریافتی قبل از انتشار اعم از رویه‌های انسانی و یا ماشینی موسوم به تعدیل محتوا یا اتصال به سامانه‌های راستی‌آزمایی مرجع ایجاد شده در کشور و همچنین مبادرت به حذف و درج توضیحات به محض اطلاع از انتشار اطلاعات، اخبار و محتوای خبری خلاف‌واقع در

۱. مصوبه ۶۸ مورخ ۱۳۹۹/۱۰/۱۶ با موضوع «الزامات پیشگیری و مقابله با نشر اطلاعات، اخبار و محتوای خبری خلاف واقع در فضای مجازی»

کنار شناسایی و حذف فوری این اطلاعات را برعهده دارند.

از طرفی دیگر، در کنار تمامی آثار مخرب انتشار اخبار جعلی، مهم‌ترین تأثیری که کشورهای عضو اتحادیه را بر آن داشت تا درباره این مهم به تفکر، تصمیم و اجرا بپردازند، مبحث تأثیر اخبار جعلی بر دموکراسی بود؛ بنابراین اتحادیه اروپا از نخستین سال‌های درک این تهدید، تصمیم به مقابله با این مسئله مهم را اتخاذ کرده و همواره کوشیده است تا به کمک روش‌های مختلف، به این موضوع رسیدگی کند. اتحادیه اروپا گام‌های متعددی را در جهت شناسایی هدف مناسب تلاش‌های تعدیل و مسئولیت‌های مختلف ذی‌نفعان مربوطه در تدوین اسنادی تحت عنوان «آیین‌نامه عملکرد در مورد اطلاعات دروغین» در سال ۲۰۱۸ و در ادامه تقویت این آیین‌نامه در سال ۲۰۲۲ برداشته است. تصویب این آیین‌نامه که به موجب آن، یک ابزار خودتنظیمی نوآورانه برای اطمینان از شفافیت و پاسخ‌گویی بیشتر پلتفرم‌های آنلاین و چارچوبی برای نظارت و بهبود سیاست‌های پلتفرم‌های آنلاین در خصوص اطلاعات دروغین طرح شد، یکی از مهم‌ترین اقدامات اتحادیه به شمار می‌رود. همچنین، در سال ۲۰۲۲، تدوین آیین‌نامه تقویت عملکرد، ارائه پیشنهاد قانون خدمات دیجیتال و قانون بازارهای دیجیتال از مهم‌ترین اقدامات اتحادیه در جهت مقابله مستقیم و غیرمستقیم با انتشار اخبار جعلی در دنیای دیجیتال به‌شمار می‌رود که البته سنجش میزان کارآمدی این اقدامات، مستلزم سپری شدن زمان بیشتری است. گروه کارشناسی سطح بالا در خصوص اخبار جعلی و اطلاعات دروغین، مشخص می‌کند که هدف تلاش‌های آن‌ها، اطلاعات دروغین است که به‌عنوان «اطلاعات غلط، نادرست یا گمراه کننده که به قصد رساندن آسیب عمومی یا کسب منفعت طراحی، ارائه و ترویج شده‌اند» در اتحادیه در نظر گرفته می‌شود (گروه تخصصی اخبار جعلی و اطلاعات دروغین آنلاین^۱، ۲۰۱۸: ۱۰). با در نظر گرفتن این تعریف، اتحادیه اروپا پنج حوزه کلیدی را برای مداخله با هدف مقابله با اطلاعات دروغین در نظر گرفته است: ۱. افزایش شفافیت؛ ۲. ارتقای سواد رسانه‌ای؛ ۳. توانمندسازی کاربران و خبرنگاران؛ ۴. حفاظت از تنوع و پایداری رسانه‌های خبری؛ ۵. ادامه تحقیقات درباره اطلاعات دروغین و اثربخشی مداخلات. با توجه به مکلف کردن پلتفرم‌ها در خصوص افزایش شفافیت و ابزارهای توانمندسازی کاربران، به‌منظور شناسایی و گزارش اخبار جعلی، اتحادیه اروپا اقداماتی همچون سانسور محتوا را توصیه نمی‌کند؛ زیرا این اقدامات برخلاف اصل آزادی بیان و دموکراسی است. اگرچه، پیشنهاد این سازمان به پلتفرم‌ها، همکاری با سازمان‌های صحت‌سنج مستقل در کنار اولویت‌بندی اطلاعات قابل اعتماد در الگوریتم‌های رتبه‌بندی و پیوند دادن کاربران به محتوای قابل اعتماد است.

تعدیلگری محتوا توسط پلتفرم‌های شبکه‌های اجتماعی

تعدیل محتوا در فضای مجازی به فرایندی اشاره دارد که در آن محتواهای آنلاین توسط افراد یا سازمان‌ها بررسی، تحلیل و ویرایش می‌شوند تا بهبود و کاهش نقاط ضعف و مشکلات ممکن در آن‌ها ایجاد شود (کارلسون^۲، ۲۰۱۸؛ هیتز و همکاران^۳، ۲۰۲۲). در واقع، تعدیل محتوا به‌منظور ارتقای کیفیت و قابلیت اعتماد محتواهای آنلاین انجام می‌شود. این

1. Tackling online disinformation: a European Approach

2. <https://digital-strategy.ec.europa.eu/en/library/communication-tackling-online-disinformation-european-approach>

3. Carlson

4. Heitz et al.

فرایند می‌تواند شامل بررسی صحت اطلاعات، تحقیق در مورد منابع و مؤلفان، حذف یا اصلاح اطلاعات نادرست یا غیرقابل اعتماد، اضافه کردن منابع و مراجع قابل اعتماد، بهبود ساختار و قالب محتوا، تصحیح املائی و قواعدی یا حذف محتوای ناسازگار و نامناسب باشد. تعدیل محتوا دارای اهداف متعددی از جمله افزایش اعتماد کاربران به محتواهای آنلاین، کاهش انتشار اطلاعات نادرست و جعلی، ارتقای کیفیت و قابلیت اعتماد محتواها، تشویق به استفاده از منابع معتبر و تأییدشده و ایجاد محتوایی متناسب با استانداردهای ارائه محتوا در فضای مجازی است (وو و لیو^۱، ۲۰۱۸). گریجل و براون^۲ (۲۰۱۹) در مقاله خود تعدیل محتوا را فرایند شناسایی مشارکتهای نامربوط، ناشایسته، غیرقانونی، مضر یا توهین‌آمیز برخلاف مشارکتهای مفید و آموزنده خوانده‌اند و این شناسایی را اغلب برای سانسور یا سرکوب دیدگاه‌های مخالف می‌دانند. هدف از تعدیل محتوا نیز، حذف یا اعمال برچسب هشدار برای محتوای مشکل‌ساز یا توانمندسازی کاربران نسبت به مسدود یا فیلتر کردن محتوا بیان شده است.

بنابراین، تعدیل محتوا در پلتفرم‌های شبکه‌های اجتماعی با رعایت قوانین و مقررات صورت می‌گیرد؛ این بدان معناست که جوامع آنلاینی که بستر تعاملات اجتماعی را برقرار می‌کنند، می‌بایست برای بررسی، غربالگری و فیلترکردن محتوا، به‌منظور ایجاد چارچوب درخصوص ارسال و دریافت محتوای نامقبول و نامناسب، قوانین و دستورالعمل‌های اساسی داشته باشند. برای شناسایی چارچوب‌های طراحی‌شده توسط پلتفرم‌های پرمخاطب، در ادامه راه‌کارهای تعدیلگری شبکه‌های اجتماعی اینستاگرام و ایکس (توییتر) شناسایی و بررسی می‌شوند.

اینستاگرام

اینستاگرام یک شبکه اجتماعی است که امکان به اشتراک گذاشتن عکس و فیلم در آن را برای کاربران خود فراهم کرده است و سرویسی به‌منظور برقراری تعاملات اجتماعی ارائه می‌دهد. این شبکه اجتماعی مانند سایر برنامه‌های این حوزه، شامل مجموعه قوانینی است که کاربران باید آن‌ها را رعایت کنند تا بتوانند عضو پایدار این شبکه اجتماعی بمانند و از قابلیت‌ها و امکانات آن بهره‌مند شوند. سخنان نفرت‌آمیز، باج‌گیری یا آزار و اذیت شخصی، تشویق افراد به خشونت، انتشار فیلم‌های منفی و خشونت‌آمیز، تشویق مردم به پذیرش خودآزاری، تهدیدکردن افراد جهت ایجاد صدمه جسمی و مالی و...، از جمله قوانین سختی هستند که اینستاگرام برای کاربران در نظر گرفته است و تمام کاربران موظف به رعایت آن‌ها هستند. در گزارش منتشر شده در وبسایت نیو امریکا^۳، در سال ۲۰۱۹، مطالعه موردی به‌منظور شناسایی و تحلیل راه‌کارهای تعدیلگری، شرکت متا تا حد زیادی بزرگ‌ترین عملیات تعدیل محتوا را دارد. متا به‌منظور ایجاد ثبات در نحوه اعمال قوانین خود در سراسر جهان، از یک رویکرد متمرکز و ترکیبی برای تعدیل محتوا استفاده کرده و دارای مجموعه‌ای جهانی از استانداردهای جامعه است. این استانداردها توسط مجموعه عظیم جهانی مدیران محتوای انسانی فعال در زمینه ایمنی و امنیت پلتفرم اعمال می‌شود. این شرکت سعی دارد مدیران را بر اساس زبان یا تخصص منطقه‌ای آن‌ها

1. Wu & Liu

2. Grygiel & Brown

3. <https://www.newamerica.org/oti/reports/everything-moderation-analysis-how-internet-platforms-are-using-artificial-intelligence-moderate-user-generated-content/>

و همچنین تخصص متفاوت آن‌ها استخدام کند و در تلاش است تا سوگیری و ذهنیت پیشین در قوانین و خط‌مشی‌های خود و همچنین ارائه قضاوت‌های ثابت را کاهش دهد. استفاده از ابزارهای خودکار، معمولاً نخستین گام بررسی هنگام شناسایی محتوای نقض‌کننده در این پلتفرم است. طبق این گزارش، هنگامی که یک عکس یا ویدئو توسط کاربر بارگذاری شود، بلافاصله در یک فرایند خودکار غربالگری الگوریتمی (با روش‌های متفاوت) می‌شود. این شرکت همچنین از ابزارهای تطبیق و اقدام پیشگیرانه برای شناسایی و حذف محتوای مطابق محتوای شناسایی شده قبلی استفاده می‌کند.

این شبکه اجتماعی همچنین در خصوص محتوای غلط و به‌اصطلاح اخبار جعلی نیز قوانین بسیاری وضع کرده است؛ زیرا در جریان انتخابات آمریکا در سال ۲۰۱۶، در دسترس قرار گرفتن اطلاعات غلط برای رأی‌دهندگان به‌واسطه این اپلیکیشن، مشکلات عدیده‌ای برای این شبکه اجتماعی به‌وجود آورد و از آن پس، مدیران خود را بر آن داشت که هرچه زودتر، قوانین جامع و مکفی در این زمینه وضع و اعمال کنند. وبسایت اینستاگرام در این رابطه می‌گوید:

«ما در اینستاگرام متعهد به کاهش گسترش اطلاعات نادرست هستیم.... برای شناسایی پُست‌ها و حساب‌هایی که ممکن است حاوی اطلاعات غلط باشند، از فناوری و بازخوردهای انجمن خود در کنار هم استفاده می‌کنیم. همچنین با صحت‌سنجی‌های شخص ثالث در سراسر جهان که مسئول بررسی محتواها در بیش از ۶۰ زبان هستند و برای کمک به شناسایی و مرور و هشتگ‌گذاری اطلاعات نادرست و جعلی، مورد تأیید شبکه صحت‌سنجی بین‌المللی قرار گرفتند، همکاری می‌کنیم.... به‌صورت شفاف‌تر، اگر اطلاعاتی که منتشر کرده‌اید نادرست باشد، اینستاگرام آن را محدود می‌کند»^۱.

ایکس (توییتر)

ایکس (توییتر سابق) یک پلتفرم شبکه اجتماعی است که در سال ۲۰۰۶ توسط جک دورسی راه‌اندازی و سپس توسط ایلان ماسک خریداری شد. ایکس نیز مانند اینستاگرام، شامل مجموعه قوانینی است که کاربران باید آن‌ها را رعایت کنند تا بتوانند عضو پایدار این شبکه اجتماعی بمانند و از قابلیت‌ها و امکانات آن بهره‌مند شوند. بنابر آنچه در خصوص لزوم وجود قوانینی جهت کاهش انتشار اخبار جعلی در فضای مجازی گفته شد، توییتر نیز خود را از این قاعده مستثنا ندانسته است و همواره می‌کوشد تا چارچوبی به‌روز و مکفی در این حیطه تدوین کند. این شرکت در کنار اقداماتی چون بهبود الگوریتم‌های تشخیص محتوای خود، به‌کارگیری ناظران، توسعه عملکردهای هشدار، مشارکت با رسانه‌های جریان اصلی و ایجاد امکان گزارش کاربران، در حال حاضر یک ابزار نظارت جمعی یا جمع‌سپاری شده برای بررسی حقایق و مقابله با اطلاعات نادرست و بررسی شده، تحت عنوان بردواج^۲ تعبیه کرده است. این قابلیت با هدف گسترش سطح واقعیت‌سنجی است؛ به‌طوری که در موضوعاتی که ایکس قصد ورود و اعمال راستی‌آزمایی را به‌طور مستقیم ندارد، این امکان را برای مخاطبان خود فراهم می‌کند تا با ارائه اطلاعات و زمینه اضافه، این کار را به عهده بگیرند. این امکان در مواردی که فراتر از خط‌مشی‌های تنظیم شده و همچنین در مرزهای مبهم و چالشی، سودمند است.

1. Combatting Misinformation on Instagram | Instagram. (2019, December 16). About.instagram.com

2. Birdwatch

روش‌شناسی پژوهش

پژوهش حاضر از حیث نوع، کیفی و از حیث ابزار گردآوری اطلاعات، از نوع کتابخانه‌ای (بررسی مستندات و مدارک موجود) است و از مصاحبه‌های نیمه‌ساختار یافته بهره برده است. همچنین، از جهت رویکرد، تطبیقی (مقایسه‌ای) است و از حیث روش بررسی و تحلیل محتوای متون و مصاحبه‌ها، از روش تحلیل مضمون استفاده کرده است. در این پژوهش ابتدا با انجام مطالعه کتابخانه‌ای، به گردآوری راه‌کارهای تعدیلگری پلتفرم‌های اینستاگرام و ایکس پرداخته و از طریق تطبیق توصیفی و گزارشی، اشتراک‌ها و تفاوت‌های راه‌کارهای این دو پلتفرم مشخص می‌شود؛ سپس با انجام مصاحبه‌های نیمه‌ساختاریافته با خبرگان این حوزه، چالش‌ها و نارسایی‌ها استخراج و مناسب‌ترین راه‌کارهای تعدیلگری برای پلتفرم‌های داخلی ارائه می‌شود. در این مقاله، از دو گروه جامعه استفاده شد. جامعه نخست را مجموعه اسناد مورد استفاده در این پژوهش شامل قوانین، اسناد و مقالات مرتبط با اخبار جعلی در پلتفرم‌های رسانه‌های اجتماعی تشکیل می‌دهند و دومین جامعه به‌منظور انجام مصاحبه نیز متشکل از کارشناسان، متخصصان رسانه و ارتباطات و همچنین خبرگان حوزه سیاست‌گذاری در بستر فضای مجازی بودند. نمونه‌گیری نیز در دو مرحله به‌صورت هدفمند انجام پذیرفت. در مرحله نخست، نمونه‌ها به‌منظور دست‌یافتن به قوانین و خط‌مشی‌های دو پلتفرم شبکه‌های اجتماعی اینستاگرام و ایکس و همچنین، سیاست‌های اتحادیه اروپا انتخاب شدند. در مرحله دوم و برای بررسی بیشتر راه‌کارهای نهایی و متناسب با پلتفرم‌های داخلی، نمونه‌های انتخابی به‌صورت هدفمند، از میان کارشناسان و متخصصان رسانه و ارتباطات و خبرگان حوزه سیاست‌گذاری در بستر فضای مجازی انتخاب شدند. در پایان، با ۱۱ نفر از خبرگان اعم از قانون‌گذاران، پژوهشگران، فعالان حوزه مقابله با اخبار جعلی که در زمینه‌های شغلی روزنامه‌نگاری، پژوهشگری، هوش مصنوعی و رگولاتوری در نهادهای گوناگونی مانند مرکز ملی فضای مجازی، سازمان تنظیم مقررات رسانه‌های صوت و تصویر فراگیر، دفتر مطالعات حکمرانی، پژوهشگاه علوم و فناوری اطلاعات ایران، پژوهشکده فرهنگ، هنر و ارتباطات و همچنین پلتفرم‌های داخلی مشغول به فعالیت و پژوهش هستند، مصاحبه صورت گرفت. این پژوهش با استفاده از روش استقرایی (تحلیل بر اساس داده‌ها از پایین به بالا) اجرا شده و فرایند آن طی سه مرحله (کدگذاری توصیفی، کدگذاری تفسیری و یکپارچه‌سازی از طریق مضامین فراگیر) بوده است؛ بدین ترتیب که پس از استخراج داده‌ها با استفاده از ابزار کتابخانه‌ای، مناسب با اهداف پژوهش، به مقایسه راه‌کارها و احصای مقوله‌های مشابه و متمایز پرداخته می‌شود و پس از آن، با انجام مصاحبه‌های نیمه‌ساختاریافته با خبرگان این حوزه، چارچوب نهایی بومی‌سازی شده ارائه می‌شود.

یافته‌های پژوهش

در این بخش، ضمن استخراج نکات مهم در مقاله‌های مختلف و شناسایی راه‌کارهای مورد استفاده اینستاگرام و ایکس (توثیر سابق) جهت تعدیلگری اخبار جعلی در محیط این شبکه‌های اجتماعی، راه‌کارهای بیان شده از جانب خبرگان این حوزه در داخل ایران نیز مورد بررسی قرار می‌گیرد. در نخستین گام، پژوهشگر با تکیه بر اسناد و مجموعه قوانین منتشر شده از هر یک از پلتفرم‌های یادشده و همچنین با تطبیق راه‌کارهای استخراج شده، می‌کوشد تا راه‌کارهای تعدیلگری اخبار جعلی را دسته‌بندی کند. در گام بعدی و با توجه به نتایج به‌دست‌آمده از مصاحبه با خبرگان در خصوص راه‌کارهای

تعدیلگری اخبار جعلی در داخل ایران، مناسب‌ترین راه‌کارها را برای تعدیلگری اخبار جعلی در بستر پلتفرم‌ها و شبکه‌های اجتماعی داخلی مشخص و چارچوب کاملی را ارائه داده است.

جدول ۱. راه‌کارهای تعدیلگری اخبار جعلی احصا شده از پیشینه و اسناد مرتبط

نویسنده	توضیحات و راه‌کارهای ارائه شده
ساعتی و آزادی (۱۴۰۰)	ایجاد قوانین به‌منظور مسئولیت‌پذیر کردن پلتفرم‌ها توسط دولت - تعبیه الگوریتم‌های تشخیص خودکار توسط پلتفرم‌های رسانه‌های اجتماعی (فیلتر محتوای مضر، تشخیص شبکه انتشار محتوا، شناسایی و حذف ربات‌ها و حساب‌های کاربری جعلی)، اقدامات مهم سازمان‌های برودکست (ارائه اخبار دقیق، صحیح و قابل اعتماد، رد خبر جعلی و آگاه‌سازی عمومی)
گائوژائو ^۱ (۲۰۲۱)	راه‌کار برچسب‌گذاری نتوانسته است موجب افزایش تفکر انتقادی شود، بلکه باعث افزایش اعتماد کورکورانه شده است؛ البته تا حدی به افزایش قدرت قضاوت افراد کمک کرده است.
آبرو، دزونینگ ناواجا و مک‌کامبز ^۲ (۲۰۲۰)	برچسب‌گذاری نقش مؤثری در تصمیم‌گیری صحیح در مورد به اشتراک‌گذاری (یا عدم به اشتراک‌گذاری) محتوا توسط کاربران دارد، اما تأثیرات ماندگار فریب اذهان عمومی توسط اخبار جعلی همچنان پابرجا و ماندگار می‌ماند.
گالو و چو ^۳ (۲۰۲۱)	بهترین روش تعدیل محتوا، استفاده هم‌زمان از عوامل انسانی و هوش مصنوعی در کنار یکدیگر است.
کامن ^۴ (۲۰۱۹)	فرایند تعدیل محتوا سه مرحله دارد: ایجاد (توصیف و توسعه شرایط و ضوابط)، اجرا (اعمال قوانین و مقررات وضع شده) و پاسخ (تصمیم‌گیری و تجدید نظر). مشکلات عدیده‌ای در اعمال قوانین توسط پلتفرم‌ها وجود دارد و همین امر موجب لزوم اعمال اصلاحات گسترده و تغییر در اولویت‌هاست.
باربرا، جوست، ناگلر، تاکر و بونو ^۵ (۲۰۲۲)	ماهیت شبکه‌ای اطلاعات نادرست موجب می‌شود تا در صورت برچسب‌گذاری آن‌ها در یک پلتفرم، در پلتفرم‌های دیگر از محبوبیت بالاتری برخوردار شوند و این امر اهمیت تعدیل محتوا در سطح اکوسیستم را نشان می‌دهد.
ریدل، ویپل و والاس ^۶ (۲۰۲۱)	در نظر گرفتن نگرش مردم و علاقه آن‌ها به وجود تعدیلگری از جانب پلتفرم یا سیاست‌گذاری از جانب دولت امری مهم در این زمینه است.
ون در لیدن و همکاران ^۷ (۲۰۲۱)	لزوم ایجاد آنتی بادی قبل از دچار شدن به ویروس بر اهمیت به‌کارگیری هشدار، پریانکینگ و تلیقیج تأکید دارد.
کوئوا و همکاران ^۸ (۲۰۲۰)	تعدیل محتوا بدون نظارت عامل انسانی خطری جدی برای حقوق بشر است؛ چرا که این فناوری هنوز قادر به درک پیچیدگی ارتباطات انسانی نیست و احتمال بروز خطرهای متعدد برای آزادی بیان و دسترسی به اطلاعات و تنوع در محیط دیجیتال را افزایش می‌دهد. بنابراین، این فناوری نباید تنها مبنای دستیابی به تصمیمات باشد.
برشیر، پنی‌کوک، برینسکی و رند ^۹ (۲۰۲۰)	ارائه مستمر اصلاحات و برچسب‌ها (قبل، حین و بعد از انتشار) می‌تواند تأثیر بسزایی بر درک درست افراد داشته باشد.
وازکز هررو، ویزوسو و لوپز گارسیا ^۱ (۲۰۱۹)	صحت‌سنجی ضروری بوده و موجب شفاف‌سازی می‌شود و مأموریت اصلی آن کشف گفتمان، بازسازی و پیگیری روند انتشار آن‌ها است.

1. Gaozhao

2. Abreu, de Zúñiga Navajas & McCombs

3. Gallo & Cho

4. Common

5. Barberá, Jost, Nagler, Tucker & Bonneau

6. Riedl, Whipple & Wallace

7. Van Der Linden et al.

8. Cueva et al.

9. Brashier, Pennycook, Berinsky & Rand

جدول ۲. مضامین راه کارهای اینستاگرام جهت تعدیلگری اخبار جعلی در اسناد ارائه شده از سوی این پلتفرم

مضامین پایه	مضامین فراگیر
اینستاگرام با سازمان های مستقل و معتبری در زمینه بررسی صحت اخبار همکاری می کند. این سازمان ها با بررسی و ارزیابی اخبار موجود در پلتفرم، سعی در تشخیص و علامت گذاری اخبار جعلی دارند.	همکاری با سازمان های تعیین صحت اطلاعات
اینستاگرام امکان گزارش کاربران را فراهم کرده است. کاربران می توانند محتوای مشکوک یا اخبار جعلی را گزارش دهند تا تیم مدیریت محتوا بررسی و اقدامات لازم را انجام دهد.	تعبیه سیستم گزارش کاربران
اینستاگرام همکاری با خبرنگاران و رسانه های معتبر را جهت ارائه محتوای موثق و قابل اعتماد ترویج می کند. این همکاری شامل ارائه امکانات و ویژگی هایی برای رسانه ها و خبرنگاران است که آن ها را قادر می سازد محتوای خبری معتبر تر را به کاربران منتقل کنند.	همکاری با خبرنگاران
اینستاگرام با خبرنگاران مستقل و سازمان های خبری معتبر همکاری می کند تا اخبار را بررسی کنند و اطلاعات صحیح و قابل اعتماد را درباره موضوعات مختلف ارائه دهند.	همکاری با خبرنگاران مستقل
اینستاگرام به منظور آموزش کاربران درباره تشخیص اخبار جعلی و افزایش آگاهی رسانه ای، منابع و ویژگی های مختلفی را در اختیار کاربران قرار می دهد. این شامل راهنماها، ویدئوها، وبینارها و منابع آموزشی می شود.	آموزش و ترویج آگاهی رسانه ای
اینستاگرام از الگوریتم ها و روش های هوش مصنوعی برای تشخیص اخبار جعلی استفاده می کند. این الگوریتم ها با تجزیه و تحلیل محتوا، الگوها و ویژگی های خاصی که در اخبار جعلی وجود دارد، سعی در شناسایی آن ها دارند.	استفاده از الگوریتم ها و هوش مصنوعی به منظور تشخیص اخبار جعلی و ارتقای مستمر آن ها
اینستاگرام سعی می کند تا الگوریتم خود را به گونه ای طراحی کند که محتوای جعلی کمتر در صفحه اصلی کاربران نمایش داده شود. این اقدام می تواند به کاهش انتشار و پخش اخبار جعلی کمک کند.	طراحی سیستم کاهش اثرات الگوریتمی و ویروسی شدن محتوای جعلی
اینستاگرام تلاش می کند تا کاربران را آموزش دهد که چگونه اخبار جعلی را تشخیص دهند و اطلاعات را از منابع قابل اعتماد بدست آورند. آموزش های مربوط به تشخیص اخبار جعلی، استفاده از تحقیقات و منابع معتبر و ارزیابی منابع اطلاعاتی می تواند به کاربران کمک کند.	آموزش کاربران جهت تشخیص اخبار جعلی
اینستاگرام اقداماتی را برای اعتبارسنجی حساب های رسمی و معتبر انجام می دهد. این اقدام می تواند به کاربران کمک کند تا حساب هایی که اطلاعات صحیح و قابل اعتماد را منتشر می کنند را از حساب های جعلی تشخیص دهند.	اعتبارسنجی حساب های معتبر و رسمی
اینستاگرام از الگوریتم ها و تیم های مدیریت محتوا برای بررسی و علامت گذاری محتوای مشکوک و احتمالاً جعلی استفاده می کند. این اقدام می تواند به کاهش انتشار اخبار جعلی و افزایش آگاهی کاربران کمک کند.	بررسی و علامت گذاری محتوای مشکوک
اینستاگرام با استفاده از الگوریتم ها و تغییرات در سیستم خود، تلاش می کند تا دسترسی و انتشار اخبار جعلی را کاهش دهد. این شامل کاهش محتوای مربوط به اخبار جعلی در صفحه اصلی کاربران و نمایش هشدارها در مورد اخباری است که به عنوان جعلی شناخته شده اند.	کاهش دسترسی و قابلیت انتشار محتوای جعلی
اینستاگرام سیاست های سخت گیرانه تری برای محدود کردن حساب هایی که به طور مداوم اخبار جعلی منتشر می کنند، اجرا می کند. این اقدام باعث می شود که حساب های مشکوک کمتر دسترسی به مخاطبان خود داشته باشند.	محدود کردن حساب های مشکوک

جدول ۳. مضامین راه‌کارهای ایکس برای تعدیلگری اخبار جعلی در اسناد ارائه شده از سوی این پلتفرم

مضامین پایه	مضامین فراگیر
توثیق از الگوریتم‌ها و تیم‌های مدیریت محتوا برای بررسی و علامت‌گذاری محتوای مشکوک و احتمالاً جعلی استفاده می‌کند. این اقدام می‌تواند به کاهش انتشار اخبار جعلی و افزایش آگاهی کاربران کمک کند.	بررسی و علامت‌گذاری محتوای مشکوک
توثیق امکان گزارش کاربران را فراهم کرده است. کاربران می‌توانند محتوای مشکوک یا اخبار جعلی را گزارش دهند تا تیم مدیریت محتوا بررسی و اقدامات لازم را انجام دهد.	طراحی سیستم گزارش کاربران
توثیق با سازمان‌های مستقل و معتبری در زمینه بررسی صحت اخبار همکاری می‌کند. این سازمان‌ها با بررسی و ارزیابی اخبار موجود در پلتفرم، سعی در تشخیص و علامت‌گذاری اخبار جعلی دارند.	همکاری با سازمان‌های صحت اطلاعات
توثیق قابلیت‌هایی چون «لحظات خبری توثیق» و «پریبان‌کنینگ» به منظور اطلاع‌رسانی در شرایط بحران و آگاه‌سازی کاربران تعبیه کرده است.	ارائه راه‌کارهای پیش‌دستانه به منظور پیش‌آگاهی دادن به کاربران
توثیق همکاری با خبرنگاران و رسانه‌های معتبر را ترویج می‌دهد. این همکاری شامل ارائه امکانات و ویژگی‌هایی برای رسانه‌ها و خبرنگاران است که آن‌ها را قادر می‌سازد محتوای خبری معتبرتر را به کاربران منتقل کنند.	همکاری با خبرنگاران و رسانه‌های معتبر
توثیق سعی می‌کند الگوریتم‌های خود را بهبود بخشیده و فیچرهایی را به آن‌ها اضافه کند که بتوانند محتوای جعلی را شناسایی کنند و برخی از آن‌ها را کاهش دهند.	ارتقای الگوریتم‌ها جهت تشخیص اخبار جعلی
توثیق تلاش می‌کند تا حساب‌های مشکوک را شناسایی کرده و محدود کند. این اقدام می‌تواند کمک کند تا اخبار جعلی کمتر در دسترس کاربران باشد.	بررسی حساب‌های مشکوک
توثیق تلاش می‌کند کاربران را آموزش دهد که چگونه اخبار جعلی را تشخیص دهند.	آموزش کاربران جهت تشخیص درست
توثیق اقداماتی را برای اعتبارسنجی حساب‌های رسمی و معتبر انجام می‌دهد. این اقدام می‌تواند به کاربران کمک کند تا حساب‌هایی را که اطلاعات صحیح و قابل اعتماد را منتشر می‌کنند، از حساب‌های جعلی تشخیص دهند.	اعتبارسنجی حساب‌های رسمی و معتبر
توثیق با تحقیق‌گران و متخصصان در زمینه اخبار جعلی همکاری می‌کند تا راه‌کارهای بهبود و تشخیص محتوای جعلی را بررسی و اجرا کند.	همکاری با تحقیق‌گران جهت بهبود راه‌کارها
توثیق تلاش می‌کند تا با ارائه اطلاعات و آموزش‌های مربوط به تشخیص اخبار جعلی به کاربران، آگاهی آن‌ها را افزایش دهد. این شامل شناسایی علائم مشترک اخبار جعلی، روش‌های بررسی صحت اطلاعات و ارزیابی منابع می‌شود.	اطلاع‌رسانی و آموزش جهت ارزیابی صحیح
توثیق سعی می‌کند منابع معتبر خبری و اطلاعاتی را ترویج دهد و کاربران را به استفاده از آن‌ها تشویق کند. این اقدام می‌تواند کمک کند تا کاربران بهتر بتوانند اخبار جعلی را از منابع قابل اعتماد تشخیص دهند.	ترویج منابع معتبر
توثیق تلاش می‌کند روندهای واکنشی کاربران به اخبار جعلی را تحلیل و مدیریت کند. این شامل کاهش انتشار اخبار جعلی توسط روبات‌ها و حساب‌های مشکوک و همچنین اطلاع‌رسانی به کاربران درباره واکنش‌های ممکن به اخبار جعلی است.	تحلیل و مدیریت روندهای واکنشی

مضامین پایه	مضامین فراگیر
توثیتر با همکاری سازمان‌های فعال در حوزه صحت سنجی در سراسر جهان، تلاش می‌کند تا اخبار جعلی را شناسایی و محدود کند. این سازمان‌ها مسئول بررسی صحت ادعاها و اخبار منتشرشده در توثیتر هستند و در صورت تشخیص اخبار جعلی، آن‌ها را علامت‌گذاری می‌کنند.	همکاری با سازمان‌های مستقل
توثیتر از الگوریتم‌ها و روش‌های هوش مصنوعی برای تشخیص اخبار جعلی استفاده می‌کند. این الگوریتم‌ها با تجزیه و تحلیل محتوا، الگوها و ویژگی‌های خاصی که در اخبار جعلی وجود دارد، سعی در شناسایی آن‌ها دارند.	استفاده از الگوریتم‌ها و هوش مصنوعی
توثیتر کاربران را تشویق می‌کند که در صورت مشاهده اخبار جعلی، آن‌ها را گزارش دهند. این امکان به کاربران داده شده تا با استفاده از ابزارها و گزینه‌های موجود، اخبار جعلی را به توثیتر اعلام کنند تا مورد بررسی قرار گیرد.	تشویق به اطلاع‌رسانی کاربران
توثیتر سعی می‌کند با شناسایی حساب‌های مشکوک و روبات‌ها، انتشار اخبار جعلی را محدود کند. این اقدام شامل ممنوعیت ثبت حساب‌های جدید از طریق روبات‌ها و ابزارهای خودکار است.	محدود کردن انتشار و دسترسی به حساب‌های مشکوک
توثیتر در تلاش است تا با تحلیل الگوهای رفتاری کاربران و تغییرات ناگهانی در فعالیت‌های آن‌ها، حساب‌های مشکوک را شناسایی کند. این اقدام می‌تواند به کاهش انتشار اخبار جعلی توسط حساب‌های نامعتبر کمک کند.	تحلیل الگوهای رفتاری
توثیتر در سال‌های اخیر قوانین خود را برای مبارزه با اخبار جعلی و محتوای مضر تشدید کرده است. این شامل قوانینی است که ممکن است حساب‌هایی که محتوای جعلی منتشر می‌کنند را محدود کند یا آن‌ها را تعلیق کند.	اعمال قوانین سخت‌تر
توثیتر ابزارهایی را برای کاربران فراهم می‌کند تا بتوانند محتوای مشکوک را گزارش دهند. این ابزارها شامل گزارش‌های مستقیم، بلاک کردن حساب‌های مشکوک و علامت‌گذاری توییت‌ها به عنوان «مشکوک» است.	ارائه ابزارهای گزارش‌دهی به کاربران
توثیتر تلاش می‌کند تا کاربران را درباره شناسایی اخبار جعلی و محتوای مضر آموزش دهد. آن‌ها منابعی را فراهم می‌کنند که کاربران می‌توانند از آن‌ها برای یادگیری روش‌های تشخیص اخبار جعلی و ارزیابی محتوا استفاده کنند.	ارائه آموزش و منابع
همچنان در تلاش برای بهبود	ارزیابی مستمر و بهبود سیاست‌ها

در مورد راه‌کارهای پلتفرم‌ها، گفتنی است که پس از شناسایی راه‌کارهای اتخاذ شده توسط پلتفرم‌های مزبور، شباهت بسیار زیاد این راه‌کارها در حالت تئوری توجه پژوهشگر را به خود جلب کرد، بنابراین وجوه افتراق این راه‌کارها در مواردی جزئی خلاصه می‌شوند که در ادامه بدان پرداخته خواهد شد. اولین مورد مقایسه، اهداف تعیین شده توسط پلتفرم‌ها برای ایجاد روند تعدیلگری و همچنین وضع مقررات و قوانین مربوط به مداخلات است؛ هدف ایکس (توثیتر) «ایجاد جهانی آگاه‌تر، مشارکت مردم در گفت‌وگوی سالم، کاهش تهدید، توانمند سازی کاربران، فعال کردن مکالمات

آزاد» نشان‌دهنده میزان اهمیت این پلتفرم به پالایش اکوسیستم خبری و تلاش برای عدم دست‌کاری ذهنیت کاربران است؛ توئیتر همچنین یکی از اهداف خود را، توانمندسازی کاربران اعلام کرده است، امری که در اتخاذ راه‌کارهای مقابله با اخبار جعلی در این پلتفرم تأثیر فراوانی گذاشته و تا حدی آن را از سایر پلتفرم‌ها برتر ساخته است. از طرفی، هدف اینستاگرام (و در کل شرکت متا) در مورد وضع قوانین و مقررات در این زمینه، تنها تعهد آن‌ها به کاهش اطلاعات نادرست بیان شده است. میزان متعهد بودن این پلتفرم از جانب اتحادیه اروپا مقبول است؛ اما اهداف دیگری همچون توانمندسازی کاربران که در دستورالعمل‌های اتحادیه اروپا بدان اشاره شده بود، در چشم‌انداز این پلتفرم شبکه اجتماعی وجود ظاهری ندارد.

فاکتور دیگر مورد مقایسه، تعریف هر یک از پلتفرم‌ها از اخبار جعلی است؛ چرا که این تعریف نقش پررنگی در اتخاذ تصمیم‌های آتی در روند صحت‌سنجی و اعمال تعدیلگری دارد. همان‌گونه که پیش از این بدان اشاره شد، اینستاگرام در تمامی اسناد و دستورالعمل‌های خود از واژه غلط استفاده می‌کند، درحالی که سایت شبکه اجتماعی ایکس (توئیتر) واژه دیگری را برای ارجاع به این حوزه مد نظر قرار داده است که آن «اطلاعات نادرست یا اطلاعات همراه‌کننده»^۱ است. با توجه به دسته‌بندی‌های ذکر شده به‌منظور تفکیک هر یک از لغات به لحاظ مفهوم آن‌ها، می‌توان بیان کرد که لغت استفاده شده توسط اینستاگرام، طیف گسترده‌تری از مفاهیم این حوزه را پوشش می‌دهد و واژه استفاده شده توسط ایکس (توئیتر) (اطلاعات نادرست یا همراه‌کننده) را در زیرمجموعه خود دربرمی‌گیرد.

مورد بعدی، نحوه اعمال اقدامات مشابهی همچون برچسب‌گذاری یا پرچم‌گذاری، اعمال جریمه‌های متناسب و... با تفاوت‌های اندکی است؛ به‌طور مثال، در ایکس، محتوای نادرست پرچم‌گذاری شده و همراه با زمینه اضافی (به تمام زبان‌های پشتیبانی شده) به کاربر ارائه می‌شود؛ حال آنکه در اینستاگرام، محتوای غلط برچسب‌گذاری شده و برچسب حاوی رتبه‌بندی محتوا بدان تعلق می‌گیرد. پیگیری روند صحت‌سنجی نیز جزو اقدامات مشابه قرار می‌گیرد؛ چراکه پیگیری اکثریت اقدامات طبق توضیح این دو پلتفرم، همانند یکدیگر صورت می‌پذیرد. اما، اتخاذ راه‌کارهای پیش‌دستانه از جمله ایجاد لحظات خبری توئیتر، همانگونه که در بالا بدان اشاره شد، یکی از اقدامات مهم ایکس در زمان بروز یک رخداد و یا رویدادی است که احتمال انتشار اطلاعات همراه‌کننده در آن زمینه است. توئیتر سعی دارد به کمک سازمان خبری جریان اصلی که پایبند به اصول هستند، محتوایی حاوی اطلاعات درست و دست‌اول را به مخاطب نشان دهد و از بروز همراهی و سردرگمی و همچنین ایجاد تأثیر منفی به سبب مشاهده اطلاعات نادرست جلوگیری کند. یکی دیگر از راه‌های جلوگیری از سردرگمی کاربران که به‌طور ویژه در توئیتر به کارگرفته می‌شود، به کاربردن قابلیت پریانک است؛ توئیتر در اقدامی هوشمندانه، در زمان رویدادهای مهم (به‌طور مثال رویدادهای سیاسی مانند انتخابات و امثال آن)، دست به انتشار پیامی می‌زند که همانند یک واکسن می‌تواند از کاربران در برابر اطلاعات همراه‌کننده محافظت کند و همچنین راه و رسم یافتن اطلاعات نادرست و همچنین تله‌های موجود را در طول زمان به آن‌ها بیاموزد.

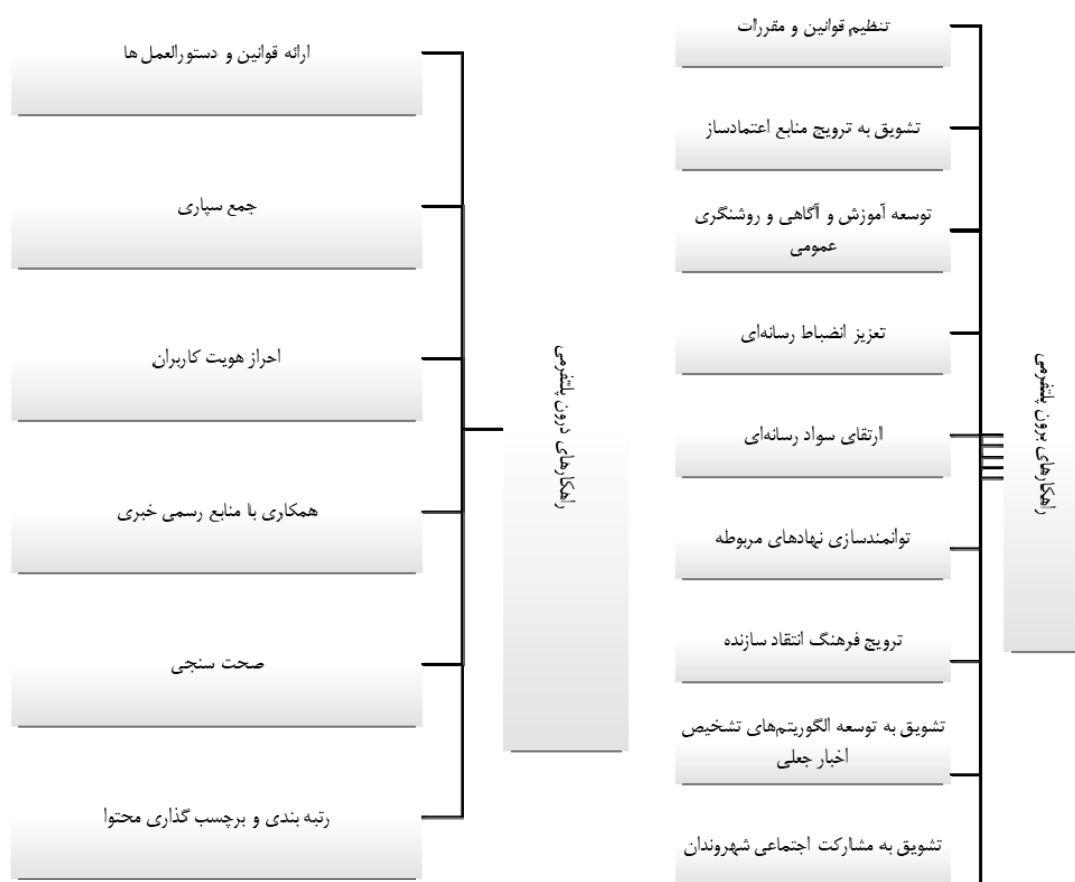
جدول ۴. مضامین پایه و فراگیر استخراج شده از مصاحبه‌ها

مضامین پایه	مضامین فراگیر
مشخص کردن مرزها و استانداردها و اطلاع‌رسانی به مخاطب در زمان ورود به پلتفرم	نوشتن دستورالعمل و قوانین مناسب پلتفرم‌ها
ایجاد سازوکار تعدیل محتوا و تدوین سند، توضیح شفاف، کامل و دقیق الگوریتم‌ها و نحوه اعمال آن‌ها	
لزوم توجه به مباحثی همچون آزادی بیان، عدم نقض حریم خصوصی	
توجه به مسائل و چارچوب‌های قانونی ایران در هنگام نگارش قوانین و دستورالعمل‌ها	
تبیین قوانین و تبصره متناسب به‌منظور برخورد با خاطیان	
تعبیه روند گزارش کاربران و تسهیل انجام آن با تعیین کردن گزینه‌های پیش فرض	رسیدگی به گزارش کاربران در اسرع وقت
درگیرسازی کاربران و درخواست مشارکت آنان در یافتن موارد مشکوک	
استفاده از تکنیک جمع‌سپاری در کنار ایجاد انگیزه برای کاربران جهت مشارکت	
اطلاع‌رسانی از روند و نتیجه اعمال گزارش کاربر به‌منظور آگاهی آن از اهمیت این عمل	
استفاده از هوش مصنوعی، الگوریتم‌ها و ربات‌ها جهت راستی آزمایی محتوا و بهره‌مندی از منابع انسانی جهت نظارت	صحت‌سنجی
طراحی مرحله «اعتراض» کاربران به نتایج صحت‌سنجی	
تعبیه سیستم هشدار به مخاطب و آگاه کردن آنان از محتوای مشکوک و در دست بررسی	برچسب‌گذاری
هشدارهای پیش‌دستانه به کاربران در مورد موضوعات جاری و هشدار به آنان	
ارجاع‌دهی کاربران به سایت‌های معتبر برای دریافت اخبار درست	
طراحی برچسب‌های مطلوب به‌منظور جلب نظر مخاطب و عدم ایجاد مزاحمت و کاهش تجربه کاربری مثبت	
دسته‌بندی و رتبه‌بندی محتوا بر اساس میزان صحت یا جعلی بودن آن	رتبه‌بندی محتوا
استفاده از واژگان و رنگ‌های مختلف به‌منظور تسهیل روند تفکیک اخبار برای کاربران	
فراهم کردن امکان و تسهیل فرایند گزارش برای کاربران و متناسب‌سازی آن	تعبیه سیستم گزارش کاربران
طراحی گزینه‌های مشخص و شفاف و در نهایت، گزینه سایر موارد به‌منظور دسترسی دقیق	
ایجاد شناسنامه‌ای برای محتوا متشکل از منبع، تاریخ و نحوه انتشار	شناسه محتوا
در معرض دید قراردادن شناسه محتوا هنگام بازنشر و پسندیدن محتوا توسط کاربران	
مخفی کردن محتوا از دید سایر کاربران در طول زمان تشخیص اصالت آن	کاهش دید یا حذف محتوا
کاهش دید و یا حذف محتوا پس از تشخیص قطعی جعلی بودن محتوا متناسب با تأثیر خبر و تعداد افراد دنبال کننده	
تعبیر الگوریتم و نشان ندادن محتوای جعلی در فهرست خوراک و جست‌وجوی محتوا	

مضامین پایه	مضامین فراگیر
لزوم تعبیه قوانین محدودسازی و محروم سازی برای نشردهندگان و بازنشردهندگان محتوای جعلی متناسب با حساسیت انتشار، دفعات تکرار خطا و تعداد افراد دنبال کننده آنها	تعبیه قوانین سخت گیرانه برای نشر و بازنشردهندگان
مسئول دانستن بازنشر دهندگان اخبار جعلی و اعمال قوانین محدودسازی فعالیت کاربر برای آنان در صورت تکرار بازنشر (به منظور ایجاد دقت و اهمیت)	
راه اندازی سیستم تشخیص حقیقی بودن کاربران و تعبیه فرایند رفع اتهام در صورت تشخیص غلط مشخص کردن منبع هر خبر به منظور ایجاد توانایی برای پیگیری آن	تشخیص هویت کاربران
دریافت مشخصات کاربران پس از تزریق حس امنیت اطلاعات آنان	
دریافت اطلاعات کاربران از طرق مختلفی چون تعبیه سیستم های خرید، پرداخت های کارتی و ...	
تعبیه راه کارهای جذاب و مخاطب پسند به منظور اتخاذ اطلاعات کاربران	
همکاری با مؤسسه های مستقل و فعال در زمینه راستی آزمایی و اصالت خبر	همکاری با موسسات فعال در حوزه صحت سنجی
استفاده بیشتر از هوش مصنوعی نسبت به منابع انسانی به دلیل حجم بالای محتوا، سرعت بسیار زیاد انتشار آن و صرفه اقتصادی پلتفرم	استفاده از هوش مصنوعی و نیروی انسانی
تعبیه نظارت ماشینی برای تمامی محتواها و در نظر گرفتن منابع انسانی به منظور نظارت بر عملکرد ماشین ها به طور رندوم و پاسخ گویی به شکایات کاربران و درخواست رفع اتهام آنها	
کمک گرفتن از رسانه های جریان اصلی داخلی به منظور تشخیص اصالت خبر و تسهیل فرایند راستی آزمایی (هرچند که مرجعیت آنان تضعیف شده باشد)، بهتر از کمک نگرفتن از آنان است.	همکاری با رسانه های جریان اصلی
آموزش مهارت های ارتباطی در عصر فضای مجازی به کاربران در قالب های متفاوت و توسط نهادها، ارگان ها و پلتفرم ها ضروری است	آموزش کاربران
آموزش سواد رسانه ای، تفکر انتقادی و ... برای تمامی کاربران، بالاخص سلبریتی ها و افراد محبوب جامعه	
شروع روند آموزش از دوران کودکی و نوجوانی به دلیل نهادینه شدن آن	
آگاه سازی و مسئولیت دادن به نهادهای دولتی مختلف جهت رسیدگی به امر آموزش کاربران برای استفاده بهینه از فضای مجازی	
نگارش نکات کلیدی به منظور تشخیص محتوای جعلی از جانب پلتفرم	
وجود مرجع عادل و بی طرف به منظور بررسی حقایق و ورود قانونی به مسائل در موارد خاص	پاسخ گویی به یک مرجع قانونی و قوانین متقن جهت رسیدگی به موضوعات
ایجاد سامانه برای سرویس دهی به سکوها و درگاه های نشر	
تصمیم گیری برای تشخیص حرکت در جهت هم تنظیم گری و اتخاذ قوانین بازدارنده یا دادن آزادی عمل به پلتفرم ها	

در مجموع مقالات مطالعه شده و همچنین سخنان خبرگان، پژوهشگران و فعالان حوزه سیاست گذاری در داخل کشور، به لزوم ارائه راه کارها و آموزه های مهارتی توسط سازمان ها، مؤسسه ها، نهادها و بنگاه های رسانه ای به عنوان شیوه ای مؤثر برای توسعه آموزش سوادهای مختلف همچون سواد رسانه ای، سواد خبری، سواد بصری و سواد اطلاعاتی

در کنار تلاش‌های مستمر پلتفرم‌ها جهت جلوگیری از ویروسی شدن این اخبار توجه شده است. در این راستا، چارچوب احصا شده به‌منظور تعدیلگری محتوای جعلی در پلتفرم‌های شبکه‌های اجتماعی داخلی (شکل ۱) را می‌توان به دو دسته درون پلتفرمی و برون پلتفرمی تقسیم کرد. راه کارهای درون پلتفرمی به تلاش سکوها جهت تعدیلگری اخبار جعلی اشاره دارد، در حالی که راه کارهای برون پلتفرمی به بسترسازی، تسهیلگری، حمایت دولتی و قانونی و همچنین ایجاد قوانین و مرجع در این حوزه اشاره می‌کنند.



شکل ۱. چارچوب تعدیلگری اخبار جعلی در پلتفرم‌های شبکه‌های اجتماعی

بحث و نتیجه‌گیری

اهمیت مقابله با اخبار جعلی، به‌عنوان یک چالش مهم در حوزه رسانه‌ها و جامعه‌شناسی، بر اساس تأملی عمیق و واقع‌گرایی مبتنی بر ارزش‌های اجتماعی، به آشکار ساختن و مبارزه با این پدیده آفت‌بار می‌پردازد. در جهانی که ارتباطات دیجیتال و رسانه‌های اجتماعی از اهمیت و سوسه‌انگیز برخوردارند، اخبار جعلی به‌عنوان ابزاری مؤثر برای ترویج اطلاعات نادرست و تضعیف اعتماد عمومی، می‌تواند پایه‌های حقیقت و صداقت را در جامعه به خطر بیندازد. تأثیر این اخبار و محتوا بر جامعه، از انقسامات اجتماعی تا فروپاشی اعتماد، درخور توجه است. آن‌ها می‌توانند نه تنها آگاهی عمومی را

تحت تأثیر قرار دهند، بلکه به خطر امنیت فردی و تمدنی نیز بیبوندند. از این رو، درک و شناخت ما از محتوای جعلی و تلاش برای مقابله با آن‌ها، به‌عنوان مأموریتی عمومی، بر تعزیز امنیت معنوی و فرهنگی جامعه تأکید می‌کند. به همین جهت، ضرورت مقابله با اخبار جعلی بر جامعه روشنگری‌هایی می‌افکند و نمادی است از توجه جامعه به قیمت‌های حقیقت، صداقت و تعهد به ارائه اطلاعات دقیق و موثق. این مبارزه، به همکاری همه اقشار جامعه نیازمند است.

در ابتدا باید به این مسئله توجه شود که محتوای جعلی در چه شرایطی بروز و ظهور می‌کنند. این محتوا زمانی که التهاب در بطن جامعه و مسائل خردوکلان در سطوح فردی، ملی و بین‌المللی فراگیر می‌شود و منابع متقن و قابل اعتماد وجود نداشته باشند یا از اظهار نظر خودداری کنند به‌وجود می‌آیند. در این شرایط اطلاع‌رسانی دیر هنگام یا محدود و همچنین عدم وجود اخبار تأیید شده موجب می‌شود که نتوان با استفاده از اخبار صحیح و تأیید شده به مبارزه محتوای جعلی پرداخت. انتشار ویروسی اخبار در ساعات اولیه و ناممکن بودن پیگیری منبع اصلی نشردهنده، در ابتدای امر نیز از مواردی هستند که به دشوارتر ساختن این جریان دامن می‌زنند. از این جهت، تعدیل محتوای جعلی امری ضروری است که به‌دلیل چندوجهی بودن آن نیازمند تلاش و سیاست‌گذاری‌های همه‌جانبه است. پژوهش حاضر در تلاش است تا با احصای تمامی نکات و راه‌کارهای ضروری از بطن قوانین سایر پلتفرم‌ها و بومی‌سازی آن‌ها طبق نظرهای خبرگان این حوزه، چارچوبی به‌منظور تعدیلگری اخبار جعلی در پلتفرم‌های شبکه‌های اجتماعی ارائه دهد و این نقطه تمایز پژوهش حاضر با مقاله‌های پیشین داخلی است؛ به‌طوری که در مقاله حسنی (۱۴۰۱)، مراحل و روش‌های بازبینی محتوا در پلتفرم‌های فیسبوک و توئیتر تحلیل و سازو کارهای آن‌ها شناسایی شد اما چارچوبی جهت استفاده پلتفرم‌های داخلی طراحی و ارائه نشد.

همان‌طور که ساعی و آزادی (۱۴۰۰) به‌درستی بدان اشاره کرد، از جمله اقدامات مهم در فرایند تعدیلگری اخبار جعلی، ایجاد قوانین توسط دولت توأم با تشخیص خودکار توسط پلتفرم‌های رسانه‌های اجتماعی است که می‌تواند موجب مسئولیت‌پذیر کردن پلتفرم‌ها و همچنین تشخیص، حذف محتوا و حساب‌های کاربری جعلی شود؛ بنابراین، در این مقاله نیز راه‌کارهای موجود در دو دسته «برون پلتفرمی» و «درون پلتفرمی» طبقه‌بندی شد، به‌طوری که در دسته اول، راه‌کارهایی همچون تنظیم قوانین و مقررات، تشویق به ترویج منابع اعتمادساز، توسعه آموزش و آگاهی و روشنگری عمومی، تعزیز انضباط رسانه‌ای، ارتقای سواد رسانه‌ای، توانمندسازی نهادهای مربوطه، ترویج فرهنگ انتقاد سازنده، تشویق به توسعه الگوریتم‌های تشخیص اخبار جعلی و تشویق به مشارکت اجتماعی شهروندان قرار گرفته است و دسته دوم نیز به راه‌کارهایی چون ارائه قوانین و دستورالعمل‌ها، جمع‌سپاری، احراز هویت کاربران، همکاری با منابع رسمی خبری، صحت‌سنجی، رتبه‌بندی و برچسب‌گذاری محتوا اشاره دارد. در خصوص راه‌کارهای برون پلتفرمی، ذکر این نکته حائز اهمیت است که نقش دولت در تعدیل محتوای پلتفرم‌ها ممکن است از نظارت، تنظیم قوانین، تشویق و همکاری تا تحریم‌ها و تنبیه‌ها متفاوت باشد و به سیاست‌ها و قوانین هر کشور بستگی دارد. در ادامه نقش دولت‌ها و لزوم اتخاذ راه‌کارهای برون پلتفرمی به‌منظور تعدیل محتوا در این حوزه، روشن و شفاف بودن برنامه‌ریزی زیرساخت‌ها و تبیین قوانین و مطالبات منطقی و عملی حکومت از پلتفرم‌ها امری ضروری است؛ به‌طور مثال، ارائه راه‌کارهای به‌روز و واضح در کنار ارائه گزارش عملکرد شفاف می‌تواند از جمله قوانین دستگاه‌های اجرایی از پلتفرم‌های فعال در داخل کشور باشد.

در همین ابتدا، ملزم دانستن پلتفرم‌ها به تبیین قوانین و اعمال آن بسیار حائز اهمیت است؛ زیرا برنامه کسب‌وکار اکثریت پلتفرم‌ها بر پایه جذب بیشتر مخاطب و فعالیت بیشتر کاربران در محیط پلتفرم است؛ از همین رو، در صورت عدم وجود پیگیری و الزامات حکومتی، پلتفرم‌ها گاهی پیگیر اعمال قوانین به‌منظور محدود ساختن کاربران نخواهند بود. درخصوص راه‌کارهای درون پلتفرمی، تعدیلگری محتوای جعلی توسط پلتفرم‌ها می‌تواند به ایجاد فضای اطلاعاتی سالم و قابل اعتماد کمک کند. این اقدام می‌تواند به حفظ اعتماد عمومی، افزایش امنیت جامعه، حمایت از اطلاعات صحیح و حقوق شهروندان و بهبود کیفیت ارتباطات آنلاین منجر شود. با اینکه تعدیلگری به تنهایی نمی‌تواند مشکل اخبار جعلی را به‌طور کامل حل کند، اما با ترکیب آن با سایر راه‌کارهای برون پلتفرمی، مانند آموزش عمومی و همکاری با منابع خبری معتبر، می‌توان به کاهش انتشار اخبار جعلی و بهبود محتوای دسترسی‌پذیر رسید. این راه‌کارها می‌بایست با در نظر گرفتن حقوق کاربران و ارزشمند شمردن نظرات و بازخوردهای آنان طراحی و اجرا شوند و البته در مورد کاربست این قوانین در پلتفرم‌های داخلی، با در نظر گرفتن عدم وجود پلتفرم شبکه‌های اجتماعی ایرانی تاکنون، مسئله تدوین سند در قدم نخست و پایه‌ریزی دستورالعمل‌های قابل اجرا امری ضروری تلقی می‌شود. استفاده از تکنیک جمع‌سپاری در کنار جلب همکاری منابع رسمی و مورد اعتماد اخبار در داخل کشور، امری مهم است. در ادامه به ذکر توصیه‌های عملی برون و درون پلتفرمی احصا شده از مقالات مختلف، نظرات خبرگان و نوآوری نویسندگان ارائه می‌شود.

توصیه‌های عملی برون پلتفرمی:

- تنظیم قوانین و مقررات
دولت می‌تواند قوانین و مقررات مناسبی را درباره محتوای جعلی در پلتفرم‌های داخلی تعیین کند. این قوانین می‌توانند شامل جریمه‌ها و تنبیه‌های قانونی برای اشخاص یا سازمان‌هایی باشند که در انتشار اخبار جعلی دخیل هستند.
- تشویق به ترویج منابع اعتمادساز
این منابع می‌توانند شامل رسانه‌های قابل اعتماد، سازمان‌های خبری مستقل، و نهادهای علمی و تحقیقاتی باشند. تشویق و حمایت از این نهادها در تولید و انتشار اخبار دقیق و موثق، در مقابله با اخبار جعلی اهمیت زیادی دارد.
- توسعه آموزش و آگاهی و روشنگری عمومی
در این راستا، برگزاری کارگاه، سمینار و تولید محتوای آموزشی منظم می‌تواند به افزایش آگاهی عمومی و مهارت‌های تشخیص اخبار جعلی کمک کند. دولت می‌تواند برنامه‌های آموزشی و روشنگری عمومی را برای افزایش آگاهی مردم درباره شناسایی اخبار جعلی و روش‌های مقابله با آن‌ها راه‌اندازی کند.
- تعزیر انضباط رسانه‌ای
قوانین و مقررات مربوط به رسانه‌ها و پلتفرم‌های دیجیتال، باید به‌گونه‌ای تعیین شود که ترویج اخبار جعلی و اطلاعات نادرست را محدود کند. این قوانین باید شامل تعیین مسئولیت‌ها و تعهدات صحیح برای رسانه‌ها،

سازوکارهای شفافیت و اعتبارسنجی در انتشار اخبار و مجازات‌های مناسب برای منتشرکنندگان اخبار جعلی باشند. تقویت نظارت و کنترل بر فعالیت رسانه‌ها و پلتفرم‌های دیجیتال نیز از اهمیت ویژه برخوردار است.

- ارتقای سواد رسانه‌ای
دولت می‌تواند برنامه‌های آموزشی و آگاهی‌بخشی را برای افزایش سواد رسانه‌ای و توانایی شناسایی اخبار جعلی در بین مردم اجرا کند. این برنامه‌ها می‌توانند شامل آموزش مهارت‌های تحلیل محتوا، تشخیص منابع معتبر و گزارش اخبار جعلی باشند.
- توانمندسازی نهادهای مربوطه
نهادهای مربوطه مانند وزارت ارتباطات و فناوری اطلاعات، پلیس فتا و نهادهای قضایی باید توانایی و منابع لازم برای مقابله با اخبار جعلی داشته باشند. تقویت زیرساخت‌های فنی و تکنولوژیک، افزایش ظرفیت افراد و تعیین وظایف و مسئولیت‌های دقیق برای هر نهاد از جمله اقداماتی است که می‌تواند در این راستا مؤثر باشد.
- ترویج فرهنگ انتقاد سازنده
تشویق مردم به ارائه انتقادهای سازنده و مبارزه با اخبار جعلی، می‌تواند به تقویت مقاومت جامعه در برابر این پدیده کمک کند. ایجاد فضایی آزاد برای بحث و تبادل نظر، تشویق به تحقیق و تأیید اطلاعات و ترویج فرهنگ منصفه‌گری و اعتماد به مردم، از اهمیت بالایی در این زمینه برخوردار است.
- تشویق به توسعه الگوریتم‌های تشخیص اخبار جعلی
دولت می‌تواند از شرکت‌های فناوری مربوطه در داخل کشور تقاضا کند تا الگوریتم‌های هوش مصنوعی و یادگیری ماشین را توسعه دهند که بتوانند اخبار جعلی را با دقت شناسایی کنند و در صورت لزوم، اقدامات لازم را انجام دهند. این گام می‌تواند با حمایت دولت از برقراری زیرساخت‌ها و تشویق عملی پلتفرم‌ها صورت پذیرد.
- تشویق به مشارکت اجتماعی شهروندان
دولت می‌تواند شهروندان را تشویق کند تا در شناسایی و گزارش اخبار جعلی به مراجع ذیصلاح مشارکت کنند. ایجاد سازوکارهای ساده و در دسترس برای گزارش اخبار جعلی می‌تواند به شناسایی و حذف سریع‌تر آن‌ها کمک کند.

توصیه‌های عملی درون پلتفرمی

- ارائه قوانین و دستورالعمل‌ها
یکی از نکات مهم در تدوین اسناد، ایجاد سازوکارهای تعدیلگری محتوا است. توضیحات شفاف و دقیق درباره نحوه برخورد الگوریتم‌های تعبیه شده با کاربران در مواقع خاص نیز بسیار ضروری است. این توضیحات باید به کاربران اطلاع دهند که الگوریتم‌ها در چه شرایطی تغییر می‌کنند، چگونه تصمیماتی اتخاذ می‌شود و آیا تداخلی با حقوق و آزادی‌های شخصی دارد یا خیر. در روند اتخاذ این قوانین، باید به اعمال محدودیت، محرومیت و حذف محتوا متناسب با بسامد خبری آن و تعداد دنبال‌کنندگان هر صفحه توجه داشت. یکی دیگر از نکات حائز اهمیت در این

موضوع، توجه به مسئول دانستن بازنشر دهندگان همانند نشردهندگان محتوای جعلی است، هرچند قوانین آسان‌تری برای برخورد با بازنشردهندگان باید وضع شود؛ اما در نهایت باید آنان را نیز مسئول پخش محتوای جعلی در یک پلتفرم و درگیرسازی عده بیشتری از مخاطبان با آن دانست. تأکید می‌شود که این کار حتی هر چه زودتر آغاز شود. تعیین قوانین و دستورالعمل و تدوین اسناد مربوطه باید به‌عنوان یک اولویت مهم در نظر گرفته شود؛ زیرا این اقدام می‌تواند به شفافیت و افزایش اعتماد کاربران به پلتفرم‌ها منجر شود. تدوین اسناد و دستورالعمل‌های دقیق و شفاف برای پلتفرم‌ها اهمیت زیادی دارد.

• جمع‌سپاری

یکی از موارد مهم در تکنیک جمع‌سپاری، درخواست مشارکت کاربران و تشویق آن‌ها به این امر است؛ به‌طوری که پلتفرم به کاربران خود اجازه می‌دهد تا در فرایند تشخیص محتوا مشارکت کرده و تأثیرگذار باشند. با این روش، کاربران می‌توانند به‌عنوان اعضای فعال در فرایند تشخیص محتوا شرکت کرده و تجربه‌ها، دیدگاه‌ها و نظرهای خود را به اشتراک بگذارند. این امر به کاربران امکان می‌دهد تا به‌طور مستقیم در ارزیابی و دسته‌بندی محتواها مشارکت داشته و نیز مواردی را اصلاح کنند که ممکن است به‌طور اشتباه تشخیص داده شوند. پلتفرم‌ها می‌بایست رسیدگی دقیق به این گزارش‌ها و اطلاع‌رسانی به گزارش‌دهندگان در مورد نحوه و نتایج ارزیابی را در دستور کار خود قرار دهند. یکی دیگر از تکنیک‌های جمع‌سپاری، همکاری و استفاده از دانش و دسترسی گسترده رسانه‌های معتبر است. با این فرض که پلتفرم‌ها رسانه نیستند و مسئولیت نشر محتوا را ندارند؛ اما چون مسئولیت پاک‌سازی بستر فراهم شده جهت نشر محتوا را به عهده دارند، نمی‌توان از آن‌ها انتظار نشر اخبار یا تکذیبیه‌های منتشر شده پس از انتشار یک خبر جعلی را داشت. از همین جهت، استفاده از ظرفیت رسانه‌های معتبر و نتایج ارزیابی آنان می‌تواند موجب تسریع روند صحت‌سنجی و روایی هر چه بیشتر آن شود.

• احراز هویت کاربران

یکی از روش‌های معمول استفاده از شماره تلفن همراه است. در این روش، کاربران برای ثبت نام یا ورود به پلتفرم، باید شماره تلفن همراه خود را وارد کنند و یک کد تأیید به آن‌ها ارسال می‌شود. با وارد کردن کد تأیید، هویت کاربر تأیید می‌شود و امکان استفاده از پلتفرم به او داده می‌شود. روش دیگر استفاده از اطلاعات هویتی از قبیل شماره ملی، کارت ملی یا اطلاعات بانکی است. با ارائه این اطلاعات، کاربر می‌تواند هویت خود را تأیید کند و از قابلیت‌های کامل پلتفرم استفاده کند. شایان ذکر است، امروزه تکنیک‌های نوینی از جمله استفاده از هوش مصنوعی و تحلیل الگوهای رفتاری کاربران و واکنش آنان در بازه‌های زمانی مشخص جهت اعتبارسنجی هویت کاربران استفاده می‌شود.

• همکاری با منابع رسمی خبری

۱. ایجاد معیارهای شناسایی و تأیید اخبار رسمی: پلتفرم‌ها و منابع رسمی خبری می‌توانند با همکاری تعیین معیارهایی برای شناسایی و تأیید اخبار رسمی ایجاد کنند. این معیارها می‌توانند شامل منابع اعتبارسنجی،

- استانداردهای خبرنگاری، و روش‌های ارائه مطالب باشند. این معیارها به پلتفرم‌ها کمک می‌کنند تا اخبار رسمی و قابل اعتماد را از میان اطلاعات نادرست و جعلی تشخیص دهند و به کاربران نشان دهند.
۲. ایجاد همکاری در فرایند تأیید اخبار: پلتفرم‌ها می‌توانند با منابع رسمی خبری در فرایند تأیید اخبار همکاری کنند. به‌طور مثال، در صورتی که یک خبر به یک پلتفرم ارسال شود، پلتفرم می‌تواند با منبع رسمی خبری همکاری کند تا قبل از انتشار، از صحت و قابل اعتماد بودن آن خبر مطمئن شود.
 ۳. ارائه منابع اعتبارسنجی: پلتفرم‌ها می‌توانند با همکاری منابع رسمی خبری اعتبارسنجی را در اخبار ارائه کنند. این منابع می‌توانند شامل لینک‌های مرتبط به خبرنگاری رسمی، سازمان‌های خبری معتبر و منابع دیگر باشند. با ارائه این منابع، پلتفرم‌ها به کاربران امکان مطالعه و بررسی بیشتری درباره خبرها و منابع اعلام می‌کنند و از آن‌ها دعوت می‌کنند که اطلاعات خود را از منابع رسمی و قابل اعتماد به دست آورند.
 ۴. تشویق به حضور منابع رسمی در پلتفرم‌ها: پلتفرم‌ها می‌توانند منابع رسمی خبری را تشویق کنند تا در خود پلتفرم‌ها حضور داشته باشند و اخبار را مستقیماً منتشر کنند. این امر به کاربران اعتماد بیشتری به اخبار رسمی و انتشار سریع‌تر آن‌ها منجر می‌شود.
 ۵. ایجاد سیستم اعلان‌ها و هشدارها: پلتفرم‌ها می‌توانند با همکاری منابع رسمی خبری سیستم اعلان‌ها و هشدارهایی را برای اخبار رسمی و مهم ایجاد کنند. این سیستم به کاربران اجازه می‌دهد تا به‌راحتی از اخبار رسمی و ضروری آگاه شوند و از اطلاعات جعلی جدا شوند.
 ۶. ارائه محتوای آموزشی: پلتفرم‌ها می‌توانند با همکاری منابع رسمی خبری محتوای آموزشی درباره تشخیص اخبار جعلی و تأیید اخبار رسمی ارائه کنند. این محتوا می‌تواند شامل راهنماها، ویدئوها و مقالات آموزشی باشد که به کاربران کمک می‌کند تا بتوانند اخبار را به درستی ارزیابی کنند و به منابع رسمی خبری رجوع کنند.

● صحت سنجی

۱. شناسایی اخبار مشکوک: ابتدا باید اخباری که ممکن است جعلی باشند و یا دارای اطلاعات نادرست باشند، شناسایی شوند. برای این کار می‌توان از الگوریتم‌ها و سیستم‌های هوش مصنوعی استفاده کرد که بر اساس معیارهای مشخصی اخبار مشکوک را تشخیص دهند. این معیارها می‌توانند شامل عناوین حساس، اطلاعات غیرقابل تأیید، منابع نامعتبر، یا تغییرات در الگوی انتشار اخبار باشند.
۲. تحلیل و ارزیابی اخبار: اخبار شناسایی شده باید تحلیل و ارزیابی شوند تا صحت و قابل اعتمادی آن‌ها بررسی شود. این مرحله شامل بررسی منابع، تعامل با خبرنگاران و منابع رسمی، بررسی اطلاعات و آمارها و ارزیابی روش‌های گزارش‌دهی است. همچنین، تحلیلگران و صفحه‌بندان می‌توانند از ابزارهای دیجیتال و پایگاه‌های داده معتبر برای تأیید صحت اخبار استفاده کنند.
۳. بررسی منابع: یکی از مراحل مهم در صحت‌سنجی اخبار، بررسی منابع است. باید بررسی شود که آیا منبع خبری قابل اعتماد است یا خیر. این بررسی می‌تواند شامل بررسی سابقه و شهرت منبع، بررسی روش‌های گزارش‌دهی و نگارش و بررسی معتبریت و اعتبارسنجی منبع باشد.

۴. تأیید محتوا با منابع رسمی: در مرحله بعد، می‌توان به منابع رسمی خبری مراجعه کرده و صحت اخبار را با آن‌ها تأیید کرد. این منابع می‌توانند شامل سازمان‌های خبری معتبر، سازمان‌های دولتی، یا منابع رسمی دیگر باشند. اما این مراحل برای صحت سنجی اخبار جعلی در پلتفرم‌ها کاربرد دارند. با استفاده از این تکنیک‌ها، امکان کاهش انتشار اخبار جعلی و اطلاعات نادرست وجود دارد. البته، همچنین لازم است که کاربران نیز آگاهی کافی درباره شناسایی اخبار جعلی و تأیید منابع داشته باشند تا بتوانند از این تکنیک‌ها بهره‌برداری کنند.

• رتبه‌بندی و برچسب گذاری محتوا

۱. راه‌اندازی الگوریتم و سامانه رتبه‌بندی: برای اجرای این تکنیک، یک الگوریتم و سامانه رتبه‌بندی نیاز است. این سامانه باید بر اساس یک سری معیارها و الگوریتم‌های مشخص، محتوا را ارزیابی کند و به آن رتبه‌بندی اعمال کند. این معیارها می‌توانند شامل معتبریت منبع، سابقه و شهرت منبع، تطابق با منابع رسمی، میزان تأیید صحت توسط سازمان‌های خبری قابل اعتماد و سایر معیارهای مشخص شده باشند.

۲. برچسب گذاری محتوا: برای اطلاع‌رسانی به کاربران، می‌توان برچسب‌های متناسب با رتبه‌بندی محتوا ارائه کرد. این برچسب‌ها می‌توانند عبارت‌هایی مانند «معتبر»، «مشکوک»، «اخبار جعلی»، «احتمالاً جعلی» و غیره باشند. این برچسب‌ها به کاربران کمک می‌کنند تا به سرعت متوجه صحت محتوا شوند و در انتخاب و مصرف اطلاعات دقت بیشتری داشته باشند. توجه به حالت گرافیکی این برچسب‌ها نیز مهم است، این برچسب‌ها باید به گونه‌ای طراحی شوند که هم جلب توجه کنند و هم موجب افت تجربه کاربران از فضای پلتفرم نشوند.

۳. بهبود الگوریتم‌ها و آموزش مداوم: الگوریتم‌ها و سیستم‌های هوش مصنوعی مورد استفاده در تکنیک رتبه‌بندی محتوا باید به طور مداوم بهبود یابند و آموزش داده شوند. با توجه به تکامل روزافزون اخبار جعلی و تغییرات روش‌های تولید آن‌ها، الگوریتم‌ها باید به روزرسانی شوند و بهبود یابند تا بتوانند به طور مؤثرتری اخبار جعلی را تشخیص دهند و آن‌ها را رتبه‌بندی کنند.

۴. شفافیت برای کاربران: پلتفرم‌ها باید به کاربران شفافیت بیشتری درباره فرایند رتبه‌بندی محتوا و برچسب گذاری ارائه دهند. کاربران باید بدانند که چگونه محتواها رتبه‌بندی می‌شوند و برچسب‌ها به چه معنا و مبتنی بر چه معیارهایی هستند. این شفافیت به کاربران کمک می‌کند تا بهتر درک کنند که چگونه محتواها ارزیابی می‌شوند و به اعتماد بیشتری در پلتفرم‌ها دست پیدا کنند.

۵. اطلاع‌رسانی به کاربران: در نهایت، اطلاع‌رسانی به کاربران درباره محتوا و باخبر کردن آن‌ها از دسته‌بندی‌های صورت گرفته درباره محتوایی که آن‌ها پیش از این منتشر یا بازنشر داده‌اند، ضروری است. این کار شاید بتواند تلنگری برای کاربران جهت یادگیری سواد رسانه‌ای و بالا بردن دقت و اهمیت محتوای بازنشر شده باشد.

در پایان ذکر این نکته ضروری این است که افزایش اعتماد عمومی به این پلتفرم‌ها و جلب توجه و رضایت مردم از آن‌ها، به عنوان یک هدف اساسی در نظر گرفته شود. برای رسیدن به این هدف، می‌توان از راه کارهای نوین و پیشرفته استفاده کرد که در سایر پلتفرم‌های معتبر و قابل اعتماد مورد استفاده قرار می‌گیرند. با توجه به موفقیت و پذیرشی که این

پلتفرم‌ها در بین مردم ایران داشته‌اند، اعمال چنین راه‌کارهایی می‌تواند به جلب توجه و اعتماد آن‌ها کمک کند. البته، برای دستیابی به اهداف موردنظر، تسریع فرایند و در اولویت قرار دادن نیاز کاربران باید در دستورکار قرار گیرد. به‌علاوه، اتخاذ راه‌کارها و قوانین قابل قبول می‌تواند نشان‌دهنده اهمیت باشد که این پلتفرم‌ها به کاربران خود می‌دهند و از طریق این اقدام می‌توانند نظر و رضایت کاربران را جلب کنند. در گذشته، اعمال فیلترینگ سخت، نتوانسته است که علاقه مردم به پلتفرم‌های خارجی را کاهش دهد و استفاده از پلتفرم‌های داخلی را به‌طور قابل توجهی افزایش دهد. بنابراین، لازم است تلاش بیشتری در این زمینه انجام شود.

امید است که راه‌کارهای ارائه شده در این مقاله نتیجه‌بخش و مفید بوده و بهبود قابل توجهی در تعدیلگری اخبار جعلی در پلتفرم‌های داخلی ایجاد کند. این مسئله نه تنها به ارتقای اعتماد عمومی به این پلتفرم‌ها کمک خواهد کرد، بلکه می‌تواند استفاده از آن و رسیدن به اقبال عمومی را به‌طور چشمگیری افزایش دهد.

منابع

- حسینی، حسین (۱۴۰۱). تحلیل سازوکارهای پلتفرم‌های رسانه‌های اجتماعی برای مقابله با اطلاعات کذب و اخبار جعلی؛ فیسبوک و توئیتر. *مطالعات هنر و رسانه*، ۴(۱)، ۱۵۵-۱۸۴.
- رحمانیان، عماد (۱۴۰۰). نوع‌شناسی اخبار جعلی شبکه‌های اجتماعی در بافت مصرف ایران، الگوی مفهومی مردم‌نگارانه. مدیریت بازرگانی، ۱۳(۳)، ۸۱۴-۸۴۴.
- ساعی، محمد حسین؛ آزادی، محمدحسین (۱۴۰۰). راه‌کارهای مقابله با اخبار جعلی در سازمان‌های خبری و رسانه‌ای ایران. *فصلنامه علمی رسانه*، ۳۲(۲)، ۳۵-۷۸.
- مجدی‌زاده، زهرا و موسوی حق‌شناس، سید میلاد (۱۳۹۹). چیستی و چرایی اخبار جعلی در عصر دیجیتال: فراتحلیل کیفی. *فصلنامه علمی مطالعات میان‌رشته‌ای ارتباطات و رسانه*، ۳(۷)، ۹۳-۱۲۴.

References

- Abreu, A. I. A., de Zúñiga Navajas, H. G., & McCombs, M. E. (2020). Orígenes y desarrollo de la teoría de la agenda setting en Comunicación. *Tendencias en España (2014-2019). El profesional de la información*, 29(4), 18.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber?. *Psychological science*, 26(10), 1531-1542.
- Barclay, D.A. (2018). *Fake News, Propaganda, and Plain Old Lies*. Rowman & Littlefield Publishers.
- Brashier, N. M., Pennycook, G., Berinsky, A. J., & Rand, D. G. (2021). Timing matters when correcting fake news. *Proceedings of the National Academy of Sciences*, 118(5), e2020043118.

- Calzada, J. & Gil, R. (2020). What do news aggregators do? Evidence from Google News in Spain and Germany. *Marketing Science*, 39(1), 134-167.
- Carlson, M. (2018). Facebook in the News: social media, Journalism, and Public Responsibility following the 2016 Trending Topics controversy. *Digital Journalism* 6(1), 4-20.
- Common, M. F. (2020). Fear the reaper: How content moderation rules are enforced on social media. *International Review of Law, Computers & Technology*, 34(2), 126-152.
- Cueva, E., Ee, G., Iyer, A., Pereira, A., Roseman, A., & Martinez, D. (2020, October). Detecting fake news on twitter using machine learning models. In *2020 IEEE MIT undergraduate research technology conference (URTC)* (pp. 1-5). IEEE.
- Evans, R., Jackson, D. & Murphy, J. (2023). Google News and machine gatekeepers: Algorithmic personalisation and news diversity in online news search. *Digital Journalism*, 11(9), 1682-1700.
- Gallo, J. A., & Cho, C. Y. (2021). Social media: Misinformation and content moderation issues for congress. *Congressional Research Service Report*, 46662.
- Gaozhao, D. (2021). Flagging fake news on social media: An experimental study of media consumers' identification of fake news. *Government Information Quarterly*, 38(3), 101591. <https://doi.org/10.1016/j.giq.2021.101591>
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Grygiel, & Brown, N. (2019). Are social media companies motivated to be good corporate citizens? Examination of the connection between corporate social responsibility and social media safety. *Telecommunications policy*, 43(5), 445-460.
- Harambam, J., Hoboken, J. & Helberger, N. (2018). Democratizing algorithmic news recommenders: How to materialize voice in a technologically saturated media ecosystem. *Philosophical Transactions of the Royal Society*, 376(2133), 1–21. <https://doi.org/10.1098/rsta.2018.0088>
- Hassani, H. (2022). Analysis of Social Media Platform Mechanism to Moderate Disinformation and Fake News; Facebook and Twitte. *Art & Media Studies*, 4(7), 155- 184. <https://civilica.com/doc/1534985> (in Persian)
- Heitz, L., Lischka, J. A., Birrer, A., Paudel, B., Tolmeijer, S., Laugwitz, L., & Bernstein, A. (2022). Benefits of diverse news recommendations for democracy: A user study. *Digital Journalism*, 10(10), 1710-1730. doi:10.1080/21670811.2021.2021804
- Kwak, k, T., Lee, S. & Lee, S. W. (2021). News and user characteristics used by personalized algorithms: The case of Korea's news aggregators, Naver News and Kakao News. *Technological Forecasting and Social Change*, 171(2), 1-12.
- Lewandowsky, S. & Van Der Linden, S. (2021). Countering misinformation and fake news through inoculation and prebunking. *European review of social psychology*, 32(2), 348-384.
- Majdizade, Z. and Mousavi Haghshenas, M. (2020). The “Why & how” of Fake News in Digital Age: A Meta Analysis. *Journal of Interdisciplinary Studies in Communication*

- and Media*, 3(7), 93-124. doi: 10.22034/jiscm.2020.222901.1099 (in Persian)
- Møller, L. A. (2022). Between personal and public interest: How algorithmic news recommendation reconciles with journalism as an ideology. *Digital journalism*, 10(10), 1794-1812.
- Nechushtai, E. & Lewis, S. C. (2019). What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations. *Computers in human behavior*, 90, 298-307
- Rahmanian, E. (2021). Typology of Fake Social Media News in the Context of Iran: An Ethnographic Approach. *Journal of Business Management*, 13(3), 814-844. doi: 10.22059/jibm.2021.319197.4066 (in Persian)
- Riedl, M. J., Whipple, K. N., & Wallace, R. (2022). Antecedents of support for social media content moderation and platform regulation: the role of presumed effects on self and others. *Information, Communication & Society*, 25(11), 1632-1649.
- Saei, M. H. & Azadi, M. H. (2021). The solutions of combating with fake news in Iranian media and news organizations. *Rasaneh*, 32(2), 35-78. doi: 10.22034/bmsp.2021.136841 (in Persian)
- Van der Linden, S., Roozenbeek, J., Maertens, R., Basol, M., Kácha, O., Rathje, S., & Traberg, C. S. (2021). How can psychological science help counter the spread of fake news?. *The Spanish journal of psychology*, 24, e25.
- Vázquez-Herrero, J., Vizoso, Á., & López-García, X. (2019). Innovación tecnológica y comunicativa para combatir la desinformación: 135 experiencias para un cambio de rumbo. *Profesional de la información*, 28(3).
- Wu, H. D. & Liu, B. F. (2018). The effects of content moderation on user engagement and community dynamics: Evidence from Reddit. *Information Systems Research*, 29(3), 631-646.