

## تأثیر مداخله ارسال پیام‌های اصلاحی بر درگیری کاربران منتشرکننده اخبار جعلی در ایکس (توییتر)

### The Effect of Correction Intervention on Engagement of Users Sharing Fake News on X (Twitter)

Mohammad Hasan Ebrahim Kani

M. A. in General Psychology, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran.

Dr. Reza Pourhosein\*

Associate Professor, Department of Cognitive Science, Faculty of Psychology and Educational Sciences, University of Tehran, Tehran, Iran.

[pourhosein@ut.ac.ir](mailto:pourhosein@ut.ac.ir)

Dr. Siyavash Salavatian

Associate Professor, Department of Media Management, Faculty of Communication & Media, IRIB University, Tehran, Iran.

محمدحسن ابراهیم کنی

کارشناسی ارشد روان‌شناسی عمومی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران.

دکتر رضا پورحسین (نویسنده مسئول)

دانشیار، گروه روان‌شناسی، دانشکده روان‌شناسی و علوم تربیتی، دانشگاه تهران، تهران، ایران.

دکتر سیاوش صلوتیان

دانشیار، گروه مدیریت رسانه، دانشکده ارتباطات و رسانه، دانشگاه صدا و سیما، تهران، ایران.

#### Abstract

This study aimed to compare the causal effect of one of the interventions to fight fake news (sending correction messages) on Persian-speaking users who spread false news by investigating the level of their engagement with this message. The study method was experimental and based on a randomized field experiment. The population was targeted by X Persian-speaking users who published specific fake news. The study sample was 334 of these users who were randomly selected within the first 48 hours of the news release and randomly assigned to 3 experimental groups. The data of the study was extracted through the application program interface (API) of X and subjected to the analysis of binomial logistic regression and chi-square test. The implementation of the intervention was that after the preparation of a neutral account by the researcher, three types of corrective messages were sent to these users and their engagement with the corrective message was studied within 24 hours. The results showed that the message that contained only the truth was significantly more effective than the message that contained the truth vs. rumors ( $p<0.05$ ). Also, the characteristics of the users did not affect their engagement with the corrective message. According to the existing research gap on the efficiency of different approaches, this study showed for the first time in the world that the approach of simply stating the truth is more effective. The results of this intervention can be applied in social media to fight fake news.

**Keywords:** Correction, Fake news, Randomized Field Experiment, Social Media, Twitter.

#### چکیده

هدف از مطالعه حاضر مقایسه تأثیر علی یکی از مداخله‌های مقابله با اخبار جعلی یعنی ارسال پیام‌های اصلاحی، بر روی کاربران فارسی‌زبان منتشرکننده اخبار نادرست از طریق برسی میزان درگیری آنها با این پیام بود. روش مطالعه از نوع آزمایشی و بر اساس طرح آزمایش میدانی تصادفی اجرا شد. جامعه مورد هدف کاربران فارسی‌زبان ایکس قرار داده شد که اقدام به انتشار یک خبر جعلی مشخص کردند. نمونه مورد مطالعه ۳۳۴ نفر از این کاربران بودند که در فاصله ۴۸ ساعت ابتدایی انتشار خبر، به صورت کاملاً تصادفی انتخاب و به صورت تصادفی به ۳ گروه آزمایشی تخصیص داده شدند. داده‌های مطالعه از طریق رابط برنامه کاربردی (API) ایکس استخراج و مورد تحلیل رگرسیون لجستیک اسمی دو وجهی و آزمون استقلال خی دو قرار گرفت. اجرای مداخله به این صورت بود که بعد از آماده‌سازی یک حساب کاربری خنثی توسط محقق، سه نوع پیام اصلاحی برای این کاربران ارسال شده و در فاصله ۲۴ ساعت درگیری آنها با پیام اصلاحی مورد مطالعه قرار می‌گرفت. نتایج نشان داد پیامی که صرفاً حاوی حقیقت بود نسبت به پیامی که حاوی حقیقت و شایعه بود به صورت معنی‌داری تأثیر بیشتری داشت ( $p<0.05$ ). همچنین ویژگی‌های کاربران تأثیری بر درگیری آنها با پیام اصلاحی نداشت. با توجه به خلاء پژوهشی موجود در موضوع کارآمدی رویکردهای مختلف اصلاح، این مطالعه برای اولین بار در دنیا نشان داد در متغیر درگیری، رویکرد صرفاً بیان حقیقت کارآمدی بیشتری دارد. نتایج این مداخله قابلیت کاربردی‌سازی در رسانه‌های اجتماعی برای مقابله با اخبار جعلی را دارد.

**واژه‌های کلیدی:** اصلاح، اخبار جعلی، آزمایش میدانی تصادفی، رسانه‌های اجتماعی، توییتر.

## مقدمه

اخبار جعلی<sup>۱</sup>، محتواهای خبری هستند که در بستر اینترنت منتشر شده و علی‌رغم اینکه شبیه به محتوای خبری اصلی و قانونی بنظر می‌رسند، ساختگی یا بسیار نادرست به حساب می‌آیند (پنیکوک و رند<sup>۲</sup>، ۲۰۲۱). انتشار اخبار جعلی و نادرست در سال‌های اخیر و با گسترش شبکه‌های اجتماعی نگرانی‌های بسیاری ایجاد کرده است (وثوقی و همکاران<sup>۳</sup>، ۲۰۱۸). این نگرانی‌ها در واقعیت چون انتخابات‌های ۲۰۱۶ (لازرو همکاران<sup>۴</sup>، ۲۰۲۰) و آمریکا (پنیکوک و رند، ۲۰۲۱b)، خروج کشور انگلستان از اتحادیه اروپا (لازرو همکاران، ۲۰۱۸) و همه‌گیری بیماری کووید-۱۹ (لومبا و همکاران<sup>۵</sup>، ۲۰۲۱) باعث توجه بیش از پیش دانشگاهیان، سیاست‌گذاران و اهالی رسانه به این موضوع شد. این اخبار تهدیدی جدی برای افراد و جوامع به حساب می‌آید (لواندوفسکی و همکاران<sup>۶</sup>، ۲۰۱۷) و می‌تواند خدمات جبران‌ناپذیری در موضوعات مختلفی چون سلامت عمومی در زمان شیوع همه‌گیری‌ها، دوقطبی شدن جامعه و کاهش اعتماد به دولتها وارد کنند (جانسون و همکاران<sup>۷</sup>، ۲۰۲۰؛ جولی و پترسون<sup>۸</sup>، ۲۰۲۰؛ واسودوا و بارکدل<sup>۹</sup>، ۲۰۲۰؛ فریمن و همکاران<sup>۱۰</sup>، ۲۰۲۰).

شبکه‌های اجتماعی امروزه در حال تبدیل شدن به محل اصلی دریافت، مصرف و انتشار اطلاعات و اخبار می‌باشند (پرین<sup>۱۱</sup>، ۲۰۱۵) و این موضوع امروزه نگرانی‌های زیادی ایجاد کرده است (استوارت و همکاران<sup>۱۲</sup>، ۲۰۱۹). در میان رویکردها و رشته‌های مختلف فعال در موضوع اخبار جعلی این روان‌شناسی، علوم شناختی و علوم رفتاری است که در سال‌های اخیر توانسته ذیل حوزه روان‌شناسی اخبار جعلی، درکی واقع‌بینانه‌تر و عمیق‌تر از این چالش جهانی ارائه دهد (پنیکوک و رند، ۲۰۲۱a؛ اکر و همکاران<sup>۱۳</sup>، ۲۰۲۲). این مطالعات در سال‌های اخیر، علاوه بر توفیقات نظری و دانشگاهی، مورد توجه دولتها و سازمان‌های مختلف در جهان نیز قرار گرفته‌اند و به صورت عملیاتی از آنها استفاده می‌شود (پولیتیکو<sup>۱۴</sup>، ۲۰۲۲؛ مرکز عالی ارتباطات استراتژیک ناتو<sup>۱۵</sup>، ۲۰۲۱ و بلاگ توییتر، ۲۰۲۲).

امروزه رویکردهای مختلفی برای مقابله با این دسته از اخبار طراحی و پیاده‌سازی شده‌اند. در یک نگاه و بصورت عمومی می‌توان این رویکردها را به دو دسته مداخلات پیشینی (اعتبارزدایی پیشینی<sup>۱۶</sup> و مداخلات واکنشی (اعتبارزدایی<sup>۱۷</sup>) تقسیم‌بندی کرد (لواندوفسکی و همکاران، ۲۰۲۰). مداخلات اعتبارزدایی پیشینی به افراد کمک می‌کنند تا بتوانند اخبار نادرست و گمراه‌کننده را تشخیص دهند و سپس در برابر تأثیر آنها مقاومت کنند؛ حتی در برابر اخبار جدیدی که در آینده منتشر می‌شوند (وندر لیندن<sup>۱۸</sup>، ۲۰۲۲). مداخلات اعتبارزدایی نیز مشتمل بر پاسخ به یک خبر نادرست مشخص هستند و نشان می‌دهند چرا آن خبر غلط است (اکر و همکاران، ۲۰۲۲). در زمان‌هایی که امکان جلوگیری از تأثیرگذاری اخبار غلط وجود ندارد، می‌بایست از مداخلات واکنشی کمک گرفت. مطالعات زیادی کارآمدی اعتبارزدایی و اصلاح<sup>۱۹</sup> مستقیم را نشان داده‌اند (چان و همکاران<sup>۲۰</sup>، ۲۰۱۷؛ والتر و مورفی<sup>۲۱</sup>، ۲۰۱۸). مهمترین موضوع در اعتبارزدایی و اصلاح، انجام درست و مؤثر آن است، چرا که در غیر اینصورت می‌تواند اثر عکس بگذارد. یافته‌های امروزی اختلاف نظرهایی بر روی ویژگی‌های اصلاح کارآمد دارند؛ عده‌ای معتقد به رویکرد «صوفا بیان حقیقت» هستند و بهترین و تنها راه درست مقابله با اخبار غلط را نادیده‌گرفتن اطلاعات نادرست و صرفا بیان حقیقت می‌دانند (شووارز و همکاران<sup>۲۲</sup>، ۲۰۱۶). این افراد معتقد‌اند اعتبارزدایی اطلاعات غلط عموماً محکوم به شکست است چرا که به آن فضای تنفس بیشتری داده، آشنایی با آن را افزایش می‌دهد و آن را بیشتر باورپذیر می‌کند

<sup>1</sup> fake news<sup>2</sup> Pennycook & Rand<sup>3</sup> Vosoughi et al.<sup>4</sup> Lazer et al.<sup>5</sup> Loomba et al.<sup>6</sup> Lewandowsky et al.<sup>7</sup> Johnson et al.<sup>8</sup> Jolley & Paterson<sup>9</sup> Vasudeva & Barkdull<sup>10</sup> Freeman et al.<sup>11</sup> Perrin<sup>12</sup> Stewart et al.<sup>13</sup> Ecker et al.<sup>14</sup> politico<sup>15</sup> north atlantic treaty organization<sup>16</sup> prebunk<sup>17</sup> debunk<sup>18</sup> Van der Linden<sup>19</sup> correction<sup>20</sup> Chan et al.<sup>21</sup> Walter & Murphy<sup>22</sup> Schwarz et al.

(لواندوفسکی و همکاران، ۲۰۱۲). مخالفت اصلی این نگاه با رویکر «حقیقت در برابر شایعه»<sup>۱</sup> است که یکی از رویکردهای اصلی و پراستفاده در اعتبارزدایی اخبار جعلی به حساب می‌آید و توسط لواندوفسکی و همکاران (۲۰۲۰) شرح داده شده است. این رویکرد معتقد است یک اصلاح ساده نمی‌تواند باعث کندeshدن خبر جعلی از ذهن شود و اعتبارزدایی در صورتی موفق است که دارای چهار جزء مشخص باشد: شروع با بیان حقیقت، یکبار اشاره به شایعه، توضیح غلط‌بودن شایعه و نهایتاً تکرار حقیقت. در کنار این دو رویکرد، برخی دیگر معتقدند یکی از مهمترین ویژگی‌های اعتبارزدایی، ارائه توضیحی از حقیقت است که در حالت ایده‌آل باید بیانگر توضیح و روایت جایگزین باشد از اینکه چرا یک چیز رخ داده است (اکر و همکاران، ۲۰۱۰). متقابلاً برخی نیز معتقدند رد ساده خبر جعلی کارآمدتر است و از رویکرد «صرف رد شایعه» سخن می‌گویند (لواندوفسکی و همکاران، ۲۰۱۲).

همانگونه که گفته شد مطالعات زیادی تا به امروز بر روی کارآمدی اصلاح و اعتبارزدایی انجام شده است، با این حال، لازمه یک اصلاح کارآمد، ایجاد درگیری<sup>۲</sup> در فرد منتشرکننده خبر جعلی (به عنوان مخاطب یک اصلاح) است چرا که یکی از چالش‌های اعتبارزدایی، نادیده‌گرفتن پیام‌های اصلاحی منتشرکننده اخبار غلط است؛ به عنوان مثال مارگولین و همکاران<sup>۳</sup> (۲۰۱۸) در مطالعه‌ای نشان دادند ۷۴ درصد اصلاح‌های اجتماعی در توییتر نادیده<sup>۴</sup> گرفته می‌شوند. از این رو یافتن راهکارهایی برای افزایش درگیری با این پیام‌ها، مقدم بر طراحی متون تأثیرگذار بوده و لازمه مقابله مؤثر با این اخبار در شبکه‌های اجتماعی است. به صورت ویژه‌تر در سال‌های اخیر، پژوهش‌های زیادی در پی یافتن راههایی برای مقابله با انتشار اخبار گمراه‌کننده و غلط در سکوهای شبکه‌های اجتماعی بودند (کوزبیرو و همکاران<sup>۵</sup>، ۲۰۲۰؛ ویتنبرگ و همکاران<sup>۶</sup>، ۲۰۲۰). مانند آنچه مارتل و همکاران<sup>۷</sup> اشاره کرده‌اند (۲۰۲۱) این مداخلات در سطح سکو<sup>۸</sup> و سطح کاربر<sup>۹</sup> قابل انجام هستند. یکی از انواع مداخلات برای مقابله با انتشار اخبار جعلی در شبکه‌های اجتماعی، اصلاح اجتماعی<sup>۱۰</sup> است (مارتل و همکاران، ۲۰۲۱) که در آن کاربران این شبکه‌ها، از طریق ارسال پاسخ (ریپلای) به پیام حاوی اطلاعات غلط، یکدیگر را اصلاح می‌کنند. مزیت این روش فراتر از مزایای عمومی اعتبارزدایی و اصلاح این است که می‌تواند علاوه بر فرد دریافت‌کننده آن، بر روی سایر مشاهده‌کنندگان اصلاح نیز اثر داشته باشد و باور آنها را نیز اصلاح کند (بود و وراگا<sup>۱۱</sup>، ۲۰۱۸) که به این اثر اصلاح مشاهده‌ای می‌گویند (وراگا و بود، ۲۰۱۷).

عمده مطالعات انجام‌شده در موضوع اعتبارزدایی و اصلاح، با استفاده از روش‌های پرسشنامه‌ای و شیوه‌سازی صورت گرفته‌اند و از این حیث محدودیت‌های جدی دارند و کارآمدی آنها در دنیای واقعی، نیازمند آزمایش‌های میدانی است (مصلح و همکاران<sup>۱۲</sup>، ۲۰۲۲). سکوهای شبکه‌های اجتماعی امروزه فرصت بی‌نظیری برای اجرای آزمایش‌های میدانی فراهم کرده‌اند؛ کاربران این شبکه‌ها را می‌توان بصورت تصادفی به گروه‌های مختلف آزمایش و کنترل تقسیم کرد و تأثیر مداخله را با تحلیل ردپای دیجیتال<sup>۱۳</sup> آنها بر روی رفتارهای شبکه‌های اجتماعی مورد نظرلت قرار داده و روابط علیٰ را استخراج کرد (مصلح و همکاران، ۲۰۲۲). این مطالعه اولین مداخله میدانی در موضوع اخبار جعلی بر روی کاربران فارسی‌زبان شبکه‌های اجتماعی محسوب می‌شود و از این حیث می‌تواند برای اولین بار یافته‌هایی جدید از رفتار دیجیتال جامعه ایران بدست دهد. تکیه بر یافتن روابط علیٰ، در مقایسه با مطالعات توصیفی، بیانگر اهمیت دیگر این طرح است. با توجه به عدم دسترسی کشور به مدیریت سکوهای پرطرفدار شبکه‌های اجتماعی و اجرای مداخله حاضر در سطح کاربر، این طرح می‌تواند چشم‌انداز جدیدی از مقابله مبتنی بر شواهد علمی با انتشار اخبار جعلی در این شبکه‌ها ارائه دهد.

بنابر موارد ذکر شده در بالا، مطالعه حاضر در پی یافتن راهی مؤثر برای مقابله با انتشار اخبار جعلی در میان کاربران فارسی‌زبان شبکه‌های اجتماعی بود. برای این هدف یک آزمایش میدانی تصادفی در سکوی ایکس (با همان توییتر) انجام و در آن با ایجاد حساب کاربری محقق‌ساخته، حالت‌های مختلف پیام‌های اصلاحی به نمونه‌ای تصادفی از کاربرانی که یک خبر سیاسی نادرست منتشر کرده‌بودند از طریق پاسخ مستقیم (ریپلای) به توییت حاوی خبر جعلی آنها ارسال شد. با توجه به خلاصه پژوهشی موجود، پیام‌های اصلاحی در سه

1 myth-versus-fact

2 engagement

3 Margolin et al.

4 ignore

5 Kozyreva

6 Wittenberg et al.

7 Martel et al.

8 platform-level

9 user-level

10 social correction

11 Bode &amp; Vraga

12 Mosleh

13 digital trace

حال مختلف ۱) صرفاً بیان حقیقت، ۲) بیان توامان حقیقت و شایعه و ۳) صرفاً رد شایعه طراحی و ارسال شدند. تأثیر علی هر یک از انواع اصلاح‌ها، بر روی رفتار درگیری این کاربران با پیام مذکور (پسند پیام اصلاحی، پاسخ به آن، بازنثر آن، حذف توییت اولیه و در مقابل نادیده‌گرفتن آن) به مدت ۲۴ ساعت مورد مطالعه قرار گرفت. این مطالعه تلاش می‌کند برای اولین بار به این خلاصه پژوهشی پرداخته و تأثیر علی هر کدام از این رویکردها در شرایط واقعی و میدانی را مورد بررسی قرار دهد. در واقع هدف این مطالعه، مقایسه تأثیر علی انواع پیام‌های اصلاحی بر درگیری کاربران فارسی‌زبان منتشرکننده اخبار غلط در سکوی ایکس با این پیام است تا موثرترین نوع پیام اصلاحی بین رویکردهای مذکور مشخص شود. سوال پژوهش حاضر این است که کدام نوع از پیام‌های اصلاحی تأثیر بیشتری در درگیری این کاربران دارد و ویژگی کاربران چه نقشی در این بین ایفا می‌کند.

## روش

پژوهش حاضر از نوع مطالعات آزمایش میدانی تصادفی بود. شرکت‌کنندگان طرح بهدلیل فعالیت روزمره خود در سکوی اجتماعی ایکس مورد مطالعه قرار گرفتند. جامعه مورد هدف طرح، کاربران فارسی‌زبان ایکس بودند که بهدلیل اتفاقات رخداده بعد از فوت آقای جواد روحی در تاریخ ۹ شهریور ۱۴۰۲ اقدام به انتشار خبر جعلی در ارتباط با آن کردند. نمونه مورد مطالعه نیز ۳۶۰ نفر از این کاربران بود که با روش نمونه‌گیری تصادفی انتخاب شدند.

برای اجرای مطالعه حاضر چهار فاز طراحی و اجرا شد: در فاز اول، با توجه به فضای ایکس، پیش از هرچیز یک حساب کاربری قدیمی ساخت سال ۲۰۲۰ تهیه و تمام ساقه آن اعم از اطلاعات حساب کاربری و فعالیت رسانه اجتماعی آن حذف و سپس حساب موجود تبدیل به یک حساب خنثی و بدون جهت‌گیری سیاسی و اجتماعی خاص شد. بعد از توافق با شرکت دیتاک برای پشتیبانی از بخش داده‌کاوی طرح، مقدمات لازم برای اجرای مداخله فراهم شد و فضای ایکس فارسی تا زمان انتشار گستره‌ده یک خبر جعلی مورد رصد قرار گرفت.



شکل ۱. حساب کاربری محقق‌ساخته با شناسه Reza\_Aqaaii

در فاز دوم، سکوی ایکس بصورت روزانه برای یافتن یک خبر جعلی با میزان انتشار بالا مورد رصد قرار گرفت. اولین خبر جعلی منتشرشده در این زمان که به میزان بسیار زیادی در همان ساعات اول انتشار پیدا کرد خبر فوت یک زندانی به نام جواد روحی در زندان نوشهر بود. طبق یافته‌های سامانه دیتاک تنها در ۲۴ ساعت اول، چیزی در حدود ۱۰ هزار توییت با استفاده از کلمات مربوط به نام زندانی زده شد. همزمان سایت‌های مختلف به راستی‌آزمایی این خبر پرداختند و رسانه قوه قضائیه نیز حقیقت ماجرا را بیان کرد (به عنوان نمونه خبرگزاری میزان، ۱۴۰۲ و خبرگزاری ایرنا، ۱۴۰۲). بعد از انتخاب خبر جعلی، با استفاده از سامانه دیتاک و جستجوی کلمات مرتبط، تمام توییت‌های مربوط به این خبر در فاصله حدوداً ۴۸ ساعت بعد از انتشار خبر اول به همراه اطلاعات عمومی حساب‌های کاربری استخراج شد. تعداد این توییت‌ها ۶۰۵۰ عدد بود. در شکل ۲ کلمات استفاده شده برای جستجو آمده است.

با استفاده از عملگرهای \* (AND)، + (OR) و ! (NOT) کوثری مناسب موضوع خود را بسازید.



شکل ۲. کوئری استفاده شده برای استخراج توابع مربوط به خبر جواد روحی

تمام این توابیت‌ها مشمول انتشار شایعات مربوط به فوت آقای جواد روحی نبودند و تعداد زیادی از آنها با استفاده از کلیدواژه‌های مربوطه به موضوع دیگری پراخته‌بودند و یا اقدام به اعتبارزدایی این خبر کرده‌بودند. همچنین تعدادی از کاربران بیش از یک توابیت زدیده‌بودند.

در فاز سوم، توابیت‌ها بصورت دستی مورد بررسی قرار گرفتند تا تنها مواردی که شامل انتشار خبر جعلی مربوطه بودند، مد نظر قرار بگیرند. جامعه هدف پژوهش تمام  $60.50$  توابیت موجود نبود و بخشی از آنها مشمول اهداف این طرح می‌شدند. با توجه به حجم بالای داده موجود، در ابتدا بصورت تصادفی حجم کوچکی از توابیت‌های موجود استخراج شد و سپس تک به تک مورد بررسی قرار گرفتند و تمام موارد نامربوط حذف شدند. از میان کاربران تکراری نیز تنها توابیت آخرشان که بیانگر موضع آخر آنها نسبت به خبر جعلی بود، نگه داشته شد. در نتیجه  $1023$  کاربر یگانه که خبر جعلی منتشر کردند، باقی ماندند و از این بین به صورت کاملاً تصادفی  $360$  نفر انتخاب و به عنوان شرکت‌کنندگان طرح وارد اجرای طرح شدند. با توجه به طرح آزمایش، نمونه به سه بازوی آزمایشی تخصیص تصادفی داده شد. سپس بر روی این شرکت‌کنندگان طرح، مداخله اصلاح اجتماعی در فاصله زمانی حدوداً  $7$  ساعته (به علت محدودیت‌های ایکس، در  $6$  نوبت زمانی) اجرا شد. در نتیجه پنج  $26$  نفر از کاربران که به دلایل مختلف چون حذف توابیت اولیه و حذف حساب کاربری امکان ارسال پیام به آنها وجود نداشت، بر روی باقی  $334$  نفر شرکت‌کننده طرح، مداخله اصلاح اجتماعی به صورت موفق اجرا شد.



### شکل ۳. طرح آزمایش

در فاز چهارم و آخر نیز نتایج اجرای مداخله بهمدت ۲۴ ساعت مورد مطالعه قرار گرفت. بدین صورت که در گیری کاربرانی که مورد اصلاح قرار گرفته بودند با پیام اصلاحی بررسی شد. منظور از در گیری در اینجا پسند، پاسخ و بازنشر پیام اصلاحی و همچنین حذف توبیت اولیه است و در مقابل این موارد نادیده گرفتن پیام تعریف شده بود. با توجه به عدم وجود گروه کنترل در طرح آزمایش، تأثیر بازوهای به صورت مقایسه‌ای با یکدیگر ارزیابی می‌شد. داده‌های این پژوهش با استفاده از نرم‌افزار SPSS-22 مورد تحلیل قرار گرفت. برای این کار از شاخص‌های آمار توصیفی، رگرسیون لجستیک اسمی، دو وجهی (BLR) و آزمون استقلال خی، دو استفاده شد.

ابزار سنجش

داده‌های مورد نیاز این پژوهش، ردپای دیجیتال کاربران و شامل اطلاعاتی چون متن توبیت‌های منتشرشده، پسندیدهای بازنمایی شده، اطلاعات حساب کاربری و غیره بود که از طریق رابط برنامه کاربردی (API)<sup>(۱)</sup> و همچنین دسترسی پژوهشگر به عنوان یک کاربر این سکو

گردآوری شدند. تمام اطلاعات مورد نیاز برای تحلیل نتایج در خود سکو در دسترس قرار داشتند و از طریق فایل اکسل خروجی داده کاوی سامانه دیتاک یا مشاهده فعالیت‌های کاربران توسط پژوهشگر در مدت انجام آزمایش به دست آمدند.

## یافته‌ها

در این مطالعه در مجموع بر روی ۳۳۴ کاربر مداخله انجام شد. توضیحات مربوط به تعداد شرکت‌کننده‌های هر بازوی پژوهشی و فراوانی و درصد درگیری در جدول ۱ آمده است.

جدول ۱. تعداد شرکت‌کننده‌گان به همراه فراوانی و درصد درگیری

مجموع	درصد نادیده‌گرفتن	فراوانی نادیده‌گرفتن	درصد درگیری	فراوانی درگیری	بازو
۱۱۱	۷۹.۲۸	۸۸	۲۰.۷۲	۲۳	بازوی ۱
۱۱۲	۹۰.۱۸	۱۰۱	۹.۸۲	۱۱	بازوی ۲
۱۱۱	۸۱.۹۸	۹۱	۱۸.۰۲	۲۰	بازوی ۳
۳۳۴	۸۳.۸۳	۲۸۰	۱۶.۱۷	۵۴	مجموع

از میان ۵۴ مورد درگیری، ۳ مورد مربوط به پسند پیام اصلاحی و ۵۱ مورد باقی‌مانده مربوط به پاسخ به پیام اصلاحی بود. در این مداخله هیچ موردی درگیری از نوع بازنگری پیام اصلاحی و حذف توبیت اولیه دیده نشد. باتوجه به ویژگی‌های سکوی ایکس، اطلاعاتی عمومی از حساب کاربری افراد وجود داشت که مواردی از آن نیز مورد تحلیل قرار گرفت. این اطلاعات در جدول ۲ نشان داده شده‌است.

جدول ۲. ویژگی‌های کاربران به همراه میانگین و انحراف استاندارد

متغیر	میانگین	انحراف استاندارد	بالاترین	پایین‌ترین
تعداد پسند توبیت خبر جعلی	۷۷.۹۰	۵۳۰.۳۰	۷۵۰.۵	.
تعداد دنبال‌کننده	۱۹۴۵۲.۵۷	۱۷۵۹۶۷.۴۱	۲۲۶۸۱۸۹	.
تعداد دنبال‌شونده	۹۱۰.۲۴	۱۲۱۶.۲۶	۷۲۶۷	.
تعداد کل پسند	۳۹۵۸۲.۲۷	۷۳۳۹۰.۰۵	۶۴۱۲۷۵	.
تعداد کل توبیت	۲۲۰۴۱.۲۵	۴۵۴۷۹.۰۲	۳۹۷۳۳۶	۳
عمر حساب کاربری	۱۹۳۱.۹۴	۳۵۲۹.۱۷	۱۹۶۰۳	۱۹
فاصله ساعتی بین توبیت خبر	۶۲.۲۴	۱۲.۸۴	۸۹	۱۹
جهلی و پیام اصلاحی				

\* باتوجه به اجرای طرح در شرایط واقعی، برخی حساب‌های کاربری مطرح با تعداد دنبال‌کننده، دنبال‌شونده و عمر و میزان فعالیت زیاد در طرح حضور داشتند و به این علت انحراف استاندارد برخی متغیرها بالا شد.

با هدف یافتن موثرترین نوع پیام اصلاحی بر درگیری کاربران از آزمون رگرسیون لجستیک اسمی دو وجهی (BLR) استفاده شد. برای این کار ابتدا از روش همزمان<sup>۱</sup> و سپس با هدف بهبود مدل و باقی‌ماندن متغیرهای اثرگذارتر، از روش حذف پس‌رو مشروط<sup>۲</sup> کمک گرفته شد. در روش همزمان، در ابتدا برای ارزیابی کل مدل رگرسیونی لجستیک، نتایج مربوط به آزمون اومنی‌بوس<sup>۳</sup> حاکی از آن بود که برآش مدل قابل قبول و در سطح کمتر از ۰.۰۵ معنی‌دار بود.

1 enter

2 backward elimination (conditional)

3 omnibus test

### جدول ۳. آزمون اومنی‌بوس برای ارزیابی کلی مدل

سطح معنی‌داری	درجه آزادی	خی دو		
۰.۰۱۳	۹	۲۰.۹۷۴	مرحله ۱	
۰.۰۱۳	۹	۲۰.۹۷۴	بلوک	
۰.۰۱۳	۹	۲۰.۹۷۴	مدل	

سپس با استفاده از نتایج جدول طبقه‌بندی<sup>۱</sup> قدرت مدل در تفکیک افراد در طبقات متغیر وابسته تبیین شد. در این بخش دقیق مدل در طبقه‌بندی افراد بالا و معادل ۸۳.۵ درصد بود.

### جدول ۴. جدول طبقه‌بندی

مورد انتظار				
درصد	درگیری	نادیده‌گرفتن (۰)	نادیده‌گرفتن (۰)	مشاهده شده
صحت	درگیری (۱)	درگیری (۱)	درگیری (۱)	مرحله ۱
۹۹.۶	۱	۲۷۹	۰	درگیری
.	۵۴	۰	۰	درگیری (۱)
۸۳.۵				درصد نهایی

در نهایت طبق نتایج جدول متغیرهای معادله (جدول ۵)، بازوی ۱ (صرف‌بیان حقیقت) بیشترین تأثیر را بر درگیری کاربران داشت و به صورت معنی‌داری از بازوی ۲ (بیان توأم‌ان حقیقت و شایعه) تأثیر بیشتری از خود نشان داد ( $p=0.021$ ). بر این اساس با توجه به اینکه معنی‌داری آماره والد<sup>۲</sup> برای مقایسه بازوی ۱ نسبت به ۲ در سطح خطای کوچکتر از  $0.005$  معنی‌دار است، در این صورت وجود این بازو در مدل مفید و دارای اثر معنی‌دار است. طبق این نتایج، بین بازوی ۱ و ۳ نیز تفاوت معنی‌داری وجود نداشت ( $p=0.530$ ). با توجه به اینکه در این مدل، بازوی ۱ پایه مقایسه دو بازوی دیگر قرار گرفت و بر این اساس نسبت دو بازوی ۲ و ۳ مشخص نبود، بار دیگر مدل با پایه قراردادن بازوی ۳ اجرا شد و طبق نتایج آن، علی‌رغم عملکرد بهتر بازوی ۳ نسبت به ۲، تفاوت معنی‌داری بین آنها مشاهده نشد ( $p=0.083$ ). همچنین هیچ‌کدام از ویژگی‌های کاربران، یعنی تعداد دنبال‌کننده، تعداد پست، تعداد کل توبیت، تعداد کل پسند، عمر حساب کاربری، فاصله بین توبیت خبر جعلی و پیام اصلاحی و تعداد دنبال‌شونده، تأثیر معنی‌داری بر درگیری کاربران نداشتند.

### جدول ۵. جدول متغیرهای معادله

آماره Exp(B)	سطح معنی‌داری	درجه آزادی	آماره والد	خطای استاندارد	آماره B	متغیرها	
۰.۳۹۴	۰.۰۶۴	۲	۵.۴۸۴		- ۰.۹۳۲	بازوی آزمایشی	مرحله ۱
۰.۸۰۲	۰.۰۲۱	۱	۵.۳۵۱	۰.۴۰۳	- ۰.۳۵۱	بازو (۲ به ۱)	
۱.۰۰۰	۰.۵۳۰	۱	۰.۳۹۴	۰.۳۵۱	- ۰.۲۲۱	بازو (۳ به ۱)	
۰.۹۹۷	۰.۱۷۸	۱	۱.۸۱۳	۰.۰۰۰	۰.۰۰۰	تعداد دنبال‌کننده	
۱.۰۰۰	۰.۵۹۵	۱	۰.۲۸۲	۰.۰۰۶	۰.۰۰۰۳	تعداد پسند	
۱.۰۰۰	۰.۲۱۹	۱	۱.۵۱۰	۰.۰۰۰	۰.۰۰۰	تعداد کل توبیت	
۱.۰۰۰	۰.۱۲۹	۱	۲.۳۰۲	۰.۰۰۰	۰.۰۰۰	تعداد کل پسند	
۱.۰۰۰	۰.۵۹۴	۱	۰.۲۸۵	۰.۰۰۰	۰.۰۰۰	عمر حساب کاربری	
۰.۹۹۰	۰.۴۰۰	۱	۰.۷۰۹	۰.۰۱۲	- ۰.۰۱۰	فاصله بین توبیت خبر جعلی و پیام اصلاحی	
۱.۰۰۰		۱			۰.۰۰۰	تعداد دنبال‌شونده	

۰.۵۶۱	۰.۴۶۷	۱	۰.۵۲۹	۰.۷۹۶	- ۰.۵۷۹	ثبت
-------	-------	---	-------	-------	---------	-----

بعد از مشاهده و تحلیل نتایج در روش همزمان، روش حذف پس‌رو مشروط اجرا شد و در این روش نیز نتایج آزمون اولمنی‌بوس حاکی از معنی‌داری نتایج بود ( $p=0.001$ ). همانطور که در جدول ۶ نشان داده شده است، نتایج جدول متغیرهای معادله در این روش حاکی از آن است که در مرحله پایانی علاوه بر متغیر بازوهای پژوهشی، متغیر تعداد دنبال‌کننده نیز علی‌رغم اینکه تأثیر معنی‌داری ندارد ( $p=0.068$ )، با این حال متغیر مفید و اثرگذاری در مدل به حساب می‌آید. در این مدل، کماکان بازوی ۱ و ۲ نیز تفاوت معنی‌داری دارند و سطح معنی‌داری آنها نیز بهبود پیدا کردند ( $p=0.019$ ).

جدول ۶. مرحله هفتم و پایانی جدول متغیرهای معادله در روش حذف پس‌رو مشروط

آماره Exp(B)	سطح معنی‌داری	درجه آزادی	آماره والد	خطای استاندارد	آماره B	متغیرها
	۰.۰۶۰	۲	۵.۶۳۷			مرحله ۷ بازوهای آزمایشی
۰.۳۹۲	۰.۰۱۹	۱	۵.۵۳۴	۰.۳۹۸	- ۰.۳۳۷	بازو (۲ به ۱)
۰.۷۹۰	۰.۴۹۶	۱	۰.۴۶۴	۰.۳۵۱	- ۰.۲۲۱	بازو (۳ به ۱)
۱.۰۰۰	۰.۰۶۸	۱	۳.۳۲۷	۰.۰۰۰	۰.۰۰۰	تعداد دنبال‌کننده
۰.۳۵۱	۰.۰۰۰	۱	۱۷.۰۲۸	۰.۷۹۶	- ۰.۵۷۹	ثبت

در پایان با هدف بررسی رابطه انواع پیام اصلاحی و انواع درگیری کاربران (شامل پسند پیام اصلاحی، پاسخ به آن، بازنثر آن و حذف توييت اوليه) از آزمون خی دو استفاده شد و با توجه به نتایج جدول ۷ رابطه معنی‌داری بین دو متغیر وجود نداشت ( $p=0.090$ ).

جدول ۷. آزمون خی دو

آماره	درجه آزادی	سطح معنی‌داری
۸.۰۳۸	۴	خی دو پیرسون
۹.۰۹۹	۴	نسبت درست‌نمایی
۷.۹۹۹		آزمون دقیق فیشر
۰.۵۵۲	۱	آزمون پیوند خطی-خطی
۳۳۴		تعداد ورودی معتبر

## بحث و نتیجه‌گیری

مطالعه حاضر با هدف مقایسه تأثیر علی ا نوع پیام‌های اصلاحی بر درگیری کاربران فارسی‌زبان منتشرکننده اخبار غلط در سکوی ایکس با این پیام صورت گرفت تا از طریق آن بتوان مؤثرترین پیام اصلاحی را برای مقابله با انتشار اخبار نادرست توسط این کاربران مشخص کرد. این مطالعه در محیط دو قطبی ایکس فارسی‌زبان‌ها و بر روی یک سوژه سیاسی در بافتی ملتهب به لحاظ اجتماعی صورت گرفته است. نتایج مطالعه حاکی از آن بود که بازوی اول پژوهش که شامل پیامی مشتمل بر صرف بیان حقیقت ماجرا بود توانست بیش از دو برابر بازوی دوم (که شامل بیان تأمین حقیقت و شایعه بود) و همچنین کمی بیشتر از بازوی سوم (که صرفاً به رد ادعای غلط پرداخته بود) در این کاربران درگیری ایجاد کند و بدین صورت نسبت به بازوی دوم تأثیر بیشتری داشته باشد. بین این دو مدل اصلاح و اعتبارزدایی بین پژوهشگران و متخصصان جهانی روان‌شناسی اخبار جعلی تفاوت نظر وجود داشت و مطالعه حاضر اولین مطالعه‌ای است که بصورت میدانی توانست تأثیر این دو مدل را با یکدیگر مقایسه کرده و نشان دهد حداقل در متغیر درگیری، مدل صرفاً بیان حقیقت از مدل چهار مرحله‌ای اعتبارزدایی کارآمدتر است.

احتمالاً تفاوت تأثیر دو بازوی ۱ و ۲ را می‌توان به این عوامل نسبت داد؛ کوتاه‌تر بودن بازوی اول و به تبع جلب توجه و درگیری بیشتر، کمتر درگیر کننده بودن بازوی دوم به علت طولانی بودن پیام و بیان حقایق کمتر شنیده شده در بازوی اول که برای مخاطب بدیع تر است.

در مقایسه با مطالعه مارتل و همکاران (۲۰۲۱) که شبیه به مطالعه حاضر بود و نتایج آن بیانگر این بود که محتوای پیام اصلاحی تأثیری بر احتمال درگیری کاربران ندارد، این مطالعه توانست خلاف آن را نشان دهد. همچنین این مطالعه بیان داشت ویژگی‌های مختلف کاربران تأثیری بر درگیری آنها با پیام‌های اصلاحی ندارد و هیچ رابطه‌ای نیز بین انواع پیام‌های اصلاحی و انواع درگیری کاربران مشاهده نشد.

به لحاظ عملی این مطالعه یعنی مداخله اصلاح اجتماعی قابلیت کاربردی‌سازی در دنیای واقعی را دارد و تمام دغدغه‌مندان، اهالی رسانه و سیاست‌گذاران بخش‌های مختلف کشور می‌توانند از یافته‌های این مطالعه برای مقابله درست و مبتنی بر شواهد علی با اخبار جعلی منتشرشده در رسانه‌های اجتماعی استفاده کنند؛ یعنی برای اعتبارزدایی موج‌های مختلف خبر جعلی و اصلاح کاربرانی که اقدام به انتشار این اخبار می‌کنند، صرفاً به بیان حقایق پردازند.

این مطالعه دارای چند محدودیت بود؛ اولاً، باتوجه به محدودیت‌های ماهه‌ای اخیر ایکس در دسترسی به اطلاعات و همچنین فعالیت کاربران و علی‌رغم اینکه در ابتداء نمونه‌ای بیش از ۱۰۰۰ نفره برای اجرای مداخله در نظر گرفته شده بود، مداخله پایانی ناچاراً بر روی نمونه‌ای کوچکتر اجرا شد و از این حیث اولین محدودیت طرح تعداد پایین نمونه بود و این موضوع احتمالاً علت معنی‌دار نشدن اثر تعدادی از متغیرهای طرح است و همچنین ممکن است قدرت آماری لازم برای نشان دادن رابطه علی بین متغیرها را فراهم نسازد. در ثانی، این مطالعه صرفاً بر روی یک خبر جعلی انجام شد و به علت تکیه بر یک خبر و یک گروه مخاطب، نتایج آن قابلیت تعمیم به سایر گروه‌های کاربران ایکس را ندارد. ثالثاً، به علت رونق‌نشاشن سایت‌های راستی‌آزمایی بی‌طرف و فعلال در کشور، امکان فعالیت بر روی تعداد بیشتری خبر جعلی و همچنین بررسی متغیرهای دیگر چون انتشار اخبار غلط وجود نداشت و راستی‌آزمایی خبر مربوط به فوت آقای جواد روحی نیز صرفاً بر اساس اعلام‌نظر دستگاه‌های رسمی کشور صورت گرفت. محدودیت آخر نیز کمبود اطلاعات بیشتر چون سن و جنسیت کاربران ایکس است که استخراج آنها نیازمند مدل‌ها و روش‌های پیچیده یادگیری ماشین.

در پایان چند پیشنهاد برای انجام مطالعات آینده ارائه می‌شود؛ اولاً، طبق نتایج این مطالعه نرخ درگیری کاربران با پیام اصلاحی کماکان بسیار پایین است و مطالعات بعدی باید با پرداخت به سایر متغیرها به دنبال افزایش کلی این عدد باشند. در اینجا می‌توان به متغیرهایی چون جنسیت و هویت حزبی کاربر اصلاح‌کننده، زمان ارسال پیام، تعداد کاربران اصلاح‌کننده و ارسال پیام به صورت عمومی یا خصوصی اشاره کرد. ثانیاً، مطالعات بعدی می‌بایست به تعداد خبر بیشتر و با گروه‌های مخاطب متفاوت پردازند تا تأثیر اصلاح اجتماعی بر گروه‌های مختلف و در موضوعات متفاوت مورد بررسی قرار گیرد. ثالثاً، علاوه بر متغیر درگیری می‌بایست متغیرهای انتشار و پذیرش مطالعه شوند تا درک کامل‌تری از باورها و رفتارهای مرتبط با اخبار جعلی حاصل شود. رابعاً، مداخله اصلاح اجتماعی در سایر سکوهای اجتماعی پرطرفدار چون اینستاگرام و همچنین پیام‌رسان‌ها اجرا شود. در پایان با اجرای این روش بر روی نمونه بزرگی از کاربران و استفاده از یادگیری ماشین می‌توان از نتایج تحلیل پیش‌بین به پیش‌بینی اثرگذاری پیام‌های اصلاحی بر گروه‌های مختلف مخاطب پرداخت.

## منابع

- خبرگزاری ایرنا (۱۴۰۲، ۹ شهریور). «جواد روحی» قبل از ورود به زندان سابقه تشنجه و بستری در بیمارستان داشته است. بازیابی شده از <https://irna.ir/xjNmht>
- خبرگزاری میزان (۱۴۰۲، ۱۰ شهریور). جواد روحی به چه جرمی در زندان بود؟ از هتاكی به قرآن تا تخريب اموال عمومي. بازیابی شده از <https://www.mizanonline.ir/00Jr5T>
- Bode, L., & Vraga, E. K. (2018). See something, say something: Correction of global health misinformation on social media. *Health Communication*, 33(9), 1131-1140. <https://doi.org/10.1080/10410236.2017.1331312>
- Chan, M. P. S., Jones, C. R., Hall Jamieson, K., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531-1546. <https://doi.org/10.1177/0956797617714579>
- Ecker, U. K., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., ... & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13-29. <https://doi.org/10.1038/s44159-021-00006-y>
- Ecker, U. K., Lewandowsky, S., & Tang, D. T. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38, 1087-1100. <https://doi.org/10.3758/MC.38.8.1087>
- Freeman, D., Waite, F., Rosebrock, L., Petit, A., Causier, C., East, A., ... & Lambe, S. (2022). Coronavirus conspiracy beliefs, mistrust, and compliance with government guidelines in England. *Psychological Medicine*, 52(2), 251-263. <https://doi.org/10.1017/S0033291720001890>

تأثیر مداخله ارسال پیام‌های اصلاحی بر درگیری کاربران منتشر کننده اخبار جعلی در ایکس (توییتر)

The Effect of Correction Intervention on Engagement of Users Sharing Fake News on X (Twitter)

- Johnson, N. F., Velásquez, N., Restrepo, N. J., Leahy, R., Gabriel, N., El Oud, S., ... & Lupu, Y. (2020). The online competition between pro-and anti-vaccination views. *Nature*, 582(7811), 230-233. <https://doi.org/10.1038/s41586-020-2281-1>
- Jolley, D., & Paterson, J. L. (2020). Pylons ablaze: Examining the role of 5G COVID-19 conspiracy beliefs and support for violence. *British Journal of Social Psychology*, 59(3), 628-640. <https://doi.org/10.1111/bjso.12394>
- Kozyreva, A., Lewandowsky, S., & Hertwig, R. (2020). Citizens versus the internet: Confronting digital challenges with cognitive tools. *Psychological Science in the Public Interest*, 21(3), 103-156. <https://doi.org/10.1177/1529100620946707>
- Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094-1096. <https://doi.org/10.1126/science.aao2998>
- Lewandowsky, S., Cook, J., Ecker, U. K. H., Albarracín, D., Amazeen, M. A., Kendeou, P., Lombardi, D., Newman, E. J., Pennycook, G., Porter, E. Rand, D. G., Rapp, D. N., Reifler, J., Roozenbeek, J., Schmid, P., Seifert, C. M., Sinatra, G. M., Swire-Thompson, B., van der Linden, S., Vraga, E. K., Wood, T. J., Zaragoza, M. S. (2020). The Debunking Handbook 2020. <https://doi.org/10.17910/b7.1182>
- Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the “post-truth” era. *Journal of Applied Research in Memory and Cognition*, 6(4), 353-369. <https://doi.org/10.1016/j.jarmac.2017.07.008>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). and successful debiasing. *Psychological science in the public interest*, 13(3), 106-131. <https://doi.org/10.1177/1529100612451018>
- Loomba, S., de Figueiredo, A., Piatek, S. J., de Graaf, K., & Larson, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour*, 5(3), 337-348. <https://doi.org/10.1038/s41562-021-01056-1>
- Margolin, D. B., Hannak, A., & Weber, I. Political fact-checking on Twitter: When do corrections have an effect? *Political Communication* 35, 196-219 (2018). <https://doi.org/10.1080/10584609.2017.1334018>
- Martel, C., Mosleh, M., & Rand, D. G. (2021). You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online. *Media and Communication*, 9(1), 120-133. <https://doi.org/10.17645/mac.v9i1.3519>
- Mosleh, M., Pennycook, G., & Rand, D. G. (2022). Field experiments on social media. *Current Directions in Psychological Science*, 31(1), 69-75. <https://doi.org/10.1177/09637214211054761>
- NATO Strategic Communications Centre of Excellence (2021). Inoculation theory and misinformation. Retrieved from: <https://stratcomcoe.org/publications/inoculationtheory-and-misinformation/217>
- Pennycook, G., & Rand, D. G. (2021a). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388-402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Pennycook, G., and Rand, D.G. (2021b) Examining false beliefs about voter fraud in the wake of the 2020 Presidential Election. *Harvard Kennedy Sch. Misinformation Rev.* 2, 1–19. <https://doi.org/10.37016/mr-2020-51>
- Perrin, A.(2015). Social media usage. *Pew Research Center* , 52-68 [https://www.secretintelligenceservice.org/wp-content/uploads/2016/02/PI\\_2015-10-08\\_Social-Networking-Usage-2005-2015\\_FINAL.pdf](https://www.secretintelligenceservice.org/wp-content/uploads/2016/02/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf)
- Politico (2022, december 3). How Ukraine Won The #LikeWar. Retrieved from <https://www.politico.com/news/magazine/2022/03/12/ukraine-russia-information-warfare-likewar00016562>
- Schwarz, N., Newman, E., & Leach, W. (2016). Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behavioral Science & Policy*, 2(1), 85-95. <https://doi.org/10.1177/237946151600200110>
- Stewart, A. J., Mosleh, M., Diakonova, M., Arechar, A. A., Rand, D. G., & Plotkin, J. B. (2019). Information gerrymandering and undemocratic decisions. *Nature*, 573(7772), 117-121. <https://doi.org/10.1038/s41586-019-1507-6>
- Twitter Blog (2022). Our approach to the 2022 US midterms. Retrieved from [https://blog.twitter.com/en\\_us/topics/company/2022/-our-approach-to-the-2022-us-midterms](https://blog.twitter.com/en_us/topics/company/2022/-our-approach-to-the-2022-us-midterms)
- Van Der Linden, S. (2022). Misinformation: susceptibility, spread, and interventions to immunize the public. *Nature medicine*, 28(3), 460-467. <https://doi.org/10.1038/s41591-022-01713-6>
- Vasudeva, F., & Barkdull, N. (2020). WhatsApp in India? A case study of social media related lynchings. *Social Identities*, 26(5), 574-589. <https://doi.org/10.1080/13504630.2020.1782730>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151. <https://doi.org/10.1126/science.aap9559>
- Vraga, E. K., & Bode, L. (2017). Using expert sources to correct health misinformation in social media. *Science Communication*, 39(5), 621-645. <https://doi.org/10.1177/107554701773177>
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423-441. <https://doi.org/10.1080/03637751.2018.1467564>
- Wittenberg, C., Berinsky, A. J., Persily, N., & Tucker, J. A. (2020). Misinformation and its correction. *Social media and democracy: The State of The Field, Prospects for Reform*, 163. [https://www.opolisci.com/wp-content/uploads/pdf-front/Social\\_Media\\_and\\_Democracy.pdf#page=183](https://www.opolisci.com/wp-content/uploads/pdf-front/Social_Media_and_Democracy.pdf#page=183)