



ویدیو تهیه شده با هوش مصنوعی درباره درگیری ایران-اسرائیل

2026-004-FB-UA

یک کاربر فیسبوک در صفحه‌ای که خود را منبع خبری معرفی می‌کرد، ۲۰۲۵ ژوئن ۱۵ در ثانیه‌ای پست کرد که بروز آسیب‌هایی به ساختمان‌ها در حیفا در طول جنگ ۲۳ ۱۲ ویدیویی روزه اسرائیل و ایران (از ۱۳ ژوئن تا ۲۵ ژوئن ۲۰۲۵) را نشان می‌داد. متن به انگلیسی با عنوان هم‌اکنون زنده - حیفا» به همراه تاریخ پست آن روی ویدیو قرار گرفته بود. ویدیو به نظر» که توسط حقیقت‌یاب‌های مستقل به عنوان تولیدشده با هوش مصنوعی تشخیص می‌رسی عباراتی به سبک-داده شده است و براساس گزارش، مبدأ آن از تیک‌تاک بود. شرح انگلیسی عنوان اخبار بیان می‌کرد که به مناقشه مرتبط بود و اصطلاحات و هشتک‌های گسسته و حمله عظیمی» از طرف ایران» بدون پیروی از روایتی روشن را بیان می‌کرد. که عبارت بود از به اسرائیل انجام شده است و اینکه اعضای کابینه جنگ اسرائیل در سنگر هستند، همچنین به تعداد موشک‌ها ارجاع می‌داد، سقوط هواپیما، شخصیت‌های سیاسی جهان، مناقشه‌های حاضر از جمله غزه، توافق هسته‌ای، آتش‌سوزی‌های طبیعی و هشتک‌هایی به حساب‌ها برای از منابع خبری اسرائیلی ذکر می‌کرد که هشدار حمله فوری لغو انسداد آن‌ها. همچنین بار دیده شده بود ۷۰۰,۰۰۰ بیش از می‌دادند. این پست گزارش بار ۹ به دلیل تروریسم، خشونت، تقلب، و کلاهبرداری، شش کاربر تخلف محتوا را کرده‌اند. با این حال، گزارش‌ها برای بازبینی انسانی اولویت‌بندی نشده‌اند. در همان روز که که پیش‌بینی می‌کرد محتوا حاوی اطلاعات Meta محتوا پست شد، یک طبقه‌بندی‌کننده نادرست است، پست را برای حقیقت‌یاب طرف سوم نشان‌گذاری کرد. حقیقت‌یاب‌های مستقل رتبه‌بندی محتوا را انجام ندادند.

یکی از کاربران گزارش‌دهنده، از «هیئت» نسبت به تصمیم متا برای نگهداشتن این محتوا در پلتفرم پژوهش‌خواهی کرده است. متا در «هیئت» تأیید کرده است که این پست استانداردهای انجمن درباره اطلاعات گمراه‌کننده» را نقض نکرده است زیرا «مستقیماً در» خطر آسیب فیزیکی فوری مشارکت نداشته است» یا «مستقیماً در عملکرد فرایند سیاسی مداخله نکرده است». با این حال، از آنجا که «هیئت» این پرونده را انتخاب کرده است، متا سه حساب مرتبط با این صفحه را به دلیل علائم درگیر بودن با سوءاستفاده و غیراصیل بودن، درعین حال-محتوا را در این پلتفرم از دسترس خارج کرد غیرفعال کرد و این صفحه و «هیئت» تصمیم گرفت تا این پرونده را به دلیل پیامدهای آن برایروال‌ها و مشکلات اجرایی و» خط‌مشی مهم پیگیری کند.



هیئت» این مورد را انتخاب کرده تا به مسئله مدیریت محتوای تولیدشده با هوش مصنوعی بپردازد که ممکن است به یکپارچگی اطلاعات آسیب بزند و اعتماد عمومی در زمینه درگیری مسلحانه را سست کند. این پرونده فرصتی به دست داد تا نظارت انسانی و خودکار متا را بر محتوای تولیدشده با هوش مصنوعی ارزیابی کنیم، ازجمله در موقعیت‌های مناقشات. همچنین به «هیئت» اجازه داد تا بهترین راه‌ها را برای توجه به تأثیر چنین مطالبی در محیط اطلاعات بررسی کند درحالی‌که به آزادی بیان و دسترسی به اطلاعات در وضعیت‌های مناقشه‌ای احترام می‌گذارد.

این پرونده تحت اولویت‌های راهبردی «وضعیت‌های مناقشه و بحران‌های هیئت» و «اجرای خودکار سیاست‌ها و دسته‌بندی و ساخت محتوا» قرار می‌گیرد.

هیئت» نظرات عموم را که این موارد را مورد توجه قرار می‌دهند تقدیر می‌کند:

- نقشی که اطلاعات نادرست/ گمراه‌کننده تولیدشده با هوش بازی کرد، ازجمله در ۲۰۲۵ مصنوعی در مناقشه ایران-اسرائیل در ژوئن رسانه‌ها و شفاف‌سازی‌های عمومی
- پژوهش درباره شیوع و تأثیر اطلاعات نادرست/ گمراه‌کننده در پلتفرم‌های رسانه‌های اجتماعی به‌طور کلی، و درطول مواجهه نظامی به‌طور خاص، و انگیزه‌ها و تمایلات برای ساخت و هم‌رسانی کردن چنین محتوایی
- چالش‌های ردیابی دقیق، برچسب‌زنی یا حقیقت‌یابی محتوای تولیدشده با هوش مصنوعی، به‌ویژه در زمینه پوشش‌های هماهنگ‌شده اطلاعات نادرست/ گمراه‌کننده، و تأثیر پاسخ‌های اجرایی، محصول و خط‌مشی
- مسئولیت‌های حقوق بشری پوشش‌های رسانه‌های اجتماعی برای توجه به هرگونه تأثیر متفاوت ارائه‌های نادرست تولیدشده با هوش مصنوعی، به‌ویژه درطول مواجهه نظامی در محیط اطلاعات، درعین‌حالی که به آزادی بیان احترام گذاشته شود و از دسترسی کاربر به اطلاعات اطمینان حاصل شود

صادر Meta هیئت» در تصمیمات خود می‌تواند توصیه‌های مربوط به سیاست‌ها را برای روز به آن‌ها پاسخ دهد ۶۰ باید ظرف Meta کند. اگرچه این پیشنهادها الزام‌آور نیستند، اما به این ترتیب، «هیئت» پذیرای نظرات عموم است که توصیه‌هایی مرتبط با این پرونده را ارائه می‌دهند.