شاپا چاپی: ۱۸۴۱-۲۷۸۳
شاپای الکترونیکی: ۱۹۲۲-۲۷۸۳

فصلنامه تمدن حقوقی

تمدن حقوقی

دوره ۹ - شماره ۲۸ - تابستان ۱۴۰۵

Legal Civilization
Vol. 9, No. 28, Summer 2026
Pages: 179-198
Article Type: Original Research
https://www.pzhfars.ir/article_246162.html
doi: 10.22034/lc.2026.581086.1777

تمدن حقوقی
دوره ۹، شماره ۲۸، تابستان ۱۴۰۵
صفحات: ۱۹۸-۱۷۹
نوع مقاله: پژوهشی

The Dual Role of Artificial Intelligence in Countering Misinformation and Deep Forgery: Challenges and Solutions

نقش دوگانه هوش مصنوعی در مقابله با اطلاعات نادرست و جعل عمیق: چالش‌ها و راهکارها

Atefeh Lorkijori
Assistant Professor, Department of Law, Faculty of Humanities, Islamic Azad University, Lahijan Branch, Lahijan, Iran

Seyed Ahmad Peyrovnaziri
Master of Science in Criminal Law and Criminology, Faculty of Humanities, Islamic Azad University, Lahijan Branch, Lahijan, Iran (Corresponding Author)

عاطفه لرکجوری
استادیار، گروه حقوق، دانشکده علوم انسانی، دانشگاه آزاد اسلامی واحد لاهیجان، لاهیجان، ایران

atefeh.lorkojuri@iau.ac.ir

سیداحمد پیروندیری
کارشناس ارشد حقوق کیفری و جرم‌شناسی، دانشکده علوم انسانی، دانشگاه آزاد اسلامی واحد لاهیجان، لاهیجان، ایران (نویسنده مسئول)

seyedahmad.peyrovnaziri@iau.ir

Abstract

The spread of misinformation has become one of the fundamental challenges of contemporary societies, its dimensions are not limited to the media sphere, but have far-reaching consequences on social, economic and political structures. The spread of this phenomenon has led to a decrease in public trust, weakening social capital, disruption of decision-making processes and increased social polarization. In such circumstances, the need to provide efficient solutions to manage and control the flow of information is felt more than ever. Meanwhile, artificial intelligence is emerging as a dual technology; in that it both provides significant capabilities to combat misinformation and can itself help to exacerbate this crisis. On the one hand, tools such as machine learning algorithms, big data analysis, and automated verification systems can be effective in identifying, monitoring, and reducing the spread of false content. On the other hand, the same technology, by enabling the production of advanced fake content such as deepfakes, enhancing user behavioral profiling, and precise information targeting, paves the way for the spread and sophistication of false information. This study, using a descriptive-analytical method and using library resources, examines the dual role of artificial intelligence in dealing with misinformation and analyzes its capacities and limitations. The research findings show that there is no single and absolute solution to control this phenomenon, but rather a combination of measures such as clarifying the functioning of algorithms, promoting media literacy, and developing effective regulatory frameworks is necessary. The research results also indicate that cooperation between governments, digital platforms, and users plays a key role in managing this challenge.

Keywords: Artificial Intelligence, Misinformation, Media Literacy.

چکیده

انتشار اطلاعات نادرست به یکی از چالش‌های اساسی جوامع معاصر تبدیل شده است که ابعاد آن صرفاً محدود به حوزه رسانه نیست، بلکه پیامدهای گسترده‌ای بر ساختارهای اجتماعی، اقتصادی و سیاسی بر جای می‌گذارد. گسترش این پدیده موجب کاهش اعتماد عمومی، تضعیف سرمایه اجتماعی، اختلال در فرایندهای تصمیم‌گیری و افزایش قطب‌بندی‌های اجتماعی شده است. در چنین شرایطی، ضرورت ارائه راهکارهای کارآمد برای مدیریت و کنترل جریان اطلاعات بیش از پیش احساس می‌شود. در این میان، هوش مصنوعی به‌عنوان یک فناوری دوگانه مطرح است؛ به‌گونه‌ای که هم ظرفیت‌های قابل توجهی برای مقابله با اطلاعات نادرست فراهم می‌کند و هم خود می‌تواند به تشدید این بحران کمک کند. ازیک‌سو، ابزارهایی نظیر الگوریتم‌های یادگیری ماشین، تحلیل کلان‌داده‌ها و سیستم‌های خودکار راستی‌آزمایی می‌توانند درشناسایی، پایش و کاهش انتشار محتوای نادرست مؤثر واقع شوند. ازسوی‌دیگر، همین فناوری با امکان تولید محتوای جعلی پیشرفته مانند دیپ‌فیک‌ها، تقویت پروفایل‌سازی رفتاری کاربران و هدف‌گیری دقیق اطلاعاتی، زمینه گسترش و پیچیده‌تر شدن اطلاعات نادرست را فراهم می‌کند. این پژوهش با روش توصیفی-تحلیلی و با استفاده از منابع کتابخانه‌ای، به بررسی نقش دوگانه هوش مصنوعی در مواجهه با اطلاعات نادرست می‌پردازد و ظرفیت‌ها و محدودیت‌های آن را تحلیل می‌کند. یافته‌های پژوهش نشان می‌دهند که هیچ راه‌حل واحد و مطلقی برای کنترل این پدیده وجود ندارد، بلکه ترکیبی از اقدامات نظیر شفاف‌سازی عملکرد الگوریتم‌ها، ارتقاء سواد رسانه‌ای و تدوین چهارچوب‌های نظارتی کارآمد ضروری است. همچنین نتایج پژوهش بیانگر آن است که همکاری میان دولت‌ها، پلتفرم‌های دیجیتال و کاربران نقش کلیدی در مدیریت این چالش ایفاء می‌کند.

واژگان کلیدی: هوش مصنوعی، اطلاعات نادرست، سواد رسانه‌ای.

Received: 2026/06/28 -Review: 2026/08/16 -Accepted: 2026/09/13

دریافت مقاله: ۱۴۰۵/۰۴/۰۷ - بازنگری مقاله: ۱۴۰۵/۰۵/۲۵ - پذیرش مقاله: ۱۴۰۵/۰۶/۲۲

ارجاع:
لرکجوری، عاطفه؛ پیروندیری، سیداحمد؛ (۱۴۰۵)، نقش دوگانه هوش مصنوعی در مقابله با اطلاعات نادرست و جعل عمیق: چالش‌ها و راهکارها، تمدن حقوقی، شماره ۲۸.

مقدمه

در سال‌های اخیر، دموکراسی‌های غربی با ترکیبی از حملات سایبری، عملیات اطلاعاتی، خرابکاری‌های سیاسی و اجتماعی، بهره‌برداری از تنش‌های موجود در جوامع خود و نفوذ مالی بدخواهانه مواجه شده‌اند. عملیات اطلاعاتی که تمرکز این پژوهش بر آن است، توسط بازیگران خارجی به منظور دستکاری در شکل‌گیری افکار عمومی، کاهش اعتماد عمومی به رسانه‌ها و نهادها، بی‌اعتبار کردن رهبری سیاسی، عمیق‌تر کردن شکاف‌های اجتماعی و تأثیرگذاری بر تصمیمات رأی‌دهندگان به کار گرفته شده است. این چالش‌ها در پس‌زمینه‌ای از اقتصاد دیجیتال در حال رشد که با ظهور و پذیرش سریع فناوری‌های جدید مانند اینترنت اشیاء، رباتیک و هوش مصنوعی یا واقعیت افزوده و مجازی همراه است، در حال وقوع هستند. اگرچه اطلاعات نادرست آنلاین پدیده جدیدی نیست، پیشرفت‌های سریع در فناوری‌های اطلاعاتی[۱] روش‌های تولید و انتشار اطلاعات[۲] را تغییر داده است (Bentzen,2018,14).

با وجود مزایای فراوان، سیستم‌های مجهز به هوش مصنوعی مجموعه‌ای از پرسش‌های اخلاقی را

۱- به ویژه استفاده از هوش مصنوعی

۲- و اطلاعات نادرست

مطرح می‌کنند و خطرات جدیدی برای حقوق بشر و فرایندهای سیاسی دموکراتیک ایجاد می‌کنند. نگرانی‌هایی که توسط جامعه کارشناسان مطرح شده است شامل عدم عدالت الگوریتمی[۳]، شخصی‌سازی محتوا که منجر به کوربینی اطلاعاتی[۴] می‌شود، نقض حریم خصوصی کاربران، احتمالاً دستکاری کاربران، یا دستکاری ویدئو و صدا بدون رضایت فرد است (Frissen,2018,87).

انتخابات ریاست جمهوری ایالات متحده امریکا در سال ۲۰۱۶ میلادی شواهدی از تأثیر تحول دیجیتال بر دموکراسی و زندگی سیاسی را نشان داد. استفاده از الگوریتم‌ها، خودکارسازی و هوش مصنوعی کارآیی و گستردگی کمپین‌های اطلاعات نادرست و فعالیت‌های سایبری مرتبط را افزایش داد که بر شکل‌گیری نظرات و تصمیمات رای‌دهی شهروندان امریکایی تأثیر گذاشت. با رشد نقش هوش مصنوعی در فناوری‌هایی که زندگی روزمره ما را تقویت می‌کنند، الگوریتم‌ها نفوذ بیشتری پیدا خواهند کرد و به بازیگران مخرب امکان می‌دهند تا به شبکه‌های دولتی و شرکتی نفوذ کرده، اطلاعات را سرقت کنند، حریم خصوصی افراد را به خطر بیندازند و انتخابات را بدون ردپای چندانی تحریف کنند.

طبق گزارش صادر شده توسط گروه کارشناسی سطح بالای کمیسیون اروپا درباره اخبار جعلی و اطلاعات نادرست آنلاین، این پژوهش اطلاعات نادرست را به‌عنوان «اطلاعات نادرست، غیر دقیق، یا گمراه‌کننده‌ای که طراحی، ارائه و ترویج شده‌اند تا عمداً به جامعه آسیب برسانند یا سود مالی کسب کنند» تعریف می‌کند.[۵] این پژوهش اطلاعات نادرست را از اطلاعات غلط[۶] و از سخنان نفرت‌انگیز متمایز می‌کند. اگرچه تعریف عمومی پذیرفته‌شده‌ای از هوش مصنوعی وجود ندارد، اما این اصطلاح می‌تواند به‌عنوان توانایی یک سیستم در انجام وظایفی که مشخصه‌های هوش انسانی مانند یادگیری و

۳- که منجر به تبعیض‌های نژادی و جنسیتی می‌شود

۴- ایجاد حباب فیلتر

5- European Commission, 2018

۶- که به اطلاعات غیرعمدی گمراه‌کننده یا نادرست اشاره دارد

تصمیم‌گیری دارند، درک شود. یادگیری ماشینی[۷] به‌طور کلی به استفاده از الگوریتم‌ها و مجموعه داده‌های بزرگ برای آموزش سیستم‌های کامپیوتری به منظورشناسایی الگوهایی که قبلاً تعریف نشده‌اند و توانایی این سیستم‌ها در یادگیری از داده‌ها و استخراج اطلاعات ارزشمند بدون برنامه‌ریزی صریح برای این کار اشاره دارد. در این پژوهش، هوش مصنوعی به تکنیک‌های یادگیری ماشینی اشاره دارد که به سمت هوش مصنوعی پیش می‌روند، مانند برنامه‌های تحلیل صوتی-تصویری که به‌صورت الگوریتمی آموزش دیده‌اند تا محتوای مشکوک و حساب‌ها راشناسایی و تعدیل کنند و به قضاوت انسانی کمک کنند.

پژوهش حاضر کاربردهای هوش مصنوعی را به‌عنوان یک شمشیر دو لبه تشخیص می‌دهد. درحالی‌که هوش مصنوعی راه‌حلی قدرتمند، مقیاس‌پذیر و مقرون‌به‌صرفه برای جلوگیری از تحریف اطلاعات آنلاین[۸] ارائه می‌دهد، اما این فناوری نیز دارای محدودیت‌ها و پیامدهای ناخواسته خود است. این پژوهش ابتدا به بررسی راه‌هایی می‌پردازد که هوش مصنوعی می‌تواند برای مقابله با اطلاعات نادرست آنلاین استفاده شود. سپس به برخی از محدودیت‌های راه‌حل‌های هوش مصنوعی و تهدیدات مرتبط با تکنیک‌های هوش مصنوعی می‌پردازد. در نهایت، پژوهش تعدادی از راه‌حل‌ها و تحولات آینده هوش مصنوعی را که می‌تواند برای مقابله با تهدید اطلاعات نادرست تحت قدرت هوش مصنوعی مطرح شود، بررسی می‌کند.

۱- راه‌حل‌های مقابله با اطلاعات نادرست توسط هوش مصنوعی

با افزایش حجم اطلاعات نادرست، بررسی دستی صحت اطلاعات به‌طور فزاینده‌ای ناکارآمد و غیرمؤثر برای ارزیابی هر قطعه اطلاعات آنلاین تلقی می‌شود. اولین پیشنهادها برای خودکارسازی بررسی صحت اطلاعات آنلاین حدود یک دهه پیش مطرح شد (Jackson,2016,76). انتخابات ترامپ علاقه‌مندی به تحقیق در مورد بررسی صحت اطلاعات با کمک هوش مصنوعی را افزایش

7- ML

۸- از طریق‌شناسایی و حذف خودکار محتوای نادرست

داد. در سال‌های اخیر، موجی از بودجه‌های اضافی برای ابتکارات بررسی صحت اطلاعات خودکار اختصاص یافته است که به متخصصان درشناسایی، تأیید و اصلاح محتوای رسانه‌های اجتماعی کمک می‌کند (Funke,2019,32). در نگاه اول، هوش مصنوعی که بر اساس قدرت محاسباتی عمل می‌کند و نه رفتار انسانی، به نظر می‌رسد که یک راهکار بی‌طرفانه برای مقابله با اطلاعات نادرست ارائه می‌دهد. با بهبود دقت و عملکرد سیستم‌های هوش مصنوعی، انتظارات از این فناوری افزایش یافته که ماشین‌ها بتوانند در جایی که انسان‌ها شکست خورده‌اند، موفق شوند؛ یعنی در غلبه بر تعصبات شخصی در تصمیم‌گیری‌ها.

در زمینه عملیات اطلاعاتی، راهکارهای هوش مصنوعی به‌ویژه در تشخیص و حذف محتوای غیرقانونی، مشکوک و نامطلوب آنلاین مؤثر بوده‌اند. تکنیک‌های هوش مصنوعی همچنین در غربالگری وشناسایی حساب‌های جعلی رباتی موفق عمل کرده‌اند؛ تکنیک‌هایی که به نام ربات‌یابی و برچسب‌گذاری ربات شناخته می‌شوند. با برچسب‌گذاری حساب‌هایی که به‌عنوان ربات‌شناسایی شده‌اند، شرکت‌های رسانه‌های اجتماعی این امکان را برای کاربران فراهم می‌کنند تا بهتر محتوای مورد نظر را درک کنند و خودشان درباره صحت آن قضاوت کنند. بااین‌حال، از نظر دقت، الگوریتم‌های تشخیص هنوز نیاز به توسعه بیشتری دارند تا قابل مقایسه با فناوری فیلتر هرزنامه‌های ایمیل شوند. شرکت‌های گوگل، فیسبوک، توییتر و سایر ارائه‌دهندگان خدمات اینترنتی از الگوریتم‌های یادگیری ماشینی برای مقابله با ترول‌ها،شناسایی و حذف حساب‌های جعلی ربات‌ها وشناسایی محتوای حساس به‌صورت پیشگیرانه استفاده می‌کنند.

در حال حاضر، بررسی صحت اخبار به‌صورت کاملاً خودکار یک هدف دوردست باقی مانده استپلتفرم‌های رسانه‌های اجتماعی همچنان به ترکیبی از هوش مصنوعی برای کارهای تکراری و بازبینی انسانی برای موارد پیچیده‌تر[۹] متکی هستند (Faesen,2019). در سال ۲۰۱۸ میلادی، فیسبوک

۹- که به‌عنوان مدل‌ها و فیلترهای هیبریدی نیز شناخته می‌شوند

هفت هزار و پانصد ناظر انسانی برای بازبینی محتوای کاربران استخدام کرد. علاوه بر این، این شرکت قصد خود را برای تأسیس یک نهاد مستقل نظارت بر محتوا تا پایان سال ۲۰۱۹ میلادی اعلام کرد که این نهاد از اعضاء خارجی به جای کارمندان فیسبوک تشکیل خواهد شد و تصمیمات بحث‌برانگیز مربوط به نظارت بر محتوای فیسبوک را بررسی خواهد کرد (Harris,2019,56).

۲- محدودیت‌های راه‌حل‌های هوش مصنوعی

چندین محدودیت در استفاده از تکنیک‌های خودکار برای‌شناسایی و مقابله با اطلاعات نادرست وجود دارد. اولین نقص مهم، خطر مسدودسازی بیش از حد محتوای قانونی و دقیق است که به آن ویژگی فراگیری بیش از حد[10] هوش مصنوعی می‌گویند. این فناوری همچنان در حال توسعه است و مدل‌های هوش مصنوعی هنوز مستعد ارائه نتایج نادرست منفی یا مثبت هستند؛ به عبارت دیگر، محتوایی را به اشتباه به‌عنوان جعلی‌شناسایی می‌کنند که در واقع جعلی نیست. این نتایج نادرست می‌توانند به آزادی بیان آسیب برسانند و منجر به سانسور محتوای مشروع و قابل‌اعتمادی شوند که به اشتباه به‌عنوان اطلاعات نادرست برچسب‌گذاری شده است. این مسئله به این دلیل است که فناوری‌های خودکار هنوز در ارزیابی دقیق عبارات فردی محدودیت دارند. سیستم‌های فعلی هوش مصنوعی تنها قادر به‌شناسایی جملات ساده اعلانی هستند و نمی‌توانند ادعاهای ضمنی یا ادعاهای پیچیده‌ای که انسان‌ها به راحتی تشخیص می‌دهند، راشناسایی کنند. همین موضوع در مورد بیاناتی که نیاز به سرنخ‌های زمینه‌ای یا فرهنگی دارند نیز صدق می‌کند (Graves,2018,163).

برخی از الگوریتم‌های خودکار ممکن است خطر تکرار و حتی خودکارسازی تعصبات و ویژگی‌های شخصیتی انسانی را داشته باشند که این امر می‌تواند به نتایجی منجر شود که برای افراد در یک گروه خاص نامطلوب باشد. به گفته برخی ناظران، هر چقدر هم که قصد داشته باشیم فناوری‌مان عینی باشد، در نهایت تحت تأثیر افرادی است که آن را می‌سازند و داده‌هایی که آن را تغذیه می‌کنند.

10- overinclusiveness

انتخاب‌های طراحی نیز می‌توانند به‌طور ذاتی دارای نقص باشند. واتس‌اپ نمونه خوبی از این است که چگونه معماری و انتخاب‌های طراحی می‌توانند بر قطب‌بندی و اطلاعات نادرست تأثیر بگذارند؟ در این پلتفرم، پیام‌ها به‌صورت انتها به انتها رمزگذاری شده‌اند و بنابراین به‌طور طراحی شده، از دسترس ناظران محتوا خارج هستند. در کشورهایی مانند هندوستان، واتس‌اپ نه‌تنها یک کانال اصلی برای تبلیغات سیاسی است، بلکه یک کانال برای گزارش‌های نادرست و سخنان نفرت‌انگیز است که مشخص شده باعث تحریک خشونت‌های گروهی و قتل‌ها شده است. از آن جا که پیام‌های فوروارد شده هیچ اطلاعاتی در مورد منبع اصلی ندارند، تمایز بین پیام‌رسانی رسمی و گسترش غیرمجاز دروغ‌ها مشخص نیست که به نوبه خود، به عاملان این امکان را می‌دهد که به‌طور معقول دخالت خود را انکار کنند. درحالی‌که رمزگذاری یک ویژگی امنیتی است، حفظ حریم خصوصی گفت‌وگوها یک انتخاب طراحی است که اساساً پلتفرم‌ها را از هرگونه مسئولیت در قبال محتوای شبکه‌هایشان بی‌بهره می‌کند (Harris,2019,38).

پیچیدگی و عدم شفافیت نیز از دیگر محدودیت‌های سیستم‌های هوش مصنوعی هستند. یادگیری ماشین شامل شبکه‌های عصبی و شبکه‌های عصبی عمیق می‌شود که به‌طور ذاتی راه‌حل‌های جعبه سیاه هستند و تکامل آن‌ها بر اساس خودآموزی، فراتر از درک توسعه‌دهندگانی است که آن‌ها را ساخته‌اند (Abdallat,2019,66). منطق پیچیده تصمیم‌گیری خودکار باعث می‌شود الگوریتم‌ها دقیق‌تر باشند، اما همین موضوع توضیح چگونگی تولید یک توصیه خاص را دشوار می‌کند. تعدادی از شرکت‌ها، به‌ویژه در سیلیکون ولی و برنامه هوش مصنوعی قابل توضیح سازمان دفاعی ایالات متحده امریکا[11] در حال توسعه تکنولوژی‌ها و سیستم‌های چهارچوبی هستند که در نهایت می‌توانند امکان صحت‌سنجی، مسئولیت‌پذیری بیشتر و شفافیت در یادگیری ماشین را فراهم کنند.

در نهایت، راه‌حل‌های هوش مصنوعی پرسش‌های مهمی را در مورد این که چه کسی بهترین جایگاه برای تعیین قانونی یا غیرقانونی بودن محتوا، مطلوب یا نامطلوب بودن آن را دارد؟ مطرح

11- DARPA

می‌کند. آیا قضاوت در مورد صحت و حذف فوری محتوای آنلاین باید در اختیار نهادهای عمومی[۱۲]، مقامات قضایی، یا پلتفرم‌های آنلاین باشد؟

۳- جعل عمیق و ارتباط آن با هوش مصنوعی

این تکنولوژی به ویژه در دنیای دیجیتال و رسانه‌های اجتماعی نگرانی‌های جدی در رابطه با اطلاعات نادرست و دستکاری‌شده ایجاد کرده است. استفاده از هوش مصنوعی در تولید محتوای صوتی و تصویری چالشی حتی بزرگ‌تر را به همراه دارد. به‌اصطلاح «دیپ فیک‌ها» محتوای صوتی یا تصویری که به‌صورت دیجیتالی دست‌کاری شده و بسیار واقعی و تقریباً غیرقابل تشخیص از محتوای واقعی است در ابتدا در صنعت فیلم‌سازی استفاده می‌شد. امروزه، این فناوری در حوزه‌های آنلاین سرگرمی، فریب مصرف‌کنندگان و حتی سیاست و امور بین‌الملل کاربرد دارد. نرم‌افزارهای تجاری و حتی رایگان در بازار آزاد در دسترس هستند. انتظار می‌رود که به زودی، تنها محدودیت عملی در توانایی یک فرد برای تولید دیپ فیک، دسترسی به مجموعه داده‌های آموزشی کافی، یعنی ویدئو و صوت از شخص مورد نظر، باشد. ویدئوها و تصاویر جعل‌شده هنوز هم دارای مصنوعات زیادی هستند که تشخیص آن‌ها را آسان می‌کند (Kietzmann,2023,14).

تصاویر و ویدئوهای بسیار واقعی و دشوار برای‌شناسایی از افراد واقعی که کارهایی را انجام می‌دهند یا چیزهایی را می‌گویند که هرگز انجام نداده‌اند یا نگفته‌اند، می‌توانند اعتبار رهبران و نهادها را خدشه‌دار کنند، باعث تحریک به خشونت و ناآرامی‌های مدنی در شهرها شوند، اختلافات موجود در جامعه را تشدید کنند، یا نتیجه انتخابات را تحت تأثیر قرار دهند. سهولت روزافزون در ساخت و به اشتراک‌گذاری محتوای جعلی ویدئویی و صوتی از طریق کامپیوترها و دستگاه‌های موبایل ممکن است فرصت‌های زیادی برای ارعاب، اخاذی و خرابکاری فراتر از حوزه سیاست و امور بین‌المللی ایجاد کند. موارد متعددی از ویدئوهای تولید شده توسط هوش مصنوعی وجود داشته است که

۱۲- چه مرتبط با دولت‌ها یا نه

سیاست‌مداران را در حال بیان جملاتی نشان می‌دهند که هرگز نگفته‌اند.

محققان در گلوبال پلاس[۱۳]، اخیراً روشی برای آموزش هوش مصنوعی جهت ایجاد سخنرانی‌های جعلی سازمان ملل متحد طراحی کردند. آن‌ها از یک مدل زبان موجود[۱۴] که بر اساس متون ویکی‌پدیا آموزش دیده بود، استفاده کردند و آن را بر روی مجموعه‌داده‌ای از سخنرانی‌های مجمع عمومی سازمان ملل متحد تنظیم کردند. در عرض سیزده ساعت، مدل هوش مصنوعی توانست سخنرانی‌های واقعی‌گونه‌ای در مورد طیف گسترده‌ای از موضوعات بحث‌برانگیز، از جمله تغییرات اقلیمی، مهاجرت و خلع سلاح هسته‌ای تولید کند. این آزمایش با هدف نشان دادن سهولت و سرعتی که هوش مصنوعی می‌تواند محتوای واقعی‌گونه تولید کند و همچنین تهدیدی که ترکیب این تکنیک با دیگر فناوری‌ها، مانند دیپ فیک‌ها، ایجاد می‌کند، انجام شد (Chadwick,2018,22).

برای مثال، می‌توان متن بحث‌برانگیزی را برای یک سخنرانی که ظاهراً توسط یک رهبر سیاسی داده شده است تولید کرد، یک ویدیوی «دیپ فیک» از آن رهبر در حال ایراد سخنرانی در مجمع عمومی سازمان ملل متحد ساخت[۱۵] و سپس این شبیه‌سازی را با تولید انبوه مقالات خبری که به‌ظاهر گزارش آن سخنرانی را می‌دهند، تقویت کرد. دیپ فیک‌ها این امکان را برای بازیگران بدخواه فراهم می‌کنند که به دو روش حقیقت را انکار کنند: نه‌تنها ویدیوهای جعلی ممکن است به‌عنوان واقعی به‌منظور ایجاد شک و تردید منتشر شوند، بلکه اطلاعات اصیل نیز ممکن است به‌عنوان جعلی تلقی شوند. با افزایش آگاهی عمومی نسبت به تهدیدات دیپ فیک‌ها، این تکنیک اخیر احتمالا بیش از پیش قابل‌باور خواهد شد.

۴- نقش هوش مصنوعی درشناسایی و مقابله با جعل عمیق

جعل عمیق یکی از پدیده‌های نوظهور در دنیای دیجیتال است که با استفاده از تکنولوژی‌های پیشرفته

13- Global Pulse
14- awd-lstm

۱۵- با استفاده از حجم زیادی از فیلم‌های این‌چنینی آموزش دیده

مانند یادگیری ماشین و شبکه‌های عصبی مصنوعی، امکان تولید محتوای جعلی و دستکاری‌شده به شکلی بسیار واقعی فراهم می‌آید. این تکنولوژی‌ها قادرند تصاویر، ویدئوها و حتی صداهای انسان‌ها را به‌طور شبیه‌سازی شده تولید کنند که این موضوع مشکلات جدی در حوزه‌های مختلف از جمله امنیت، حریم خصوصی و حقوق افراد ایجاد کرده است.

در این زمینه، هوش مصنوعی می‌تواند به‌عنوان ابزاری مهم و کارآمد درشناسایی و مقابله با جعل عمیق عمل کند. یکی از راهکارهای مؤثر در این حوزه استفاده از مدل‌های یادگیری عمیق است که به تشخیص تغییرات غیرطبیعی در تصاویر و ویدئوها می‌پردازد. این مدل‌ها با تحلیل ویژگی‌های ظریف و غیرقابل مشاهده توسط انسان‌ها، قادر به‌شناسایی جعل‌های عمیق حتی در مقیاس‌های کوچک و پیچیده هستند. یکی از تکنیک‌های رایج در این زمینه، استفاده از شبکه‌های عصبی متمایزکننده است که برای تشخیص تفاوت‌های بین محتوای واقعی و جعلی به‌کار می‌روند. این شبکه‌ها با بررسی ویژگی‌های بیولوژیکی مانند حرکت چشم، تغییرات در پوست صورت و نحوه تعامل اجزای صورت با نور، می‌توانند به‌طور دقیق‌ترین تفاوت‌ها راشناسایی کنند.

در نتیجه، با توجه به پیشرفت‌های روزافزون در زمینه فناوری‌های مبتنی‌بر هوش مصنوعی، این فناوری‌ها نه‌تنها می‌توانند به‌شناسایی سریع‌تر و دقیق‌تر جعل‌های عمیق کمک کنند، بلکه می‌توانند در حفاظت از حقوق افراد و جلوگیری از استفاده نادرست از این تکنولوژی‌ها در زمینه‌های مختلف، نقشی اساسی ایفاء نمایند. جعل عمیق نه‌تنها مشکلات فنی ایجاد می‌کند، بلکه می‌تواند تهدیدات جدی برای امنیت اجتماعی و سیاست‌گذاری‌ها به همراه داشته باشد. برای مثال، جعل عمیق می‌تواند در تخریب اعتبار عمومی یا تحت تأثیر قرار دادن انتخابات استفاده شود، چرا که امکان ساخت ویدئوهایی از شخصیت‌های سیاسی که کاملاً واقعی به نظر می‌رسند، می‌تواند به گمراه کردن افکار عمومی منجر شود.

این الگوریتم‌ها با استفاده از اطلاعات بیشتر، از جمله حرکت‌های غیرطبیعی بدن، شتاب‌های غیرمعمول و یا ناهماهنگی‌های صوتی و تصویری، می‌توانند به‌شناسایی جعل‌های پیچیده‌تر کمک

کنند. یکی دیگر از رویکردهای نوین درشناسایی جعل عمیق، استفاده از تکنیک‌های ترکیبی است که به تحلیل داده‌های تصویری و صوتی به‌طور هم‌زمان می‌پردازد. به‌عنوان مثال، با ترکیب پردازش تصویر با پردازش زبان طبیعی، می‌توان به بررسی تناقضات موجود در متن‌های همراه با ویدئوهای جعلی پرداخت. این نوع روش‌ها می‌توانند به‌شناسایی ویدئوهایی که در آن‌ها صدا و تصویر با هم تطابق ندارند، کمک کنند.

علاوه بر پیشرفت‌های فنی در حوزه‌شناسایی، آموزش عمومی و آگاهی‌بخشی نیز برای مقابله با جعل عمیق ضروری است. درحالی‌که هوش مصنوعی می‌تواند در تشخیص جعل‌ها مؤثر باشد، اما استفاده از این ابزار باید به‌طور گسترده در سطح جامعه گسترش یابد تا افراد بتوانند به‌راحتی محتوای جعلی راشناسایی کرده و از آن اجتناب کنند. به این ترتیب، توجه به ساخت و توسعه پلتفرم‌های آموزشی و ابزارهای ضدجعل که به کاربران کمک می‌کند تا توانایی تشخیص جعل‌های عمیق را بیاموزند، نقش مهمی در مقابله با این تهدید خواهد داشت.

در نهایت، استفاده از پروتکل‌ها و قوانین قانونی به‌منظور مقابله با جعل عمیق در کنار توسعه فناوری‌های‌شناسایی ضروری است (Dolhansky,2023). باید قوانین جدیدی تصویب شود که استفاده از تکنولوژی‌های جعل عمیق را محدود کرده و مجازات‌های سختی برای تولیدکنندگان محتوای جعلی تعیین کند. این اقدامات باید در کنار پیشرفت‌های فنی به‌کار گرفته شوند تا جامعه بتواند به‌طور مؤثری با تهدیدات ناشی از جعل‌های عمیق مقابله کند.

۶- آینده جعل عمیق و خطرات ناشی از آن

یکی از نگرانی‌های عمده‌ای که در مورد جعل عمیق وجود دارد، تهدیداتی است که این تکنولوژی برای جامعه ایجاد می‌کند. اولین خطر بزرگ این است که جعل عمیق می‌تواند به طرز چشمگیری توانایی تشخیص واقعیت از غیرواقعیت را در میان کاربران کاهش دهد. جعل‌های عمیق می‌توانند در رسانه‌ها، شبکه‌های اجتماعی و حتی در انتخابات سیاسی استفاده شوند تا تصویری نادرست از واقعیت ارائه دهند.

برای مثال، می‌توان از این تکنولوژی برای ساخت ویدئوهایی استفاده کرد که در آن سیاست‌مداران به‌صورت غیرواقعی در حال بیان اظهاراتی هستند که هیچ‌گاه در واقعیت انجام نداده‌اند. چنین جعل‌هایی می‌توانند به سرعت پخش شده و افکار عمومی را تحت تأثیر قرار دهند (Franks et al.,2023).

خطر دیگر، استفاده از جعل عمیق در کلاهبرداری و جعل هویت است. این تکنولوژی می‌تواند از طریق تولید ویدئو یا صوت جعلی، دسترسی‌های غیرمجاز به اطلاعات حساس، مانند حساب‌های بانکی، فایل‌های شخصی و اطلاعات خصوصی دیگر ایجاد کند. به‌عنوان مثال، با جعل صدای یک فرد و ترکیب آن با ویدئوهای جعلی، می‌توان کلاهبرداری‌هایی انجام داد که قربانیان حتی متوجه ساختگی بودن آن نشوند (Wang,2019).

در پاسخ به این چالش‌ها، تلاش‌هایی برای‌شناسایی جعل عمیق انجام شده است. الگوریتم‌های یادگیری ماشین که بر پایه شبکه‌های عصبی کانولوشنی طراحی شده‌اند، می‌توانند تغییرات غیرطبیعی در ویدئوها و تصاویر راشناسایی کنند. این الگوریتم‌ها به‌صورت مداوم بهبود می‌یابند تا با پیشرفت‌های پیچیده‌تر در تولید جعل عمیق مقابله کنند (Dolhansky et al.,2020). علاوه بر این، بلاک‌چین به‌عنوان یک راهکار بالقوه برای تأیید اصالت محتوا مطرح شده است. این فناوری می‌تواند منبع و تاریخچه محتوا را رهگیری کرده و تضمین کند که تغییراتی در محتوای دیجیتال اعمال نشده است (Yang et al.,2023).

در نهایت، جعل عمیق یک چالش چندوجهی است که نیازمند ترکیب فناوری‌های پیشرفته، چهارچوب‌های قانونی و آگاهی عمومی برای مقابله با سوءاستفاده‌های احتمالی است. اگرچه این فناوری می‌تواند در زمینه‌های مثبت مانند سرگرمی و آموزش تحول‌آفرین باشد، اما بدون مدیریت صحیح می‌تواند تهدیدی جدی برای امنیت فردی و اجتماعی باشد (Franks,2023).

۷- راهکارها و توصیه‌ها برای مقابله با اطلاعات نادرست و سایر تهدیدات آنلاین

سیاست‌گذاران، جامعه کاربران، حقیقت‌سنجان، پلتفرم‌های رسانه‌های اجتماعی، روزنامه‌نگاران و سایر

ذینفعان با چالشی پیچیده مواجه هستند که نمی‌توان آن را با یک راه‌حل واحد و همه‌جانبه حل کرد. برای قانون‌گذار، استفاده از هوش مصنوعی برای مقابله با اطلاعات نادرست و سایر تهدیدات آنلاین، پرسش‌های زیادی در حوزه تنظیم مقررات به وجود می‌آورد.

۷-۱- کاهش تأکید و اصلاح محتوای نادرست

شرکت‌های رسانه‌های اجتماعی می‌توانند الگوریتم‌های خبررسانی خود را به‌روزرسانی کنند تا تأکید بر اطلاعات نادرست کاهش یابد. علاوه بر نشانه‌گذاری و کاهش رتبه محتوای نادرست، برای پلتفرم‌ها مهم است که اصلاحات مؤثر محتوای به‌وضوح نادرست یا گمراه‌کننده که به‌صورت آنلاین ظاهر شده است را نمایش دهند. همچنین، انتشار پیام‌های متقابل مبتنی‌بر واقعیت از اهمیت بالایی برخوردار است. اگرچه انتساب در کمپین‌های اطلاعات نادرست آنلاین پیچیده است، در صورت وجود شواهد کافی، مهم است که عاملان اطلاعات نادرست را علناً محکوم کرده و انتساب و پاسخ را هماهنگ کنند.

۷-۲- ترویج مسئولیت‌پذیری و شفافیت بیشتر

ممکن است تعصبات موجود در سیستم‌های تصمیم‌گیری الگوریتمی با انجام حسابرسی از سیستم‌های هوش مصنوعی جبران شود. حسابرسی باعث افزایش نظارت بر داده‌ها و فرایندهایی می‌شود که برای ایجاد مدل‌ها با استفاده از داده‌ها به کار رفته‌اند.[۱۶]

۷-۳- راه‌حل‌های تکنولوژیکی برای مقابله با جعل عمیق

درخصوص راه‌حل‌های مقابله با دیپ فیک‌ها، استادان حقوق رابرت چسنی و دانیل سیترون سه راه‌حل تکنولوژیکی پیشنهاد می‌کنند. اولین راه‌حل مربوط به تشخیص بهتر مواد جعلی با استفاده از ابزارهای جنایی است. از آن جا که اکثر مجموعه‌های آموزشی فاقد تصاویری از صورت‌ها با چشم‌های بسته هستند، تکنیک‌هایی برای جست‌وجوی الگوهای غیرعادی حرکت پلک‌ها توسعه یافته است تا

16- European Commission, 2019

تشخیص دیپ فیک‌ها بهبود یابد. اما با توجه به این که تکنولوژی دیپ فیک‌ها بر اساس یک دینامیک ویروس یا ضدویروس توسعه می‌یابد، به محض این که این تکنیک جنایی عمومی شد، نسل جدید دیپ فیک‌ها به سرعت به آن تطبیق پیدا کرد.

دومین راه‌حل تکنولوژیکی شامل احراز هویت محتوا قبل از انتشار آن است.[۱۷] اگر محتواهای صوتی، تصویری و ویدیویی در لحظه ایجاد، به‌صورت دیجیتالی واترمارک شوند، این شواهد می‌تواند بعداً به‌عنوان مرجعی برای مقایسه با دیپ فیک‌های مشکوک استفاده شود. سومین رویکرد تکنولوژیکی که بیشتر نظری است، به خدمات آلبی معتبر مربوط می‌شود که تمام فعالیت‌ها، حرکات و مکان‌های فرد را نظارت و ذخیره می‌کند تا ثابت کند که فرد در هر زمان کجا بوده و چه گفته یا کرده است. اگرچه خدمات آلبی و ثبت فعالیت‌های زندگی ممکن است برای افراد برجسته با شهرت شکننده مانند سلبریتی‌ها و سیاست‌مداران بسیار ارزشمند باشد، اما این‌ها پیامدهای جدی منفی برای حریم خصوصی فردی دارند. پیشنهاد دیگر این است که از همان ابزارهایی که دیپ فیک‌ها را تولید می‌کنند برای‌شناسایی آن‌ها استفاده شود. یک جنبه حیاتی از این راه‌حل، تعیین نقش‌ها، مسئولیت‌ها و مسئولیت‌پذیری ذینفعان مختلف است. بازتنظیم این اکوسیستم نباید به معنای تعیین پلتفرم‌های آنلاین به‌عنوان قاضی و هیئت منصفه در تعیین حقیقت باشد.

این می‌تواند به سانسور بیش از حد منجر شود، زیرا پلتفرم‌ها ممکن است به دلیل احتیاط و ترس از جریمه‌ها، محتوای قانونی را نیز حذف کنند. این خطر باعث بروز جنجال در مورد قانون آلمان درباره اخبار جعلی شد که از اول ژانویه ۲۰۱۸ میلادی به اجرا درآمد. این قانون محدودیتی بیست و چهار ساعته را برای پلتفرم‌ها تعیین می‌کند تا اخبار جعلی و سخنان نفرت‌انگیز را حذف کنند که برای پلتفرم‌های بزرگ عملی نیست و برای شرکت‌های کوچک غیرقابل اجرا است. این موضوع برای شرکت‌های بزرگ اینترنتی نیز صدق می‌کند. درحالی‌که جامعه کاربران فیس‌بوک بزرگ‌تر از

۱۷- به اصطلاح راه‌حل منشأ دیجیتال

جمعیت چین و هندوستان است، تعداد کارکنان آن که به امنیت و ایمنی محتوای آنلاین پرداخته‌اند، به زحمت به تعداد نیروی پلیس بلژیک[۱۸] می‌رسد. یک چهارچوب کارآمدتر برای تنظیم محیط آنلاین در زمینه محتوای آلوده نیازمند حمایت بیشتری از شرکت‌ها رسانه‌های اجتماعی است که با مسئله‌ای روبرو هستند که بزرگ‌تر از خودشان شده است.

به‌طور مشابه، قانون اخیر روسیه انتشار اطلاعات نادرست آنلاین، از جمله اظهاراتی که به بی‌احترامی به دولت می‌پردازند، را جرم‌انگاری کرده است. این زبان به‌طور شگفت‌انگیزی مبهم می‌تواند امکان سانسور سیاسی را برای خاموش کردن مخالفان فراهم کند. توجه به این نکته مهم است که تلاش‌ها برای تنظیم مقررات و تدوین سیاست برای فناوری‌ای که تعاریف، ریسک‌ها، چالش‌ها و زمینه‌های آن متغیر است، نیازمند احتیاط و گفت‌وگوی مداوم با تمام ذینفعان است. این باید یک همکاری مشترک بین صنعت، دانشگاه و دولت باشد و رویکردهای نظارتی معمولی لزوماً در یک محیط با سرعت بالا کارآمد نخواهد بود. تنوع شرکت‌های آنلاین نیازمند قوانین و استانداردهای متنوع و مناسبی برای پاسخگویی است. پیشنهاد اجرای یکسان و کلی الزاماًت برای همه، اشتباه خواهد بود.

۷-۴- تکنوپلوماسی در کنار دیپلماسی

به دنبال پیشروی دانمارک، کشورها می‌توانند با ایجاد «هیئت‌های فناوری» یا «سفیران فناوری» و یا با واگذاری برخی از این مسئولیت‌ها به مقامات ملی مرتبط موجود، تعامل و اعتماد میان ذینفعان مختلف را افزایش دهند. با شناخت این که شرکت‌های فناوری بازیگران مهمی در تأثیرگذاری بر سیاست‌های جهانی هستند، تکنوپلوماسی پتانسیل ایجاد مسیرهای جدید برای گفت‌وگو و همکاری میان صنعت فناوری و دولت‌ها را دارد. تصمیم‌گیرندگان کشورها می‌توانند از این رابطه برای بحث در مورد مسائلی مانند دخالت در انتخابات، اطلاعات نادرست و محتوای مضر، امنیت سایبری یا جمع‌آوری شواهد الکترونیکی برای تحقیقات سیاسی استفاده کنند. تکنوپلوماسی همچنین می‌تواند به اطمینان از

۱۸- سی هزار نفر

این که شرکت‌های فناوری به اندازه تأثیری که دارند، مسئولیت‌های خود را بر عهده می‌گیرند، کمک کند. در سطح ملی، تکنوپلوماسی می‌تواند به فرد یا نهادی واگذار شود که مسئول نظارت و هماهنگی تلاش‌های سفیران سایبری و دیپلمات‌های موجود است. اطمینان از این که این مقام ملی بتواند به‌طور مؤثری از این مسئولیت‌ها حمایت کند، باید یک مأموریت مشخص و حمایت قوی سیاسی به آن داده شود.

۷-۵- سواد رسانه‌ای و دیجیتال

در مبارزه با اطلاعات نادرست، راه‌حل‌های فناورانه به تنهایی برای مقابله با این مشکل کافی نیستند. همان‌طور که دکتر الکساندر کلیمبورگ، مدیر برنامه سیاست و تاب‌آوری سایبری در مرکز مطالعات استراتژیک لاهه، بیان می‌کند: «حمله به بدن سایبر[۱۹] تنها یک مسیر انحرافی برای حمله به ذهن انسان است»؛ بنابراین، پاسخ‌ها باید فراتر از تکنولوژی باشند و بر بعد روان‌شناختی تمرکز کنند. در نهایت، دلیل موفقیت اطلاعات نادرست این است که برای آن‌ها مخاطب وجود دارد. افزایش سواد رسانه‌ای و دیجیتال ممکن است یکی از کارآمدترین و قدرتمندترین ابزارها برای بازگرداندن رابطه سالم با اطلاعات و افزایش تاب‌آوری دموکراسی‌های ما در برابر اطلاعات نادرست آنلاین باشد. آموزش سواد دیجیتال و رسانه‌ای باید از کودکی تشویق شود. تمرکز نباید فقط بر کودکان باشد، بلکه باید شامل مسئولان انتخابات، شهروندان مسن و گروه‌های حاشیه‌نشین و اقلیت‌ها نیز شود. در واقع، شهروندان مسن که با بزرگ‌ترین شکاف در زمینه سواد دیجیتال مواجه هستند، بیشترین احتمال را برای شرکت در انتخابات ملی دارند.

نتیجه

در دنیای امروز، فناوری هوش مصنوعی به ابزاری دوگانه تبدیل شده است که می‌تواند هم در مقابله با اطلاعات نادرست و تقویت فرایندهای نظارتی مؤثر باشد و هم برای ایجاد چالش‌های جدید از جمله

۱۹- لایه‌های فنی

جعل عمیق مورد سوءاستفاده قرار گیرد. این فناوری با توانایی شبیه‌سازی محتوا، تهدیدی جدی برای حریم خصوصی، امنیت اطلاعات و حتی ثبات جوامع دموکراتیک به شمار می‌رود. ازسوی‌دیگر، پتانسیل آن درشناسایی و حذف اطلاعات نادرست نشان می‌دهد که می‌توان از آن برای ایجاد جامعه‌ای آگاه‌تر و پایدارتر بهره گرفت.

یکی از چالش‌های اصلی در این زمینه، ابعاد حقوقی فناوری جعل عمیق است. این فناوری نه‌تنها مسئولیت حقوقی سازندگان و کاربران محتوای جعلی را به یک مسئله پیچیده تبدیل کرده است، بلکه با نقض حریم خصوصی و ایجاد آسیب‌های روانی و اقتصادی برای افراد، نظم حقوقی موجود را نیز به چالش می‌کشد. درعین‌حال، قوانین و مقررات فعلی در بسیاری از کشورها قادر به همگامی با سرعت پیشرفت این فناوری نیستند و شکاف‌های قانونی موجود موجب دشواری در پاسخگویی مناسب به سوءاستفاده‌ها می‌شود.

بااین‌حال، مقابله با این چالش‌ها نیازمند یک رویکرد جامع است. در سطح فردی، سرمایه‌گذاری در سواد دیجیتال و رسانه‌ای برای توانمندسازی شهروندان در تشخیص و تحلیل اطلاعات نادرست ضروری است. در سطح نهادی، همکاری نزدیک میان دولت‌ها، پلتفرم‌های فناوری و سازمان‌های بین‌المللی برای تدوین چهارچوب‌های قانونی شفاف و هماهنگ و توسعه فناوری‌های پیشرفته‌شناسایی محتوا اهمیت دارد. علاوه بر این، حفظ تعادل میان مقابله با محتوای جعلی و احترام به آزادی بیان یکی از الزامات حیاتی است. وضع قوانین سخت‌گیرانه بدون توجه به جنبه‌های خلاقانه و مشروع فناوری‌های هوش مصنوعی می‌تواند به سرکوب نوآوری منجر شود. در همین راستا، سازمان‌های بین‌المللی نظیر او اس سی اس ای[۲۰] و اتحادیه اروپا می‌توانند نقش کلیدی در ترویج استانداردهای اخلاقی و همکاری جهانی ایفاء کنند.

در نهایت، برای کاهش پیامدهای منفی فناوری‌های مبتنی‌بر هوش مصنوعی و بهره‌گیری از ظرفیت‌های مثبت آن، نیاز به بازنگری در قوانین، تقویت تحقیق و توسعه، و ترویج اخلاق در توسعه و

20- OSCE

استفاده از این فناوری‌ها است. تنها با یکپارچگی تلاش‌ها در سطح فردی، نهادی و بین‌المللی می‌توان از موج پیچیده اطلاعات نادرست و آسیب‌های جعل عمیق کاست و از هوش مصنوعی به‌عنوان ابزاری برای تقویت ارزش‌های دموکراتیک و حفاظت از حقوق بشر استفاده کرد.

چالش‌های حقوقی جعل عمیق، از جمله نقض حریم خصوصی، آسیب به اعتبار افراد و ایجاد دشواری در تعیین مسئولیت حقوقی، تنها بخشی از پیچیدگی‌های این پدیده است. این فناوری به بازیگران بدخواه این امکان را می‌دهد تا در پوشش ناشناس، عملیات اطلاعاتی خود را گسترش دهند و موجب بی‌ثباتی در جوامع شوند. ازسوی‌دیگر، نبود چهارچوب‌های قانونی جامع در بسیاری از نظام‌های حقوقی، فرایند پیگرد قانونی و اعمال مجازات را دشوارتر می‌کند. این موارد نشان می‌دهند که نیاز به بازنگری در ساختارهای قانونی و تنظیم مقررات شفاف و منسجم، بیش از هر زمان دیگری احساس می‌شود.

به موازات این چالش‌ها، جنبه‌های اجتماعی نیز اهمیت بسیاری دارند. سرمایه‌گذاری در آموزش عمومی، به ویژه در زمینه سواد دیجیتال و رسانه‌ای، می‌تواند نقش مؤثری در تقویت توانایی شهروندان برای تشخیص محتوای جعلی و تحلیل انتقادی اطلاعات ایفاء کند. جامعه‌ای که بتواند در برابر اطلاعات نادرست تاب‌آور باشد، احتمالا کمتر در معرض آسیب‌های ناشی از موج اطلاعات غلط قرار می‌گیرد. ازسوی‌دیگر، همکاری‌های بین‌المللی و مشارکت میان دولت‌ها، پلتفرم‌های فناوری و سازمان‌های بین‌المللی برای مقابله با سوءاستفاده از هوش مصنوعی ضروری است. ایجاد استانداردهای جهانی برای استفاده مسئولانه از این فناوری و ترویج اخلاق در توسعه آن، گامی کلیدی در جهت کاهش آسیب‌ها و تقویت کاربردهای مثبت هوش مصنوعی است. در سطح فنی، توسعه الگوریتم‌های بی‌طرف، کارآمد و دقیق که بتوانند محتوای جعلی را به‌طور مؤثرشناسایی کنند، یکی از الزامات است. اما این توسعه باید با رعایت حقوق بنیادین افراد، از جمله حریم خصوصی و آزادی بیان، همراه باشد. به همین دلیل، یکپارچگی میان سیاست‌گذاران، مهندسان و محققان اهمیت ویژه‌ای دارد تا راه‌حل‌هایی تدوین شود که نه‌تنها کارآمد باشند، بلکه با ارزش‌های انسانی نیز همسو باشند.

در نهایت، هوش مصنوعی و ابزارهایی مانند جعل عمیق، نمایانگر دو روی یک سکه هستند. ازیک‌سو، تهدیدی برای حقوق، ارزش‌ها و ساختارهای دموکراتیک و ازسوی‌دیگر، فرصتی برای بهبود فرایندهای اجتماعی و ایجاد جامعه‌ای آگاه‌تر. برای بهره‌برداری بهینه از این فناوری و کاهش خطرات آن، باید میان تمام ذینفعان در سطوح مختلف، از فردی تا بین‌المللی، یکپارچگی و همکاری ایجاد شود. تنها از این طریق می‌توان آینده‌ای پایدار، مسئولانه و همسو با ارزش‌های انسانی برای جوامع تضمین کرد.

ملاحظات اخلاقی: موارد مربوط به اخلاق در پژوهش و نیز امانتداری در استناد به متون و ارجاعات مقاله تماماً رعایت گردیده است.

تعارض منافع: تعارض منافع در این مقاله وجود ندارد.

تأمین اعتبار پژوهش: این پژوهش بدون تأمین اعتبار مالی نگارش یافته است.

منابع

- Abdallat, A. J., 2019, Explainable AI: Why we need to open the black box. Forbes.
- Bentzen, N., 2018, Computational propaganda techniques. European Parliamentary Research Service (EPRS).
- Chadwick, P., 2018, The liar’s dividend, and other challenges of deep-fake news. The Guardian.
- Dolhansky, B., Howes, C., & Encarnacion, P., State of deepfake detection: Challenges and solutions. ACM Computing Surveys, 55 (4).
- European Commission., 2018, A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation. Brussels: European Commission.
- European Commission., 2019, Code of practice against disinformation: Commission recognises platforms’ efforts ahead of the European elections.
- Faesen, L., et al., 2019, Understanding the strategic and technical significance of technology for security: Implications of AI and machine learning for cybersecurity. The Hague Security Delta (HSD).
- Franks, B., Smith, J., & Taylor, L., 2023, The ethics of deepfakes: Navigating truth and deception in the digital age. Journal of Media Ethics, 38 (2).
- Frissen, V., Lakemeyer, G., & Petropoulos, G., 2018, Ethics and artificial intelligence.

Bruegel.
- Funke, D., 2019, These fact-checkers won $2 million to implement AI in their newsrooms. Poynter.
- Graves, L., 2018, Understanding the promise and limits of automated fact-checking. The Reuters Institute for the Study of Journalism at the University of Oxford.
- Harris, J., 2019, Is India the frontline in big tech's assault on democracy? The Guardian.
- Jackson, J., 2016, Fake news clampdown: Google gives €150,000 to fact-checking projects. The Guardian.
- Kietzmann, J., Pitt, L., & Berthon, P., 2023, Deepfakes: Opportunities, challenges, and ethical dilemmas. Business Horizons, 66 (3).
- Wang, Y., Li, X., & Zhou, F., 2023, Deepfake detection and mitigation: Recent advances and future directions. IEEE Access, 11.
- Yang, X., Guo, R., & Liu, Z., 2023, Blockchain-based verification for deepfake content: A comprehensive review. Future Generation Computer Systems, 145.

Legal Civilization

ISSN: 2873-1841
ISSN: 2873-1922

No.28- Summer 2026