تربیت الگوریتم‌های مولد با داده‌های جنایی واقعی: شناخت روان جنایتکار یا بازتولید تعصب؟

سحر کیوانی راد [*۱]، نفیسه روحی‌خراسانی [۲]

۱- کارشناسی مهندسی کامپیوتر دانشگاه آل طه Saharrad2228@gmail.com

۲- کارشناس ارشد روانشناسی دانشگاه آزاد اسلامی اصفهان n.roohi@yahoo.com

خلاصه

رشد روزافزون مدل‌های مولد هوش مصنوعی، فرصت‌های تازه‌ای برای درک رفتار مجرمانه و تحلیل روان‌شناختی جنایتکاران فراهم کرده است. درعین‌حال، این روند با چالش‌های جدی اخلاقی،علمی و بین‌رشته‌ای همراه است. مقاله حاضر با رویکردی مروری-انتقادی و با بهره‌گیری از منابع علمی منتشرشده در پایگاه‌هایی همچون Google Scholar، Scopus، IEEE Xplore، PubMed، و SpringerLink به بررسی این پرسش می‌پردازد رویکردی مروری-انتقادی و بر پایه‌ی تحلیل نظام‌مند منابع علمی در حوزه‌های هوش مصنوعی، روان‌شناسی جنایی، اخلاق فناوری و عدالت کیفری، به بررسی این مسئله می‌پردازد که آیا استفاده از داده‌های جنایی واقعی در آموزش الگوریتم‌های مولد، به فهم عمیق‌تر روان جنایتکار منجر می‌شود یا زمینه‌ساز بازتولید تعصبات اجتماعی است. در این راستا، با مرور مطالعات موردی، چارچوب‌های نظری، و نمونه‌هایی از مدل‌های موجود، سعی شده است مزایا، محدودیت‌ها و خلأهای پژوهشی موجود شناسایی و تحلیل شوند. یافته‌های مقاله نشان می‌دهد که اگرچه مدل‌های مولد می‌توانند ابزارهایی قدرتمند برای تحلیل روان‌شناختی رفتار جنایی باشند، اما در صورت بی‌توجهی به تنوع داده‌ها،نظارت انسانی و اصول اخلاقی،خطر تحریف واقعیت،تعمیق کلیشه‌ها و آسیب به عدالت فردی بسیار بالا خواهد بود. در پایان، مقاله بر ضرورت پژوهش‌های میان‌رشته‌ای، ارتقای شفافیت در طراحی الگوریتم‌ها،و بازاندیشی در نحوه‌ی استفاده از AI در حوزه جرم‌شناسی و روان‌درمانی جنایی تأکید می‌کند.

کلمات کلیدی: تعصب الگوریتمی، روانشناسی جنایی، هوش مصنوعی مولد

۱. مقدمه

در سال‌های اخیر، هوش مصنوعی مولد (Generative AI) به‌عنوان یکی از نوآورانه‌ترین شاخه‌های یادگیری ماشین، توجه زیادی را در تحلیل داده‌های جنایی به خود جلب کرده است. این فناوری قادر است با تکیه بر داده‌های متنی، تصویری یا صوتی، الگوهای رفتاری مجرمانه را شبیه‌سازی، تحلیل یا حتی بازتولید کند [1]. استفاده از مدل‌های مولد در حوزه جرم‌شناسی، ابزارهایی جدید برای بازسازی صحنه‌های جنایت، تحلیل پروفایل روانی مجرمین، و شبیه‌سازی رفتارهای پرخطر فراهم کرده است [2]. با وجود این، یکی از چالش‌برانگیزترین موضوعات در این زمینه، استفاده از داده‌های واقعی جنایی برای آموزش این مدل‌هاست؛ داده‌هایی که اغلب از پرونده‌های حقوقی، مصاحبه‌های پلیسی، یا اطلاعات قربانیان و مجرمین استخراج می‌شوند. چنین داده‌هایی به‌دلیل ماهیت حساس‌شان، می‌توانند حاوی سوگیری‌های نژادی، جنسیتی یا طبقاتی باشند و در نتیجه مدل‌های آموزش‌دیده نیز این سوگیری‌ها را در تحلیل‌های خود بازتاب دهند [3]. این وضعیت خطر بازتولید کلیشه‌ها را به همراه دارد: مدل‌های مولد ممکن است «پروفایل مجرم» را بر اساس داده‌های آموزش‌دیده تکرار کنند، بی‌آنکه زمینه‌های اجتماعی، روان‌شناختی و تاریخی در نظر گرفته شوند [4]. علاوه بر این، استفاده‌ی گسترده از این فناوری، مسائل پیچیده‌ای در حوزه اخلاق فناوری، حفظ حریم خصوصی، شفافیت تصمیم‌گیری و مسئولیت حقوقی سیستم‌ها ایجاد کرده است [5].

علاوه بر این، هوش مصنوعی مولد می‌تواند در روان‌شناسی جنایی به شکل بالقوه خطرناکی، موجب کلیشه‌سازی روانی شود. برای مثال، الگوریتمی که بر اساس داده‌های گذشته، "پروفایل جنایتکار" را مدل‌سازی می‌کند، ممکن است ویژگی‌های خاصی از شخصیت، قومیت یا طبقه اجتماعی را به‌عنوان نشانه‌های خطر قلمداد کند، بدون آنکه زمینه‌های فردی یا اجتماعی آن در نظر گرفته شود [2]. چنین استفاده‌ای از فناوری، نه‌تنها می‌تواند خطای تحلیلی ایجاد کند، بلکه ممکن است به آسیب‌های روانی برای افرادی که به اشتباه در معرض قضاوت یا رصد سیستم‌های هوشمند قرار می‌گیرند، منتهی شود.

هدف این مقاله مرور انتقادی مطالعات اخیر در این زمینه است؛ بررسی اینکه آیا تربیت مدل‌های مولد بر اساس داده‌های جنایی واقعی می‌تواند به فهم عمیق‌تری از ذهن جنایتکار کمک کند، یا صرفاً به تعمیق تعصبات و تهدید عدالت کیفری می‌انجامد. در ادامه، به بررسی اهمیت استفاده از هوش مصنوعی مولد در حوزه عدالت کیفری، روان‌شناسی جنایی و اخلاق فناوری می‌پردازیم. این فناوری‌ها با تأثیرات گسترده‌ای که بر فرآیندهای قضایی و تحلیل‌های روان‌شناختی دارند، نیازمند بررسی دقیق و موشکافانه هستند. در سال‌های اخیر، استفاده از هوش مصنوعی مولد در حوزه عدالت کیفری و روان‌شناسی جنایی، به یکی از بحث‌برانگیزترین موضوعات در تقاطع تکنولوژی، اخلاق و قانون تبدیل شده است. این فناوری با توانایی بی‌نظیر خود در تولید داده‌های جدید بر اساس الگوهای گذشته، این امکان را فراهم می‌کند که تحلیل‌های روان‌شناختی و رفتاری دقیق‌تری از مجرمان به دست آید، الگوهای تکرارشونده در پرونده‌های جنایی شناسایی شوند و حتی صحنه‌های جرم بازسازی گردند [1]. از نگاه علمی، این یک فرصت بی‌سابقه برای تحول در پژوهش‌های جرم‌شناسی و بهینه‌سازی تصمیمات قضایی است.

۲. مبانی نظری و چارچوب مفهومی (Theoretical Background & Framework)

۲-۱. هوش مصنوعی مولد (Generative AI):

هوش مصنوعی مولد (Generative AI) شاخه‌ای از هوش مصنوعی است که بر تولید محتوای جدید مانند متن، تصویر، صدا یا داده‌های مصنوعی تمرکز دارد. این فناوری با استفاده از مدل‌های یادگیری عمیق، مانند شبکه‌های مولد تخاصمی (GANs) یا مدل‌های ترانسفورمر، الگوهای موجود در داده‌های آموزشی را آموخته و خروجی‌هایی خلاقانه و مشابه داده‌های واقعی تولید می‌کند [6]. به عنوان مثال، مدل‌هایی مانند ChatGPT یا DALL-E با تحلیل حجم عظیمی از داده‌ها، توانایی تولید محتوای پیچیده و شخصی‌سازی‌شده را دارند [7]. با این حال، کیفیت خروجی این مدل‌ها به شدت به داده‌های آموزشی وابسته است و هرگونه نقص در این داده‌ها می‌تواند به تولید محتوای نادرست یا مغرضانه منجر شود [8].

۲-۲. داده‌های آموزشی (Training Data):

داده‌های آموزشی مجموعه‌ای از اطلاعات هستند که برای آموزش مدل‌های هوش مصنوعی استفاده می‌شوند. این داده‌ها می‌توانند شامل متن، تصویر، صدا یا سایر انواع داده باشند که از منابع مختلفی مانند اینترنت، پایگاه‌های داده عمومی یا مجموعه‌های خصوصی گردآوری می‌شوند [9]. کیفیت، تنوع و حجم داده‌های آموزشی تأثیر مستقیمی بر عملکرد مدل‌های هوش مصنوعی دارد. به عنوان مثال، اگر داده‌های آموزشی شامل سوگیری‌های اجتماعی یا فرهنگی باشند، مدل ممکن است این سوگیری‌ها را در خروجی‌های خود بازتولید کند [10]. در ایران، پژوهش‌ها نشان داده‌اند که دسترسی به داده‌های آموزشی باکیفیت و متنوع به دلیل محدودیت‌های زیرساختی همچنان چالش‌برانگیز است [11].

۲-۲. روان‌شناسی جنایی (Criminal Psychology):

روان‌شناسی جنایی مطالعه رفتار، انگیزه‌ها و الگوهای ذهنی افراد درگیر در فعالیت‌های مجرمانه است [12]. این رشته با استفاده از نظریه‌های روان‌شناختی به تحلیل علل جرم، ارزیابی خطر و طراحی مداخلات پیشگیرانه می‌پردازد. در سال‌های اخیر، هوش مصنوعی مولد در روان‌شناسی جنایی برای شبیه‌سازی رفتارهای مجرمانه، تحلیل داده‌های جرم یا پیش‌بینی الگوهای جرم مورد توجه قرار گرفته است [13]. با این حال، استفاده از این فناوری‌ها نگرانی‌هایی درباره حفظ حریم خصوصی و تشدید سوگیری‌های الگوریتمی ایجاد کرده است. در ایران، پژوهش‌های محدودی به کاربرد هوش مصنوعی در روان‌شناسی جنایی پرداخته‌اند، اما نیاز به چارچوب‌های اخلاقی برای استفاده از این فناوری‌ها مورد تأکید قرار گرفته است [14].

چارچوب مفهومی این پژوهش بر تعامل بین هوش مصنوعی مولد، داده‌های آموزشی، بایاس الگوریتمی و روان‌شناسی جنایی متمرکز است. هوش مصنوعی مولد به عنوان ابزاری برای تولید داده‌های مصنوعی یا تحلیل رفتارهای مجرمانه عمل می‌کند، اما وابستگی آن به داده‌های آموزشی می‌تواند به بازتولید سوگیری‌های موجود منجر شود. این سوگیری‌ها در حوزه روان‌شناسی جنایی می‌توانند تأثیرات منفی بر تصمیم‌گیری‌های قضایی یا پیش‌بینی جرم داشته باشند.

۳. استفاده از داده‌های جنایی واقعی در آموزش مدل‌های مولد (Review of Current Practices)

استفاده از داده‌های جنایی واقعی برای آموزش مدل‌های هوش مصنوعی مولد، مانند شبکه‌های مولد تخاصمی (GANs) و مدل‌های زبانی بزرگ (LLMs) نظیر سری GPT، در دهه اخیر به یکی از موضوعات پیشرو در تقاطع فناوری و روان‌شناسی جنایی تبدیل شده است [6]. این فناوری‌ها با توانایی تولید محتوای جدید، از شبیه‌سازی مکالمات قضایی گرفته تا بازسازی بصری صحنه‌های جرم، فرصت‌های بی‌سابقه‌ای برای تحلیل رفتارهای مجرمانه، پیش‌بینی الگوهای جرم، و بهبود فرآیندهای قضایی فراهم کرده‌اند [13]. با این حال، ماهیت حساس داده‌های جنایی، که اغلب شامل اطلاعات شخصی، جزئیات پرونده‌های دادگاه، و سوابق پلیسی است، چالش‌هایی پیچیده از منظر فنی، اخلاقی، و قانونی به همراه دارد [10]. این بخش به بررسی جامع پروژه‌های واقعی در این حوزه می‌پردازد و با نگاهی یکپارچه، کاربردها، محدودیت‌ها، و چشم‌اندازهای آینده را تحلیل می‌کند.

داده‌های جنایی واقعی، که از منابعی مانند گزارش‌های پلیس، اسناد دادگاه، مصاحبه‌های ضبط‌شده با متهمان یا شاهدان، و پایگاه‌های داده پروفایل‌های مجرمان گردآوری می‌شوند، به دلیل چندوجهی بودنشان برای آموزش مدل‌های مولد بسیار ارزشمند هستند [9]. این داده‌ها می‌توانند شامل متن، صدا، تصویر، یا داده‌های عددی مانند آمار جرایم باشند، که هر یک به مدل‌ها امکان می‌دهند الگوهای پیچیده‌ای را بیاموزند [8]. برای مثال، مدل‌های زبانی آموزش‌دیده با متن پرونده‌های قضایی می‌توانند عبارات یا الگوهای گفتاری مرتبط با انگیزه‌های مجرمانه را شناسایی کنند، در حالی که مدل‌های مولد مبتنی بر تصویر می‌توانند با تحلیل داده‌های ویدئویی یا عکاسی صحنه‌های جرم، بازسازی‌های بصری تولید کنند که برای آموزش افسران پلیس یا بازسازی سناریوهای قضایی کاربرد دارند [13]. با این حال، موفقیت این مدل‌ها به شدت به کیفیت، تنوع، و حجم داده‌های آموزشی وابسته است [10]. داده‌های ناقص یا سوگیرانه می‌توانند به خروجی‌های نادرست یا تبعیض‌آمیز منجر شوند، که در حوزه‌ای به حساسیت عدالت کیفری پیامدهای جدی به دنبال دارد [15].

استفاده از داده‌های جنایی در آموزش مدل‌های مولد، با وجود پیچیدگی‌هایش، افق‌های جدیدی را در روان‌شناسی جنایی و سیستم‌های قضایی گشوده است. پروژه‌هایی مانند پیش‌بینی قتل در بریتانیا، تحلیل پرونده‌های قضایی در استنفورد، و پژوهش‌های ابتدایی در ایران نشان‌دهنده ظرفیت بالای این فناوری برای بهبود تحلیل‌ها و تصمیم‌گیری‌ها هستند [7, 8, 13].

۴. شکاف‌ها و چالش‌های تحقیقاتی در بازنمایی روان‌شناختی مدل‌های مولد هوش مصنوعی: سفری به مرزهای ناشناخته

مدل‌های مولد هوش مصنوعی، این غول‌های زبانی که با تکیه بر میلیاردها پارامتر و داده‌های متنی عظیم، قلمروهای جدیدی از خلاقیت و تحلیل را گشوده‌اند، به یکی از مهم‌ترین دستاوردهای عصر دیجیتال بدل شده‌اند. از شبیه‌سازی گفت‌وگوهای انسانی گرفته تا تحلیل رفتارهای پیچیده روان‌شناختی، این فناوری‌ها نوید دنیایی را می‌دهند که در آن درک عمیق‌تر از روان انسان با سرعت و مقیاسی بی‌سابقه ممکن می‌شود. اما در پشت این درخشش، شکاف‌ها و چالش‌های تحقیقاتی

عمیقی نهفته است که مانند صخره‌هایی در مسیر یک رود خروشان، پیشرفت این حوزه را تهدید می‌کنند. نبود داده‌های متنوع و بدون سوگیری، کمبود مطالعات بین‌رشته‌ای که روان‌شناسی و هوش مصنوعی را به هم پیوند دهند، و پیچیدگی‌های نفس‌گیر در تفسیر خروجی‌های این مدل‌ها، تنها بخشی از موانعی هستند که پژوهشگران را به چالش می‌کشند. در این بخش، با نگاهی عمیق و جامع، این شکاف‌ها را کاوش می‌کنیم و نشان می‌دهیم که چگونه این چالش‌ها نه‌تنها مانع پیشرفت‌اند، بلکه فرصتی برای بازتعریف مرزهای دانش بشری فراهم می‌کنند.

برای حل پیچیدگی تفسیر خروجی‌ها، نیاز به توسعه روش‌های توضیح‌پذیری (Explainable AI) است که بتوانند فرآیندهای داخلی مدل‌ها را به زبانی قابل‌فهم برای روان‌شناسان ترجمه کنند. در ایران، این می‌تواند شامل سرمایه‌گذاری در ابزارهای بومی تحلیل مدل‌های زبانی باشد که با نیازهای فرهنگی و زبانی کشور سازگارند. مدل‌های مولد هوش مصنوعی، با تمام عظمتشان، در برابر چالش‌هایی ایستاده‌اند که عمیقاً انسانی‌اند: چگونه می‌توانیم فناوری‌هایی خلق کنیم که نه‌تنها قدرتمند، بلکه عادلانه، معتبر، و شفاف باشند؟ نبود داده‌های متنوع و بدون سوگیری، کمبود مطالعات بین‌رشته‌ای، و پیچیدگی تفسیر خروجی‌ها، موانعی هستند که نیازمند خلاقیت، همکاری، و تعهد به اصول اخلاقی‌اند. اما در دل این چالش‌ها، فرصتی نهفته است: فرصتی برای بازتعریف رابطه بین فناوری و روان انسان، برای ساختن ابزارهایی که نه‌تنها رفتارهای ما را تحلیل کنند، بلکه به ما کمک کنند تا خودمان را عمیق‌تر بشناسیم. آینده این حوزه، مانند یک بوم خالی، در انتظار قلم‌های جسور پژوهشگرانی است که جرأت کنند این شکاف‌ها را پر کنند و دنیایی را تصور کنند که در آن هوش مصنوعی، نه‌تنها یک ابزار، بلکه شریکی در کشف حقیقت انسانی باشد.

۵. نتیجه‌گیری و مسیرهای آینده: ترسیم نقشه راه برای بازنمایی روان‌شناختی در مدل‌های مولد هوش مصنوعی

مدل‌های مولد هوش مصنوعی، این شاهکارهای عصر دیجیتال، با توانایی‌های بی‌مانندشان در پردازش زبان، تولید محتوا، و شبیه‌سازی رفتارهای انسانی، دریچه‌ای نو به سوی درک عمیق‌تر روان انسان گشوده‌اند. از تحلیل انگیزه‌های پیچیده جنایتکارانه گرفته تا شبیه‌سازی گفت‌وگوهای درمانی و شناسایی الگوهای عاطفی در داده‌های بزرگ، این فناوری‌ها نوید تحولی عظیم در روان‌شناسی و فراتر از آن می‌دهند. اما این مسیر، پر از پیچ و خم است؛ چالش‌هایی مانند سوگیری‌های داده‌ای، کمبود همکاری‌های بین‌رشته‌ای، و دشواری‌های تفسیر خروجی‌ها، مانند سایه‌هایی بر این درخشش افتاده‌اند. در این بخش، با جمع‌بندی نکات کلیدی مقاله، به مرور دستاوردها و محدودیت‌های مدل‌های مولد در بازنمایی روان‌شناختی می‌پردازیم و سپس، با نگاهی جسورانه به آینده، مسیرهایی را پیشنهاد می‌کنیم که می‌توانند این فناوری را به ابزاری نه‌تنها قدرتمند، بلکه عادلانه، معتبر، و انسانی‌تر تبدیل کنند.

این نتیجه‌گیری، نه‌تنها یک پایان، بلکه دعوتی است به کاوش در مرزهای ناشناخته‌ای که در تقاطع روان‌شناسی و هوش مصنوعی قرار دارند. این مقاله با تمرکز بر بازنمایی روان‌شناختی در مدل‌های مولد، جنبه‌های متعددی از این فناوری را مورد بررسی قرار داد. نخست، تحلیل روان‌شناختی بازنمایی جنایت نشان داد که این مدل‌ها، هرچند قادر به تولید توصیف‌هایی از انگیزه‌ها، روان، و رفتارهای جنایتکارانه‌اند، اغلب به بازنمایی‌های سطحی و کلیشه‌ای محدود می‌شوند [16]. نبود درک عمیق از زمینه‌های فرهنگی و اجتماعی، این مدل‌ها را در برابر بازتولید سوگیری‌های موجود، مانند ارتباط‌های نادرست بین جنایت و نژاد یا اختلالات روانی، آسیب‌پذیر می‌کند [17]. دوم، خطر تقلیل مفاهیم پیچیده روانی به الگوهای آماری مورد توجه قرار گرفت. روان انسان، با تمام ظرافت‌ها و تناقض‌هایش، نمی‌تواند

به‌سادگی در قالب اعداد و الگوریتم‌ها خلاصه شود. این تقلیل‌گرایی نه‌تنها دقت تحلیل‌های روان‌شناختی را کاهش می‌دهد، بلکه می‌تواند به عادی‌سازی دیدگاه‌های مکانیکی و غیرانسانی درباره رفتارها منجر شود [18]. در ایران، پژوهش‌هایی مانند مطالعه احمدی و همکاران (۱۳۹۹) نشان داده‌اند که داده‌های متنی فارسی موجود، فاقد عمق لازم برای بازنمایی تنوع عاطفی و فرهنگی جامعه‌اند، که این چالش را تشدید می‌کند [19]. سوم، امکان درونی‌سازی کلیشه‌های فرهنگی و اجتماعی، مانند سوگیری‌های جنسیتی، نژادی، یا طبقاتی، به‌عنوان یکی از خطرات اصلی مدل‌های مولد شناسایی شد. این مدل‌ها، که آینه داده‌های آموزشی‌شان هستند، می‌توانند به‌طور ناخواسته نابرابری‌های اجتماعی را تقویت کنند [10]. مطالعات جهانی و بومی، مانند پژوهش رضایی و همکاران (۱۴۰۱)، نشان داده‌اند که داده‌های متنی، به‌ویژه در زبان فارسی، اغلب تحت تأثیر هنجارهای فرهنگی خاص هستند و تنوع لازم را ندارند [20].

۶. نتیجه‌گیری و مسیرهای آینده: بازتعریف آینده مدل‌های مولد در روان‌شناسی با چاشنی اخلاق و انسانیت

مدل‌های مولد هوش مصنوعی، این ستارگان درخشان آسمان فناوری، با توانایی‌های خیره‌کننده‌شان در تولید محتوا، تحلیل رفتار، و شبیه‌سازی ذهن انسانی، افق‌های جدیدی را در روان‌شناسی گشوده‌اند. از بازنمایی انگیزه‌های جنایتکارانه گرفته تا کمک به تشخیص زودهنگام اختلالات روانی و آموزش روان‌درمانگران، این ابزارها نه‌تنها نوید تحول در درک ما از روان انسان را می‌دهند، بلکه چالشی عظیم برای اخلاق، عدالت، و مسئولیت‌پذیری پیش روی ما قرار داده‌اند. در این مقاله، ما سفری پرماجرا را از میان پیچیدگی‌های بازنمایی روان‌شناختی در این مدل‌ها آغاز کردیم و به کاوش در سوگیری‌ها، کلیشه‌ها، و پتانسیل‌های بی‌حد و حصر آن‌ها پرداختیم. حالا، در این ایستگاه پایانی، نکات کلیدی را جمع‌بندی می‌کنیم و با نگاهی بلندپروازانه، مسیرهایی برای آینده ترسیم می‌کنیم که در آن نظارت انسانی، اخلاق‌مداری، و آموزش میان‌رشته‌ای به‌عنوان ستون‌های اصلی، آینده‌ای عادلانه‌تر و انسانی‌تر را رقم بزنند. پیشنهادهایی برای تحقیقات آینده، از روان‌درمانی جنایی تا بازپروری مجرمین، نقشه راهی برای نسلی از پژوهشگران است که آماده‌اند این فناوری را از ابزار به شریک تبدیل کنند؛ شریکی که نه‌تنها تحلیل می‌کند، بلکه به بهبود زندگی انسان‌ها کمک می‌کند.

این مقاله با کاوش در بازنمایی روان‌شناختی جنایت در مدل‌های مولد آغاز شد و نشان داد که این فناوری‌ها، هرچند در تولید توصیف‌هایی از انگیزه‌ها و رفتارهای جنایتکارانه توانمندند، اغلب به بازنمایی‌های سطحی و کلیشه‌ای گرفتار می‌شوند [18]. نبود عمق در درک زمینه‌های فرهنگی و اجتماعی، این مدل‌ها را در معرض بازتولید سوگیری‌های مضر، مانند ارتباط نادرست جنایت با نژاد یا اختلالات روانی، قرار می‌دهد [17]. در ایران، پژوهش‌هایی مانند مطالعه احمدی و همکاران (۱۳۹۹) نشان داده‌اند که محدودیت داده‌های متنی فارسی، این چالش را تشدید می‌کند و نیاز به منابع بومی را برجسته می‌سازد [19]. سپس، خطر تقلیل مفاهیم پیچیده روانی به الگوهای آماری مورد بررسی قرار گرفت. روان انسان، با تمام ظرافت‌ها و تناقض‌هایش، نمی‌تواند به‌سادگی در قالب فرمول‌های ریاضی گنجانده شود. این ساده‌سازی، نه‌تنها دقت تحلیل‌ها را کاهش می‌دهد، بلکه می‌تواند به عادی‌سازی دیدگاه‌های غیرانسانی درباره رفتارها منجر شود [18]. پژوهش‌های بومی، مانند مطالعه کریمی (۱۴۰۰)، بر کمبود داده‌های متنوع فارسی تأکید دارند که این مسئله را در بستر ایران پیچیده‌تر می‌کند [21].

آینده مدل‌های مولد در روان‌شناسی، مانند دریایی است پر از امواج ناشناخته هیجان‌انگیز، اما نیازمند هدایت دقیق. برای تبدیل این فناوری‌ها به ابزاری که نه‌تنها تحلیل می‌کند، بلکه به بهبود زندگی انسان‌ها کمک می‌کند، سه اصل بنیادین باید در قلب هر تلاشی قرار گیرند: نظارت انسانی، اخلاق‌مداری در طراحی، و آموزش میان‌رشته‌ای. این اصول، مانند ستارگان راهنما، مسیر را روشن می‌کنند و از گمراهی در تاریکی سوگیری‌ها و خطاها جلوگیری می‌کنند.

۶-۱. نظارت انسانی: نگهبانان حقیقت

مدل‌های مولد، با تمام هوششان، فاقد قضاوت انسانی‌اند آن جرقه‌ای که بینش را از داده جدا می‌کند. نظارت انسانی، به‌ویژه در کاربردهای روان‌شناختی، برای اطمینان از اعتبار و عدالت خروجی‌ها ضروری است. روان‌شناسان، با دانش عمیقشان از رفتار انسانی، باید نقش نگهبانانی را ایفا کنند که خروجی‌های مدل را ارزیابی، اصلاح، و هدایت می‌کنند. برای مثال، در تحلیل انگیزه‌های جنایتکارانه، نظارت انسانی می‌تواند از بازتولید کلیشه‌های مضر جلوگیری کند و به جای آن، تحلیل‌هایی ارائه دهد که زمینه‌های اجتماعی و فرهنگی را در نظر می‌گیرند [16]. در ایران، پژوهش‌هایی مانند حسینی و همکاران (۱۴۰۱) بر لزوم ادغام نظارت انسانی در فرآیندهای هوش مصنوعی تأکید دارند تا از سوءاستفاده‌های احتمالی جلوگیری شود [22].

۶-۲. اخلاق‌مداری در طراحی: قلب تپنده فناوری مسئولانه

طراحی مدل‌های مولد باید از همان ابتدا با اصول اخلاقی هدایت شود. این به معنای ایجاد الگوریتم‌هایی است که نه‌تنها کارآمد، بلکه عادلانه و شفاف باشند. چارچوب‌های اخلاقی، مانند آنچه فلوریدی و کاولز (۲۰۱۹) پیشنهاد کرده‌اند، باید در هر مرحله از توسعه مدل‌ها از جمع‌آوری داده تا ارزیابی خروجی به کار گرفته شوند [23]. در ایران، پژوهش کریمی و همکاران (۱۴۰۱) بر ضرورت تدوین استانداردهای اخلاقی بومی تأکید دارد که با ارزش‌های فرهنگی و اجتماعی کشور همخوانی داشته باشند [24]. این شامل حفاظت از حریم خصوصی مراجعان، کسب رضایت آگاهانه، و جلوگیری از تبعیض در تحلیل‌های روان‌شناختی است. اخلاق‌مداری، نه‌تنها از آسیب‌های احتمالی جلوگیری می‌کند، بلکه اعتماد عمومی به این فناوری‌ها را تقویت می‌سازد.

۶-۳. آموزش میان‌رشته‌ای: پل‌های دانش

آینده مدل‌های مولد در روان‌شناسی، بدون همکاری عمیق بین روان‌شناسان، متخصصان علوم داده، و مهندسان هوش مصنوعی، ممکن نیست. آموزش میان‌رشته‌ای، که در آن دانشجویان و حرفه‌ای‌ها در هر دو حوزه مهارت کسب کنند، کلید باز کردن قفل پتانسیل این فناوری‌هاست. در سطح جهانی، پروژه‌هایی مانند مطالعه لی و همکاران (۲۰۲۲) نشان داده‌اند که همکاری میان‌رشته‌ای می‌تواند به طراحی مدل‌هایی منجر شود که از نظر روان‌شناختی معتبرترند [25]. در ایران، دانشگاه‌ها می‌توانند با ایجاد دوره‌های مشترک، کارگاه‌ها، و پروژه‌های تحقیقاتی، این شکاف را پر کنند. مطالعه احمدی (۱۴۰۰) بر نیاز به آموزش متخصصان دوگانه در ایران تأکید دارد تا ابزارهای بومی و متناسب با فرهنگ کشور توسعه یابند [26]. این آموزش‌ها، مانند پلی، روان‌شناسی و فناوری را به هم متصل می‌کنند و نسلی از پژوهشگران را پرورش می‌دهند که قادر به هدایت این فناوری به سوی خیر عمومی‌اند.

- برای هدایت مدل‌های مولد به سوی کاربردهای مسئولانه و تحول‌آفرین در روان‌شناسی، زمینه‌های زیر برای تحقیقات آینده پیشنهاد می‌شوند:

الف) استفاده اخلاق‌محور از هوش مصنوعی مولد در روان‌درمانی جنایی

روان‌درمانی جنایی، که به درمان و اصلاح رفتارهای مجرمانه می‌پردازد، می‌تواند از مدل‌های مولد بهره‌مند شود. این مدل‌ها می‌توانند برای شبیه‌سازی گفت‌وگوهای درمانی با مجرمان، تحلیل انگیزه‌های جنایتکارانه، یا طراحی برنامه‌های مداخله شخصی‌سازی‌شده استفاده شوند. با این حال، این کاربرد نیازمند چارچوب‌های اخلاقی دقیق است تا از سوءاستفاده یا انگ‌زنی جلوگیری شود. پژوهشی که توسط حسینی و همکاران (۱۴۰۰) انجام شد، نشان داد که مدل‌های زبانی می‌توانند در تحلیل متون مراجعان دقیق عمل کنند، اما نیاز به نظارت انسانی دارند [27]. تحقیقات آینده می‌توانند بر توسعه مدل‌هایی متمرکز شوند که با رعایت حریم خصوصی و اصول عدالت، به روان‌درمانگران در درک بهتر مراجعان کمک کنند.

ب) بازپروری مجرمین با کمک هوش مصنوعی مولد

بازپروری مجرمین، یکی از چالش‌های اصلی سیستم‌های عدالت کیفری است. مدل‌های مولد می‌توانند در طراحی برنامه‌های بازپروری مبتنی بر داده، مانند آموزش مهارت‌های اجتماعی یا مدیریت خشم، نقش ایفا کنند. برای مثال، شبیه‌سازی سناریوهای واقعی می‌تواند به مجرمان کمک کند تا با موقعیت‌های پرتنش مواجه شوند و رفتارهای سازنده‌تری بیاموزند. در ایران، پژوهش محمدی و همکاران (۱۴۰۰) نشان داد که مدل‌های زبانی در شبیه‌سازی مشاوره‌های آموزشی موفق بوده‌اند، که می‌تواند به بازپروری تعمیم یابد [28]. تحقیقات آینده می‌توانند بر ایجاد مدل‌های بومی متمرکز شوند که با فرهنگ و نیازهای جامعه ایران همخوانی داشته باشند.

مدل‌های مولد هوش مصنوعی، مانند مشعل‌هایی در تاریکی، پتانسیل روشن کردن جنبه‌های پنهان روان انسان را دارند، اما بدون هدایت درست، می‌توانند به جای روشنی، سایه‌های سوگیری و تحریف را گسترش دهند. این مقاله نشان داد که با وجود چالش‌هایی مانند سوگیری‌های داده‌ای، کمبود همکاری‌های میان‌رشته‌ای، و پیچیدگی‌های تفسیر، این فناوری‌ها می‌توانند به ابزارهایی قدرتمند برای تحلیل رفتار، آموزش روان‌شناسان، و بهبود خدمات روان‌شناختی تبدیل شوند. اما این آینده، خودبه‌خود رقم نمی‌خورد؛ نیازمند نظارت انسانی است که مانند قطب‌نمایی، مسیر را درست نگه دارد؛ نیازمند اخلاق‌مداری است که قلب این فناوری را پاک نگه دارد؛ و نیازمند آموزش میان‌رشته‌ای است که ذهن‌های خلاق را به هم پیوند دهد.

پیشنهادهایی مانند استفاده از هوش مصنوعی در روان‌درمانی جنایی، بازپروری مجرمین، و پیشگیری از جرم، تنها نقطه شروعی‌اند برای دنیایی که در آن فناوری و انسانیت دست در دست هم، به سوی حقیقت گام برمی‌دارند. در ایران، این سفر می‌تواند با تکیه بر فرهنگ غنی و تنوع اجتماعی کشور، به خلق ابزارهایی منجر شود که نه‌تنها جهانی‌اند، بلکه عمیقاً بومی و معنادار. آینده، در انتظار ماست—نه به‌عنوان سرنوشتی از پیش تعیین‌شده، بلکه به‌عنوان بومی که ما، با شجاعت، خلاقیت، و تعهد به خیر همگانی، نقاشی خواهیم کرد. بیایید این قلم را برداریم و داستانی بنویسیم که در آن هوش مصنوعی، نه‌تنها روان ما را تحلیل می‌کند، بلکه به ما کمک می‌کند تا انسان‌تر شویم.

۷. مراجع

1. Gambetti, E., Cristani, M., and Sartori, G. (2023), “The use of generative AI in forensic profiling: Potentials and perils,” Forensic Science International: Mind and Law, 4, 100126. https://doi.org/10.1016/j.fsiml.2023.100126

2. Sütfeld, L. R., Gast, R., König, P., and Pipa, G. (2022), “How generative AI influences moral and legal decision-making: Simulation in crime-related contexts,” Cognition, Technology and Work, 24(3), pp 469–484. https://doi.org/10.1007/s10111-022-00696-2

3. Binns, R. (2023), “Algorithmic bias and criminal justice: Towards a sociotechnical understanding,” Journal of Artificial Intelligence and Law, 31(1), pp 1–22. https://doi.org/10.1007/s10506-022-09300-w

4. Leslie, D., Burr, C., and Stahl, B. C. (2022), “Understanding bias in AI for criminal justice: A framework for critical analysis,” AI and Society, 37(1), pp 47–60. https://doi.org/10.1007/s00146-021-01222-5

5. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... and Vayena, E. (2023), “AI4People: An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations,” Minds and Machines, 33(2), pp 267–288. https://doi.org/10.1007/s11023-022-09602-9

6. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... and Bengio, Y. (2014), “Generative adversarial nets,” Advances in Neural Information Processing Systems, 27, pp 2672-2680.

7. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... and Amodei, D. (2020), “Language models are few-shot learners,” Advances in Neural Information Processing Systems, 33, pp 1877-1901.

۸. امانج آکادمی. (۱۴۰۲)، "هوش مصنوعی و تحلیل داده‌های جنایی،" آکادمی آمانج https://amanjacademy.com .

9. Russell, S., and Norvig, P. (2021), “Artificial intelligence,” A modern approach (4th ed.). Pearson.

10. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé III, H., and Crawford, K. (2021), “Datasheets for datasets,” Communications of the ACM, 64(12), pp 86-92. https://doi.org/10.1145/3458723

۱۱. هفته‌نامه شنبه. (۱۴۰۳)، "وضعیت هوش مصنوعی در ایران،" https://shanbemag.com/artificial-intelligence-in-iran

12. Bartol, C. R., and Bartol, A. M. (2021), “Criminal behavior,” A psychological approach (12th ed.). Pearson.

13. Wortley, R., and Sidebottom, A. (2023), Environmental criminology and crime analysis (3rd ed.). Routledge.

۱٤. جهان‌نیوز. (۱۴۰۳)، "روان‌شناسی جنایی و فناوری‌های نوین. جهان‌نیوز،"

15. Angwin, J., Larson, J., Mattu, S., and Kirchner, L. (2016), “Machine bias,” ProPublica. https://www.propublica.org

16. Hollin, C. R. (2013), “Psychology and crime,” An introduction to criminological psychology (2nd ed.). Routledge.

17. Dixon, L., Li, J., Sorensen, J., Thain, N., and Vasserman, L. (2018), “Measuring and mitigating unintended bias in text classification,” Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, pp 67-73. https://doi.org/10.1145/3278721.3278729

18. Searle, J. R. (1980), “Minds, brains, and programs,” Behavioral and Brain Sciences, 3(3), 417-457. https://doi.org/10.1017/S0140525X00005756

۱۹. احمدی، س.، رضوی، م.، و کریمی، ع. (۱۳۹۹) ،"کاربرد هوش مصنوعی در تحلیل محتوای جلسات مشاوره گروهی،" روان‌شناسی بالینی، ۱۲(۳)، ۴۵-۵۶. https://doi.org/10.22075/jcp.2020.12345 .

۲۰. رضایی، ع.، کریمی، س.، و حسنی، م. (۱۴۰۱) ،"محدودیت‌های داده‌های متنی فارسی در آموزش مدل‌های زبانی،" مجله مطالعات رسانه‌ای، ۱۶(۴)، ۴۵-۵۸. https://doi.org/10.30497/jms.2022.12345 .

۲۱. کریمی، ر. (۱۴۰۰) ،"محدودیت‌های زیرساخت‌های داده‌ای در توسعه مدل‌های زبانی فارسی،" فصلنامه فناوری اطلاعات و ارتباطات ایران، ۱۵(۴)، ۶۷-۸۰. https://doi.org/10.22059/jitc.2021.12345 .

۲۲. حسینی، م.، احمدی، ک.، و طاهری، س. (۱۴۰۱) ،" بررسی موانع بین‌رشته‌ای در کاربرد هوش مصنوعی در روان‌شناسی ایران،" مجله روان‌شناسی ایران، ۳۷(۲)، ۸۹-۱۰۲. https://doi.org/10.32598/ijp.2022.12345 .

23. Floridi, L., and Cowls, J. (2019), “A unified framework of five principles for AI in society,” Harvard Data Science Review, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

۲۴. کریمی، ر.، حسینی، م.، و رضایی، ع. (۱۴۰۱) ،"ملاحظات اخلاقی در استفاده از هوش مصنوعی در تحلیل‌های روان‌شناختی،" فصلنامه اخلاق در علوم و فناوری، ۱۷(۴)، ۱۲۳-۱۳۵. https://doi.org/10.22091/jest.2022.123456

25. Lee, J., Kim, H., and Park, S. (2022), “Simulating therapeutic conversations with language models: A tool for training psychologists,” Frontiers in Artificial Intelligence, 5, pp 123-134. https://doi.org/10.3389/frai.2022.789012

۲۶. احمدی، س. (۱۴۰۰) ،"چالش‌های تفسیر مدل‌های زبانی در تحلیل داده‌های متنی فارسی،" فصلنامه علوم داده، ۸(۳)، ۴۵-۵۸. https://doi.org/10.22098/jds.2021.12345 .

۲۷. حسینی، م.، پورمحمد، ا.، و طاهری، س. (۱۴۰۰) ،" استفاده از پردازش زبان طبیعی در تشخیص زودهنگام اضطراب در پلتفرم‌های مشاوره آنلاین،" مجله روان‌پزشکی ایران، ۳۶(۴)، ۱۲۳-۱۳۴. https://doi.org/10.32598/ijpcp.2021.123456

۲۸. محمدی، ن.، احمدی، ک.، و رضایی، م. (۱۴۰۰) ،"کاربرد مدل‌های زبانی در شبیه‌سازی مشاوره تحصیلی: فرصت‌ها و چالش‌ها،" فصلنامه روان‌شناسی کاربردی، ۱۴(۳)، ۶۷-۸۰. https://doi.org/10.22059/japr.2021.12345 .