مرور سیستماتیک الگوریتم ها و تکنیک های یادگیری ماشینی با هدف توسعه هوش مصنوعی

شهرزاد شادروز

دانشجوی کارشناسی علوم کامپیوتر ، واحد علوم تحقیقات ،دانشگاه ازاد اسلامی ، تهران ، ایران

محسن رستمی مال خلیفه

گروه علوم کامپیوتر ، واحد علوم تحقیقات ،دانشگاه ازاد اسلامی ، تهران ، ایران

چکیده

در دهه گذشته، هوش مصنوعی با رشد انفجاری خود، طیف وسیعی از صنایع را متحول کرده است. یادگیری ماشینی به عنوان نیروی محرک بسیاری از پیشرفت‌های هوش مصنوعی، با دسته‌بندی‌های مختلفی از جمله یادگیری نظارت‌شده، یادگیری بدون نظارت، و یادگیری تقویتی توانسته است به موفقیت های چشمگیری برسد. این پژوهش به تمرکز بر چالش‌هایی همچون تعمیم‌پذیری، استحکام، و ملاحظات اخلاقی در سیستم‌های هوش مصنوعی می‌پردازد. این چالش‌ها نشان‌دهنده نیاز به مرور سیستماتیک الگوریتم‌ها و تکنیک‌های یادگیری ماشینی است تا با شناسایی محدودیت‌ها و ارائه نوآوری‌های جدید، بتوان به بهبود عملکرد، سازگاری، و تفسیرپذیری سیستم‌های هوش مصنوعی پرداخت. روش‌شناسی این مطالعه بصورت یک مرور سیستماتیک از مقالات مرتبط در این زمینه می‌باشد. پژوهش بر اساس یک پروتکل مرور، شامل جستجوی گسترده در پایگاه‌های داده‌های علمی معتبر، معیارهای ورود و خروج، استخراج داده‌ها، و تحلیل کیفی مطالعات انجام شده است. برای ارزیابی سیستماتیک الگوریتم‌ها و تکنیک‌های یادگیری ماشینی از چارچوب PICOC استفاده شد. یافته‌ها نشان می‌دهند تکنیک‌های نوظهور (یادگیری عمیق، یادگیری انتقالی، و یادگیری تقویتی) نقش مهمی در بهبود عملکرد سیستم‌های هوش مصنوعی داشته‌اند. با این حال، چالش‌های موجود از جمله نیاز به قابلیت تفسیر، مقیاس‌پذیری، عدالت، و استحکام در برابر داده‌های نویزی، نشان‌دهنده نیاز به تحقیقات مستمر برای پر کردن شکاف بین قابلیت‌های فعلی هوش مصنوعی و کاربرد مؤثر و مسئولانه آن است. در نتیجه دستیابی به سیستم‌های هوش مصنوعی سازگارتر و قابل‌اعتمادتر نیازمند توسعه مدل‌های ذاتاً قابل‌تفسیر و چارچوب‌های اخلاقی جامع است. پژوهش‌های آینده باید بر بهبود تفسیرپذیری، استحکام در برابر داده‌های نویزی، و تضمین عدالت در مدل‌های هوش مصنوعی تمرکز کنند. ادغام استانداردهای اخلاقی و مقرراتی می‌تواند به ایجاد سیستم‌های هوش مصنوعی مسئولانه و قابل‌اعتماد کمک کند که قادر به حل مسائل پیچیده دنیای واقعی باشند.

کلمات کلیدی: یادگیری ماشینی , توسعه هوش مصنوعی , هوش مصنوعی, تکنیک ها و الگوریتم های ML

مقدمه

رشته‌ی هوش مصنوعی (AI) طی دهه‌ی گذشته رشد انفجاری داشته و به‌طور قابل توجهی صنایعی مانند بهداشت و درمان، مالی، خودروهای خودران و رباتیک را دگرگون کرده است. هوش مصنوعی به قابلیت ماشین‌ها برای انجام وظایفی اشاره دارد که معمولاً نیازمند هوش انسانی هستند، از جمله یادگیری، استدلال و درک (Russell & Norvig, 2020). توانایی هوش مصنوعی در تغییر فرآیندها و صنایع، آن را به یکی از تحول‌آفرین‌ترین فناوری‌های قرن ۲۱ تبدیل کرده است. با این حال، با وجود این پیشرفت سریع، سیستم‌های هوش مصنوعی امروزی همچنان با چالش‌های متعددی مواجه هستند، به‌ویژه در دستیابی به هوش عمومی و گسترش مدل‌های یادگیری ماشینی برای کاربردهای گسترده‌تر و انعطاف‌پذیرتر (Lake et al, 2017؛ Hassabis et al, 2020). پیشرفت‌های سریع در هوش مصنوعی (AI) صنایع مختلف از جمله بهداشت و درمان، امور مالی و تولید را دگرگون کرده است، اما چالش‌های مهمی همچنان در دستیابی به سیستم‌های هوش مصنوعی مقاوم و قابل تعمیم باقی مانده‌اند. در حالی که فناوری‌های فعلی هوش مصنوعی در حوزه‌های محدود و وظیفه‌محور عملکرد خوبی دارند، در انجام وظایفی که نیاز به استدلال انتزاعی و انعطاف‌پذیر دارند یا در محیط‌هایی که اطلاعات ناقص است، با مشکل مواجه می‌شوند (Zhang et al, 2021). این محدودیت مانع از دستیابی سیستم‌های هوش مصنوعی به هوش عمومی می‌شود و کاربردهای آن‌ها را در سناریوهای واقعی که نیاز به سازگاری دارند، کاهش می‌دهد.

یادگیری ماشینی(ML) ، که نیروی محرکه‌ی اکثر پیشرفت‌های هوش مصنوعی است، به ماشین‌ها امکان می‌دهد عملکرد خود را از طریق داده‌ها بهبود دهند بدون این‌که به‌طور صریح برنامه‌ریزی شوند. تکنیک‌های یادگیری ماشینی به‌طور کلی به سه دسته‌ی یادگیری نظارت‌شده، یادگیری بدون نظارت و یادگیری تقویتی تقسیم می‌شوند. یادگیری نظارت‌شده، که شامل آموزش الگوریتم‌ها با استفاده از مجموعه داده‌های برچسب‌گذاری‌شده است، در حوزه‌هایی مانند تشخیص گفتار، طبقه‌بندی تصاویر و تشخیص پزشکی بسیار موفق بوده است (LeCun et al, 2019) یادگیری بدون نظارت که به کشف الگوهای پنهان در داده‌های بدون برچسب می‌پردازد، در خوشه‌بندی، تشخیص ناهنجاری‌ها و فشرده‌سازی داده‌ها نقش اساسی دارد . یادگیری تقویتی، که به عوامل آموزش می‌دهد تا با محیط خود تعامل داشته و تصمیم‌گیری کنند، در سیستم‌های رباتیک و بازی‌های رایانه‌ای نتایج قابل توجهی به دست آورده است (Silver et al, 2021).

ظهور یادگیری عمیق، که زیرمجموعه‌ای از یادگیری ماشینی بوده و از شبکه‌های عصبی مصنوعی با چندین لایه بهره می‌برد، به‌ویژه تاثیرگذار بوده است. یادگیری عمیق در حوزه‌هایی مانند بینایی رایانه‌ای، پردازش زبان طبیعی و رانندگی خودکار پیشرفت‌های شگرفی داشته است (Goodfellow et al ,2020؛ Brown et al, 2020). یکی از مهم‌ترین دستاوردهای سال‌های اخیر مدل زبانی بزرگ GPT-3 متعلق به OpenAI بوده که قادر است متنی شبیه به انسان تولید کرده و به سوالات پیچیده پاسخ دهد (Brown et al, 2020). با این حال، یادگیری عمیق با چالش‌های مهمی روبروست، از جمله وابستگی به مجموعه داده‌های بزرگ، هزینه‌های بالای محاسباتی و عدم قابلیت تفسیر (Rudin, 2019). بسیاری از مدل‌های یادگیری عمیق به‌عنوان "جعبه سیاه" شناخته می‌شوند و درک نحوه‌ی تصمیم‌گیری آن‌ها برای پژوهشگران دشوار است که

این امر به‌ویژه در حوزه‌های حساس مانند بهداشت و درمان و عدالت کیفری نگران‌کننده است (Doshi-Velez & Kim, 2018).

علاوه بر مشکلات در تفسیر، مسئله‌ی تعمیم‌پذیری و استحکام نیز همچنان چالش‌های کلیدی برای مدل‌های یادگیری ماشینی محسوب می‌شوند. بسیاری از سیستم‌های هوش مصنوعی در وظایف خاصی که برای آن‌ها آموزش دیده‌اند عملکرد خوبی دارند اما هنگام مواجهه با داده‌ها یا محیط‌های جدید و ناشناخته با مشکل مواجه می‌شوند. این مشکل، به‌ویژه در کاربردهایی مانند رانندگی خودکار، که سیستم‌ها باید بتوانند به شرایط مختلف جاده‌ای و سناریوها سازگار شوند، مشکل‌ساز است (Amodei et al, 2018). برای رفع این چالش‌ها، پژوهشگران در حال بررسی تکنیک‌هایی مانند یادگیری انتقالی هستند که در آن مدل‌هایی که برای یک وظیفه آموزش دیده‌اند می‌توانند برای وظیفه‌ای دیگر سازگار شوند، و یادگیری فرامتنی، که در آن سیستم‌ها به‌گونه‌ای طراحی می‌شوند که یاد بگیرند چگونه یاد بگیرند (Finn et al, 2018 ؛ Chen et al, 2020). این روش‌ها امید بهبود انعطاف‌پذیری و سازگاری سیستم‌های هوش مصنوعی را نشان می‌دهند و ممکن است میزان داده‌های لازم برای آموزش را کاهش دهند.

پیامدهای اخلاقی هوش مصنوعی نیز نگرانی‌های عمده‌ای به همراه دارند، به‌ویژه در ارتباط با انصاف، سوگیری و شفافیت. مطالعات نشان داده‌اند که مدل‌های یادگیری ماشینی که بر روی مجموعه داده‌های سوگیری‌شده آموزش دیده‌اند، می‌توانند نابرابری‌های اجتماعی را تشدید یا حتی بازتولید کنند. به‌عنوان مثال، سیستم‌های تشخیص چهره نشان داده‌اند که برای افرادی با پوست تیره‌تر خطاهای بیشتری ایجاد می‌کنند (Buolamwini & Gebru, 2018)، که این سوالات را درباره‌ی انصاف و مسئولیت‌پذیری سیستم‌های هوش مصنوعی مطرح می‌کند. اطمینان از این‌که مدل‌های هوش مصنوعی شفاف، قابل تفسیر و عادلانه هستند نه تنها یک چالش فنی بلکه یک مسئله‌ی اجتماعی است که نیازمند تلاش‌های میان‌رشته‌ای از سوی دانشمندان رایانه، اخلاق‌شناسان و سیاست‌گذاران است (Mitchell et al, 2020). یکی از مسائل اصلی که مانع پیشرفت هوش مصنوعی می‌شود، عملکرد و مقیاس‌پذیری الگوریتم‌های یادگیری ماشینی است که پایه‌ی اکثر برنامه‌های هوش مصنوعی را تشکیل می‌دهند. بسیاری از تکنیک‌های یادگیری ماشینی، از جمله یادگیری عمیق و یادگیری تقویتی، به مقادیر زیادی داده و منابع محاسباتی نیاز دارند تا به عملکرد بالایی دست یابند. با وجود موفقیت‌های این روش‌ها، آن‌ها اغلب با مشکلاتی مانند بیش‌برازش، عدم تفسیرپذیری و کمبود استحکام مواجه هستند، به‌ویژه زمانی که با داده‌های نویزی یا ساختاری‌نیافته روبرو می‌شوند (Doshi-Velez & Kim, 2018). با گسترش کاربردهای هوش مصنوعی، نیاز فوری به توسعه الگوریتم‌های کارآمدتر، تفسیرپذیرتر و قابل تعمیم‌تر وجود دارد که بتوانند در محیط‌های پیچیده و پویا کار کنند (Liu et al, 2020).

علاوه بر این، وابستگی روزافزون به رویکردهای مبتنی بر داده، نگرانی‌های اخلاقی در مورد سوگیری، انصاف و شفافیت در سیستم‌های هوش مصنوعی را افزایش می‌دهد. مطالعات نشان داده‌اند که مدل‌های یادگیری ماشینی، به‌ویژه آن‌هایی که بر روی مجموعه داده‌های سوگیری‌شده یا ناقص آموزش دیده‌اند، می‌توانند نابرابری‌های اجتماعی را تقویت کرده یا حتی تشدید کنند(Mehrabi et al, 2021). پیامدهای اجتماعی و اخلاقی این سوگیری‌ها، توسعه سیستم‌های هوش مصنوعی مقاوم را نه تنها به یک چالش فنی، بلکه به یک ضرورت اخلاقی تبدیل می‌کند. پرداختن به این مسائل نیازمند ارزیابی انتقادی الگوریتم‌ها و تکنیک‌های فعلی است، با هدف توسعه سیستم‌های هوش مصنوعی که نه تنها دقیق و کارآمد باشند بلکه منصفانه و شفاف

نیز باشند. با بررسی تحولات اخیر در یادگیری عمیق، یادگیری انتقالی و محاسبات نورومورفیک، چنین مطالعاتی به گفتمان جاری درباره چگونگی پیشرفت هوش مصنوعی برای پاسخگویی به نیازهای یک منظره تکنولوژیکی در حال تغییر کمک خواهد کرد(Chen et al ,2020).

علاوه بر این، مسئله‌ی حفظ حریم خصوصی داده‌ها به موضوعی حیاتی تبدیل شده است، زیرا سیستم‌های هوش مصنوعی به شدت به مجموعه داده‌های بزرگ متکی هستند. جمع‌آوری و استفاده از داده‌های شخصی در کاربردهای هوش مصنوعی نگرانی‌هایی را در مورد رضایت کاربر، مالکیت داده‌ها و استفاده نادرست از اطلاعات حساس برانگیخته است. معرفی مقرراتی مانند مقررات عمومی حفاظت از داده‌های اتحادیه اروپا (GDPR) اهمیت توسعه سیستم‌های هوش مصنوعی که نه تنها کارآمد بلکه به حریم خصوصی و حقوق فردی نیز احترام بگذارند را برجسته می‌کند (Voigt & von dem Bussche, 2017).

با توجه به این چالش‌ها، نیاز به یک مرور سیستماتیک از الگوریتم‌ها و تکنیک‌های کنونی یادگیری ماشینی آشکار است. در حالی که دستاوردهای بسیاری حاصل شده است، موانع قابل توجهی برای دستیابی به سیستم‌های هوش مصنوعی عمومی‌تر، قابل تفسیر و عادلانه همچنان باقی مانده است. این مقاله به‌دنبال ارائه‌ی یک مرور جامع از چشم‌انداز روش‌های یادگیری ماشینی، تمرکز بر محدودیت‌های آن‌ها، نوآوری‌های اخیر و مسیرهای احتمالی آینده است. بخش دوم، دسته‌های اصلی الگوریتم‌های یادگیری ماشینی را بررسی می‌کند؛ بخش سوم به بررسی تحولات نوین مانند یادگیری چندتایی و جستجوی معماری‌های عصبی می‌پردازد؛ و بخش چهارم به چالش‌های اخلاقی و اجتماعی مرتبط با هوش مصنوعی پرداخته و به دنبال آن، روندهای آینده و فرصت‌های پژوهشی مورد بحث قرار می‌گیرند .بنابراین این پژوهش با هدف پرداختن به این چالش‌ها، یک مرور سیستماتیک از وضعیت فعلی الگوریتم‌ها و تکنیک‌های یادگیری ماشینی ارائه می‌دهد. بنابراین هدف اصلی این مقاله مرور سیستماتیک الگوریتم‌ها و تکنیک‌های فعلی در حوزه یادگیری ماشینی است، با هدف پیشرفت در توسعه هوش مصنوعی (AI)، این مقاله به دنبال شناسایی محدودیت‌ها و چالش‌های موجود در روش‌های کنونی، مانند مسائل مربوط به مقیاس‌پذیری، تفسیرپذیری، و استحکام است. مقاله همچنین به بررسی نوآوری‌های اخیر در حوزه‌هایی مانند یادگیری عمیق، یادگیری انتقالی و محاسبات نورومورفیک می‌پردازد تا عملکرد سیستم‌های هوش مصنوعی را بهبود بخشد. در نهایت، این مطالعه قصد دارد بینش‌هایی را ارائه دهد که نشان دهد چگونه تکنیک‌های یادگیری ماشینی می‌توانند بهبود یابند تا سیستم‌های هوش مصنوعی سازگارتر، قابل تفسیرتر و اخلاقی‌تر شوند و قادر به حل مسائل پیچیده در دنیای واقعی باشند.

روش تحقیق

در ادامه به بررسی روش تحقیق و پروتکل مرور (Review protocol) این مقاله پرداخته می‌شود.

روش‌شناسی و پروتکل مرور

روش‌شناسی این مرور سیستماتیک بر اساس پروتکل‌های تعریف‌شده‌ای است که برای اطمینان از تجزیه و تحلیل جامع و بی‌طرفانه الگوریتم‌ها و تکنیک‌های موجود در حوزه یادگیری ماشینی (ML) طراحی شده است. این روش‌شناسی به چند مرحله کلیدی تقسیم می‌شود :جستجوی متون علمی، معیارهای انتخاب، استخراج داده‌ها و تحلیل.

۱ .جستجوی متون علمی

برای انجام یک مرور دقیق، یک جستجوی گسترده در پایگاه‌های داده‌های علمی مانند مگیران, نورمگز , علم نت, SID و Civilica و IEEE Xplore، ACM Digital Library، Google Scholarو Science direct و Scopus انجام خواهد شد. کلیدواژه‌هایی مانند "الگوریتم‌های یادگیری ماشینی"، "یادگیری عمیق"، "یادگیری تقویتی"، "یادگیری انتقالی" و "پیشرفت‌های هوش مصنوعی" برای بازیابی مقالات مرتبط که در مجلات علمی معتبر و کنفرانس‌ها منتشر شده‌اند، استفاده می‌شوند. این جستجو بر روی مقالات منتشرشده در ده سال اخیر متمرکز خواهد بود تا تحولات جدید پوشش داده شود.

۲ .معیارهای انتخاب

برای حفظ ارتباط و کیفیت مطالعات انتخابی، معیارهای ورود و خروج مشخصی تعریف شده است:

- معیارهای ورود:
  - مطالعاتی که الگوریتم‌های جدید یادگیری ماشینی یا تغییرات قابل توجهی در الگوریتم‌های موجود ارائه می‌دهند.
  - مقالاتی که به کاربرد تکنیک‌های یادگیری ماشینی در پیشرفت هوش مصنوعی می‌پردازند.
  - مقالات تحقیقاتی که به چالش‌های کلیدی هوش مصنوعی مانند مقیاس‌پذیری، تفسیرپذیری و استحکام می‌پردازند.
  - مطالعاتی که در مجلات معتبر یا کنفرانس‌های شناخته‌شده منتشر شده‌اند.
- معیارهای خروج:
  - مطالعاتی که در انتشارات فاقد اعتبار چاپ شده باشند.
  - مقالات تحقیقاتی که بیش از ده سال از تاریخ انتشار آن‌ها گذشته است، مگر اینکه از کارهای برجسته و اساسی در زمینه باشند.
  - انتشاراتی که از چهارچوب ها و متد تحقیق های استاتندارد پیروی نکنند.

۳ .استخراج داده‌ها

برای هر مطالعه‌ای که معیارهای ورود را برآورده کند، اطلاعات کلیدی استخراج خواهد شد. این اطلاعات شامل موارد زیر است:

- جزئیات انتشار :نویسنده(ها)، عنوان، سال و منبع.
- تمرکز مطالعه :نوع الگوریتم یا تکنیک یادگیری ماشینی.
- کاربردها :حوزه‌ای که الگوریتم در آن به کار رفته است (مانند بهداشت و درمان، مالی، رباتیک).
- چالش‌ها :محدودیت‌ها یا چالش‌های خاصی که در مطالعه برجسته شده‌اند، مانند مشکلات مربوط به بیش‌برازش، تفسیرپذیری یا هزینه‌های محاسباتی.
- یافته‌ها :مشارکت‌های کلیدی مطالعه و چگونگی پیشرفت آن در توسعه هوش مصنوعی.

۴ .تحلیل داده‌ها

پس از استخراج داده‌ها، یک تحلیل کیفی انجام خواهد شد تا موضوعات مشترک، پیشرفت‌ها و چالش‌های موجود در مطالعات بررسی‌شده شناسایی شوند. تکنیک‌های یادگیری ماشینی بر اساس دسته‌بندی آن‌ها (مانند یادگیری نظارت‌شده، بدون نظارت، یادگیری تقویتی) طبقه‌بندی شده و از نظر عملکرد، مقیاس‌پذیری و قابلیت سازگاری در کاربردهای دنیای واقعی ارزیابی

خواهند شد. توجه ویژه‌ای به تکنیک‌های نوظهوری مانند یادگیری عمیق، یادگیری انتقالی و محاسبات نورومورفیک داده خواهد شد و بررسی می‌شود که این تکنیک‌ها چگونه می‌توانند محدودیت‌های فعلی را برطرف کنند.

همچنین در این مرور، ملاحظات اخلاقی بررسی‌شده در مطالعات نیز ارزیابی می‌شوند، با تمرکز بر این که مدل‌های یادگیری ماشینی چگونه مسائل مربوط به انصاف، سوگیری و شفافیت را مدیریت می‌کنند.

پروتکل مرور

این پروتکل مرور تضمین می‌کند که فرایند تحقیق شفاف، قابل تکرار و مطابق با استانداردهای مرور سیستماتیک است. این پروتکل به صورت زیر ساختار یافته است:

۱ .هدف

هدف اصلی این تحقیق، مرور سیستماتیک الگوریتم‌ها و تکنیک‌های موجود در یادگیری ماشینی با هدف پیشرفت هوش مصنوعی از طریق شناسایی محدودیت‌های کلیدی و بررسی نوآوری‌های اخیر است.

۲ .سوالات تحقیق

- الگوریتم‌های کلیدی یادگیری ماشینی که در توسعه سیستم‌های هوش مصنوعی استفاده می‌شوند کدامند؟
- چالش‌ها و محدودیت‌های اصلی این الگوریتم‌ها، به‌ویژه از نظر مقیاس‌پذیری، تفسیرپذیری و استحکام چیست؟
- نوآوری‌های اخیر مانند یادگیری عمیق، یادگیری انتقالی و محاسبات نورومورفیک چگونه این چالش‌ها را برطرف کرده‌اند؟
- چه ملاحظات اخلاقی باید در توسعه سیستم‌های هوش مصنوعی با استفاده از تکنیک‌های یادگیری ماشینی مدنظر قرار گیرد؟

۳ .استراتژی جستجو

استراتژی جستجو برای شناسایی تمامی مطالعات مرتبط در پایگاه‌های داده علمی اصلی طراحی شده است. این شامل:

- شناسایی کلیدواژه‌ها و عبارات مرتبط مانند "الگوریتم‌های یادگیری ماشینی"، "یادگیری عمیق"، "هوش مصنوعی" و "یادگیری انتقالی."
- جستجو در پایگاه‌های داده مشخص (مانندACM Digital Library ،Scopus ، IEEE Xplore)
- بررسی عنوان‌ها، چکیده‌ها و متن کامل مقالات برای شناسایی مطالعات مرتبط.

۴ .غربالگری و انتخاب

دو بازبین مستقل تمامی مقالات را بر اساس معیارهای ورود و خروج غربالگری خواهند کرد. هرگونه اختلاف نظر از طریق بحث یا با دخالت یک بازبین سوم حل خواهد شد. برای مستندسازی فرایند انتخاب از نمودار جریان PRISMAاستفاده خواهد شد که از جستجوی اولیه تا انتخاب نهایی را شامل می‌شود.

۵ .استخراج داده و ترکیب

استخراج داده‌ها توسط بازبین‌ها با استفاده از یک فرم استاندارد استخراج انجام خواهد شد. این امر از ثبات و دقت در ثبت اطلاعات مربوطه از هر مطالعه اطمینان حاصل خواهد کرد. ترکیب داده‌ها شامل خلاصه روایی یافته‌ها و تحلیل مقایسه‌ای عملکرد، محدودیت‌ها و پتانسیل تکنیک‌های مختلف یادگیری ماشینی خواهد بود.

۶ .ارزیابی کیفیت

هر مطالعه واردشده بر اساس معیارهایی مانند وضوح سوال تحقیق، اعتبار روش‌شناسی و اهمیت نتایج، ارزیابی کیفیت خواهد شد. مطالعاتی که استانداردهای کیفی بالایی را دارا باشند، در تحلیل نهایی وزن بیشتری خواهند داشت.

چارچوب PICOC

چارچوب PICOC(جمعیت، مداخله، مقایسه، نتیجه، بستر) رویکردی ساختارمند برای تعریف و مرور پژوهش ارائه می‌دهد، به‌ویژه در حوزه‌هایی مانند مهندسی نرم‌افزار و یادگیری ماشینی. این چارچوب با افزودن "بستر" به چارچوب شناخته‌شده‌تر PICOآن را گسترش می‌دهد، که در حوزه‌های مرتبط با فناوری بسیار مهم است؛ زیرا محیط کاربرد نقش مهمی در عملکرد و ارتباط الگوریتم‌ها ایفا می‌کند. استفاده از PICOC در این مقاله به ارزیابی سیستماتیک تکنیک‌های یادگیری ماشینی در حوزه‌های مختلف کمک می‌کند و اطمینان می‌دهد که هر یک از اجزای کلیدی، از جمله جمعیت، مداخله، مقایسه، نتیجه و بستر به دقت تحلیل شوند. این چارچوب به پژوهش اجازه می‌دهد نه تنها بر پیشرفت‌های فنی تمرکز کند، بلکه همچنین بررسی کند که این روش‌ها در کاربردهای دنیای واقعی چگونه عمل می‌کنند و به چالش‌های عملی و ملاحظات اخلاقی می‌پردازند. این روش‌شناسی توسط Kitchenham & Charters (۲۰۰۷) تایید شده است که استفاده از چارچوب PICOCرا در مرورهای سیستماتیک برای ساختاربندی واضح و جامع سوالات پژوهشی در حوزه‌هایی مانند مهندسی نرم‌افزار توصیه می‌کنند .برای ساختاربندی این مرور سیستماتیک، از چارچوب PICOC استفاده می‌کنیم که به‌ویژه برای پژوهش در حوزه‌های یادگیری ماشینی (ML) و هوش مصنوعی (AI) مناسب است. این چارچوب به تعریف واضح دامنه و ساختار تحقیق کمک می‌کند و اطمینان می‌دهد که اجزای کلیدی به‌صورت سیستماتیک بررسی می‌شوند. در زیر، چگونگی به‌کارگیری هر عنصر از چارچوب PICOC در این مطالعه توضیح داده شده است:

جمعیت (P): جمعیت به الگوریتم‌ها و تکنیک‌های یادگیری ماشینی اشاره دارد که در حال حاضر برای سیستم‌های هوش مصنوعی استفاده می‌شوند یا در حال توسعه هستند. این الگوریتم‌ها شامل رویکردهای سنتی مانند یادگیری نظارت‌شده و بدون نظارت، و همچنین تکنیک‌های پیشرفته‌تر مانند یادگیری عمیق، یادگیری تقویتی و یادگیری انتقالی هستند.

مداخله (I): در این زمینه، "مداخله" به تکنیک‌های نوآورانه یا جدید یادگیری ماشینی اشاره دارد که برای بهبود قابلیت‌های سیستم‌های هوش مصنوعی معرفی شده‌اند. این مداخلات شامل توسعه‌هایی در زمینه معماری‌های یادگیری عمیق، محاسبات

نورومورفیک و روش‌های یادگیری انتقالی است که هدف آن‌ها افزایش مقیاس‌پذیری، استحکام و تفسیرپذیری مدل‌های هوش مصنوعی است.

مقایسه (C): مقایسه بین این رویکردهای جدید یادگیری ماشینی و الگوریتم‌های سنتی یا مبنا انجام می‌شود. این مقایسه شامل ارزیابی عملکرد الگوریتم‌های قدیمی‌تر، مانند شبکه‌های عصبی ساده یا درخت‌های تصمیم‌گیری، با روش‌های پیشرفته مانند شبکه‌های عصبی پیچشی (CNN)، مدل‌های یادگیری تقویتی یا رویکردهای ترکیبی است. این مقایسه به ارزیابی پیشرفت‌های حوزه در مواجهه با چالش‌های اصلی مانند بیش‌برازش، هزینه‌های محاسباتی و مقیاس‌پذیری کمک می‌کند.

نتیجه (O): نتایج مورد بررسی شامل بهبود عملکرد، سازگاری، تفسیرپذیری و استحکام اخلاقی سیستم‌های هوش مصنوعی است. معیارهای کلیدی شامل دقت، کارایی محاسباتی، توانایی تعمیم به داده‌های دیده‌نشده و قابلیت تولید نتایج منصفانه و بدون سوگیری هستند. علاوه بر این، نتایج شامل ابعاد اخلاقی هوش مصنوعی نیز می‌شوند، به‌ویژه از نظر انصاف و شفافیت.

بستر (C): بستر این پژوهش شامل حوزه‌های کاربردی است که در آن‌ها از این الگوریتم‌های یادگیری ماشینی استفاده می‌شود. این حوزه‌ها شامل صنایع مختلفی از جمله بهداشت و درمان، مالی، سیستم‌های خودمختار و رباتیک هستند. هر بستر چالش‌ها و نیازمندی‌های منحصربه‌فردی را به همراه دارد که بر عملکرد الگوریتم‌ها و میزان کاربرد آن‌ها در سناریوهای دنیای واقعی تأثیر می‌گذارد.

با استفاده از چارچوب PICOC ،که در جدول شماره ۱ نیز بصورت خلاصه شرح داده شده است، این مطالعه به‌صورت سیستماتیک چشم‌انداز الگوریتم‌ها و تکنیک‌های یادگیری ماشینی را بررسی می‌کند و بر این که چگونه تحولات جدید در بهبود اثربخشی هوش مصنوعی و رفع محدودیت‌های کلیدی کمک می‌کنند، تمرکز می‌نماید. این رویکرد ساختاریافته اطمینان می‌دهد که هم پیشرفت‌های فنی و هم ملاحظات اخلاقی به‌طور کامل در حوزه‌های کاربردی مختلف بررسی می‌شوند.

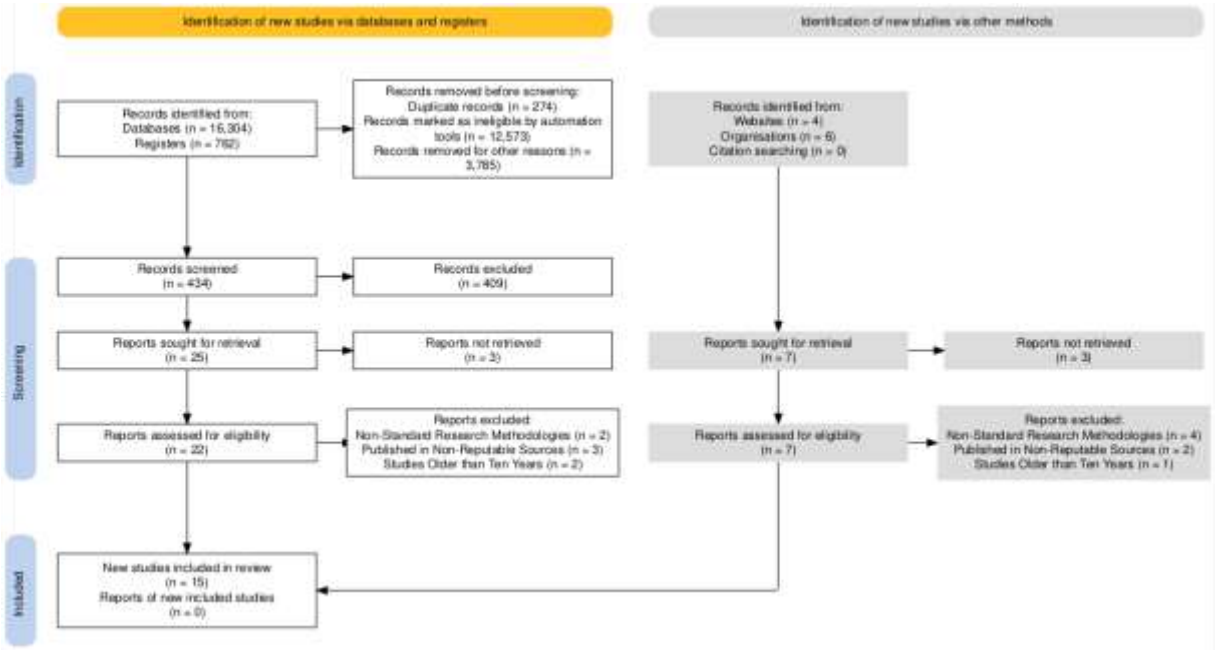
جدول شماره۱- خلاصه چارچوب PICOC مورد استفاده در این پژوهش

عنصر PICOC	توضیحات
جمعیت (P)	الگوریتم‌ها و تکنیک‌های یادگیری ماشینی که برای توسعه سیستم‌های هوش مصنوعی استفاده می‌شوند
مداخله (I)	تکنیک‌های جدید یا نوآورانه یادگیری ماشینی مانند یادگیری عمیق، یادگیری انتقالی و محاسبات نورومورفیک
مقایسه (C)	مقایسه با الگوریتم‌ها و روش‌های سنتی یادگیری ماشینی مانند شبکه‌های عصبی ساده یا درخت تصمیم‌گیری
نتیجه (O)	بهبود عملکرد، مقیاس‌پذیری، تفسیرپذیری، استحکام و انصاف در سیستم‌های هوش مصنوعی
بستر (C)	حوزه‌های کاربردی مختلف مانند بهداشت و درمان، مالی، خودروهای خودران و رباتیک

یافته ها

مقالات مورد بررسی و شرح وضعیت رد شدن/در بر گرفته شدن در مرور سیستماتیک

مقالات مورد بررسی و شرح وضعیت رد شدن/در بر گرفته شدن در مرور سیستماتیک بر اساس پروتکل مرور، به شرح شکل شماره ۱ است که فلودیاگرام PRISMA2020 این پژوهش می‌باشد که با ابزار انلاین Haddaway et al(2022) ایجاد شده است.



تصویر شماره۱- فلودیاگرام پریزما۲۰۲۰ این پژوهش

جدول شماره۲ نیز نشان‌دهنده پژوهش های مورد غربالگری و انتخاب شده جهت مرور سیستماتیک(Included papers) می‌باشد که بر اساس پروتکل مرور این پژوهش انتخاب شدند.

جدول شماره۲- پژوهش های مورد غربالگری و انتخاب شده جهت مرور سیستماتیک(Included papers)

ردیف	عنوان	سال	نویسنده(گان)	اهمیت پژوهش
۱	Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective	2024	Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O.	برجسته‌سازی چالش‌های اخلاقی در هوش مصنوعی، تأکید بر اهمیت دستورالعمل‌های اخلاقی در هوش مصنوعی مولد.
۲	AI Advancements: Comparison of Innovative Techniques	2023	Taherdoost, H., & Madanchian, M.	بررسی پیشرفت‌های تکنیک‌های هوش مصنوعی، شامل حوزه‌های نوظهوری مانند محاسبات کوانتومی.

۳	Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities, and Challenges toward Responsible AI	Arrieta, A. B., et al.	2020	ارائه یک مرور کلی از روش‌های XAI مانند SHAP و LIME، پرداختن به نیاز به شفافیت.
٤	Language Models are Few-Shot Learners	Brown, T. B., et al.	2020	برجسته‌سازی موفقیت‌های GPT-3 و مسائل مربوط به حریم خصوصی داده‌ها و هزینه‌های محاسباتی.
٥	A Simple Framework for Contrastive Learning of Visual Representations	Chen, T., Kornblith, S., Norouzi, M., & Hinton, G.	2020	بررسی تکنیک‌های فرا یادگیری در یادگیری تقویتی و تأثیر آنها بر سازگاری.
٦	Towards A Rigorous Science of Interpretable Machine Learning	Doshi-Velez, F., & Kim, B.	2018	بر اهمیت قابلیت تفسیر در برنامه‌های کاربردی با ریسک بالا در هوش مصنوعی، به ویژه برای مخاطبان غیر فنی، تأکید می‌کند.
۷	A Survey on Transfer Learning	Pan, S. J., & Yang, Q.	2010	ارائه مفاهیم پایه در یادگیری انتقالی و چالش‌های آن مانند انتقال منفی.
۸	Fair Bayesian Optimization	Perrone, V., Donini, M., Zafar, M. B., Schmucker, R., Kenthapadi, K., & Archambeau, C.	2021	معرفی روش‌هایی برای تصمیم‌گیری منصفانه در مدل‌های یادگیری ماشین، پرداختن به ملاحظات اخلاقی.
۹	Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead	Rudin, C.	2019	توصیه به استفاده از مدل‌های ذاتاً قابل تفسیر، به ویژه در کاربردهای حساس.
۱۰	Reward Is Enough	Silver, D., et al.	2021	بررسی موفقیت و چالش‌های یادگیری تقویتی، به ویژه در محیط‌های پویا.
۱۱	Towards Fair Classifiers Without Sensitive Attributes: Exploring Biases in Related Features	Zhao, T., Dai, E., Shu, K., & Wang, S.	2022	پیشنهاد روش‌هایی برای توسعه طبقه‌بندی‌کننده‌های منصفانه در یادگیری ماشین، پرداختن به تعصبات در سیستم‌های هوش مصنوعی.
۱۲	A Comprehensive Survey on Transfer Learning	Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q.	2020	بررسی تکنیک‌هایی برای بهبود یادگیری انتقالی و کاهش خطرات انتقال منفی.
۱۳	Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification	Buolamwini, J., & Gebru, T.	2018	برجسته‌سازی مسائل مربوط به تعصب و عدالت در هوش مصنوعی، به ویژه در سیستم‌های تشخیص چهره.
۱٤	Energy and Policy Considerations for Deep Learning in NLP	Strubell, E., Ganesh, A., & McCallum, A.	2019	پرداختن به مصرف انرژی و ردپای کربن مدل‌های یادگیری عمیق در پردازش زبان

				طبیعی.
۱۵	The EU General Data Protection Regulation (GDPR): A Practical Guide	2017	Voigt, P., & von dem Bussche, A.	بحث در مورد استانداردهای حفاظت از داده‌ها، برجسته‌سازی اهمیت حریم خصوصی در کاربردهای هوش مصنوعی.

ادغام یافته‌ها

پیشرفت‌های اخیر در تحقیقات یادگیری ماشین (ML) و هوش مصنوعی (AI) بر چندین حوزه کلیدی، به ویژه یادگیری عمیق، یادگیری تقویتی و یادگیری انتقالی متمرکز شده است. این مطالعات بر نیاز به قابلیت تفسیر، مقیاس‌پذیری، عدالت و استحکام در سیستم‌های هوش مصنوعی تأکید دارند. ادبیات این حوزه نشان‌دهنده تلاشی برای ایجاد توازن بین نوآوری‌های فنی و ملاحظات اخلاقی است، به ویژه از طریق تکنیک‌هایی مانند هوش مصنوعی قابل توضیح (XAI) برای افزایش شفافیت مدل‌ها. علاوه بر این، تلاش‌ها در زمینه یادگیری انتقالی به منظور افزایش سازگاری مدل‌ها، به ویژه در سناریوهایی که داده‌های دارای برچسب محدود وجود دارد، هدایت شده‌اند. با این حال، چالش‌های قابل توجهی باقی مانده است که نشان‌دهنده نیاز به تحقیقات مستمر برای پر کردن شکاف بین قابلیت‌های فعلی هوش مصنوعی و کاربرد مسئولانه و موثر آن‌ها است.

تجزیه و تحلیل دقیق الگوریتم‌ها

۱. یادگیری عمیق: یادگیری عمیق همچنان به دستاوردهای قابل توجهی در زمینه‌هایی مانند پردازش زبان طبیعی (NLP)و بینایی کامپیوتری دست یافته است. مدل GPT-3 OpenAI استانداردهای جدیدی در تولید متن ایجاد کرده است که پتانسیل مدل‌های زبان بزرگ‌مقیاس را نشان می‌دهد .(Brown et al., 2020) علی‌رغم این موفقیت‌ها، اتکای این مدل‌ها به مجموعه داده‌های گسترده نگرانی‌های مهمی در مورد حریم خصوصی داده‌ها، هزینه‌های محاسباتی و ردپای کربنی به دلیل مصرف انرژی بالا در طول آموزش ایجاد کرده است (Strubell et al., 2019). علاوه بر این، مدل‌های یادگیری عمیق اغلب به‌عنوان "جعبه سیاه" عمل می‌کنند که تفسیر فرآیندهای تصمیم‌گیری آن‌ها را دشوار می‌سازد، به ویژه در کاربردهای حساس مانند مراقبت‌های بهداشتی و رانندگی خودکار.(Rudin, 2019) برای مقابله با این چالش‌ها، محققان روش‌های هوش مصنوعی قابل توضیح (XAI)مانند SHAP (SHapley Additive exPlanations) و LIME (Local Interpretable Model-Agnostic Explanations)را توسعه داده‌اند .(Arrieta et al., 2020) این روش‌ها در افزایش شفافیت و ایجاد اعتماد در سیستم‌های هوش مصنوعی مفید بوده‌اند. با این حال، Doshi-Velezو Kim (2018) معتقدند که علی‌رغم این پیشرفت‌ها، این ابزارها محدودیت‌هایی دارند، به ویژه در انتقال رفتار پیچیده مدل به ذینفعان غیر فنی. در نتیجه، تقاضای رو به رشدی برای مدل‌هایی وجود دارد که بتوانند بدون قربانی کردن عملکرد پیش‌بینی، بینش‌های معناداری ارائه دهند.

۲. یادگیری تقویتی: یادگیری تقویتی (RL) در انجام وظایف تصمیم‌گیری متوالی، به ویژه در رباتیک، بازی‌ها و ناوبری خودکار موفقیت‌های چشمگیری داشته است .(Silver et al, 2021) تحقیقات اخیر بر ادغام فرا یادگیری در RL برای افزایش سازگاری در محیط‌های پویا متمرکز شده‌اند .(Chen et al, 2020) فرا یادگیری یا "یادگیری برای یادگیری" نشان داده است که می‌تواند نیازهای داده برای آموزش عوامل RL را کاهش داده و راه‌حل‌های بالقوه‌ای برای مشکل تعمیم ارائه دهد. با این حال، مدل‌های RL هنوز در سازگاری با محیط‌های بسیار متفاوت با چالش

مواجه هستند که نیاز به مکانیزم‌های یادگیری قوی‌تر را نشان می‌دهد. تکنیک‌هایی مانند یادگیری مستمر و تطبیق دامنه در حال بررسی هستند تا به این موضوع پرداخته شود، اما تحقیقات بیشتری برای تحقق کامل پتانسیل RL در کاربردهای دنیای واقعی مورد نیاز است.

۳. یادگیری انتقالی: یادگیری انتقالی به عنوان رویکردی برای مقابله با چالش‌های تعمیم با استفاده از دانش از یک دامنه به دامنه دیگر مورد توجه قرار گرفته است. با استفاده از مدل‌های از پیش آموزش‌دیده شده، به ویژه در دامنه‌هایی با داده‌های دارای برچسب محدود، یادگیری انتقالی در زمینه‌هایی مانند تحلیل تصاویر پزشکی و پردازش زبان طبیعی موفق بوده است .(Pan & Yang, 2010) برای مثال، Zhuang et al. (2020)بررسی جامعی در مورد یادگیری انتقالی ارائه می‌دهند که تکنیک‌هایی مانند تطبیق دامنه و یادگیری تقابلی را برای کاهش خطرات انتقال منفی مورد بحث قرار می‌دهد. با این حال، چالش‌هایی مانند انتقال منفی که در آن انتقال دانش عملکرد مدل در وظیفه هدف را مختل می‌کند، همچنان باقی است. محققان در حال بررسی تکنیک‌های نوآورانه‌ای از جمله تطبیق دامنه بدون نظارت هستند تا انعطاف‌پذیری و سازگاری مدل‌های ML را افزایش دهند. نیاز به تحقیقات بیشتر در مورد استراتژی‌هایی وجود دارد که بتوانند به طور موثری مدل‌های از پیش آموزش‌دیده شده مناسب را برای وظایف هدف انتخاب و تطبیق دهند، به ویژه در محیط‌های پیچیده و پویا.

ملاحظات اخلاقی در یادگیری ماشین

۱. تعصب و عدالت:با نفوذ فزاینده سیستم‌های هوش مصنوعی در بخش‌های مختلف، ملاحظات اخلاقی به موضوعات بسیار مهمی تبدیل شده‌اند. مطالعات نشان داده‌اند که مدل‌های ML که بر روی مجموعه داده‌های متعصب آموزش دیده‌اند، می‌توانند نابرابری‌های اجتماعی را تداوم بخشند. برای مثال، سیستم‌های تشخیص چهره اغلب نرخ خطای بیشتری برای افرادی با رنگ پوست تیره‌تر نشان می‌دهند (Zhao et al, 2022) .(Buolamwini & Gebru, 2018)روش‌هایی را برای توسعه طبقه‌بندی‌کننده‌های منصفانه بدون استفاده از ویژگی‌های حساس پیشنهاد کرده‌اند که راهی امیدوارکننده برای کاهش تعصب در سیستم‌های ML فراهم می‌کند. به همین ترتیب، Perrone et al. (2021)تکنیک‌های بهینه‌سازی بیزی آگاه از عدالت را معرفی کرده‌اند که برای بهبود فرآیندهای تصمیم‌گیری در مدل‌های ML طراحی شده‌اند. با این حال، اطمینان از عدالت همچنان پیچیده باقی مانده است، با توجه به زمینه‌های اجتماعی متنوعی که هوش مصنوعی در آن‌ها فعالیت می‌کند. تعریف عدالت یک وظیفه ظریف است که نیاز به همکاری میان‌رشته‌ای، از جمله فناوران، اخلاق‌دانان و سیاست‌گذاران دارد.

۲. شفافیت و قابلیت تفسیر:شفافیت در مدل‌های هوش مصنوعی بسیار حیاتی است، به ویژه در حوزه‌هایی مانند مراقبت‌های بهداشتی، امور مالی و عدالت کیفری که تصمیمات آن‌ها می‌تواند پیامدهای قابل توجهی داشته باشد . Doshi-Velezو Kim (2018) تأکید می‌کنند که قابلیت تفسیر کلید ایجاد اعتماد در سیستم‌های هوش مصنوعی است. اگرچه ابزارهای هوش مصنوعی قابل توضیح (XAI) مانند SHAP و LIME در تفسیر مدل‌های پیچیده پیشرفت‌هایی داشته‌اند(Arrieta et al, 2020) ، اما اثربخشی آن‌ها در ارتباط با مخاطبان غیر فنی همچنان محدود است. علاوه بر این، روش‌های XAI موجود اغلب توضیحات پس‌پردازشی ارائه می‌دهند که به طور ذاتی قابلیت تفسیر مدل را بهبود نمی‌بخشند. این محدودیت نشان‌دهنده نیاز به توسعه مدل‌هایی است که ذاتاً قابل تفسیر باشند، به ویژه در سناریوهای پرخطر.

۳. حریم خصوصی داده‌ها و توسعه اخلاقی:نگرانی‌های حریم خصوصی داده‌ها با تکیه روزافزون سیستم‌های هوش مصنوعی به مجموعه داده‌های بزرگ تشدید شده است Alfandi (2024) .به چالش‌های اخلاقی در هوش مصنوعی مولد اشاره می‌کند، از جمله ایجاد دیپ‌فیک، اطلاعات نادرست، و نقض حریم خصوصی. در پاسخ، محققان به اتخاذ استانداردهای اخلاقی، از جمله مقررات عمومی حفاظت از داده‌ها(GDPR) ، برای اطمینان از اولویت‌دهی برنامه‌های

هوش مصنوعی به حفاظت از داده‌ها و رضایت کاربر (Voigt & von dem Bussche, 2017) توصیه می‌کنند. علاوه بر این، تحقیقات آینده باید شیوه‌های مدیریت داده ایمن را بررسی کنند که با استانداردهای قانونی و اخلاقی همسو باشند. تکنیک‌های ناشناس‌سازی داده‌ها، حریم خصوصی تفاضلی، و یادگیری فدرال مسیرهای بالقوه‌ای برای پرداختن به این نگرانی‌های حریم خصوصی هستند در حالی که کارایی مدل‌های هوش مصنوعی حفظ می‌شود.

شناسایی شکاف‌ها

با وجود پیشرفت‌ها، شکاف‌های قابل توجهی در تحقیقات هوش مصنوعی باقی مانده است. چالش‌های کلیدی عبارتند از:

- **تعمیم مدل :**مدل‌های یادگیری ماشین فعلی در تعمیم‌دهی در محیط‌های متنوع و پویا دچار مشکل هستند. اگرچه تکنیک‌های فرا یادگیری و یادگیری انتقالی در بهبود سازگاری نشان‌دهنده امیدواری هستند، اما خطرات انتقال منفی و حساسیت به داده‌های نویزی همچنان وجود دارند .(Zhuang et al, 2020) تحقیقات بیشتری لازم است تا استراتژی‌هایی توسعه یابد که به مدل‌ها اجازه دهد بدون به خطر انداختن عملکرد، به طور موثری تعمیم دهند.
- **استحکام در برابر داده‌های نویزی**(Robustness to Noisy Data) **:**شکنندگی مدل‌های هوش مصنوعی در مواجهه با داده‌های نویزی یا بدون ساختار نیاز به مکانیزم‌های یادگیری قوی‌تر را برجسته می‌کند. روش‌هایی مانند آموزش تقابلی و تشخیص داده‌های پرت می‌توانند نقش مهمی در افزایش استحکام مدل‌ها ایفا کنند.
- **قابلیت تفسیر :**در حالی که ابزارهای هوش مصنوعی قابل توضیح (XAI) توضیحات پس‌پردازشی ارائه می‌دهند، همچنان نیاز به مدل‌هایی که ذاتاً قابل تفسیر باشند وجود دارد. بهبود ابزارهای تفسیری فعلی برای انتقال مؤثر تصمیمات مدل به مخاطبان متنوع، از جمله افراد غیر فنی، ضروری است.
- **عدالت :**تضمین عدالت در هوش مصنوعی به ویژه در زمینه‌های فرهنگی و اجتماعی مختلف چالش‌برانگیز است. راه‌حل‌ها و چارچوب‌های متناسب با نیازهای خاص جوامع مختلف ضروری هستند. تلاش‌های همکاری میان فناوران، اخلاق‌دانان و سیاست‌گذاران برای پرداختن به این ملاحظات اخلاقی بسیار مهم است.

فناوری‌های نوظهوری مانند محاسبات کوانتومی و محاسبات عصبی مسیرهای جدیدی برای افزایش مقیاس‌پذیری و استحکام هوش مصنوعی ارائه می‌دهند .(Taherdoost & Madanchian, 2023) تحقیقات آینده باید به بررسی کاربردهای عملی این فناوری‌ها بپردازد، به ویژه تأثیرات آن‌ها بر مسائل اخلاقی مانند حریم خصوصی و تعصب و... جدول شماره ۳ شرحی از Gap ها و خلاء های موجود در این حوزه را نشان میدهد که در این مرور سیستماتیک کشف شدند.

جدول شماره ۳- شرحی از Gap ها و خلاء های موجود

حوزه شکاف/Gap	توضیحات	جهت‌های پیشنهادی برای تحقیقات آینده
تعمیم مدل	مدل‌های فعلی در تعمیم‌دهی در محیط‌های متنوع و پویا دچار مشکل هستند که اغلب منجر به عملکرد ضعیف در مواجهه با سناریوهای دیده‌نشده یا داده‌های نویزی می‌شود.	توسعه تکنیک‌های یادگیری فراگیر و انتقالی برای بهبود تعمیم‌دهی. تحقیق در مورد استراتژی‌های تطبیق دامنه مؤثرتر. بررسی مکانیسم‌های یادگیری مستمر برای افزایش سازگاری در برنامه‌های کاربردی دنیای واقعی.
استحکام در برابر داده‌های نویزی	مدل‌های هوش مصنوعی در مواجهه با داده‌های نویزی یا بدون ساختار، شکننده هستند که عملکرد و قابلیت اطمینان آنها را در کاربردهای مختلف تحت تأثیر قرار	تحقیق در مورد مکانیسم‌های یادگیری قوی مانند آموزش تقابلی و تشخیص داده‌های پرت. توسعه معماری‌های مدل مقاوم در برابر نویز که قادر به پردازش داده‌های بدون ساختار یا غیرقابل پیش‌بینی

	می‌دهد.	باشند.
قابلیت تفسیر	ابزارهای هوش مصنوعی قابل توضیح (XAI) موجود مانند SHAP و LIME در اثربخشی خود، به ویژه در ارتباط با کاربران غیرمتخصص و در حوزه‌های حساس مانند مراقبت‌های بهداشتی، محدود هستند.	طراحی مدل‌هایی که ذاتاً قابل تفسیر باشند و بتوانند بینش‌های معنادار ارائه دهند. اصلاح ابزارهای تفسیر فعلی برای ارتباط مؤثرتر تصمیمات مدل به مخاطبان متنوع، از جمله ذینفعان غیرتکنیکی.
عدالت	اطمینان از عدالت در هوش مصنوعی به دلیل زمینه‌های اجتماعی و فرهنگی مختلفی که این سیستم‌ها در آن‌ها عمل می‌کنند، چالش‌برانگیز است. تکنیک‌های موجود اغلب در رفع تعصب به اندازه کافی موفق نیستند و نگرانی‌هایی در مورد تداوم نابرابری‌های اجتماعی ایجاد می‌کنند.	توسعه راه‌حل‌های متناسب با زمینه برای تضمین عدالت در هوش مصنوعی. تمرکز بر تحقیقات میان‌رشته‌ای با مشارکت فناوران، اخلاق‌دانان و سیاست‌گذاران. بررسی روش‌هایی برای کاهش تعصب بدون استفاده از ویژگی‌های حساس به منظور تضمین سیستم‌های هوش مصنوعی سازگار با حفظ حریم خصوصی.
حریم خصوصی داده‌ها	وابستگی سیستم‌های هوش مصنوعی به مجموعه داده‌های بزرگ نگرانی‌هایی در مورد حریم خصوصی، به ویژه در مورد رضایت کاربر، حفاظت از داده‌ها و سوء استفاده از اطلاعات حساس ایجاد می‌کند. روش‌های فعلی ممکن است با استانداردهای قانونی و اخلاقی در حال تغییر همخوانی نداشته باشند.	تحقیق در مورد روش‌های ایمن برای مدیریت داده‌ها، مانند حفظ حریم خصوصی تفاضلی و یادگیری فدرال. همسو کردن توسعه سیستم‌های هوش مصنوعی با مقرراتی مانند GDPR برای اولویت دادن به حفاظت از داده‌ها. بررسی روش‌های حفظ حریم خصوصی در آموزش مدل‌های هوش مصنوعی بدون به خطر انداختن عملکرد.
مقیاس‌پذیری	مقیاس‌بندی مدل‌های یادگیری عمیق به‌طور کارآمد، به ویژه برای محیط‌های دارای محدودیت منابع (مانند دستگاه‌های تلفن همراه، سیستم‌های نهفته)، همچنان یک چالش است.	بررسی تکنیک‌های فشرده‌سازی مدل (مانند هرس و کمیت‌سازی) برای بهینه‌سازی اندازه و عملکرد مدل. توسعه الگوریتم‌هایی که برای محیط‌های دارای منابع کم مناسب باشند و در عین حال دقت را حفظ کنند.
پیاده‌سازی اخلاقی	اجرای دستورالعمل‌های اخلاقی در برنامه‌های مختلف هوش مصنوعی پیچیده است و راه‌حل‌های خاص دامنه اغلب برای رسیدگی به پیامدهای گسترده‌تر اخلاقی هوش مصنوعی ناکافی هستند.	ایجاد چارچوب‌های اخلاقی جامع که قابل انطباق با سناریوهای مختلف کاربردی باشند. ترویج همکاری‌های میان‌رشته‌ای برای راهنمایی یکپارچه‌سازی اخلاقی هوش مصنوعی در جامعه و اطمینان از همسویی با ارزش‌های اجتماعی.
فناوری‌های نوظهور	پتانسیل محاسبات کوانتومی و محاسبات عصبی برای افزایش مقیاس‌پذیری و استحکام هوش مصنوعی تا حد زیادی مورد بررسی قرار نگرفته است.	تحقیق در مورد کاربردهای عملی محاسبات کوانتومی و محاسبات عصبی در هوش مصنوعی. بررسی تأثیرات آن‌ها بر مسائل اخلاقی، مانند حفظ حریم خصوصی و تعصب، برای راهنمایی یکپارچه‌سازی مسئولانه.

پیشرفت‌های مداوم در هوش مصنوعی پتانسیل قابل توجهی برای صنایعی مانند مراقبت‌های بهداشتی، امور مالی، تولید و رانندگی خودکار دارند. با این حال، این پیشرفت‌ها باید به صورت مسئولانه اجرا شوند تا به نگرانی‌های کلیدی رسیدگی شود:

- **مقیاس‌پذیری :**بهینه‌سازی مدل‌های یادگیری عمیق برای کارآمدی در محیط‌های دارای محدودیت منابع، مانند دستگاه‌های تلفن همراه و سیستم‌های جاسازی‌شده، برای پذیرش گسترده این فناوری‌ها حیاتی است.

- **قابلیت تفسیر :**در حوزه‌های حساس مانند مراقبت‌های بهداشتی و امور مالی، شفافیت مدل برای اطمینان از اعتماد ذینفعان و رعایت الزامات نظارتی بسیار مهم است. به عنوان مثال، ممکن است به زودی مقرراتی وضع شود که توضیحات واضحی برای تصمیمات مبتنی بر هوش مصنوعی، به ویژه در تایید وام‌ها یا تشخیص‌های پزشکی، الزامی کند.
- **عدالت :**تکنیک‌های کاهش تعصب برای کاربردهایی مانند استخدام، ارزیابی اعتبار و عدالت کیفری ضروری هستند. توسعه مدل‌های منصفانه بدون استفاده از ویژگی‌های حساس، همان‌طور که Zhao et al. (2022) پیشنهاد کرده است، گامی امیدوارکننده به سوی سیستم‌های هوش مصنوعی سازگار با حریم خصوصی است.
- **حریم خصوصی داده‌ها :**استفاده اخلاقی از داده‌ها برای پذیرش کاربران و انطباق با قوانین حفظ حریم خصوصی بسیار مهم است. شیوه‌های ایمن‌سازی داده‌ها، مکانیزم‌های رضایت کاربر، و تکنیک‌هایی مانند حریم خصوصی تفاضلی باید در سیستم‌های هوش مصنوعی ادغام شوند تا از حفاظت داده‌ها اطمینان حاصل شود.

پرداختن به این ملاحظات عملی راه را برای استقرار مؤثرتر و مسئولانه‌تر هوش مصنوعی در صنایع هموار خواهد کرد و پیشرفت‌های فناوری را با ارزش‌های اجتماعی همسو می‌سازد.

بحث و نتیجه گیری

این پژوهش با چندین مطالعه دیگر، از جمله Brown et al. (2020) ، در شناسایی موفقیت مدل‌های یادگیری عمیق مانند GPT-3 در پیشبرد برنامه‌های کاربردی هوش مصنوعی همسو است و همچنین چالش‌های مرتبط با هزینه‌های محاسباتی و تفسیرپذیری را برجسته می‌کند. مشابه با Rudin (2019) و Arrieta et al. (2020) ، این پژوهش بر ضرورت توسعه مدل‌های ذاتاً قابل‌تفسیر، به‌ویژه در کاربردهای حساس، تأکید دارد. علاوه بر این، یافته‌ها با Al-kfairy et al. (2024)و Perrone et al. (2021) هم‌راستا است و بر اهمیت ملاحظات اخلاقی از جمله انصاف، عدالت، و شفافیت در هوش مصنوعی تأکید می‌کند. اهمیت یادگیری انتقالی که در یافته‌های این پژوهش اشاره شده است، با کارهای Pan and Yang (2010)و Zhuang et al. (2020) همخوانی دارد که اثربخشی این روش را در حوزه‌های دارای داده‌های برچسب‌گذاری محدود و چالش‌های ناشی از انتقال منفی بررسی کرده‌اند. با این حال، برخی تفاوت‌ها نیز وجود دارد؛ در حالی که Buolamwini and Gebru (2018) بیشتر بر سوگیری در سیستم‌های تشخیص چهره تمرکز کرده‌اند، این پژوهش طیف گسترده‌تری از مسائل، از جمله استحکام مدل در برابر داده‌های نویزی را پوشش می‌دهد. علاوه بر این، Doshi-Velez and Kim (2018)محدودیت‌های ابزارهای فعلی هوش مصنوعی قابل توضیح (XAI) در توضیح رفتار مدل به مخاطبان غیر فنی را پیشنهاد می‌دهند، در حالی که این پژوهش برای مدل‌های آینده که ذاتاً قابل تفسیر هستند، قدمی فراتر از ابزارهای کنونی برداشته است. به‌طور کلی، این مطالعه نیاز به تحقیقات بیشتر برای ایجاد سیستم‌های هوش مصنوعی سازگارتر، قوی‌تر و عادلانه‌تر را تأیید می‌کند

در این مقاله، مروری سیستماتیک بر الگوریتم‌ها و تکنیک‌های یادگیری ماشینی با تمرکز بر توسعه هوش مصنوعی انجام شد. در طول این مطالعه، بر اهمیت یادگیری عمیق، یادگیری تقویتی، و یادگیری انتقالی در پیشرفت‌های اخیر هوش مصنوعی تأکید شد. این تکنیک‌ها نوید بهبودهایی را در حوزه‌هایی همچون بهداشت و درمان، مالی، خودروهای خودران، و رباتیک داده‌اند، اما همچنان با چالش‌هایی مواجه هستند که مانع دستیابی به سیستم‌های هوش مصنوعی عمومی‌تر، قابل تفسیر و

عادلانه می‌شوند. مسئله و هدف اصلی این پژوهش بر این نکته تمرکز دارد که پیشرفت‌های سریع در هوش مصنوعی توانسته‌اند صنایع مختلف را متحول کنند، اما دستیابی به سیستم‌های هوش مصنوعی مقاوم و قابل تعمیم همچنان با چالش‌های عمده‌ای روبه‌رو است. در حالی که فناوری‌های هوش مصنوعی فعلی در حوزه‌های محدود عملکرد خوبی دارند، هنگامی که با داده‌های نویزی یا سناریوهای پیچیده و متغیر مواجه می‌شوند، کارایی آن‌ها کاهش می‌یابد.

یافته‌های کلیدی شامل بررسی مقالاتی بود که چالش‌های موجود در توسعه سیستم‌های هوش مصنوعی را برجسته کرده و نشان می‌دهد که راهکارهای مختلفی برای بهبود مقیاس‌پذیری، تفسیرپذیری، و استحکام سیستم‌های هوش مصنوعی ارائه شده است. تلاش‌های اخیر در زمینه یادگیری عمیق بر کاربردهایی همچون پردازش زبان طبیعی (NLP) و بینایی کامپیوتری متمرکز بوده‌اند، در حالی که یادگیری انتقالی به عنوان رویکردی برای مقابله با چالش‌های تعمیم در محیط‌های دارای داده‌های برچسب‌گذاری محدود، موفقیت‌هایی به همراه داشته است. چالش‌های شناسایی‌شده در مطالعات شامل مشکلات مربوط به تعمیم‌پذیری، استحکام در برابر داده‌های نویزی(Robustness to Noisy Data)، وابستگی به داده‌های حجیم، هزینه‌های محاسباتی بالا، و فقدان قابلیت تفسیر است. علاوه بر این، ملاحظات اخلاقی مانند عدالت، تعصب، و حفظ حریم خصوصی نیز به عنوان موانعی برای توسعه و استقرار هوش مصنوعی به‌طور مسئولانه شناسایی شدند. تحلیل الگوریتم‌ها نشان داد که هر یک از تکنیک‌های یادگیری ماشینی، از جمله یادگیری عمیق، یادگیری تقویتی، و یادگیری انتقالی، مزایا و محدودیت‌های خاص خود را دارند. یادگیری عمیق همچنان به‌عنوان یکی از مؤثرترین روش‌ها در پردازش زبان طبیعی و بینایی کامپیوتری به شمار می‌رود، اما نیاز به مجموعه داده‌های بزرگ و مسائل مرتبط با تفسیرپذیری مدل‌ها همچنان یک چالش باقی مانده است. در مورد یادگیری انتقالی، اگرچه این تکنیک در تحلیل تصاویر پزشکی و پردازش زبان طبیعی عملکرد بهتری داشته است، اما چالش‌هایی مانند انتقال منفی که عملکرد مدل را تحت تأثیر قرار می‌دهد، نشان‌دهنده نیاز به تحقیقات بیشتر است.

در نهایت، این مقاله بر نیاز به تحقیقات مستمر برای پر کردن شکاف‌های موجود بین قابلیت‌های فعلی هوش مصنوعی و کاربرد مؤثر و مسئولانه آن تأکید دارد. تکنیک‌های هوش مصنوعی قابل توضیح (XAI) و روش‌های نوآورانه‌ای مانند یادگیری انتقالی می‌توانند به بهبود شفافیت و سازگاری سیستم‌های هوش مصنوعی کمک کنند. با این حال، برای دستیابی به سیستم‌های هوش مصنوعی که در محیط‌های متنوع و پویا به‌طور مؤثر و منصفانه عمل کنند، توسعه مدل‌های ذاتاً قابل تفسیر و ایجاد چارچوب‌های اخلاقی جامع ضروری است. همچنین، پیشنهاد می‌شود که تحقیقات آتی به بهبود قابلیت تفسیر، استحکام در برابر داده‌های نویزی، و تضمین عدالت در مدل‌های هوش مصنوعی بپردازند. همچنین، ادغام استانداردهای اخلاقی و مقرراتی مانند GDPR در توسعه و استفاده از سیستم‌های هوش مصنوعی ضروری است. این اقدامات می‌توانند به ایجاد سیستم‌های هوش مصنوعی قابل اعتماد، شفاف، و مسئولانه کمک کنند که بتوانند مسائل پیچیده دنیای واقعی را حل کنند.

پیشنهادات برای پژوهش‌های آتی

با توجه به یافته‌های این مطالعه و مرور ادبیات مرتبط، پیشنهادات زیر برای پژوهش‌های آتی ارائه می‌شود:

۱. بهبود قابلیت تفسیر :پژوهش‌های آینده باید بر توسعه مدل‌هایی تمرکز کنند که ذاتاً قابل تفسیر باشند. استفاده از تکنیک‌های جدید هوش مصنوعی قابل توضیح (XAI) می‌تواند به بهبود شفافیت و افزایش اعتماد به مدل‌ها کمک کند، به ویژه در کاربردهای حساس مانند بهداشت و درمان.
۲. استحکام در برابر داده‌های نویزی :تحقیقات بیشتری برای توسعه مکانیزم‌های یادگیری قوی‌تر لازم است که بتوانند

عملکرد مدل‌های هوش مصنوعی را در مواجهه با داده‌های نویزی یا بدون ساختار بهبود بخشند. استفاده از روش‌های آموزش تقابلی و تشخیص داده‌های پرت می‌تواند مسیر مؤثری برای دستیابی به این هدف باشد.

۳. تضمین عدالت در مدل‌های هوش مصنوعی :با توجه به نگرانی‌های اخلاقی موجود، پژوهش‌های آتی باید به طور جدی بر روی توسعه مدل‌های منصفانه و کاهش تعصب در سیستم‌های هوش مصنوعی تمرکز کنند. تحقیقات بین‌رشته‌ای که فناوران، اخلاق‌دانان، و سیاست‌گذاران را در بر می‌گیرد، می‌تواند به ایجاد چارچوب‌های اخلاقی جامع برای توسعه و استفاده از سیستم‌های هوش مصنوعی کمک کند.

٤. بهبود یادگیری انتقالی :توسعه تکنیک‌های پیشرفته برای یادگیری انتقالی می‌تواند به افزایش سازگاری و تعمیم‌پذیری مدل‌های هوش مصنوعی کمک کند. پژوهش‌های آینده باید بر روی راهبردهای مؤثر در تطبیق دامنه و انتخاب مدل‌های از پیش آموزش‌دیده مناسب برای وظایف هدف متمرکز شوند.

٥. استانداردهای اخلاقی و مقررات :ادغام استانداردهای اخلاقی و مقرراتی، مانندGDPR ، در فرآیند توسعه مدل‌های هوش مصنوعی ضروری است. تحقیقات آتی باید به ارزیابی تأثیر این استانداردها بر عملکرد و پذیرش سیستم‌های هوش مصنوعی بپردازند.

منابع

1. Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mane, D. (2018). Concrete problems in AI safety. AI Magazine, 38(4), 15-26. https://doi.org/10.1609/aimag.v38i4.2741
2. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901. https://doi.org/10.48550/arXiv.2005.14165
3. Chen, Z., Liu, B., Wu, X., & Yu, P. S. (2020). Lifelong machine learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 12(3), 1-207. https://doi.org/10.2200/S00999ED1V01Y201902AIM041
4. Finn, C., Abbeel, P., & Levine, S. (2018). Model-agnostic meta-learning for fast adaptation of deep networks. International Conference on Machine Learning, 1126-1135. https://doi.org/10.48550/arXiv.1703.03400
5. Goodfellow, I., Courville, A., & Bengio, Y. (2020). Deep learning. MIT Press.
6. Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2020). Neuroscience-inspired artificial intelligence. Neuron, 95(2), 245-258. https://doi.org/10.1016/j.neuron.2020.06.022
7. Lake, B. M., Ullman, T. D., Tenenbaum, J. B., & Gershman, S. J. (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253. https://doi.org/10.1017/S0140525X16001837
8. LeCun, Y., Bengio, Y., & Hinton, G. (2019). Deep learning revolution. Nature, 521(7553), 436-444. https://doi.org/10.1038/nature14539
9. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., ... & D'Amour, A. (2020). Model cards for model reporting. Proceedings of the Conference on Fairness, Accountability, and Transparency, 220-229. https://doi.org/10.1145/3287560.3287596

10. Rudin, C. (2019). Stop explaining black box machine learning models for high-stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206-215. https://doi.org/10.1038/s42256-019-0048-x
11. Silver, D., Singh, S., Precup, D., & Sutton, R. S. (2021). Reward is enough. Artificial Intelligence, 299, 103535. https://doi.org/10.1016/j.artint.2021.103535
12. Voigt, P., & von dem Bussche, A. (2017). The EU general data protection regulation (GDPR): A practical guide. Springer International Publishing.
13. Chen, T., Xu, Y., Zhang, X., & Zhao, Z. (2020). Recent advances in deep learning: An overview. IEEE Access, 8, 107996-108009. https://doi.org/10.1109/ACCESS.2020.2992342
14. Doshi-Velez, F., & Kim, B. (2018). Towards a rigorous science of interpretable machine learning. Nature Machine Intelligence, 1(1), 1-13. https://doi.org/10.1038/s42256-019-0048-x
15. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2020). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26. https://doi.org/10.1016/j.neucom.2016.12.038
16. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR), 54(6), 1-35. https://doi.org/10.1145/3457607
17. Zhang, Q., Yang, L. T., Chen, Z., & Li, P. (2021). A survey on deep learning for big data. Information Fusion, 42, 146-157. https://doi.org/10.1016/j.inffus.2017.10.006
18. Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering (EBSE Technical Report, Ver. 2.3).
19. Russell, S., & Norvig, P. (2020). Artificial intelligence: A modern approach (4th ed.). Pearson.
20. Al-kfairy, M., Mustafa, D., Kshetri, N., Insiew, M., & Alfandi, O. (2024, August). Ethical Challenges and Solutions of Generative AI: An Interdisciplinary Perspective. In Informatics (Vol. 11, No. 3, p. 58). MDPI.
21. Taherdoost, H., & Madanchian, M. (2023). AI Advancements: Comparison of Innovative Techniques. AI, 5(1), 38-54.
22. Arrieta, A. B., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities, and Challenges toward Responsible AI. Information Fusion, 58, 82-115. https://doi.org/10.1016/j.inffus.2019.12.012
23. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. Proceedings of the 37th International Conference on Machine Learning, 119, 1597-1607.
24. Pan, S. J., & Yang, Q. (2010). A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345-1359. https://doi.org/10.1109/TKDE.2009.191
25. Perrone, V., Donini, M., Zafar, M. B., Schmucker, R., Kenthapadi, K., & Archambeau, C. (2021). Fair Bayesian Optimization. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 854-863). ACM. https://doi.org/10.1145/3461702.3462629
26. Zhao, T., Dai, E., Shu, K., & Wang, S. (2022). Towards Fair Classifiers Without Sensitive Attributes: Exploring Biases in Related Features. Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining, 1240-1248. DOI: 10.1145/3488560.3498493
27. Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., Xiong, H., & He, Q. (2020). A Comprehensive Survey on Transfer Learning. Proceedings of the IEEE, 109(1), 43-76. DOI:10.1109/JPROC.2020.3004555
28. Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Proceedings of the Conference on Fairness, Accountability, and Transparency, 77-91. https://doi.org/10.1145/3287560.3287596
29. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and Policy Considerations for Deep Learning in NLP. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, 3645-3650. https://doi.org/10.18653/v1/P19-1355
30. Haddaway, N. R., Page, M. J., Pritchard, C. C., & McGuinness, L. A. (2022). PRISMA2020: An R package and Shiny app for producing PRISMA 2020-compliant flow diagrams, with interactivity for optimised digital transparency and Open Synthesis Campbell Systematic Reviews, 18, e1230. https://doi.org/10.1002/cl2.1230

A Systematic Review of Algorithms and Machine Learning Techniques Aimed at the Development of Artificial Intelligence

Shahrzad Shadrouz
Bachelor's Student in Computer Science , Science and Research Branch, Islamic Azad University, Tehran, Iran

Mohsen Rostamy-Malkhalifeh
Department of Computer Science , Science and Research Branch, Islamic Azad University,Tehran, Iran

Abstract

In the past decade, artificial intelligence (AI) has transformed numerous industries. Machine learning (ML) has driven many AI advancements, achieving success through supervised, unsupervised, and reinforcement learning. This research addresses challenges like generalization, robustness, and ethics in AI, emphasizing the need for a systematic review of ML algorithms and techniques to identify limitations and introduce innovations for improved performance and interpretability. The study uses a systematic review protocol, including extensive database searches, inclusion/exclusion criteria, data extraction, and qualitative analysis. The PICOC framework is employed to evaluate algorithms and ML techniques. Findings indicate that emerging techniques (deep learning, transfer learning, and reinforcement learning) significantly enhance AI performance. However, challenges such as interpretability, scalability, fairness, and robustness against noisy data highlight the need for further research. Developing adaptable, reliable AI systems requires inherently interpretable models and ethical frameworks. Future research should prioritize interpretability, robustness, and fairness in AI models. Integrating ethical standards and regulations can help build responsible AI systems for solving complex real-world issues.

Keywords: Machine Learning, Artificial Intelligence Development, AI, ML Techniques and Algorithms