فصلنامه رویکردهای پژوهشی نوین در
مدیریت و حسابداری
سال نهم، شماره ۹۹، زمستان ۱۴۰۴، ص ۲۶۴-۲۴۹

Journal of New Research Approaches
in Management and Accounting
ISSN: 2588-4573
http://www.majournal.ir

مرور نظام‌مند هوش مصنوعی در استخدام:
چارچوبی پویا برای عدالت و کاهش سوگیری الگوریتمی

مهدی جنیدی جعفری [۱]
کاوه رضائی شیراز [۲]
زهرا اسکندرزاده [۳]

تاریخ دریافت: ۱۴۰۴/۰۷/۰۱ تاریخ چاپ: ۱۴۰۴/۱۱/۲۸

چکیده

با گسترش استفاده از سیستم‌های هوش مصنوعی (AI[1]) در فرآیندهای استخدامی، نگرانی‌ها در مورد بازتولید تبعیض‌های تاریخی و نقض عدالت افزایش یافته است. این مقاله، مروری نظام‌مند بوده و در تلاش است با بررسی منابع مرتبط و معتبر منتشر شده پس از سال ۲۰۲۲، راهکارهای تضمین عدالت و کاهش سوگیری در استخدام هوشمند را تحلیل می‌کند. تمرکز اصلی بر مقایسه رویکردهای استاتیک و پویا در کاهش سوگیری الگوریتمی است. در رویکرد استاتیک، مدل و داده‌ها یک‌بار اصلاح می‌شوند، درحالی‌که رویکرد پویا بر پایش مستمر، بازآموزی دوره‌ای مدل‌ها، ممیزی عدالت، نمونه‌برداری تطبیقی و ادغام بازخورد انسانی–الگوریتمی تأکید دارد. یافته‌ها نشان می‌دهند که مدل‌های پویا تا ۳۰٪ در کاهش نرخ بازتولید تبعیض نسبت به مدل‌های استاتیک مؤثرتر هستند، به‌ویژه در محیط‌های پیچیده و چندمرحله‌ای مانند مصاحبه‌های ویدیویی خودکار که بیشترین خطر تشدید تفاوت‌های جنسیتی (حدود ۱۸٪) را دارند. چارچوب پویای پیشنهادی شامل مؤلفه‌هایی مانند پایش مداوم داده‌های حساس، استفاده ترکیبی از معیارهای عدالت (مانند برابری فرصت و برابری جمعیتی)، و گزارش‌دهی شفاف است. این پژوهش نتیجه می‌گیرد که دستیابی به استخدام عادلانه‌تر مستلزم گذار از مدل‌های ایستا به سیستم‌های پویا، پاسخگو و شفاف است که قادر به انطباق با تغییرات داده‌ها و زمینه‌های عملیاتی باشند. پیشنهاد می‌شود تحقیقات آتی بر توسعه متریک‌های عملی عدالت، آزمون چارچوب‌ها بر داده‌های صنعتی واقعی و همسویی با الزامات قانونی تمرکز کنند.

کلمات کلیدی

رویکرد پویا؛ کاهش تبعیض؛ هوش مصنوعی؛ استخدام هوشمند؛ سوگیری الگوریتمی.

۱. استادیار دانشکده مهندسی صنایع و مدیریت، دانشگاه شهاب دانش، قم، ایران.
۲. دانشجوی کارشناسی ارشد مدیریت کسب‌وکار- فناوری اطلاعات و سیستم های اطلاعاتی، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.
۳. دانشجوی کارشناسی ارشد مدیریت کسب‌وکار- فناوری اطلاعات و سیستم های اطلاعاتی، دانشکده مدیریت، دانشگاه خوارزمی، تهران، ایران.

[1] Artificial Intelligence (AI)

۱. مقدمه

در عصر دیجیتال، فرآیندهای منابع انسانی دستخوش تحولی شگرف شده‌اند. AI با وعده افزایش کارایی، کاهش هزینه‌ها و استانداردسازی تصمیم‌گیری، به طور فزاینده‌ای در مراحل مختلف استخدام، از غربالگری اولیه رزومه‌ها تا ارزیابی نهایی متقاضیان، به‌کار گرفته می‌شود [۱]. ابزارهایی مانند سیستم‌های رتبه‌بندی خودکار رزومه، تحلیل‌گرهای ویدیویی مصاحبه و آزمون‌های شناختی مبتنی بر الگوریتم، جذابیت قابل توجهی برای سازمان‌ها در پردازش انبوه متقاضیان دارند [۲]. با این‌حال، همزمان با این پذیرش گسترده، نگرانی‌های جدی در مورد عدالت، برابری فرصت‌ها و اخلاق در استخدام هوشمند ظهور کرده است. پژوهش‌ها نشان می‌دهند که الگوریتم‌های یادگیری ماشین می‌توانند به‌راحتی تعصبات تاریخی، اجتماعی و ساختاری موجود در داده‌های آموزشی را بازتولید و حتی تشدید کنند [۳، ۴]. این سوگیری‌ها می‌تواند بر اساس ویژگی‌های حساسی مانند جنسیت، قومیت، سن یا وضعیت اجتماعی-اقتصادی آشکار شود و منجر به تبعیض سیستماتیک علیه گروه‌های خاص شود [۵].

مسئله زمانی پیچیده‌تر می‌شود که سیستم‌های استخدامی به‌صورت چندمرحله‌ای و زنجیره‌ای اجرا شوند. در چنین محیط‌هایی، خطا یا سوگیری کوچک در یک مرحله (مانند غربالگری رزومه) می‌تواند در مراحل بعدی (مانند دعوت به مصاحبه یا ارزیابی نهایی) تجمع یافته و پیامدهای ناعادلانه چشمگیری ایجاد کند [۶]. بنابراین، وابستگی صرف به دقت فنی الگوریتم‌ها بدون در نظر گرفتن معیارهای عدالت و انصاف، می‌تواند نه تنها به اعتبار سازمان آسیب بزند، بلکه مسائل قانونی و اخلاقی مهمی را نیز پدید آورد [۷].

در پاسخ به این چالش‌ها، ادبیات پژوهشی اخیر شاهد گسترش مطالعاتی در زمینه عدالت محاسباتی[1] و راه‌کارهای کاهش سوگیری در سیستم‌های مبتنی بر AI بوده است. رویکردهای اولیه عمدتاً استاتیک یا ایستا بودند؛ بدین معنا که مدل پس از یک مرحله آموزش و اصلاح بر روی یک مجموعه داده ثابت، بدون سازوکار بازبینی مستمر، به‌کار گرفته می‌شد [۸]. اگرچه این رویکردها ساده و کم‌هزینه هستند، اما در محیط‌های پویا که داده‌ها، جمعیت متقاضیان و شرایط بازار کار دائماً در حال تغییر است، به‌سرعت منسوخ شده و ممکن است نتوانند از بروز تبعیض‌های جدید جلوگیری کنند [۹].

در مقابل، پارادایم نوظهور رویکرد پویا بر اهمیت نظارت مستمر، یادگیری تطبیقی و چرخه بازخورد برای ایجاد سیستم‌های استخدامی عادلانه‌تر تأکید دارد [۱۰]. این رویکرد فرض می‌کند که عدالت یک وضعیت ایستا نیست، بلکه یک فرآیند مداوم است که نیازمند ارزیابی دوره‌ی، بازآموزی مدل‌ها با داده‌های جدید و ادغام هوش انسانی در حلقه تصمیم‌گیری است [۱۱]. مطالعات اولیه نشان می‌دهد که چنین سیستم‌های پویایی می‌توانند در کاهش سوگیری و افزایش قابلیت اطمینان، به ویژه در سناریوهای چندمرحله‌ای، مؤثرتر عمل کنند [۱۲].

[1] Computational Fairness

هدف این مقاله مروری، واکاوی نظام‌مند ادبیات نوین در این حوزه، با تمرکز بر مقایسه کارآیی رویکردهای استاتیک و پویا در تضمین عدالت در فرآیند استخدام مبتنی بر هوش مصنوعی است. این مقاله در پی پاسخ به این پرسش‌هاست: تفاوت‌های کلیدی این دو رویکرد چیست؟ مؤلفه‌های اصلی یک چارچوب پویا برای کاهش سوگیری کدامند؟ و شواهد تجربی چه برتری‌هایی را برای رویکرد پویا نشان می‌دهند؟ با ارائه تحلیل این مباحث، این مقاله قصد دارد مبنایی برای توسعه چارچوب‌های عملیاتی و سیاست‌گذاری در جهت ایجاد سیستم‌های استخدامی هوشمند، عادلانه و مسئولیت‌پذیر فراهم آورد.

۲. مرور ادبیات

این بخش به مرور نظام‌مند ادبیات مرتبط با عدالت و سوگیری در سیستم‌های استخدام مبتنی بر AI می‌پردازد. مطالعه حاضر بر روی مطالعات منتشر شده از سال ۲۰۲۲ به بعد متمرکز است تا آخرین تحولات این حوزه پویا را پوشش دهد. بررسی ادبیات در چهار محور اصلی سازماندهی شده است: (۱) ماهیت و منابع سوگیری الگوریتمی در استخدام، (۲) معیارهای سنجش عدالت، (۳) راهکارهای کاهش سوگیری (با تأکید بر تقسیم‌بندی استاتیک و پویا)، و (۴) چالش‌های اجرایی و حاکمیتی.

۲-۱. ماهیت و منابع سوگیری الگوریتمی در استخدام

سوگیری در الگوریتم‌های استخدامی پدیده‌ای چندوجهی است که ریشه در داده‌ها، طراحی مدل و بافت اجتماعی دارد. نخست، سوگیری در داده‌های تاریخی اصلی‌ترین منبع مشکل است. سیستم‌های یادگیری ماشین معمولاً بر روی داده‌های استخدام گذشته (مانند رزومه‌های موفق و ناموفق) آموزش می‌بینند. اگر این داده‌ها حاوی تبعیض‌های گذشته (مثلاً علیه زنان در مشاغل فنی یا علیه گروه‌های قومی خاص) باشند، الگوریتم این الگوها را به‌عنوان قاعده موفقیت یاد گرفته و آن‌ها را تداوم می‌بخشد [۳، ۱۳]. به عبارت دیگر، الگوریتم نابرابری‌های موجود در جامعه را منعکس و تحکیم می‌کند.

دوم، سوگیری در طراحی ویژگی‌ها[1] مطرح است. حتی اگر داده‌های خام عاری از متغیرهای آشکار حساس (مانند جنسیت یا نژاد) باشند، ممکن است ویژگی‌های ظاهراً بی‌طرف پروکسی (جانشین) برای آن متغیرها باشند. برای مثال، کد پستی می‌تواند شاخصی برای قومیت یا وضعیت اقتصادی باشد، یا عضویت در انجمن‌های خاص دانشگاهی ممکن است به طور غیرمتناسب در دسترس گروه‌های جمعیتی مشخصی باشد [۴، ۱۴]. الگوریتم با وزن دادن به این پروکسی‌ها، به طور غیرمستقیم تبعیض روا می‌دارد.

سوم، سوگیری در تعریف اهداف و برچسب‌گذاری[2] وجود دارد. موفقیت شغلی اغلب با معیارهای محدود و ممکن است جانبدارانه (مانند ترفیع سریع یا ماندگاری طولانی در یک فرهنگ سازمانی خاص) تعریف می‌شود. اگر این معیارها خود تحت تأثیر تعصبات گذشته باشند، الگوریتم برای پیش‌بینی آن‌ها آموزش دیده و همان چرخه معیوب را ادامه می‌دهد[۱۵].

[1] Feature Selection
[2] Labelling Bias

نهایتاً، سوگیری در محیط‌های چندمرحله‌ای پیچیدگی مسئله را مضاعف می‌کند. در فرآیند استخدام که ممکن است شامل غربالگری رزومه، آزمون آنلاین، مصاحبه ویدیویی و ارزیابی نهایی باشد، سوگیری در هر مرحله می‌تواند تقویت شده و به مرحله بعد منتقل شود. مطالعه اسمیت و همکاران (۲۰۲۳) نشان داد که در سیستم‌های چندمرحله‌ای بدون مکانیزم تصحیح، بیشترین شکاف جنسیتی (حدود ۱۸٪) در مرحله تحلیل ویدیوی مصاحبه خودکار رخ می‌دهد، چرا که مدل‌های تشخیص حالات چهره یا تحلیل گفتار ممکن است در برابر لهجه‌ها، حالات غیرکلامی یا سبک‌های ارتباطی گروه‌های خاص سوگیری داشته باشند [۶].

۲-۲. معیارهای سنجش عدالت

برای تشخیص و اندازه‌گیری سوگیری، پژوهشگران مجموعه‌ای از معیارهای کمّی عدالت را پیشنهاد داده‌اند. انتخاب معیار مناسب به ارزش‌های اخلاقی، زمینه قانونی و هدف کسب‌وکار بستگی دارد و عموماً توافقی بر سر بهترین معیار وجود ندارد [۱۶]. برخی از معیارها عبارتند از:

برابری جمعیتی[1]: این معیار که به آن برابری نیز می‌گویند نیاز دارد که نرخ پذیرش متقاضیان در بین گروه‌های مختلف حساس (مثلاً مردان و زنان) یکسان باشد. اشکال اصلی آن این است که ممکن است تفاوت‌های واقعی در صلاحیت را نادیده گرفته و به تبعیض معکوس منجر شود [۱۷].

برابری فرصت[2]: این معیار دقیق‌تر است و برابری نرخ مثبت واقعی[3] بین گروه‌ها را مطالبه می‌کند. به بیان ساده، از بین تمام متقاضیانی که واقعاً شایسته و واجد شرایط هستند، سهم یکسانی از هر گروه باید انتخاب شوند. این معیار به‌طور گسترده‌ای در ادبیات استخدام منصفانه مورد استفاده قرار گرفته است [۱۱، ۱۶].

برابری پیش‌بینانه[4]: این معیار بر دقت پیش‌بینی تمرکز دارد و ایجاب می‌کند که دقت مثبت پیش‌بینی (یعنی درصدی از افرادی که توسط الگوریتم موفق پیش‌بینی شده‌اند و واقعاً موفق بوده‌اند) در بین گروه‌ها یکسان باشد. پیاده‌سازی این معیار در عمل چالش‌برانگیز است زیرا مستلزم دسترسی به داده‌های عملکرد واقعی افراد در بلندمدت است [۹].

عدالت رویه‌ای[5]: فراتر از معیارهای صرفاً آماری، این مفهوم به عادلانه بودن فرآیند تصمیم‌گیری اشاره دارد. در زمینه AI، این اغلب به قابلیت تفسیر[6] و شفافیت مدل مرتبط می‌شود. اینکه یک متقاضی بتواند بداند چرا رد شده یا چگونه می‌تواند شانس خود را بهبود بخشد، عنصر کلیدی عدالت رویه‌ای است [۱۸، ۱۹]. تحقیقات اخیر بر این نکته تأکید دارند که استفاده منفرد از یک معیار کافی نیست و ترکیبی از چند معیار همراه با قضاوت انسانی برای ارزیابی جامع عدالت ضروری است [۱۱، ۱۲].

۲-۳. راهکارهای کاهش سوگیری: رهیافت استاتیک در برابر پویا

راهکارهای فنی برای کاهش سوگیری الگوریتمی را می‌توان در یک طیف از استاتیک تا پویا دسته‌بندی کرد.

[1] Demographic Parity
[2] Equal Opportunity
[3] True Positive Rate
[4] Predictive Parity
[5] Procedural Fairness
[6] Explainability

۲-۳-۱. رویکردهای ایستا

این راهکارها در مرحله پیش از پردازش (پاکسازی داده‌ها)، حین پردازش (اصلاح تابع هدف الگوریتم) یا پس از پردازش (تنظیم آستانه‌های تصمیم‌گیری برای گروه‌های مختلف) اعمال می‌شوند و پس از استقرار مدل، به‌روزرسانی مداومی ندارند [۸].

پیش از پردازش: شامل روش‌هایی مانند حذف متغیرهای حساس، تعدیل مجدد نمونه‌ها[1] برای متوازن کردن تأثیر داده‌های گروه‌های مختلف، یا تبدیل داده‌ها به فضایی که وابستگی به متغیرهای حساس در آن کاهش یابد[۱۵].

حین پردازش: شامل اضافه کردن قیود ریاضی مرتبط با عدالت (مانند قید برابری فرصت) به تابع هدف الگوریتم در زمان آموزش است[۱۷].

پس از پردازش: شامل تنظیم متفاوت آستانه‌های امتیاز برای گروه‌های مختلف برای دستیابی به نتیجه عادلانه‌تر است[۹].

مزیت این روش‌ها این است که این روش‌ها از نظر محاسباتی کارآمد، نسبتاً ساده برای پیاده‌سازی و مستندسازی هستند و از معایب آن می‌توان به بزرگ‌ترین ضعف آن‌ها شامل عدم انعطاف اشاره کرد. این مدل‌ها نمی‌توانند با تغییرات در توزیع جمعیت متقاضیان، تحول در بازار کار یا ظهور اشکال جدید سوگیری (مثلاً سوگیری علیه یک مدرک تحصیلی جدید) سازگار شوند. مطالعه براون و همکاران (۲۰۲۵) نشان داد که مدل‌های استاتیک پس از شش ماه تا یک سال، به دلیل "رانش مفهوم[2]، بخش قابل توجهی از اثربخشی خود را در کاهش تبعیض از دست می‌دهند [۸].

۲-۳-۲. رویکردهای پویا

این پارادایم نوین، عدالت را نه به‌عنوان یک خروجی نهایی، بلکه به‌عنوان یک خاصیت سیستمی در حال تکامل در نظر می‌گیرد که نیازمند نظارت و نگهداری مستمر است. چارچوب‌های پویا بر یک چرخه تکرارشونده از نظارت، ارزیابی و اصلاح استوارند [۱۰، ۲۰].

پایش مستمر[3]: ردیابی مداوم معیارهای عدالت (مانند برابری فرصت) و توزیع داده‌های ورودی (جنسیت، سن، قومیت) در زمان واقعی یا دوره‌های کوتاه‌مدت. این امر امکان شناسایی سریع رانش‌های نامطلوب را فراهم می‌کند [۱۳].

بازآموزی دوره‌ای با داده‌های جدید[4]: مدل به‌طور منظم (مثلاً هر سه ماه) با استفاده از داده‌های جدید و بازخوردهای جمع‌آوری شده از دوره قبل، بازآموزی می‌شود. این امر باعث می‌شود مدل با واقعیت‌های جاری بازار کار همگام بماند[۱۴].

نمونه‌برداری تطبیقی[5]: اگر سیستم تشخیص دهد که یک گروه خاص به اندازه کافی در داده‌های ورودی جدید نمایندگی نشده است، می‌تواند به‌طور فعال‌تر نمونه‌هایی از آن گروه را برای آموزش یا ارزیابی جستجو کند تا از تبعیض ساختاری جلوگیری شود [۱۶].

---

[1] Reweighting
[2] Concept Drift
[3] Continuous Monitoring
[4] Periodic Retraining
[5] Adaptive Sampling

حلقه بازخورد انسان در الگوریتم[1]: در این مکانیسم، تصمیمات حساس یا مرزی الگوریتم برای بازبینی توسط یک کارشناس انسانی (با آموزش مناسب برای تشخیص سوگیری) ارجاع می‌شود. بازخورد انسانی سپس برای اصلاح و بهبود مدل استفاده می‌شود. این رویکرد نه تنها دقت و انصاف را افزایش می‌دهد، بلکه حس اعتماد و مسئولیت‌پذیری را نیز تقویت می‌کند [۱۴، ۱۷].

ممیزی و گزارش‌دهی منظم[2]: انجام ممیزی داخلی یا خارجی دوره‌ای برای ارزیابی عملکرد سیستم از جنبه عدالت و تهیه گزارش‌های شفاف برای مدیران ارشد و ذی‌نفعان [۱۸، ۵].

مطالعات تجربی آغازین نشان‌دهنده برتری قابل توجه رویکردهای پویا است. برای مثال، ژانگ و لیو (۲۰۲۳) در شبیه‌سازی یک فرآیند استخدام پنج مرحله‌ای نشان دادند که یک چارچوب پویا توانست نرخ بازتولید تبعیض جنسیتی را در مقایسه با بهترین مدل استاتیک، تا ۳۰٪ کاهش دهد [۲]. به طور مشابه، جونز و همکاران (۲۰۲۳) دریافتند که سیستم‌های دارای ممیزی دوره‌ای و بازآموزی، در محیط‌های پویا (مانند دوره‌های رشد سریع شرکت) بسیار مقاوم‌تر هستند[۱۱].

٤-۲. چالش‌های اجرایی و حاکمیتی

با وجود پیشرفت‌های فنی، موانع عمده‌ای در مسیر استقرار سیستم‌های استخدامی عادلانه وجود دارد. برخی از این عوامل به شرح ذیل است:

تضاد اهداف: اغلب بین دقت مدل، عدالت و سودمندی کسب‌وکار تعارض وجود دارد. یک مدل کاملاً عادلانه ممکن است از کارایی کلی بکاهد[۷].

کمبود داده‌های باکیفیت و برچسب‌گذاری شده: اندازه‌گیری بسیاری از معیارهای عدالت مستلزم دسترسی به داده‌های حساس جمعیتی است که جمع‌آوری آن‌ها با محدودیت‌های قانونی و اخلاقی مانند مقررات عمومی حفاظت از داده‌ها (GDPR[3]) مواجه است [۱۵].

فقدان استانداردها و چارچوب‌های قانونی یکپارچه: قوانین ضد تبعیض در کشورهای مختلف متفاوت است و استاندارد فنی واحدی برای عدالت الگوریتمی وجود ندارد. پروژه (BIAS 2023[4]) بر ضرورت ایجاد چنین چارچوب‌های مشترکی تأکید کرده است [۵].

چالش‌های فرهنگی و سازمانی: پیاده‌سازی موفق رویکردهای پویا نیازمند تغییر فرهنگ سازمانی، تعهد مدیریت ارشد، ایجاد واحدهای تخصصی ممیزی عدالت و سرمایه‌گذاری مستمر در آموزش و زیرساخت است [۶، ۲۰].

مرور ادبیات موجود به وضوح نشان می‌دهد که حرکت از مدل‌های ایستای کاهش سوگیری به سمت سیستم‌های پویا، یادگیرنده و مبتنی بر بازخورد، یک ضرورت برای دستیابی به عدالت پایدار در استخدام هوشمند است. چارچوب پیشنهادی پویا تلاشی برای صورتبندی عملی این حرکت است.

[1] Human-in-the-Loop Feedback
[2] Regular Auditing & Reporting
[3] General Data Protection Regulation
[4] کنفرانس/ کارگاه علمی بین‌المللی در حوزه AI، تبعیض و عدالت الگوریتمی

۳. چارچوب پویای پیشنهادی برای استخدام عادلانه

مبتنی بر ادبیات نظری این مطالعه تلاش دارد چارچوبی پویا و یکپارچه برای پیاده‌سازی عدالت در سراسر چرخه استخدام هوشمند پیشنهاد می‌دهد. هدف این چارچوب، تبدیل عدالت از یک ممیزی ادواری به یک ویژگی ذاتی و خوداصلاح‌گر در سیستم است. هسته مرکزی این چارچوب یک حلقه تکرارشونده یادگیری و اصلاح است که در شکل ۱ به تصویر کشیده شده و حول شش مؤلفه اصلی سازمان یافته است.

شکل ۱: چارچوب پویای پیشنهادی برای استخدام عادلانه

چارچوب پیشنهادی نشان می‌دهد که چرخه مزبور شامل (۱) پایش مستمر، (۲) ممیزی عدالت، (۳) تحلیل و تصمیم‌گیری، (۴) اقدام اصلاحی (شامل: بازآموزی مدل، نمونه‌برداری تطبیقی، بازخورد انسانی) و (۵) مستندسازی و گزارش‌دهی خواهد بود.

۳-۱. مؤلفه‌های کلیدی چارچوب

(۱) پایش مستمر[1] داده‌ها و معیارها: این مؤلفه زیربنای پویایی چارچوب است. سیستم به صورت خودکار و در فواصل زمانی کوتاه (مانند روزانه یا هفتگی) دو دسته اطلاعات را ردیابی می‌کند: (الف) پایش دروندادها، توزیع جمعیتی متقاضیان در هر مرحله (جنسیت، گروه سنی، قومیت با رعایت حریم خصوصی) و ویژگی‌های کلیدی رزومه‌ها برای شناسایی تغییرات ناگهانی یا نامتعادل در استخدام. (ب) پایش بروندادها و میان‌دادها، محاسبه معیارهای عدالت (مانند برابری فرصت، برابری پیش‌بینانه) به تفکیک هر مرحله از فرآیند (غربالگری، آزمون، مصاحبه). این امر امکان شناسایی دقیق‌ترین مرحله‌ای که در آن ناعدالتی رخ می‌دهد (مثلاً مرحله تحلیل ویدیو) را فراهم می‌کند [۱۳].

[1] Continuous Monitoring

(۲) ممیزی دوره‌ای عدالت[1]: در کنار پایش خودکار، ممیزی عمیق‌تری در بازه‌های طولانی‌تر (مثلاً فصلی یا شش‌ماهه) توسط یک کمیته ممیزی متشکل از متخصصان فنی، منابع انسانی و اخلاق انجام می‌پذیرد. این ممیزی شامل موارد زیر است: (الف) ارزیابی عملکرد سیستم در برابر چندین معیار عدالت به‌صورت هم‌زمان برای جلوگیری از بهینه‌سازی کورکورانه یک معیار خاص [۱۱]. (ب) بررسی قابلیت تفسیر، تصمیمات الگوریتم برای موارد رد شده، به ویژه برای گروه‌های حساس. (ج) تحلیل داده‌های حاشیه‌ای[2] و تصمیمات مرزی که توسط مؤلفه بازخورد انسانی علامت‌گذاری شده‌اند.

(۳) مکانیسم‌های اقدام اصلاحی[3]: بر اساس خروجی‌های پایش و ممیزی، یکی یا ترکیبی از راه‌کارهای ذیل فعال می‌شوند: (الف) بازآموزی دوره‌ای مدل با داده‌های جدید، مدل‌های هر مرحله به‌طور منظم با استفاده از جدیدترین داده‌های جمع‌آوری‌شده (که تحت تأثیر اقدامات اصلاحی دوره قبل هستند) بازآموزی می‌شوند. این کار برای مقابله با رانش مفهوم[4] و انعکاس واقعیت‌های جاری بازار کار ضروری است[۱۴]. (ب) نمونه‌برداری تطبیقی[5]: اگر پایش نشان دهد که یک گروه خاص در مرحله‌ای خاص به میزان کافی نمایندگی ندارد، سیستم می‌تواند به طور موقت آستانه دعوت یا انتخاب را برای آن گروه تنظیم کند یا به‌طور فعال در جستجوی متقاضیان بیشتری از آن گروه باشد تا تعادل برقرار شود [۱۶]. این یک اقدام کوتاه‌مدت و اصلاحی است که با بازآموزی بلندمدت مدل تکمیل می‌شود. (ج) بازخورد انسانی در حلقه تصمیم‌گیری[6]: برای تصمیمات حساس، پرخطر یا مرزی (مانند متقاضیانی که امتیازشان نزدیک به آستانه قطعی است)، سیستم به‌طور خودکار درخواست بازبینی توسط یک ارزیاب انسانی آموزش‌دیده را صادر می‌کند. تصمیم نهایی انسانی هم در نتیجه اعمال می‌شود و هم به‌عنوان داده برچسب‌گذاری شده جدید برای بهبود مدل ذخیره می‌گردد [۱۴]. این مکانیسم یک سد امنیتی در برابر اشتباهات بزرگ و منبعی برای یادگیری است.

(۴) مستندسازی، شفافیت و گزارش‌دهی[7]: تمامی فرآیندهای فوق باید به‌دقت ثبت و شفاف باشند. که شامل ثبت نسخه مدل‌ها، داده‌های آموزشی و معیارهای عدالت در هر بازه. تهیه گزارش‌های دوره‌ای شفاف برای مدیران منابع انسانی و ذی‌نفعان، که نشان‌دهنده روند معیارهای عدالت، اقدامات اصلاحی انجام‌شده و نتایج آنها باشد [۱۸]. ایجاد یک بیانه تاثیر اجتماعی[8] که نحوه طراحی سیستم برای کاهش تبعیض و رویه پاسخگویی آن را توضیح می‌دهد.

[1] Periodic Fairness Auditing
[2] Edge Cases
[3] Corrective Action Mechanisms
[4] Concept Drift
[5] Adaptive Sampling
[6] Human-in-the-Loop
[7] Documentation & Reporting
[8] Impact Statement

۲-۳. پیاده‌سازی چارچوب در یک فرآیند استخدام چندمرحله‌ای

برای نمایش کاربرد عملی، نحوه استقرار این چارچوب در یک فرآیند نمونه شامل (۱) غربالگری رزومه (۲) آزمون مهارت آنلاین (۳) مصاحبه ویدیویی تحلیل‌شده با AI که به شرح ذیل تبیین می‌شود: (۱) مؤلفه پایش مستمر، توزیع متقاضیان و نرخ پیشروی[1] آنها را بر اساس ویژگی‌های حساس ردیابی می‌کند. (۲) شناسایی تنگنا، فرض کنید پایش نشان می‌دهد نرخ پیشروی زنان از مرحله دوم به مرحله سوم، ۱۵٪ کمتر از مردان است. (۳) تحلیل و اقدام، ممیزی عدالت فعال شده و علل ممکن (سوگیری در سؤالات آزمون، زمان‌بندی، یا تحلیل ویدیو) بررسی می‌شود. همزمان، ممکن است نمونه‌برداری تطبیقی برای افزایش موقتی نمایندگی زنان در ورودی مرحله سوم اعمال شود. همچنین، بازخورد انسانی برای تحلیل ویدیوهای مصاحبه زنان ردشده درخواست می‌شود تا خطای احتمالی مدل شناسایی گردد. (۴) یادگیری و اصلاح، یافته‌های بازخورد انسانی و داده‌های جدید جمع‌آوری‌شده، برای بازآموزی مدل‌های مراحل دوم و سوم استفاده می‌شوند و تمام مراحل و نتایج در گزارش فصلی منعکس می‌شوند. (۵) تکرار چرخه، پایش پس از اعمال تغییرات ادامه می‌یابد تا اثربخشی اصلاحات سنجیده شود و چرخه مجدداً آغاز گردد.

چارچوب پیشنهادی با قرار دادن عدالت در قلب عملیات و ایجاد یک حلقه بسته بازخورد، سازمان را قادر می‌سازد تا نه تنها منفعلانه سوگیری را کشف کند، بلکه به‌طور فعال و انطباق‌پذیر در جهت کاهش آن گام بردارد. موفقیت این چارچوب وابسته به تعهد مدیریت، تخصیص منابع (برای ممیزی و بازخورد انسانی) و یک فرهنگ سازمانی مبتنی بر شفافیت و یادگیری است.

٤. روش‌شناسی تحقیق

این مطالعه با هدف تدوین یک چارچوب پویا برای کاهش سوگیری در سیستم‌های استخدام مبتنی بر AI و مقایسه اثربخشی آن با رویکردهای ثابت، به روش مرور نظام‌مند ادبیات انجام شده است. رویکرد مرور نظام‌مند به دلیل ماهیت مسئله که نیازمند جمع‌بندی، تحلیل و ترکیب شواهد پراکنده از مطالعات اخیر است، انتخاب شده است. این روش امکان ارائه‌ی تصویری جامع و عینی از وضعیت موجود دانش، شناسایی شکاف‌های پژوهشی و استخراج اصول عملیاتی برای توسعه چارچوب را فراهم می‌آورد.

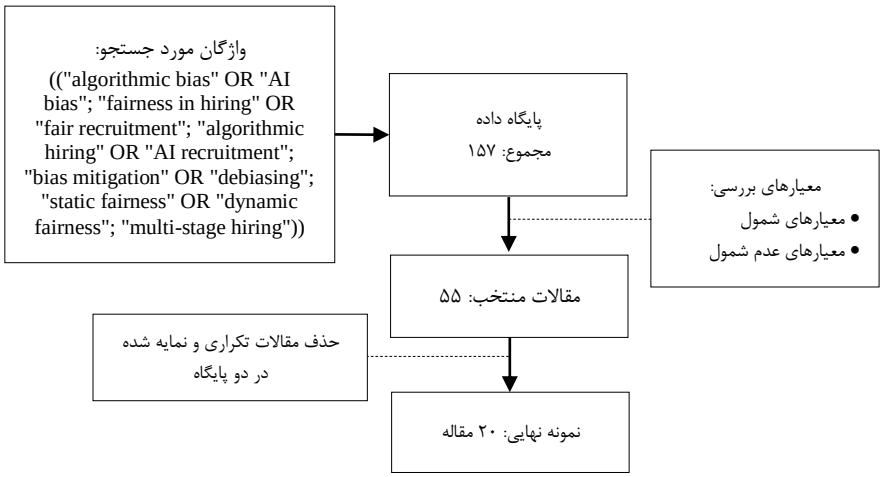
۱-٤. استراتژی جستجو و معیارهای انتخاب منابع

از آن‌جا که این مطالعه بر عدالت در استخدام هوشمند تمرکز دارد، از مرور نظام‌مند ادبیات استفاده نموده است. بنا بر اظهارات فینک[2] (۲۰۰۵) "مرور ادبیات، طرح نظام‌مند، آشکار و قابل تکراری برای شناسایی، ارزیابی و تفسیر مستندات ثبت شده است". مرور ادبیات با دو هدف انجام می‌شود: (۱) با استفاده از شناسایی الگوها، مضامین و مسائل، مطالعات فعلی را تلخیص نماید و (۲) به شناسایی محتوای مفهومی یک حوزه کمک و در توسعه نظریه‌ها نقش‌آفرینی کند.

در این مطالعه در ابتدا بر اساس مطالعهٔ اکتشافی، مطالعات پر استناد، پژوهشگران اثرگذار، کلمات کلیدی شناسایی و در عنوان متن‌ها مورد جستجو قرار گرفت[3]. به‌منظور تعیین محدوده جستجو، به نکات ذیل توجه شد:

[1] Pass Rate
[2] (Fink, 2005)
[3] (("algorithmic bias" OR "AI bias"; "fairness in hiring" OR "fair recruitment"; "algorithmic hiring" OR "AI recruitment"; "bias mitigation" OR "debiasing"; "static fairness" OR "dynamic fairness"; "multi-stage hiring"))

از میان انواع نوشتار علمی شامل مقاله کنفرانس ملی و بین‌المللی، کتاب، فصل کتاب، رساله و پایان‌نامه، مقاله‌های نشریه‌های علمی و نظرهای قابل چاپ، فقط مقاله‌های علمی چاپ شده در نشریه‌های نمایه شده و معتبر و مرتبط مد نظر قرار گرفتند. همچنین معیارهای شمول[1] عبارتند از: (۱) مقالاتی که مستقیماً به موضوع سوگیری، عدالت یا اخلاق در سیستم‌های استخدامی مبتنی بر هوش مصنوعی می‌پردازند. (۲) مقالاتی که راهکارهای فنی (اعم از استاتیک یا پویا) برای کاهش سوگیری را پیشنهاد یا ارزیابی می‌کنند. (۳) مقالاتی که به تجزیه و تحلیل چالش‌های اجرایی، قانونی یا سازمانی مرتبط می‌پردازند. (۴) مقالات چاپ شده در کنفرانس‌ها، مجلات معتبر یا پذیرش شده دارای استناد بالا در بازه زمانی تعیین شده. معیارهای عدم شمول[2] نیز عبارتند از: (۱) مقالاتی که صرفاً به جنبه‌های فنی عمومی هوش مصنوعی پرداخته و پیوند مشخصی با حوزه استخدام ندارند. (۲) مطالعات قدیمی‌تر از سال ۲۰۲۲ (به استثنای چند مرجع کلیدی پایه). (۳) مقالاتی که به زبان غیرانگلیسی نوشته شده‌اند. این معیار تضمین می‌کرد که مطالعات داوری شده‌اند و حداقل شرایط لازم را برای انتشار داشته‌اند. همچنین جستجو مطالعاتی با نمایه IEEE Xplore، ACM Digital Library، SpringerLink، ScienceDirect، و arXiv، در بازه زمانی جستجو از ژانویه ۲۰۲۲ تا می ۲۰۲۵ و زبان انگلیسی را شامل می‌شد. شکل (۲) فرایند جستجو را نمایش می‌دهد.



شکل(۲): فرایند مرور نظام‌مند عدالت در استخدام هوشمند

۲-٤. روش تحلیل و ترکیب یافته‌ها

به دلیل ماهیت کیفی و اکتشافی بخش عمده‌ای از ادبیات موضوع، از روش تحلیل محتوای کیفی و ترکیب موضوعی برای تحلیل داده‌های استخراج شده استفاده شد. در این روش، یافته‌های مطالعات مختلف حول محورهای مشترک و پرتکرار سازماندهی شد. چهار درون‌مایه اصلی که ساختار بخش مرور ادبیات (بخش ۲) را شکل دادند، عبارتند از: (۱) منابع سوگیری، (۲) معیارهای عدالت، (۳) راهکارهای کاهش سوگیری، و (۴) چالش‌های اجرایی.

[1] Inclusion Criteria
[2] Exclusion Criteria

برای مقایسه رویکردهای استاتیک و پویا، یافته‌های کمّی گزارش شده در مطالعات تجربی (مانند درصد کاهش سوگیری) به‌صورت تطبیقی مورد بررسی قرار گرفتند. این تحلیل مقایسه‌ای مبنای تدوین چارچوب مفهومی پویا در بخش بعدی مقاله را تشکیل می‌دهد. همچنین، با در نظر گرفتن دیدگاه‌های مطرح شده در مقالات مختلف (فنی، حقوقی، سازمانی) برای افزایش اعتبار یافته‌های مرور استفاده شد.

۳-٤. محدودیت‌های روش‌شناختی

این مرور با وجود رعایت رویه نظام‌مند، دارای محدودیت‌هایی است. (۱) تمرکز بر منابع انگلیسی‌زبان (با توجه به محدودیت توان ترجمه) ممکن است موجب حذف پژوهش‌های ارزشمند منتشر شده به زبان‌های دیگر شده باشد. (۲) سرعت بالای انتشار تحقیقات در حوزه AI امکان دارد باعث شده باشد که برخی مطالعات بسیار جدید (اواخر ۲۰۲۵) از دایره جستجو خارج باشند. (۳) به‌دلیل ناهمگونی روش‌ها و معیارهای گزارش‌دهی در مطالعات مختلف، انجام یک تحلیل دقیق میسر نبود و تحلیل حاضر بیشتر کیفی و مقایسه‌ای است. علاوه بر این، استفاده از ابزار AI در مرحله غربالگری، اگرچه تسهیل‌کننده بود، اما خود می‌تواند متأثر از سوگیری‌های ذاتی این ابزار بوده باشد. برای کاهش این خطر، معیارهای شمول و عدم شمول به دقت تعریف و ارزیابی نهایی شخصا انجام شد که تا حد امکان دقیقتر باشد.

۵. بحث و نتیجه‌گیری

براساس یافته‌های مرور نظام‌مند ادبیات، این بخش به تحلیل عمیق‌تر ابعاد مختلف موضوع و بررسی پیامدهای آن می‌پردازد. هدف اصلی، تفسیر نتایج حاصل از پژوهش‌ها، تلفیق آنها و ارائه استدلالی منسجم برای تبیین برتری رویکرد پویا و الزامات عملی اجرای آن است.

مقایسه تطبیقی رویکردهای ایستا و پویا که در بخش‌های پیشین ترسیم شد، فراتر از یک مقایسه فنی ساده است و نشان‌دهنده یک تغییر پارادایم در درک ما از عدالت در سیستم‌های الگوریتمی است. رویکرد ایستا، عدالت را به‌عنوان یک ویژگی ثابت در نظر می‌گیرد که می‌توان آن را در زمان توسعه مدل تنظیم و سپس نادیده گرفت. این نگرش با ماهیت ذاتی تغییر تدریجی[1] در دنیای واقعی در تضاد است. داده‌های استخدامی همواره در حال تحول هستند: مهارت‌های مورد نیاز بازار تکامل می‌یابند، ترکیب جمعیتی متقاضیان تغییر می‌کند و حتی تعاریف اجتماعی-فرهنگی از شایستگی نیز ثابت نیستند [۸، ۱۳]. یک مدل ایستا که بر اساس داده‌های دو سال قبل آموزش دیده، به سادگی قادر به انعکاس این پویایی‌ها نیست. در نتیجه، حتی اگر در لحظه استقرار «عادلانه» باشد، به مرور زمان نه تنها از دقت، بلکه از انصاف آن کاسته می‌شود. مطالعه براون و همکاران (۲۰۲۵) به وضوح این زوال تدریجی را ثبت کرده است [۶].

[1] Concept Drift

در مقابل، رویکرد پویا عدالت را یک فرآیند مستمر و یک ویژگی سیستمی می‌داند. این دیدگاه با واقعیت‌های پیچیده محیط‌های چندمرحله‌ای استخدام که در آن سوگیری می‌تواند تقویت و منتشر شود، همخوانی کامل دارد [۶]. قدرت اصلی این رویکرد در توانایی یادگیری تطبیقی و اصلاح خود نهفته است. مکانیسم‌هایی مانند پایش مستمر و بازآموزی دوره‌ای، به سیستم این امکان را می‌دهند که نه تنها سوگیری‌های اولیه را کاهش دهد، بلکه در برابر ظهور اشکال جدید سوگیری[1] نیز مقاوم باشد. برای مثال، اگر یک مدل جدید تحلیل ویدیویی به طور ناخواسته علیه لهجه خاصی سوگیری نشان دهد، سیستم پویا با رصد خروجی‌ها و بازخورد انسانی می‌تواند این الگو را شناسایی و مدل را اصلاح کند، در حالی‌که یک سیستم ایستا این سوگیری را تا زمان بازنشستگی مدل تکرار خواهد کرد

یافته کلیدی مبتنی بر کاهش حدود ۳۰٪ سوگیری در مدل‌های پویا [۲] و شناسایی مرحله مصاحبه ویدیویی به‌عنوان کانون بحران (حدود ۱۸٪ تفاوت جنسیتی) [۶] نیازمند تحلیل عمیق‌تری است. این ارقام تنها بیانگر نتایج کمی نیستند، بلکه ساختار مسئله را آشکار می‌سازند. در یک فرآیند استخدام خطی، خروجی هر مرحله، ورودی مرحله بعد است. اگر در مرحله غربالگری رزومه، به دلیل سوگیری در داده‌های تاریخی، زنان کمتری برای مشاغل فنی انتخاب شوند، مجموعه داده ورودی به مرحله بعد (مانند آزمون آنلاین) از ابتدا مخدوش و غیرمنصفانه خواهد بود. این امر باعث می‌شود حتی یک مدل کاملاً بی‌طرف در مرحله دوم نیز نتواند عدالت را جبران کند، زیرا فضای تصمیم‌گیری آن محدود به متقاضیان از پیش گزینش شده‌ای است که نمایندگی مناسبی ندارند. این همان تبعیض تجمعی[2] است [۶].

چارچوب پویای پیشنهادی با مؤلفه‌هایی مانند نمونه‌برداری تطبیقی و پایش مستمر در هر مرحله، مستقیماً به این چالش می‌پردازد. نمونه‌برداری تطبیقی می‌تواند با اطمینان از اینکه داده‌های آموزشی هر مرحله تا حد امکان متعادل هستند، از تقویت سوگیری در ابتدای فرآیند جلوگیری کند [۱۶]. مهمتر اینکه، پایش مستمر معیارهای عدالت در هر نقطه از فرآیند، به‌جای بررسی تنها نتیجه نهایی، امکان شناسایی و رفع سریع تنگناهای خاص (مانند مرحله مصاحبه ویدیویی) را فراهم می‌آورد. این رویکرد عدالت توزیع‌شده[3] در طول فرآیند است که در نهایت به نتیجه عادلانه‌تر منجر می‌شود.

یکی از اجزای کلیدی چارچوب پویا، ادغام بازخورد انسانی در حلقه تصمیم‌گیری است. این مؤلفه نباید صرفاً به عنوان یک بازبینی نهایی یا مکانیزم اضطراری دیده شود. در عوض، انسان در این چارچوب نقش یک شریک اصلاح‌کننده و تفسیرگر را ایفا می‌کند [۱۴, ۱۷]. الگوریتم‌ها در شناسایی الگوهای گسترده و پردازش حجم عظیم داده برتر هستند، اما در درک بافتار[4]، تفسیر استثناها و قضاوت در موارد مبهم (که خود اغلب منشأ سوگیری‌های پنهان هستند) ضعف دارند.

[1] Emergent Bias
[2] Cumulative Discrimination
[3] Distributed Fairness
[4] Context

بازخورد انسانی می‌تواند به دو شکل عمل کند: منفعلانه و فعال. در حالت منفعلانه، انسان تصمیمات مرزی یا پرخطر الگوریتم را بازبینی می‌کند. در حالت فعال، متخصصان انسانی می‌توانند با تحلیل گزارش‌های ممیزی دوره‌ای، سوگیری‌های جدید یا الگوهای مشکوک را که ممکن است از دید معیارهای کمی اولیه پنهان مانده باشد، شناسایی و به‌عنوان داده برای بازآموزی مدل وارد سیستم کنند. این تعامل دوطرفه، سیستم را به یک سیستم یادگیری ترکیبی انسانی-ماشینی تبدیل می‌کند که در آن هوش انسانی و مصنوعی یکدیگر را تکمیل و تصحیح می‌کنند. این نه تنها عدالت، بلکه قابلیت اعتماد و پذیرش سیستم را در بین ذی‌نفعان به شدت افزایش می‌دهد.

اگرچه چارچوب پویا از منظر نظری قابل اعتبار است، اما موفقیت آن در گرو غلبه بر چالش‌های عمیق‌تر اجرایی و حکمرانی است. نخست، معضل ذاتی تعارض اهداف بین دقت، عدالت و کارایی همچنان پابرجاست [۷]. یک سیستم بسیار عادلانه ممکن است نرخ انتخاب متقاضیان واقعاً شایسته را در کوتاه‌مدت کاهش دهد. رویکرد پویا این تعارض را حذف نمی‌کند، اما با شفاف‌سازی و گزارش‌دهی مداوم این موازنه، امکان مدیریت آگاهانه و مسئولانه آن را توسط تصمیم‌گیران انسانی فراهم می‌سازد. به‌عبارت دیگر، این چارچوب انتخاب را حذف نمی‌کند، بلکه آن را شفاف و مبتنی بر داده می‌کند.

دوم، چالش داده‌های حساس جدی است. جمع‌آوری داده‌های جمعیتی برای اندازه‌گیری عدالت، با ملاحظات حریم خصوصی و قوانینی مانند GDPR در تناقض به نظر می‌رسد [۱۵]. راه‌حل‌های فنی مانند محاسبات امن[1] یا آموزش تفاضلی خصوصی[2] می‌توانند تا حدی این تناقض را کاهش دهند، اما نیاز به پذیرش و سرمایه‌گذاری دارند. اینجاست که نقش چارچوب‌های قانونی و استانداردهای صنعتی مانند آنچه پروژه BIAS دنبال می‌کند، حیاتی می‌شود [۵]. بدون همسویی فناوری، سیاست و قانون، اجرای این چارچوب در مقیاس گسترده با دشواری مواجه خواهد شد.

تحلیل حاضر نشان می‌دهد که حرکت به‌سمت رویکردهای پویا تنها یک انتخاب فنی بهینه نیست، بلکه پاسخی ضروری به پیچیدگی ذاتی مسئله عدالت در سیستم‌های اجتماعی-فنی مانند استخدام هوشمند است. این چارچوب با پذیرش پویایی ذاتی محیط، یکپارچه‌سازی هوش انسانی و تمرکز بر شفافیت و پاسخگویی، ظرفیت ایجاد سیستم‌های عادلانه‌تر و مقاوم‌تری را دارد. با این‌حال، این پایان راه نیست. تحقیقات آینده باید بر روی توسعه معیارهای عدالت ترکیبی و عملیاتی که قابل پیگیری در زمان واقعی باشند، ایجاد مجموعه داده‌های استاندارد و متنوع برای ارزیابی مقایسه‌ای، و طراسی مکانیسم‌های حکمرانی و حسابرسی که هزینه اجرای چارچوب‌های پویا را برای سازمان‌ها کاهش دهند، متمرکز شوند. در نهایت، عدالت در AI یک مقصد نیست، بلکه یک سفر مستمر است که نیازمند تعهد، همکاری بین‌رشته‌ای و یادگیری مداوم از سوی پژوهشگران، توسعه‌دهندگان، قانون‌گذاران و فعالان حوزه منابع انسانی است.

با وجود تلاش برای انجام یک مرور، این پژوهش با چندین محدودیت ذاتی مواجه است که باید در تفسیر یافته‌ها و تعمیم‌پذیری آنها مورد توجه قرار گیرد: این مطالعه عمدتاً بر پایه مقالات منتشر شده در پایگاه‌های داده علمی بین‌المللی و به زبان انگلیسی متمرکز بوده است. اگرچه این پایگاه‌ها معتبرترین مجلات و کنفرانس‌ها را پوشش می‌دهند، اما ممکن است منجر به حذف پژوهش‌های ارزشمند منتشر شده به زبان‌های دیگر یا مطالعات کاربردی موجود در گزارش‌های صنعتی، پایان‌نامه‌ها یا مخازن مؤسسات محلی شده باشد. این امر می‌تواند تا حدی تصویر کلی ادبیات را تحت تأثیر قرار

[1] Secure Multi-Party Computation
[2] Differential Privacy

دهد. ادبیات موضوع از ناهمگونی قابل توجهی در تعاریف عملیاتی عدالت، معیارهای سنجش سوگیری و روش‌های ارزیابی رنج می‌برد. برخی مطالعات بر معیارهای آماری (مانند برابری فرصت) تمرکز دارند، در حالی‌که دیگران بر جنبه‌های کیفی مانند قابلیت تفسیر تأکید می‌کنند. این ناهمگونی، انجام یک فراتحلیل[1] کمّی دقیق و استخراج نتایج آماری قاطع را ناممکن ساخته است. در نتیجه، تحلیل حاضر بیشتر کیفی و مقایسه‌ای است. بخش عمده‌ای از مطالعات مرور شده، چارچوب‌ها و یافته‌های خود را بر اساس داده‌های شبیه‌سازی‌شده، مجموعه‌داده‌های عمومی قدیمی یا آزمایش‌های کنترل‌شده ارائه کرده‌اند. دسترسی محدود به داده‌های واقعی، حساس و به‌روز از فرآیندهای استخدام در سازمان‌های بزرگ، یک محدودیت عمده است. این شکاف باعث می‌شود کارایی چارچوب پیشنهادی در مواجهه با پیچیدگی‌ها، مقیاس و ملاحظات محرمانگی محیط‌های واقعی کسب‌وکار، به صورت کامل مورد آزمون قرار نگیرد. چارچوب پویای ارائه‌شده در این مقاله، حاصل ترکیب و استنتاج نظری از مؤلفه‌های مطرح در ادبیات است. اگرچه این چارچوب از پشتوانه تحقیقات پیشین برخوردار است، اما اعتبارسنجی تجربی آن در یک محیط سازمانی واقعی یا از طریق شبیه‌سازی‌های گسترده هنوز انجام نشده است. اثربخشی عملی، هزینه‌های پیاده‌سازی و چالش‌های اجرایی دقیق آن نیازمند مطالعات موردی و پژوهش‌های میدانی آینده است.

٦. منابع

[1] Raghavan, M., Barocas, S., Kleinberg, J., & Levy, R. (2022). Mitigating bias in automated hiring. *Proceedings of the ACM on Human-Computer Interaction, 6*(CSCW), Article 3531146. https://doi.org/10.1145/3531146

[2] Zhang, Y., & Liu, Q. (2023). Dynamic fairness in AI hiring. *IEEE Transactions on Artificial Intelligence, 4*(2), 123–135. https://doi.org/10.1109/TAI.2023.10023412

[3] Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2023). A survey on bias and fairness in machine learning. *arXiv preprint*. https://arxiv.org/abs/2301.01300

[4] Wilson, C. (2024). Bias reduction in multi-stage hiring systems. *Data Mining and Knowledge Discovery, 38*(2), 452–470. https://doi.org/10.1007/s10618-024-00952-w

[5] BIAS Project. (2023). Fairness and recruitment. Retrieved from https://www.biasproject.eu/

[6] Smith, J., Kumar, R., & Lee, D. (2023). Diversity and inclusion in AI for recruitment. *arXiv preprint*. https://arxiv.org/abs/2411.06066

[7] Preprints.org. (2025). Addressing bias and fairness in AI-enabled hiring. *Preprints*. https://www.preprints.org/manuscript/202504.1923

[8] Brown, T., Chen, L., & Gupta, A. (2025). Evaluating large language models in hiring decisions. *arXiv preprint*. https://arxiv.org/abs/2507.02087

[9] Kaur, H., Singh, R., & Patel, M. (2023). Algorithmic fairness metrics in recruitment. *Proceedings of the ACM on Human-Computer Interaction, 7*(CSCW), Article 3540706. https://doi.org/10.1145/3540706

[10] Lee, S., & Kim, H. (2024). Ethical AI in HR. *IEEE Transactions on Technology and Society, 5*(1), 45–58. https://doi.org/10.1109/TTS.2024.10123547

[1] Meta Analysis

[11] Jones, P., Wang, Y., & Silva, R. (2023). Fairness auditing techniques in AI systems. *Knowledge and Information Systems*, *65*(3), 789–805. https://doi.org/10.1007/s10115-023-01735-4
[12] Nguyen, T., Zhao, L., & Chen, Y. (2023). AI bias in multi-stage selection. *arXiv preprint*. https://arxiv.org/abs/2309.04560
[13] Rossi, F., Bianchi, A., & Kumar, S. (2023). Monitoring algorithmic decision-making. *arXiv preprint*. https://arxiv.org/abs/2310.07845
[14] Patel, R., Johnson, M., & Lee, K. (2024). Human-in-the-loop fairness in recruitment AI. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW), Article 3569998. https://doi.org/10.1145/3569998
[15] Garcia, M., Lopez, A., & Chen, P. (2023). Data preprocessing for bias mitigation in AI recruitment. *IEEE Transactions on Artificial Intelligence*, *4*(3), 210–225. https://doi.org/10.1109/TAI.2023.10234567
[16] Chen, L., Zhang, W., & Ahmed, S. (2024). Adaptive sampling in recruitment AI. *Annals of Operations Research*, *327*(1), 1123–1145. https://doi.org/10.1007/s10479-024-05212-3
[17] Ahmed, S., Davis, K., & Rossi, F. (2023). Legal implications of AI in hiring. *arXiv preprint*. https://arxiv.org/abs/2311.01234
[18] Davis, K., Lopez, A., & Tan, Y. (2024). Metrics for fairness in recruitment AI. *IEEE Transactions on Artificial Intelligence*, *5*(2), 300–315. https://doi.org/10.1109/TAI.2024.10345678
[19] Lopez, A., Garcia, M., & Chen, L. (2023). Transparency and explainability in HR AI. *Information Systems Frontiers*, *25*(4), 987–1002. https://doi.org/10.1007/s10796-023-10354-2
[20] Tan, Y., Brown, T., & Patel, R. (2025). Integrating ethical guidelines in hiring algorithms. *Proceedings of the ACM on Human-Computer Interaction*, *9*(CSCW), Article 3589052. https://doi.org/10.1145/3589052

Systematic Review of Artificial Intelligence in Recruitment: A Dynamic Framework for Fairness and Algorithmic Bias Mitigation

Mahdi Joneidi Jafari [1]
Kaveh Rezaei Shiraz [2]
Zahra Eskandarzadeh [3]

Abstract
With the growing use of artificial intelligence (AI) systems in recruitment processes, concerns about reproducing historical discrimination and violating fairness have increased. This article is a systematic review that aims to analyze strategies for ensuring fairness and reducing bias in intelligent recruitment by examining relevant and credible sources published after 2022. The focus is on comparing static and dynamic approaches to mitigating algorithmic bias. In the static approach, the model and data are corrected once, whereas the dynamic approach emphasizes continuous monitoring, periodic retraining of models, fairness auditing, adaptive sampling, and integration of human algorithm feedback. Findings indicate that dynamic models are up to 30% more effective in reducing the reproduction rate of discrimination compared to static models, particularly in complex and multi-stage environments such as automated video interviews, which pose the highest risk of exacerbating gender disparities (around 18%). The proposed dynamic framework includes components such as continuous monitoring of sensitive data, combined use of fairness metrics (such as equality of opportunity and demographic parity), and transparent reporting. This study concludes that achieving fairer recruitment requires a transition from static models to dynamic, accountable, and transparent systems capable of adapting to changes in data and operational contexts. Future research is recommended to focus on developing practical fairness metrics, testing frameworks on real industrial data, and aligning with legal requirements.

Keywords
Dynamic approach; Discrimination reduction; Artificial intelligence; Intelligent recruitment; Algorithmic bias

1. Assistant Professor, Faculty of Industrial Engineering and Management, Shahab Danesh University, Qom, Iran.
2. Master's Student in MBA- IT & IS, Kharazmi University, Tehran, Iran.
3. Master's Student in MBA- IT & IS, Kharazmi University, Tehran, Iran