

مطالعه حقوقی تبعیض الگوریتمی

باقر انصاری^{*۱}

تاریخ دریافت: ۱۴۰۰/۱۱/۱۵ تاریخ پذیرش: ۱۴۰۱/۰۱/۱۴

نوع مقاله: پژوهشی

DOI: 10.22034/qjplk.2022.1507.1408

چکیده

استفاده از الگوریتم‌ها به‌ویژه هوش مصنوعی برای اخذ تصمیمات مختلف، مسائل و پرسش‌های حقوقی متعددی را برانگیخته است. یکی از این مسائل، تبعیض‌آمیز بودن تصمیمات الگوریتمی است. این تبعیض ممکن است در خصوص یک فرد یا گروه یا طبقه‌ای از افراد جامعه صورت گیرد. با توجه به وظیفه دولت‌ها برای مقابله با انواع تبعیض، شناسایی مصادیق و علل و تبعیض‌های الگوریتمی و ارائه راهکارهای مقابله با آن‌ها به یکی از موضوعات حقوقی روز تبدیل شده است. این مقاله، به بررسی این پرسش پرداخته است که تبعیض الگوریتمی چیست، چه عللی دارد و چگونه باید با آن مقابله شود؟ بدین‌منظور، ضمن تعریف و دسته‌بندی الگوریتم‌ها و تصمیمات الگوریتمی، دو دسته علل (مرتبط با داده‌های ورودی و مرتبط با طراحی و عملکرد) را برای تبعیض‌های الگوریتمی شناسایی کرده و تجربیات کشورهای پیشرو در زمینه مقابله با چنین تبعیض‌هایی را نیز دسته‌بندی، تبیین و نقد کرده است. با توجه به نبود قواعد خاصی برای حمایت از اشخاص در برابر چنین تصمیماتی در کشور ما، استفاده از نقاط قوت مجموع راهکارهای معرفی‌شده در مقاله، برای رفع این خلأ قانونی پیشنهاد شده است.

واژگان کلیدی: تبعیض، تبعیض الگوریتمی، تصمیم‌گیری الگوریتمی، سوگیری، هوش مصنوعی.

۱. دانشیار گروه حقوق بشر و محیط زیست، دانشکده حقوق، دانشگاه شهید بهشتی، تهران، ایران.

* Email: b_ansari@sbu.ac.ir



مقدمه

روند فزاینده استفاده از سیستم‌های الگوریتمی از جمله هوش مصنوعی برای تصمیم‌گیری، چه در بخش عمومی و چه در بخش خصوصی، مسائل و چالش‌های حقوقی گوناگونی را درباره ابعاد مختلف این نوع تصمیم‌گیری مطرح کرده است. این الگوریتم‌ها در سطوح مختلف از وابستگی به انسان می‌توانند طراحی شوند و عمل کنند. برخی تحت نظارت انسان تصمیم می‌گیرند و بعضی دیگر، به صورت مستقل و خارج از نظارت.

همه سیستم‌های الگوریتمی، برای فعالیت خود به داده‌های زیادی از منابع گوناگون نیاز حیاتی دارند. در مورد سیستم‌های هوش مصنوعی، وجود حجم زیادی از داده‌های کافی و باکیفیت (مربوط، به‌روز، صحیح) در وهله اول برای یادگیری آن‌ها لازم است؛ بدون دسترسی به داده‌های با کیفیت بالا، الگوریتم‌ها نمی‌توانند خوب یاد بگیرند و قوی‌ترین تکنیک‌های هوش مصنوعی با بهترین سخت‌افزارها، نمی‌توانند تصمیمات درست بگیرند؛ چنانچه حجم داده‌ها کافی نباشد یا داده‌ها از تنوع کافی برخوردار نباشند یا کیفیت داده‌ها پایین باشد، نتایج و بروندهای سیستم‌های هوش مصنوعی، گمراه‌کننده یا خطا خواهند بود (European Commission Artificial Intelligence Act., 2021, para. 44). با فرض دسترسی سیستم‌های الگوریتمی به داده‌های کافی و باکیفیت، ممکن است مدل‌های پردازش داده به‌گونه‌ای باشد که به تصمیمات نادرست و تبعیض‌آمیز منتج شود. در واقع، سیستم‌های الگوریتمی از یک سو، به دلیل عمل کردن براساس قواعد منطقی و ریاضی، از احساسات و حب و بغض‌های انسانی عاری است و می‌توانند در مقایسه با انسان‌ها، رفتارهای بی‌طرفانه و مبتنی بر داده‌های عینی داشته باشند و در نتیجه، با رویه‌های تبعیض‌آمیز مقابله کنند. (Institut Montaigne, 2020: 17-18) از سوی دیگر، ممکن است به دلایل گوناگون از جمله وجود ایرادهای مختلف در داده‌های ورودی یا در طراحی و عملکرد درونی سیستم، نتایج تبعیض‌آمیزی را رقم بزنند.

با توجه به وظیفه همه دولت‌ها در زمینه گسترش اصل برابری و مبارزه با تبعیض، چه در بخش عمومی و چه در بخش خصوصی، نگرانی از شیوع تبعیض در نتیجه تصمیمات الگوریتمی - که به تصمیمات خودکار نیز موسوم‌اند- و پیدا کردن راهکارهای مقابله با آن، به دغدغه عمومی در کشورهای جهان تبدیل شده است.^۱ هم در سطح بین‌المللی نسبت به چنین تبعیض‌هایی هشدار داده شده است (Kaye, 2018, para. 36-38).

https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx و هم در سطوح ملی برخی کشورها، تصویب قوانین و مقررات خاص برای پیشگیری از تبعیض الگوریتمی در دستور کار قرار گرفته است. (The Law Library of Congress, 2019: 16-132). با وجود این، مفهوم و مصادیق این نوع تبعیض و روش‌های مقابله با آن در نظام‌های حقوقی مختلف چندان روشن نیست. از این رو پرسش آن است که مفهوم تبعیض الگوریتمی چیست و چگونه می‌توان با آن مقابله کرد؟

با توجه به نوپا بودن سیستم‌های الگوریتمی از جمله هوش مصنوعی در کشور ما و روند فزاینده استفاده از این سیستم‌ها در تصمیم‌گیری‌های بخش عمومی و خصوصی، لازم است مفهوم تبعیض الگوریتمی و علل آن تبیین شده و راهکارهای لازم برای پیشگیری از چنین تبعیض‌هایی اتخاذ شود. بدین منظور این مقاله، نخست تبعیض الگوریتمی و تصمیمات خودکار را تعریف می‌کند، سپس علل تبعیض‌آمیز شدن تصمیمات الگوریتمی را تبیین کرده و سرانجام، راهکارهای مقابله با تبعیض‌های الگوریتمی را معرفی می‌کند.

۱. مفهوم تبعیض الگوریتمی

برای فهم تبعیض الگوریتمی، نخست لازم است مفهوم الگوریتم و تصمیمات الگوریتمی روشن شود.

۱-۱. تعریف الگوریتم

تعاریف گوناگونی از الگوریتم وجود دارد. طبق تعریفی، «الگوریتم‌ها، مجموعه‌ای از دستورالعمل‌ها هستند که برای حل یک یا چند مشکل طراحی شده‌اند. الگوریتم‌ها، داده‌های ورودی را براساس محاسبات خاص به یک خروجی دلخواه تبدیل می‌کنند. این دستورالعمل‌ها، هم مشکل را مشخص می‌کنند و هم گام‌هایی را که با طی آن‌ها، آن مشکل باید حل شود. بنابراین «مجموعه‌ای از گام‌هایی هستند که برای حل یک مسئله خاص یا تحصیل یک نتیجه مشخص، برداشته می‌شوند» (Committee of experts on internet intermediaries, 2018: 5). در گزارش دولت فرانسه به کنگره این کشور، الگوریتم چنین تعریف شده است: «الگوریتم عبارت است از توصیف یک‌سری از مراحل (یا دستورالعمل‌ها) محدود و بدون ابهام که اجازه می‌دهد نتیجه‌ای از عناصر ارائه‌شده به‌عنوان ورودی،

به دست آید. این تعریف گسترده شامل تمام توالی عملیاتی است که پردازش اطلاعات را تشکیل می‌دهد» (Lucchesi, Chignard, 2019: 1).

با توجه به تعاریف بالا، می‌توان سیستم‌های الگوریتمی را به‌عنوان برنامه‌هایی تعریف کرد که اغلب با استفاده از تکنیک‌های بهینه‌سازی ریاضی، یک یا چند کار مانند جمع‌آوری، ترکیب، متمایز کردن، مرتب‌سازی، طبقه‌بندی و استنباط، انتخاب، اولویت‌بندی، پیش‌بینی، توصیه و تصمیم‌گیری انجام می‌دهند. (Guidelines on addressing the human rights impacts of algorithmic systems, 2020). سیستم‌های گوناگونی را می‌توان به‌عنوان مصادیق سیستم‌های الگوریتمی نام برد. سیستم‌های داده‌کاوی و تحلیل داده^۲، کلان‌داده^۳ و سیستم‌های هوش مصنوعی^۴ که از یادگیری ماشین^۵ و دیگر فناوری‌های هوشمند مانند سیستم‌های خبره^۶، زبان طبیعی^۷ و بینایی رایانه‌ای^۸ بهره می‌گیرند، همگی براساس الگوریتم‌ها فعالیت می‌کنند.

سیستم‌های الگوریتمی از نظر سطح دخالت در تصمیم‌گیری به دو دسته تقسیم می‌شوند:

قاعده‌بنیان^۹ یا ثابت و مبتنی بر یادگیری ماشین^{۱۰} یا پویا. (Gerards, Xenidis, 2021: 32-36)

الگوریتم‌های قاعده‌بنیان، الگوریتم‌های ساده‌اند که آیین‌های تصمیم‌گیری آن‌ها در قالب منطقی «اگر این، پس آن»^{۱۱} طراحی شده است. برای مثال، یک قاعده حقوقی ممکن است مقرر کند که «رانندگی با سرعت بالای ۱۰۰ کیلومتر با هر وسیله موتوری ممنوع است و متخلف به ۱۰۰ هزار تومان محکوم خواهد شد». قالب منطقی این قاعده که به‌صورت دستورالعمل رایانه‌ای قابل اجرا باشد، این است که «اگر شخصی با سرعت بیشتر از ۱۰۰ کیلومتر مشاهده شود، پس به پرداخت ۱۰۰ هزار تومان محکوم خواهد شد». الگوریتم‌های قاعده‌بنیان بسیار پیش‌بینی‌پذیرند، زیرا مجموعه دستورالعمل‌ها و قواعد آن‌ها ثابت است و همه متغیرها و نتایج احتمالی در الگوریتم برنامه‌ریزی شده‌اند.

الگوریتم‌های یادگیری ماشین پیچیده است و با توانایی «یادگیری» یعنی انطباق، تکامل و بهینه‌سازی خودکار، نتایج را براساس داده‌های ورودی و بدون برنامه‌ریزی صریح برای این کار، بهینه‌سازی و مشخص می‌کنند. برخلاف الگوریتم‌های قاعده‌بنیان که ثابت‌اند (چون قواعد آن‌ها فقط از طریق برنامه‌نویسی تغییرپذیر است)، الگوریتم‌های یادگیری ماشین پویا هستند، چون قواعد آن‌ها بسته به داده‌های ورودی تغییر می‌کند. این الگوریتم‌ها از ابزارها و تکنیک‌های مختلف تحلیلی استفاده می‌کنند که همگی بر یافتن همبستگی‌ها و الگوها در داده‌های با حجم

زیاد متمرکزند. این الگوریتم‌ها از نظر نوع یادگیری به چهار دسته تقسیم می‌شوند: یادگیری نظارت‌شده^{۱۲}، یادگیری بدون نظارت^{۱۳}، یادگیری تقویت‌شده^{۱۴} و یادگیری عمیق^{۱۵}.

الگوریتم‌های نظارت‌شده، نوعی هوش مصنوعی در سطح ضعیف‌اند که می‌توانند کارهای مشخص و به‌نسبت ساده‌ای مانند جست‌وجو در اینترنت یا راندن یک خودرو را انجام دهند. الگوریتم‌های یادگیری عمیق را می‌توان در دسته هوش مصنوعی قوی قرار داد که می‌توانند کارهای تشخیصی دقیق و پیچیده‌ای در سطح گسترده انجام دهند که تاکنون صرفاً از سوی انسان انجام‌شده‌اند بوده است. در هوش مصنوعی ضعیف، انسان نقش محوری در طراحی، توسعه و استفاده از فناوری دارد، اما در هوش مصنوعی قوی، هوش مصنوعی از استقلال نسبی در تشخیص، تصمیم و عمل برخوردار است. همچنین می‌توان گفت که هر سیستم هوش مصنوعی، سیستم الگوریتمی است، اما تنها نوعی از سیستم‌های الگوریتمی، هوش مصنوعی‌اند. به هر حال، هر دو، تصمیم‌گیری الگوریتمی دارند.

۱-۲. تعریف تصمیم الگوریتمی

منظور از تصمیمات الگوریتمی، تصمیماتی هستند که اولاً آثاری بر افراد دارند (اعم از آثار حقوقی یا تأثیرات غیرحقوقی) و ثانیاً با مداخله الگوریتم گرفته می‌شوند. الگوریتم‌ها در سطوح مختلف می‌توانند در مراحل و فرایندهای تصمیم‌گیری نقش داشته باشند. گاهی صرفاً به‌عنوان ابزار کمکی انسان و در چارچوب ازپیش‌تعریف‌شده و به‌نحو پیش‌بینی‌پذیر عمل می‌کنند و زمانی نقش محوری و تعیین‌کننده در اتخاذ تصمیم نهایی توسط انسان‌ها دارند یا خودشان به‌صورت مستقل، تصمیم نهایی را می‌گیرند. در همه مواردی که استنباط‌ها، پیش‌بینی‌ها و توصیه‌های الگوریتمی تأثیر مهم بر تصمیم‌گیری دارند، برای مثال چنانچه برای تصمیم‌گیری در مورد خاصی سه گزینه پیش رو باشد، گزینه‌ای انتخاب شود که الگوریتم آن را توصیه کرده است یا یکی از آن گزینه‌ها کاملاً خودکار توسط هوش مصنوعی انتخاب شود، با تصمیمات الگوریتمی روبه‌رو هستیم (Orwat, 2020: 21-22).

تصمیمات الگوریتمی هم در بخش عمومی وجود دارند و هم در بخش خصوصی. برای مثال، تصمیم درباره پذیرش یا رد درخواست استخدام، پذیرش در مدرسه یا دانشگاه، اعطا یا عدم اعطای آزادی مشروط به یک زندانی، جرمه به‌سبب یک تخلف، صدور یا عدم صدور مجوز ورود به کشور یا خروج از کشور، از مصادیق تصمیماتی هستند که در بخش عمومی ممکن است با کمک

الگوریتم‌ها یا توسط الگوریتم‌ها گرفته شوند. در بخش خصوصی نیز دامنه گوناگونی از تصمیمات متصور است، مانند به‌کارگیری نیروی انسانی کارآمد، تعیین قیمت برخی محصولات یا خدمات خاص، کشف تقلب در قراردادها، بهینه‌سازی محاسبات خطر در صنعت بیمه، ارسال تبلیغات شخصی‌شده، تشخیص اخبار جعلی یا سخنان تفرآمیز در رسانه‌های اجتماعی و تشخیص برخی بیماری‌ها (Panel for the Future of Science and Technology, 2019: 19-25).

۱-۳. تعریف تبعیض الگوریتمی

تمایز و تفاوت قائل شدن بین افراد متفاوت براساس ویژگی‌هایی مشخص، از کارویژه‌های اصلی همه سیستم‌های الگوریتمی است (Xenidis, Senden, 2020: 5). از این‌رو الگوریتم‌ها در پایین‌ترین سطح، به‌نوعی کارگزينش و غربالگری^{۱۶} انجام می‌دهند و از میان افراد و گزینه‌های بسیار زیاد، یک یا تعداد محدودی را انتخاب می‌کنند و در سطح بالاتر، براساس مدل ناشی از یادگیری، تصمیم‌گیری می‌کنند (Kleinberg et al., 2018: 215-216). نکته اصلی در این خصوص آن است که باید در مورد افراد و گروه‌ها (برای مثالذنان و مردان، سفیدپوستان و سیاه‌پوستان، کم‌درآمدها و ثروتمندان)، تصمیم یکسان و در مورد افراد متفاوت، تصمیم متفاوت داشته باشند، در غیر این صورت، تبعیض واقع می‌شود.

ممنوعیت تبعیض در مورد افراد بر مبنای عواملی همچون رنگ پوست، نژاد و خاستگاه اجتماعی، در معاهدات حقوق بشری آمده است. ولی تبعیض الگوریتمی ممکن است براساس عواملی مانند وضعیت اقتصادی صورت گیرد که در فهرست علل مذکور نیست. از این‌رو زمینه‌های تبعیض الگوریتمی گسترده‌تر است. اشاره به برخی از نمونه‌های واقعی از چنین تصمیماتی می‌تواند مفهوم و اهمیت توجه به تبعیض الگوریتمی را بهتر نشان دهد:

مثال اول: پژوهشی در سال ۲۰۱۵ نشان داد که بسیاری از الگوریتم‌های مورد استفاده از سوی شرکت‌های بیمه در ایالات متحده برای تعیین مبلغ سهم بیمه خودرو، بیش از اینکه به سوابق راننده در امر رانندگی توجه کنند، به ارزیابی وضعیت اعتبارسنجی (مالی) اتکا دارند. در نتیجه، برای مثال در فلوریدا، فردی با سابقه رانندگی خوب و بدون نمره منفی اما با نمره اعتبار مالی ضعیف، در مقایسه با راننده‌ای با سابقه محکومیت رانندگی در حال مستی ولی دارای نمره اعتبار مالی خوب، باید مبلغ بیشتری برای بیمه خودروی خودش بپردازد (Rovatsos et al., 2019: 19).

مثال دوم، مطالعه‌ای در مورد دقت سه سیستم خودکار تشخیص چهره، اختلاف‌های شایان توجهی را در عملکرد آنها آشکار کرد و نشان داد که داده‌های یاددهنده سیستم‌های مورد مطالعه به‌طور گسترده‌ای از افراد دارای پوست روشن‌تر تشکیل شده است. در نتیجه، تشخیص چهره زنان دارای پوست تیره بیش از ۳۰ درصد خطا دارد، درحالی‌که میزان خطا برای مردان دارای پوست روشن‌تر ۰/۸ درصد است.

(International Working Group on Data Protection in Telecommunications, 2018 para. 28).

مثال سوم: پژوهشگران دانشگاه ام‌آی‌تی متوجه شدند که سه مورد از جدیدترین برنامه‌های تشخیص جنسیت که توسط شرکت‌های آی‌بی‌ام، مایکروسافت و مگی‌وی‌آی^{۱۷} طراحی شده‌اند، می‌توانند در ۹۹ درصد موارد جنسیت فرد را از روی عکس‌ها به‌نحو صحیح تشخیص دهند، اما فقط برای مردان سفیدپوست. در مورد زنان غیرسفیدپوست (سیاه، آسیایی و اقلیت‌های دیگر)، صحت نتایج ۳۵ درصد است (Revell, 2018).

همچنین می‌توان به برخی ارزیابی‌های میدانی اشاره کرد که نشان می‌دهند الگوریتم‌ها چگونه در کشورهای مختلف منشأ انواع تصمیمات تبعیض‌آمیز شده‌اند. برای مثال در فرانسه، در خریدهای آنلاین، تعیین قیمت شخصی‌شده^{۱۸} برای کالاها و خدمات به تبعیض علیه زنان به‌دلیل افزایش قیمت محصولات خاص برای آنان منجر شده است. در ایتالیا نیز تبلیغات رفتاری شخصی‌شده^{۱۹} در بازار املاک و مستغلات منشأ تبعیض براساس نژاد و طبقه اجتماعی شده و به ایجاد محله‌های متمایز فقیرنشین و مرفه‌نشین دامن زده است. الگوریتم‌های کاریابی، در بلژیک در پیدا کردن فرصت‌های کاری برای افراد زن و مرد و در کرواسی در پیدا کردن فرصت‌های کاری برای زنان دارای دو بچه کمتر و بیشتر، تبعیض قائل شده‌اند (Gerards, Xenidis, 2021: 86).

۲. علل تبعیض الگوریتمی

به‌طور کلی، چند عامل می‌تواند به تبعیض و سوگیری در تصمیم‌گیری‌های الگوریتمی منجر شود. این عوامل را در دسته‌بندی کلی می‌توان به دو دسته تقسیم کرد: علل مرتبط با طراحی و عملکرد درونی این سیستم‌ها و عوامل مرتبط با ورودی داده‌ها به سیستم‌های الگوریتمی.

۲-۱. علل مرتبط با طراحی و عملکرد درونی این سیستم‌ها

سوگیری‌های شناختی از علل مرتبط با طراحی درونی این سیستم‌هاست، توضیح آنکه انسان‌ها نقشی اساسی و حیاتی در برنامه‌نویسی، آموزش و استفاده از الگوریتم‌ها دارند. دخالت

انسان‌ها در مورد تصمیم‌گیری الگوریتمی، به‌ویژه از منظر برابری و عدم تبعیض، خطرهای خاصی را به‌همراه دارد. بدیهی است که استدلال انسانی همواره در معرض انواع پیش‌فرض‌ها، سوگیری‌ها، اشتباهات منطقی و چه‌بسا مغالطه‌هاست و همین مسئله ممکن است در برنامه‌نویسی الگوریتم‌ها تأثیر بگذارد (Gerards & Xenidis, 2021: 41). برنامه‌نویسان ممکن است افرادی با نگاه‌های تبعیض‌آمیز و پایبند به برخی کلیشه‌های اجتماعی باشند و در نتیجه، سوگیری و رفتار تبعیض‌آمیز را در سیستم الگوریتمی نیز جاسازی و منعکس کنند (Zuiderveen & Borgesius, 2018: 13). اینان ممکن است خواسته یا ناخواسته انواع سوگیری‌ها داشته باشند: سوگیری‌های شناختی^{۲۰} می‌توانند به تحریف روش پردازش اطلاعات منجر شوند، برای مثال سوگیری‌های پیش‌بینی و تأیید^{۲۱} می‌توانند برنامه‌نویسان را به سمت طرفداری از دیدگاه‌های مورد قبول خودشان در مورد پدیده‌های جهان سوق دهند، حتی اگر داده‌های موجود این دیدگاه‌ها را رد کرده باشند. سوگیری همبستگی‌های واهی^{۲۲} برنامه‌نویس را به سمت شناسایی همبستگی بین دو متغیری سوق می‌دهد که در واقع، ارتباطی با یکدیگر ندارند. سوگیری کلیشه^{۲۳} موجب می‌شود که فردی به‌جای تکیه بر توانایی‌ها و صلاحیت شخصی خود، با ارجاع خود به یک گروه اجتماعی که خودش را با آن گروه معرفی می‌کند (و با آن گروه به خودش هویت می‌دهد)، عمل کند. همه این سوگیری‌ها می‌توانند سبب انتخاب متغیرهایی شوند که برداشت خاصی از پدیده‌ها را ترجمه کرده‌اند الگوریتم را به سمت تصمیمات خطا و تبعیض‌آمیز راهنمایی می‌کنند (Bertail et al., 2019: 10).

نحوه تعریف «متغیر هدف»^{۲۴} و «برچسب طبقه»^{۲۵} از دیگر عواملی مرتبط با طراحی سیستم الگوریتمی است که می‌تواند به بروز نتایج تبعیض‌آمیز منجر شود. توضیح آنکه، تحلیل مسئله و برنامه‌ریزی، نخستین مرحله در استفاده از الگوریتم‌هاست. پیش از هر کاری، باید هدف (متغیر) خاصی برای استفاده از الگوریتم توسط یک شرکت یا نهاد عمومی تعیین شود. به‌کارگیری افراد کارآمدتر، کشف و تعقیب جرائم و تخلفات، تعیین قیمت محصولات یا خدمات خاص، کشف وضعیت عرضه و تقاضا در بازار، بهینه‌سازی محاسبات خطر، کشف تقلب در بیمه‌های اجتماعی یا تجاری، تشخیص اخبار جعلی یا سخنان تفرآمیز و تشخیص یک بیماری، از جمله اهداف شناخته‌شده در این زمینه است (Gerards & Xenidis, 2021: 38).

پس از مشخص شدن هدف، باید متناسب با آن، بهترین نوع الگوریتم برای دستیابی به آن استفاده شود. باید مشخص شود که آیا الگوریتم موردنظر را می‌توان در جریان کارها و فرایندهای پیشین قرار داد یا نه و چگونه، همین‌طور آیا الگوریتم موردنظر می‌تواند یا باید به الگوریتم‌های دیگر یا فرایندهای خودکار متصل شود یا نه و چگونه. یکی از عواملی که باید در این فرایند تصمیم‌گیری مورد توجه قرار گیرد، این است که از خروجی الگوریتم چگونه استفاده خواهد شد. بر این اساس، افزایش کارایی در تصمیم‌گیری خودکار، گاهی ممکن است مستلزم استفاده از الگوریتم‌های قاعده‌بنیان باشد، درحالی‌که اگر هدف موردنظر نیازمند پیش‌بینی رفتار انسان یا پروفایل‌سازی باشد، الگوریتم‌های خودآموز می‌توانند مفیدتر باشند. هنگام انتخاب نیز در نظر گرفتن ویژگی‌های خاص هر الگوریتم مهم است. ویژگی‌های قاعده‌بنیان به‌طور معمول بسیار پیش‌بینی‌پذیرند، بدین ترتیب که تمامی شاخص‌ها، متغیرها و گزینه‌های مربوط می‌توانند به‌عنوان بخشی از فرایند توسعه، از پیش تعیین شوند. پس از آماده شدن الگوریتم، همان نتایجی رقم خواهد خورد که پیش‌بینی شده است. با وجود این، باید هوشیار بود که این نوع الگوریتم‌ها نسبتاً خشک و غیرمنعطف‌اند و نمی‌توانند به‌طور مستقل، شرایط و احوال متغیر را در نظر بگیرند و تصمیمات و پیشنهاد‌های خود را بهینه سازند. برای مثال، پیشنهادها یا توصیه‌های جدیدی درباره قیمت قابل قبول یا جریمه مناسب ارائه دهند. این الگوریتم‌ها، برای اهداف جدید، باید دوباره برنامه‌ریزی شوند.

به همین ترتیب، الگوریتم‌های یادگیری نظارت‌شده، آموزش می‌بینند که به‌طور معمول یک وضعیت را در یک لحظه خاص منعکس کنند. در نتیجه، اگر عوامل زمینه‌ای تغییر کنند (برای مثال تغییر نظرها و ترجیحات فردی یا رواج استفاده از اصطلاحات و اصطلاحات خاص در رسانه‌های اجتماعی)، ممکن است کارایی خود را از دست بدهند و خیلی سریع، منسوخ شوند. بنابراین، چنین الگوریتم‌های باید به‌طور پیوسته به‌روز شده و اعتبارسنجی شوند و ممکن است برای کار در زمینه‌های بسیار پویا، کمتر مناسب باشد. برعکس، برای وضعیت‌های متغیر، استفاده از الگوریتم‌های پویا که براساس یادگیری عمیق عمل می‌کنند، مناسب‌تر است. گاهی نیز ممکن است ترکیبی از انواع الگوریتم‌ها برای تصمیم‌گیری بهتر مورد نیاز باشد (Gerards, Xenidis, 2021: 38-39).

در هر حال، الگوریتم‌های یادگیری ماشین برای اتخاذ تصمیمات خاص تعلیم می‌بینند و براساس مدلی که در نتیجه یادگیری می‌سازند، همبستگی را در مجموعه داده‌ها پیدا می‌کنند. برای مثال شرکتی می‌خواهد یک سیستم هوش مصنوعی را برای یافتن کارمندان «خوب» به کار گیرد؛ چگونه یک کارمند خوب تعریف می‌شود؟ به عبارت دیگر، برچسب‌های طبقه چه باید باشد؟ آیا کارمندی خوب است که محصولات بیشتری را می‌فروشد؟ یا کسی که هرگز در کار تأخیر ندارد؟ آیا کارمندی خوب است که بیشترین ساعت حضور در محل کار را دارد؟ برخی از متغیرهای هدف و برچسب‌های طبقه ممکن است تأثیرات کم یا بیش نامطلوبی بر طبقات یا گروه‌های خاصی داشته باشند. برای مثال فرض کنید افراد فقیر به‌ندرت در مرکز شهر زندگی می‌کنند و باید بیشتر از سایر کارکنان برای به رسیدن به محل کار خود سفر کنند. بنابراین، افراد فقیر به دلیل مشکلات ترافیکی یا مشکلات موجود در وسایل حمل‌ونقل عمومی دیرتر از سایرین در محل کار حاضر می‌شوند. این شرکت می‌تواند برچسب «به‌ندرت برای کارش تأخیر می‌کند» را برای انتخاب کارمند خوب به کار ببرد. حال در مورد افراد دارای سابقه مهاجرت که به‌طور معمول، فقیرند و دورتر از محل کارشان زندگی می‌کنند (حاشیه‌نشین هستند)، این انتخاب برچسب طبقه سبب می‌شود افراد مذکور (دارای سابقه مهاجرت) حتی اگر در سایر جنبه‌ها از دیگر کارکنان بهتر باشند، به‌عنوان کارمند خوب شناسایی نشوند و در معرض تبعیض و آسیب قرار گیرند. به‌طور کلی، حسب اینکه یک سازمان چگونه متغیرهای هدف و برچسب‌های طبقه را تعریف کند، تبعیض می‌تواند از یک سیستم هوش مصنوعی ناشی شود (Zuiderveen Borgesius, 2018: 11).

۲-۲. عوامل مرتبط با ورودی داده‌ها به سیستم‌های الگوریتمی

سیستم‌های الگوریتمی از فرایندهای پیچیده‌ای استفاده می‌کنند که این فرایندها به داده‌های ورودی نیاز دارند. به همین دلیل، تصمیم برای استفاده از داده‌های خاص و کنار گذاشتن سایر داده‌ها نیز می‌تواند به خروجی‌های تبعیض‌آمیز منجر شود. گاهی داده‌های مورد استفاده به دلیل ناکافی یا ناقص بودن نمی‌توانند نمایندگی همه اعضای یک طبقه یا گروه را داشته باشند و زمانی که در ترکیب با سایر عوامل برای نتیجه‌گیری کلی به کار می‌روند، به نتایج تبعیض‌آمیز منجر می‌شوند؛ یعنی به تصمیماتی منجر می‌شوند که درباره گروهی از افراد گرفته می‌شود،

بدون اینکه تفاوت‌های افراد آن گروه در نظر گرفته شود. در نتیجه، با افرادی که مستحق یک تصمیم هستند و افرادی که مستحق آن نیستند، یکسان رفتار می‌شود.

دلایل مختلفی ممکن است برای ناکافی بودن داده‌های جمع‌آوری شده به منظور یادگیری هوش مصنوعی وجود داشته باشد. یک دلیل این است که افراد داخل در یک گروه یا طبقه، ممکن است آگاهانه یا ناخواسته در مجموعه داده‌ها حضور نداشته باشند یا حضور کمتری داشته باشند. برای مثال برخی افراد ممکن است بسیار مراقب باشند که چه اطلاعاتی را در مورد خودشان در فضای عمومی منتشر می‌کنند و احتیاط آن‌ها در عدم افشای داده‌های شخصی، به نبود داده‌های آن‌ها در مجموعه داده‌های مورد استفاده سیستم‌های هوش مصنوعی منجر شود (International Working Group on Data Protection in Telecommunications, 2018, para. 30). در چنین مواردی، اگر داده‌های یاددهنده از طریق جست‌وجو در منابع داده‌های موجود تأمین شده باشد، به سبب در دسترس نبودن داده‌های این افراد در منابع مذکور، ویژگی‌ها و تفاوت‌های خاص آن‌ها در یادگیری لحاظ نخواهد شد، درحالی‌که تصمیماتی که براساس داده‌های موجود گرفته می‌شود، ممکن است به آن‌ها نیز تعمیم یابد.

دلیل دیگر این است که افرادی ممکن است به فناوری‌هایی که داده‌های مربوط به فعالیت‌ها و رفتارهایشان را تولید می‌کند و در دسترس سیستم‌های هوش مصنوعی قرار می‌دهد، دسترسی یا تسلط نداشته باشند. در نتیجه، داده‌های مربوط به آن‌ها در تصمیم‌گیری‌ها لحاظ نمی‌شود. برای مثال افراد فقیر ممکن است در یک مجموعه داده‌ای که در صورت استفاده از گوشی تلفن همراه آیفون جمع‌آوری می‌شود، حضور کمتری داشته باشند. استفاده از برنامه کاربردی Street Bump در شهر بوستون آمریکا چنین چالشی را آشکار کرده است. این برنامه، یک برنامه کاربردی در بستر تلفن هوشمند است که از ویژگی‌هایی مانند جی‌پی‌اس برای گزارش شرایط و مشکلات جاده‌ها به مراجع ذی‌ربط از جمله شهرداری‌ها استفاده می‌کند. چون تعداد کاربران تلفن‌های هوشمند در بین افراد فقیر کمتر از افراد ثروتمند است، مشکلات راه‌های مورد استفاده افراد فقیر کمتر گزارش می‌شوند و در نتیجه، بودجه کمتری نیز برای رفع آن مشکلات اختصاص می‌یابد. با کشف این چالش، مقامات اعلام می‌کنند که درصدد اصلاح این اشکالات در جمع‌آوری داده‌ها هستند (Zuiderveen Borgesius, 2018: 12).

همچنین افرادی ممکن است به دلایلی از نظر داده‌ای کمتر مورد توجه جمع‌آوری‌کننده داده‌ها قرار بگیرند (برای مثال ممکن است در یک طبقه اقتصادی خاص نباشند) و به عدم درج اطلاعات آن‌ها در مجموعه داده‌های مربوط منجر شود. به هر حال، چنانچه مجموعه داده‌های ورودی به یک مدل، نماینده کل یک جمعیت نباشد، به نتایجی منجر می‌شود که گروه‌های خاصی را نسبت به گروه‌های دیگر ترجیح می‌دهد. برای مثال در یک برنامه کاربردی مربوط به تاکسی اینترنتی (مانند اوبر و اسنپ) اگر داده‌های سرعت فقط از افرادی که تلفن‌های هوشمند دارند جمع‌آوری شده باشد، دقت نتایج سیستم به احتمال زیاد برای جمعیت ثروتمند با تراکم بالای تلفن‌های هوشمند، بیشتر و در مناطق فقیرنشین که تراکم تلفن‌های هوشمند پایین است، کمتر خواهد بود. به هر حال، میزان در دسترس بودن داده‌های جمعیت موردنظر، دسترسی آن‌ها به فناوری و مشارکت آن‌ها در اکوسیستم دیجیتال، به دلایل گوناگون از جمله به سبب موانع اقتصادی، اجتماعی، ساختاری و زبانی، به طور چشمگیری متفاوت است و این عوامل، حضور آن‌ها در مجموعه داده‌های مورد استفاده در تصمیمات الگوریتمی را قویاً تحت تأثیر قرار می‌دهد. این نکته اگر مورد توجه کافی قرار نگیرد، می‌تواند داده‌های برخی از گروه‌های جمعیت را بیشتر از حد متناسب نشان دهد و داده‌های برخی گروه‌ها را کمتر از حد تناسب و در نتیجه، الگوهای تبعیض‌آمیز موجود را تقویت کند (Kochling & Claus Wehner, 2020: 800-801).

گاهی نیز ناکافی بودن داده‌های یاددهنده، ناشی از انتخاب روش نمونه‌گیری و استفاده از داده‌های موجود است. برای مثال هنگام جمع‌آوری داده‌های جرائم، اگر داده‌ها از یکی از واحدهای پلیس گرفته شود که در گذشته افراد بیشتری را به سبب سابقه مهاجرت بازداشت کرده است، نمونه‌ها بی‌شک تبعیض‌آمیز خواهند بود. اگر پلیس توجه خود را به گروه‌های قومی خاص و محله‌های خاص معطوف کرده باشد، به احتمال زیاد، سوابق پلیس به طور نظام‌مند، داده‌های آن گروه‌ها و محله‌ها را بیش‌ازحد نشان خواهد داد و سیستمی که با چنین داده‌هایی آموزش ببیند، یاد خواهد گرفت که مثلاً افراد دارای سابقه مهاجرت، بیشتر مرتکب جرم می‌شوند. در واقع، اگر از داده‌های مغرضانه برای آموزش مدل‌های پیش‌بینی استفاده شود، مدل‌ها همان سوگیری‌ها را بازتولید خواهند کرد. فرض کنید پلیس به محله‌ای که مهاجران زیادی دارد، حساسیت بیشتری نشان می‌دهد، درحالی‌که این محله از نظر آمار ارتکاب جرم در سطح متوسط قرار دارد. بدیهی است که پلیس در آن محله جرائم بیشتری را نسبت به

محلله‌های دیگر ثبت می‌کند. از آنجا که این اعداد نشان می‌دهد جرم بیشتری در آن محله واقع شده است و ممکن است نتیجه‌گیری شود که پس جرم بیشتری در آن محله اتفاق می‌افتد و آن محله جرم‌خیز است؛ حتی ممکن است تعداد نیروهای پلیس در آنجا افزایش پیدا کند (Zuiderveen Borgesius, 2018: 11).

گاهی نیز تبعیض‌آمیز بودن داده‌های ورودی به نتایج تبعیض‌آمیز منجر می‌شود. در واقع، در صورت استفاده از داده‌های یاددهنده تبعیض‌آمیز برای یادگیری ماشین، تصمیم‌گیری هوش مصنوعی نیز می‌تواند نتایج تبعیض‌آمیز نسبت به افراد یا گروه‌های مختلف در پی داشته باشد. برخی از داده‌های یاددهنده ممکن است حاوی سوگیری ذاتی باشند. این داده‌ها، گاهی از این حیث سوگیری دارند که از تصمیمات تبعیض‌آمیز انسان‌ها ناشی شده‌اند. در چنین وضعیت‌هایی، حتی اگر تلاش‌های آگاهانه برای پیشگیری از هرگونه سوگیری در انتخاب داده‌ها انجام گیرد، داده‌ها منعکس‌کننده انواع سوگیری‌هایی خواهند بود که در بستر واقعیت‌های اجتماعی به صورت رسمی و غیررسمی وجود دارند. در این جوامع یا در زمینه‌های این‌چنینی، سیستم‌های الگوریتمی به نحو اجتناب‌ناپذیری نتایج تبعیض‌آمیز را بازتولید خواهند کرد (Cofone, 2019: 1404-1406; International Working Group on Data Protection in Telecommunications, 2018, para. 32). برای مثال چنین تبعیضی در دهه ۱۹۸۰ در یک دانشکده پزشکی در انگلیس واقع شد. این دانشکده با تعداد زیادی متقاضی رشته‌های پزشکی مواجه شد که ظرفیت پذیرش همه آن‌ها را نداشت؛ بنابراین یک برنامه رایانه‌ای برای کمک به مرتب‌سازی تقاضاها ایجاد کرد. داده‌های یاددهنده برنامه رایانه‌ای، پرونده‌های پذیرش متقاضیان در سال‌های قبل بود که متقاضیان توسط افراد انتخاب می‌شدند. داده‌های یاددهنده، به برنامه رایانه‌ای یاد داد که کدام ویژگی (داده ورودی) با خروجی مطلوب (پذیرفته‌شده در دانشکده پزشکی) ارتباط دارد و رایانه آن سیستم انتخاب را بازتولید کرد. بعدها معلوم شد که این برنامه رایانه‌ای، علیه زنان و افرادی که سابقه مهاجرت دارند، تبعیض قائل شده است. ظاهراً در سال‌هایی که داده‌های یاددهنده از آن سال‌ها گرفته شده بود، افرادی که دانشجویان را انتخاب می‌کردند، نسبت به زنان و افرادی که سابقه مهاجرت داشتند، سوگیری داشته‌اند. همان‌طور که مجله پزشکی انگلیس متذکر شد، «این برنامه حاوی سوگیری جدیدی نبود، بلکه صرفاً منعکس‌کننده موارد موجود در سیستم بود» (Zuiderveen Borgesius, 2018: 11).

همچنین تحقیقات نشان می‌دهد هنگامی که شرکت‌ها بر «استخدام افراد متناسب با فرهنگ شرکت» در روال‌های شغلی خود تأکید می‌کنند، اگر فرهنگ فعلی آن‌ها اغلب براساس یک مجموعه خاص و محدودی از تجربیات شکل گرفته باشد، ناخواسته می‌توانند الگوی نادرست استخدام را تداوم بخشند. برای مثال در محل کاری که در درجه اول مردان جوان سفیدپوست مشغول کارند، یک سیستم الگوریتمی که اساساً برای استخدام متناسب با فرهنگ آن محل طراحی شده است (بدون در نظر گرفتن سایر اهداف استخدام، مانند تنوع تجربه و چشم‌انداز)، ممکن است برای سازگاری با فرهنگ موجود، استخدام بیشتر مردان جوان سفیدپوست را توصیه کند (Executive Office of the President, 2016: 8-10).

۳. راهکارهای مقابله با تبعیض الگوریتمی

به‌عنوان قاعده کلی، دولت‌ها موظف‌اند ضمن ممنوع کردن تبعیض، از یک‌سو، خودشان مرتکب رفتارهای تبعیض‌آمیز نشوند و از سوی دیگر، مانع از چنین رفتارهایی از سوی بخش خصوصی شوند. تبعیض‌های الگوریتمی مصادیق جدید و در عین حال پیچیده از رفتارهای تبعیض‌آمیزند که دولت‌ها باید ضمن شناسایی مصادیق جدید و علل آن‌ها، راهکارهای قانونی برای مقابله با آن‌ها را نیز پیش‌بینی کنند. تحقق این هدف، نیازمند ابتکارهای قانونی، اداری و قضایی است. از این‌رو اگر قوانینی برای مقابله با تبعیض وجود دارد، آن‌ها را باید در حدی که منطبق حقوقی و اصول تفسیر اجازه می‌دهد، به‌نحوی تفسیر کرد که شامل تبعیض‌های الگوریتمی نیز شوند و اگر احکام قانونی خاصی در این باره ندارند، باید خلأهای قانونی مربوط به مقابله با این نوع تبعیض را با تصویب احکام قانونی متناسب برطرف کرد.

در نظام حقوقی ایران، قوانین و مقررات مختلفی در ممنوعیت تبعیض وجود دارد. قانون اساسی، در اصول متعددی به این ممنوعیت تصریح دارد. اصل ۳ (بند ۹) قانون اساسی، «رفع تبعیضات ناروا و ایجاد امکانات عادلانه برای همه در تمام زمینه‌های مادی و معنوی» را از وظایف دولت برشمرده است. همچنین در اصل ۱۹ مساوات در برخورداری از حقوق و در اصل ۲۰ نیز مساوات همه مردم در برخورداری از حمایت قانون را پیش‌بینی کرده است.

بند ۲ ماده ۸ قانون رسیدگی به تخلفات اداری (مصوب ۱۳۷۲) «تبعیض یا اعمال غرض در اجرای قوانین و مقررات نسبت به اشخاص» را از مصادیق تخلفات اداری دانسته است. همچنین ماده ۵ مصوبه شورای عالی اداری با عنوان حقوق شهروندی در نظام اداری (مصوب

۱۳۹۵)، موارد زیر را در مقام بیان مصادیق حق مصون بودن از تبعیض در نظام‌ها، فرایندها و تصمیمات اداری مقرر کرده است:

۱. دستگاه‌های اجرایی باید فرایند و رویه مشخص و اعلام‌شده‌ای برای ارائه خدمات خود داشته باشند و آن را به‌طور یکسان نسبت به همه مراجعان رعایت کنند؛

۲. مدیران و کارکنان دستگاه‌های اجرایی باید تصمیمات و اقدام‌های خود را مستند به قوانین و مقررات مربوط انجام دهند و از هرگونه تبعیض یا اعمال سلیقه در اجرای قوانین و مقررات اجتناب ورزند؛

۳. کارکنان دستگاه‌های اجرایی در همه سطوح باید در اعمال صلاحیت‌ها و اختیارات اداری خود مانند احراز صلاحیت‌ها، جذب نیرو، صدور مجوزها و نظایر آن بدون تبعیض عمل کنند.

با وجود این ممنوعیت‌های قانونی و با اینکه استفاده از سیستم‌های الگوریتمی در حال شیوع است، مصوبه یا تدبیر خاصی در زمینه چگونگی مقابله تبعیض‌های الگوریتمی مشاهده نمی‌شود. قانون تجارت الکترونیک (مصوب ۱۳۸۲) نیز که تنها قانون ناظر بر حمایت از داده‌های شخصی در کشور ماست، برای این منظور کارایی ندارد. در چنین وضعیتی، قوانینی که از مطلق تبعیض منع کرده‌اند، می‌توانند در برابر مصوبات و اقدام‌هایی که در بخش عمومی منشأ تبعیض‌های الگوریتمی‌اند، قابلیت استناد داشته باشند، اما ضرورت توجه خاص به این موضوع را منتفی نمی‌کنند، زیرا اولاً تبعیض‌های الگوریتمی در بخش خصوصی از قلمرو این نظارت‌ها خارج می‌مانند و ثانیاً پیشگیری از اقدام‌ها و رویه‌های منشأ تبعیض‌های مذکور چه در بخش عمومی و چه در بخش خصوصی، مستلزم مقررات خاصی است که باید از سوی نهادهایی مانند شورای عالی فضای مجازی یا مجلس شورای اسلامی تصویب شود. در این میان، راهکارهای سایر کشورها نیز می‌تواند قابل استفاده باشد.

راهکارهای خاصی را که برخی کشورها برای مقابله با تبعیض‌های الگوریتمی به‌کار برده‌اند، می‌توان به چند دسته تقسیم کرد: ممنوعیت کلی استفاده از تصمیمات الگوریتمی خودکار با استثنای موردی (مانند آلمان)، تمایز بین انواع تصمیم‌گیری الگوریتمی خودکار (مانند فرانسه)، مشروط کردن هر نوع تصمیم الگوریتمی خودکار به رعایت ضوابط خاص (مانند بریتانیا)، شفاف‌سازی چگونگی اتخاذ تصمیمات مذکور برای عموم جامعه (مانند کانادا)

و آمریکا). این دسته‌بندی را با معیار دیگری نیز می‌توان صورت داد؛ از جمله اینکه می‌توان به راهکارهای مبتنی بر قوانین حمایت از داده‌های شخصی (در مورد کشورهای اروپایی) و راهکار تصویب قوانین و مقررات خاص (کانادا و آمریکا)، دسته‌بندی کرد.

۳-۱. منع کلی تصمیمات الگوریتمی خودکار با استثنای بخشی

به موجب ماده ۲۲ مقررات عمومی اتحادیه اروپا (۲۰۱۶)، شخص موضوع داده باید این حق را داشته باشد که موضوع تصمیم‌گیری‌های صرفاً مبتنی بر پردازش خودکار، از جمله پروفایل‌سازی، که آثار حقوقی برای وی ایجاد می‌کند یا به‌نحو مشابهی به‌طور چشمگیری وی را تحت تأثیر قرار می‌دهد^{۲۶}، قرار نگیرد. استثنای این قاعده کلی نیز در سه مورد و در ماده مذکور آمده است:

۱. تصمیم برای انعقاد یا اجرای قرارداد بین شخص موضوع داده و کنترلگر داده ضروری باشد؛
 ۲. به موجب قانون اتحادیه یا کشور عضوی که کنترلگر تابع آن است و تدابیر مناسبی را برای حفاظت از حقوق و آزادی‌ها و منافع مشروع شخص موضوع داده مقرر کرده مجاز باشد؛
 ۳. براساس رضایت صریح موضوع داده باشد.
- تصمیمات خودکار تنها در صورتی می‌تواند براساس داده‌های حساس باشد که شخص موضوع داده به‌صراحت برای پردازش این داده‌ها رضایت داده باشد یا پردازش به دلایل منافع عمومی قابل ملاحظه، ضروری باشد و تدابیر مناسب برای حفاظت از حقوق و آزادی‌های شخص موضوع داده و منافع مشروع اتخاذ شده باشند.

در آلمان https://www.sciencedirect.com/science/article/pii/S0267364918303753-cit_60

به‌موجب ماده ۳۸ قانون فدرال حمایت از داده‌ها^{۲۷} علاوه بر استثنای مقرر در ماده مذکور، حق افراد در خصوص اینکه موضوع تصمیمات صرفاً مبتنی بر پردازش خودکار قرار نگیرند، در زمینه ارائه خدمات به موجب قرارداد بیمه اعمال نمی‌شود. البته این مسئله فقط در دو صورت مجاز است: در صورتی که شخص موضوع داده از چنین پردازشی پاسخ مثبت دریافت کند (۱) و اگر پاسخ منفی باشد (برای مثال عدم ارائه خدمات)، تصمیم خودکار فقط با رعایت موارد زیر مجاز خواهد بود (۲):

۱. مبتنی بر قواعد الزام‌آور در زمینه پرداخت حق‌الزحمه درمان باشد و اگر

۲. کنترلگر ۲۸ در صورت عدم پذیرش درخواست به طور کامل، تدابیر مناسبی را برای حمایت از منافع قانونی شخص موضوع داده اتخاذ کند و این تدابیر را حداکثر پیش از اعلام عدم پذیرش کامل درخواست شخص موضوع داده، به او اطلاع دهد. این تدابیر عبارت‌اند از:

- (۱) حق درخواست از کنترلگر برای مداخله انسانی (در تصمیم‌گیری)؛
- (۲) حق اعلام نظر از سوی شخص موضوع داده (راجع به تصمیم)؛
- (۳) حق اعتراض به تصمیم.

۳. اگر تصمیم براساس داده‌های سلامت باشد، کنترلگر باید تدابیر مناسب و خاصی را برای حفاظت از منافع شخص موضوع داده مطابق با ماده ۲۲ مقررات اتحادیه اروپا، اتخاذ کند. بنابراین دو نکته مهم در مورد پردازش خودکار داده‌های شخص به موجب قانون فدرال حمایت از داده‌ها این است که نخست، رویه‌های تصمیم‌گیری خودکار فقط در صورت برآورده نشدن درخواست شخص موضوع داده، باید با تضمین‌های مناسب همراه باشند و نه در فرض تأیید درخواست مشتری. نکته دیگر عبارت است از محدودیت اختیارات تخییری کنترلگر، بدین معنا که هنگام طراحی الگوریتم، باید قواعد الزام‌آور جبران هزینه‌های درمان را رعایت کنند.

۲-۳. تمایز بین انواع تصمیم‌گیری الگوریتمی خودکار

در فرانسه، «قانون شماره ۷۸-۱۷ ششم ژانویه ۱۹۷۸ در خصوص پردازش داده‌ها، پرونده‌ها و آزادی‌ها»^{۲۹} با در نظر گرفتن سه فرض مختلف، تصمیم‌گیری خودکار را تنظیم می‌کند: تصمیم‌گیری خودکار در زمینه قضایی^{۳۰}، تصمیم خودکار و نیمه‌خودکار اداری و سایر تصمیمات خودکار دارای آثار حقوقی یا تأثیرات مهم بر افراد (Malgieri, 2019: 13). در مورد تصمیمات قضایی به موجب ماده ۴۷ قانون مذکور، در صورتی که پردازش الگوریتمی خودکار برای ارزیابی جنبه‌های شخصیتی در نظر گرفته شود، تصمیم خودکار یا نیمه‌خودکار، کلاً ممنوع است.^{۳۱}

همین ماده، در مورد تصمیمات اداری^{۳۲} بین تصمیمات خودکار و نیمه‌خودکار تمایز قائل شده است. تصمیمات کاملاً خودکار در پژوهش‌خواهی اداری ممنوع است. انواع دیگر تصمیمات اداری، حتی اگر به طور کامل یا جزئی خودکار باشد، با رعایت شرایط زیر مجازند:

(الف) شامل داده‌های حساس نباشند؛

ب) فرایندهای اداری را محترم بشمارند؛

ج) تصمیمی که براساس پردازش الگوریتمی اتخاذ شده باشد، باید حاوی یک اطلاعیه / تذکری باشد که شخص مربوط را از چنین امری مطلع کند؛

د) اداره مربوط، قواعد تعریف پردازش خودکار داده‌ها و ویژگی‌های اصلی اجرای آن را به شخص مربوط در صورت درخواست او اعلام کند؛

ه) کنترلگر، کنترل خود بر پردازش الگوریتمی و توسعه‌های آن را تضمین کند تا بتواند با جزئیات و به‌صورت قابل فهم، چگونگی پردازش انجام گرفته را به شخص مربوط توضیح دهد.

در مورد تصمیمات بخش خصوصی نیز هر تصمیمی که آثار حقوقی در مورد یک شخص داشته باشد یا به‌نحو معناداری وی را تحت تأثیر قرار دهد^{۳۳}، صرفاً براساس پردازش خودکار داده‌های شخصی، از جمله بر مبنای پروفایل‌سازی، قابل اتخاذ نیست، مگر در موارد زیر:

۱. موارد مذکور در ماده ۲۲ مقررات اتحادیه اروپا و
۲. به‌شرطی که قواعد تعریف پردازش داده‌ها و ویژگی‌های اصلی اجرای آن به استثنای اسرار مورد حمایت قوانین، به درخواست شخص مربوط و توسط کنترلگر اطلاعات به او اعلام شود.

همچنین حکم خاصی در مورد تصمیمات خودکار در زمینه آموزش وجود دارد که به موجب آن، کمیته اخلاق و علمی خاص (ذکرشده در کد آموزش) هر ساله، در پایان آیین ملی پیش ثبت‌نام و قبل از دسامبر، گزارشی را به پارلمان در مورد این آیین و رویه‌های بررسی برنامه‌های مؤسسات آموزش عالی ارائه کند. کمیته ممکن است به همین مناسبت پیشنهادی را برای بهبود شفافیت این آیین ارائه دهد (Malgieri, 2019: 14).

در مقام مقایسه با مفاد ماده ۲۲ مقررات اتحادیه اروپا، ممنوعیت تصمیم‌گیری خودکار از نظر قلمرو اعمال، فقط به تصمیماتی که دارای آثار حقوقی‌اند یا به‌طور مشابه بر شخص موضوع داده تأثیر دارند، محدود نمی‌شود. در ماده ۴۷ هیچ اشاره‌ای به تأثیرات شایان توجه «مشابه» با تأثیرات حقوقی نمی‌شود، بلکه هر نوع «تأثیر معنادار» موردنظر است. بر این اساس، تصمیماتی که تأثیراتشان به اندازه «آثار حقوقی» مهم نباشد نیز باید مقررات تصمیم‌گیری خودکار مذکور در ماده ۴۷ قانون فرانسه را رعایت کنند. کلمه «به‌طور مشابه» که در ماده ۲۲ مقررات اتحادیه اروپا به‌کار رفته، بحث گسترده‌ای را در مورد قلمرو حق مصون بودن از

تصمیمات خودکار دامن زده است. به نظر می‌رسد که قانون فرانسه رویکرد تفسیر موسع را پذیرفته و کنترل‌گرها را ملزم کرده تا در هرگونه تصمیم‌گیری خودکار و تولید «تأثیر معنادار»، تضمین‌های خاصی را رعایت کنند.

در مورد درجات مختلف حمایت، سخت‌ترین محدودیت‌ها برای تصمیمات قضایی است که جنبه‌های شخصیتی افراد را ارزیابی می‌کنند. قانونگذار فرانسوی احتمالاً تصمیم‌گیری قضایی را از نظر تأثیرات احتمالی بیشتر بر افراد، حساس‌تر از سایر زمینه‌ها می‌داند. البته تصمیمات قضایی نمی‌تواند مبتنی بر پردازش کاملاً یا نیمه‌خودکاری باشد که برای ارزیابی جنبه‌های شخصیتی در نظر گرفته شده است. انواع دیگر تصمیمات خودکار که دارای آثار حقوقی‌اند نیز ممنوع هستند، مگر اینکه برای اهداف قراردادی انجام گرفته باشند و شخص موضوع داده امکان ارائه اظهارات را داشته باشد یا اینکه تأثیر مثبت (موافقت) بر درخواست شخص موضوع داده داشته باشند. به عبارت دیگر، تمایز بین تصمیمات قضایی (ارزیابی جنبه‌های شخصیتی) و سایر تصمیمات دوازده سال پیش از تصویب مقررات اتحادیه اروپا پذیرفته شده بود. تصمیمات اداری در سطح دوم از نظر شدت محدودیت‌ها قرار دارند و برای تصمیمات خصوصی محدودیت‌های کمتری وجود دارد. نگرانی سنتی فرانسویان این است که پردازش داده‌های شخصی از سوی ادارات دولتی، دارای تأثیرات ناخواسته و چه‌بسا مشکل‌ساز برای افراد داشته باشد و اداره باید به اصول سختگیرانه حقوق اداری یعنی بی‌طرفی، برابری، قانونی بودن و عدم تبعیض احترام بگذارد. درحالی‌که چنین الزاماتی لزوماً در مورد بخش خصوصی وجود ندارد (Malgieri, 2019: 16).

۳-۳. مشروط کردن تصمیمات الگوریتمی خودکار به رعایت ضوابط خاص

قانون حمایت از داده‌های انگلستان مصوب ۳۴۲۰۱۸ رویکرد کاملاً متفاوتی نسبت به تصمیم‌گیری خودکار دارد. ماده ۱۴ این قانون، استثنای ماده ۲۲ مقررات اتحادیه اروپا را «تصمیمات مهم» و تصمیماتی را که مشمول ماده مذکور نیستند، «در حکم تصمیمات مهم» دانسته و هر دو نوع را در شرایط خاصی مجاز اعلام کرده است. قانون مذکور به هر کنترل‌گر داده اجازه می‌دهد تا هر نوع تصمیم‌گیری خودکار را مشروط به رعایت الزامات کلی زیر صورت دهد:

۱. کنترلگر باید در سریع‌ترین زمان ممکن، کتباً به شخص موضوع داده اطلاع دهد که تصمیمی صرفاً براساس پردازش خودکار گرفته شده است، و
 ۲. موضوع داده می‌تواند ظرف یک ماه از تاریخ دریافت اطلاعیه از کنترلگر بخواهد تا یا در تصمیم‌گیری تجدیدنظر کند یا تصمیم جدیدی بگیرد که صرفاً مبتنی بر پردازش خودکار نباشد. در صورتی که موضوع داده این درخواست را انجام دهد، کنترلگر - بدون تأخیر باید:
 ۱. درخواست را از جمله هرگونه اطلاعات ارائه‌شده توسط شخص موضوع داده مورد توجه قرار دهد،
 ۲. درخواست را اعمال کند و با اطلاع‌رسانی کتبی به موضوع داده و او را از اقدام‌های انجام‌گرفته برای رعایت درخواست و نتیجه مطابقت با درخواست آگاه کند.
- در مورد تضمین‌های خاص نیز قانون انگلستان مانند قانون آلمان، کنترلگر را موظف می‌داند تا کتباً مطلع کند که تصمیمی صرفاً براساس پردازش خودکار گرفته شده است. افزون‌بر این، موضوع داده دارای دو تضمین دیگر است، او می‌تواند از کنترلگر بخواهد در تصمیم‌گیری تجدیدنظر کند یا تصمیم جدیدی بگیرد که فقط مبتنی بر پردازش خودکار نباشد.
- ظاهراً این حقوق با فهرست تضمین‌های مقرر در بند ۳ ماده ۲۲ مقررات اتحادیه اروپا (اعتراض به تصمیم‌گیری، بیان دیدگاه خود و مداخله انسانی) متفاوت است. در واقع، هیچ اشاره مستقیمی به حق دریافت اطلاعات یا درخواست توضیح از تصمیمات اخذشده به‌صورت خودکار یا از منطق الگوریتم که در بند ۷۱ مقررات اتحادیه اروپا ذکر شده است، وجود ندارد. با وجود این، مفسران معتقدند حق اعتراض به تصمیم (یا به چالش کشیدن آن) به‌طور ضمنی و بالملازمه با حق درخواست تجدیدنظر در تصمیم‌گیری از کنترلگر وجود دارد. همچنین حق دستیابی به مداخله انسانی در ضمن حق درخواست تصمیم جدید که صرفاً مبتنی بر پردازش خودکار مبتنی نباشد، قرار دارد. در مورد حق شخص موضوع داده در خصوص بیان نظر خود نیز می‌توان آن را از این الزام که «کنترلگر داده باید درخواست را بررسی کند»، از جمله هرگونه اطلاعات ارائه‌شده توسط موضوع داده را که مربوط باشد، استنباط کرد. بر این اساس، شخص موضوع داده نه تنها حق درخواست تصمیم جدید، بلکه حق ارائه اطلاعاتی را دارد که ممکن است برای تجدیدنظر در تصمیم اول قابل استفاده باشد.

به‌طور خلاصه، به‌نظر می‌رسد قانون حمایت از داده‌های انگلستان جایگزینی مشخص برای ارائه توضیح از الگوریتم بر مبنای سه مرحله ارائه می‌دهد: الف) اطلاع‌رسانی به شخص موضوع داده، ب) درخواست موضوع داده و ج) توضیح مراحل و نتیجه مطابقت با درخواست فردی.

۳-۴. شفافیت و پاسخگویی الگوریتمی (کانادا و آمریکا)

راهکار دیگری که از سوی برخی دولت‌ها از جمله کانادا و آمریکا برای قاعده‌مند کردن تصمیم‌گیری خودکار دنبال شده است، شفافیت و پاسخگویی الگوریتمی است. در این روش، نظارت‌های نهادهای تنظیم‌گر به موازات نظارت عمومی بر فعالیت سیستم‌های تصمیم‌گیری الگوریتمی، تأمین و اعمال می‌شود.

همان‌طور که اشاره شد، یکی از مشکلات سیستم‌های هوش مصنوعی، جعبه سیاه بودن و عدم شفافیت آن‌هاست. این ویژگی، کشف تبعیض را دشوار می‌کند. در نتیجه، شفاف کردن نحوه عملکرد این سیستم‌ها هم می‌تواند به کشف علل و نتایج تبعیض‌آمیز کمک کرده و هم مقابله با آن‌ها را تسهیل کند. بدین‌منظور دولت‌ها می‌توانند با وضع قانون یا الزام دیگری مقرر کنند که تمامی سیستم‌های الگوریتمی مورد استفاده در بخش عمومی به‌گونه‌ای طراحی و توسعه یابند که امکان ممیزی و تبیین آن‌ها فراهم باشد.

برای بخش خصوصی نیز می‌توان الزامات مشابهی را مقرر کرد. مسئله مهم در مورد بخش خصوصی آن است که طراحان و استفاده‌کنندگان این الگوریتم‌ها یا با ادعای حقوق مالکانه بر داده‌های مورد استفاده در هوش مصنوعی خود یا با ادعای اسرار تجاری در مورد الگوریتم‌هایی که از آن داده‌ها برای تصمیم‌گیری استفاده می‌کند، از افشای جزئیات آن‌ها خودداری می‌کنند. در چنین وضعی، الزام استفاده‌کنندگان از این سیستم‌ها به افشای عمومی کل الگوریتم‌ها یا کدهای نرم‌افزاری مورد استفاده نمی‌تواند راه‌حل مناسبی برای شفاف‌سازی باشد، زیرا به حقوق و منافع خصوصی آن‌ها به‌شدت لطمه وارد می‌کند، اما این امکان وجود دارد که از آن‌ها خواسته شود تا بخشی از اطلاعات کلیدی درباره الگوریتم‌های مورد استفاده خود را برای عموم عرضه کنند؛ برای مثال اعلام کنند از کدام متغیرها استفاده کرده‌اند؛ الگوریتم‌ها برای تحقق کدام اهداف بهینه‌سازی شده‌اند؛ داده‌های یاددهنده و انحراف‌های استاندارد از نتایج تولیدشده چیستند؛ نوع داده‌هایی که از سوی الگوریتم‌ها پردازش می‌شود چیست و میزان آن‌ها چقدر است. البته، صرف افشای بیشتر داده‌ها نمی‌تواند برای شفافیت مفید باشد و کلیدی بودن

نیز به معنای افشای عمومی همه داده‌ها نیست، بلکه عرضه داده‌هایی است که بتواند به شفافیت و کارآمدی منجر شود. در واقع امر مورد نیاز عبارت است از تبیین واقعی چگونگی تولید نتایج و اتخاذ تصمیمات توسط الگوریتم‌های مورد استفاده.

۳-۴-۱. دستورالعمل تصمیم‌گیری خودکار در کانادا^{۳۵}

در مقدمه این دستورالعمل آمده است که هدف آن عبارت است از حصول اطمینان از اینکه استقرار سیستم‌های تصمیم‌گیری خودکار به روشی است که خطرهای موجود در کانادا و مؤسسات فدرال را کاهش دهد و به تصمیمات کارآمدتر، دقیق‌تر، سازگار و قابل تفسیر مطابق با قوانین کانادا منجر شود. بدین منظور نتایج مورد انتظار از این مصوبه عبارت‌اند از اینکه:

- تصمیماتی که توسط ادارات دولت فدرال گرفته می‌شود، مبتنی بر داده‌ها، پاسخگویی و رویه‌های منصفانه و الزامات دادرسی منصفانه است؛

- تأثیرات الگوریتم‌ها بر تصمیمات اداری ارزیابی شده و در صورت مواجهه بودن با تأثیرات منفی، به کاهش آن‌ها اقدام می‌شود؛

- داده‌ها و اطلاعات مربوط به استفاده از سیستم‌های تصمیم‌گیری خودکار در مؤسسات فدرال، در صورت لزوم، در دسترس عموم قرار گیرد.

براساس اهداف مذکور، احکام مذکور در این مصوبه را می‌توان در سه دسته کلی قرار گیرد: کیفیت داده‌ها و عملکرد، ارزیابی تأثیرات و شفافیت.

منظور از کیفیت داده‌ها و عملکرد عبارت است از اطمینان از اینکه داده‌های جمع‌آوری شده و مورد استفاده در سیستم تصمیم‌گیری خودکار، مربوط، دقیق، به‌روز و مطابق با خط‌مشی خدمات و دیجیتال و قانون حریم خصوصی‌اند.

بدین منظور لازم است پیش از شروع به کار (سیستم) فرایندهایی طراحی شود تا داده‌ها و اطلاعات استفاده‌شده توسط سیستم‌های تصمیم‌گیری خودکار از نظر سوگیری ناخواسته داده‌ها و دیگر عواملی که ممکن است ناعادلانه بر نتایج تأثیر بگذارد، آزمایش شوند.

همچنین باید فرایندهایی برای نظارت بر نتایج سیستم‌های تصمیم‌گیری خودکار طراحی شود تا در برابر نتایج ناخواسته حفاظت کند و انطباق با قوانین و این دستورالعمل را به‌صورت برنامه‌ریزی‌شده، تصدیق کند.

از نظر ارزیابی تأثیرات نیز این مصوبه، انجام ارزیابی تأثیرات الگوریتمی را با رعایت الزامات مقرر در پیوست مصوبه، پیش از راه‌اندازی هر سیستم تصمیم‌گیری خودکار لازم دانسته است (بند ۶). همچنین مقرر کرده است در صورت تغییر عملکرد سیستم یا دامنه سیستم تصمیم‌گیری خودکار، ارزیابی تأثیرات الگوریتمی باید به‌روزرسانی شود. نتایج نهایی ارزیابی تأثیرات الگوریتمی باید در شکل قابل دسترس از طریق وبسایت‌های دولت کانادا و سایر سرویس‌های تعیین‌شده توسط دبیرخانه هیأت خزانهداری کانادا طبق دستورالعمل دولت باز، منتشر شود.

در خصوص شفافیت تصمیم‌گیری الگوریتمی نیز این مصوبه مقرر کرده است که باید پیش از تصمیم‌گیری، از طریق تمامی کانال‌های ارائه خدمات در حال استفاده اطلاع داده شود که تصمیم گرفته‌شده به‌طور کامل یا جزئی توسط یک سیستم تصمیم‌گیری خودکار که در پیوست C تجویز شده است، اتخاذ خواهد شد. همچنین پس از تصمیم‌گیری، ارائه توضیح معنادار برای افراد تحت تأثیر در مورد چگونگی و دلیل تصمیم‌گیری طبق پیوست C ضروری است. افزون‌بر این، دسترسی به اجزا و عناصر نرم‌افزارهای مورد استفاده نیز در شرایطی مجاز است: در بخش عمومی، باید برای اجزای نرم‌افزار مطابق با الزامات مشخص‌شده در دستورالعمل خدمات و دیجیتال^{۳۶}، مجوزهای لازم پیش‌بینی شود. کد منبع متعلق به دولت کانادا طبق شرایط مشخص‌شده در دستورالعمل مذکور، منتشر خواهد شد، مگر اینکه کد منبع برای پردازش داده‌های طبقه‌بندی‌شده مورد استفاده باشد. در صورت استفاده از نرم‌افزار اختصاصی (مالکانه) باید تمام نسخه‌های منتشرشده از اجزای نرم‌افزار اختصاصی مورد استفاده برای سیستم‌های تصمیم‌گیری خودکار به واحد پیش‌بینی‌شده در دستورالعمل مذکور تحویل داده شوند و از آن‌ها حفاظت شود.

دولت کانادا حق دسترسی و آزمایش سیستم تصمیم‌گیری خودکار از جمله تمامی نسخه‌های منتشرشده از اجزای نرم‌افزار اختصاصی را در صورتی که برای یک ممیزی خاص، تحقیق، بازرسی، معاینه، اقدام‌های اجرایی یا روند قضایی لازم باشد، ضمن حفاظت از آن‌ها در برابر افشای غیرمجاز، برای خود محفوظ می‌داند. به‌عنوان بخشی از این دسترسی، دولت مجاز است به اشخاص بیرونی (خارج از اداره) اجازه دهد در صورت لزوم اجزا و عناصر را بررسی و ممیزی کنند.

۳-۴-۲. پاسخگویی الگوریتمی در آمریکا (لایحه ۳۷۲۰۱۹)

در آمریکا، به جای تصویب قوانین خاص راجع به تبعیض الگوریتمی، قوانین موجود در زمینه مقابله با تبعیض و حمایت برابر از افراد در برابر قوانین، برای مقابله با تبعیض‌های الگوریتمی نیز مورد استناد قرار می‌گیرند. با این حال، اعمال قوانین سنتی ضد تبعیض در الگوریتم‌هایی که نتایج تبعیض‌آمیز به‌ویژه برای گروه‌ها ایجاد می‌کنند، دشوار است.

قوانین ضد تبعیض اغلب بر دو حالت از تبعیض یعنی رفتار متفاوت و تأثیر متفاوت، تمرکز دارند. ممنوعیت رفتار متفاوت، رفتارهای نابرابر را بر اساس ویژگی‌های حمایت‌شده (نژاد، جنسیت، سن و غیره) منع می‌کند، درحالی‌که منع تأثیر متفاوت، رفتارهای موجهی را که تأثیرات نامطلوب بر افراد دارای ویژگی‌های حمایت‌شده دارند، منع می‌کند. نکته مهم این است که هیچ‌یک از این ممنوعیت‌ها نمی‌توانند به‌طور مؤثر تبعیض الگوریتمی را دفع و رفع کنند. بر همین اساس، لایحه پاسخگویی الگوریتمی در سال ۲۰۱۹ در کنگره آمریکا ثبت شده و با آنکه هنوز تصویب نشده است، مورد توجه بسیار واقع شده و تأثیرات خود را داشته است.

این لایحه، ارزیابی تأثیرات سیستم‌های تصمیم‌گیری خودکار را مهم‌ترین راهکار برای پاسخگویی الگوریتمی و نیز مقابله با تبعیض الگوریتمی دانسته است. منظور از «ارزیابی تأثیرات سیستم تصمیم‌گیری خودکار» در این لایحه عبارت است از مطالعه هر سیستم تصمیم‌گیری خودکار و روند توسعه این سیستم از جمله چگونگی طراحی و آموزش الگوریتم‌های سیستم، برای احراز تأثیرات آن‌ها بر صحت، انصاف، سوگیری، تبعیض، حریم خصوصی و امنیت و حداقل شامل موارد زیر می‌شود:

الف) شرح جزئیات سیستم تصمیم‌گیری خودکار، طراحی آن، آموزش آن، داده‌های مورد استفاده و هدف آن؛

ب) ارزیابی منافع و هزینه‌های نسبی سیستم تصمیم‌گیری خودکار در پرتو اهداف آن، با در نظر گرفتن عوامل مربوط، از جمله:

(۱) رویه‌های به حداقل رساندن داده‌ها؛

(۲) مدت زمان ذخیره اطلاعات شخصی و نتایج سیستم تصمیم‌گیری خودکار؛

(۳) میزان اطلاعاتی که در مورد سیستم تصمیم‌گیری خودکار در دسترس مصرف‌کنندگان است؛

۴) میزان دسترسی مصرف‌کنندگان به نتایج سیستم تصمیم‌گیری خودکار و اینکه آیا می‌توانند نتایج آن را اصلاح کرده یا مورد اعتراض قرار دهند یا نه؛

۵) دریافت‌کنندگان نتایج سیستم تصمیم‌گیری خودکار؛

ج) ارزیابی خطرهای ناشی از سیستم تصمیم‌گیری خودکار برای حریم خصوصی یا امنیت اطلاعات شخصی مصرف‌کنندگان و خطرهای منتج شدن سیستم تصمیم‌گیری خودکار به تصمیم‌گیری‌های نادرست، ناعادلانه، مغرضانه، تبعیض‌آمیز در مورد مصرف‌کنندگان؛ و

د) تدابیری که نهاد مشمول برای به حداقل رساندن خطرهای توصیف‌شده در زیر بند «ج»، از جمله در زمینه پروتکل‌های فنی و فیزیکی، به کار خواهد گرفت.

نتیجه‌گیری

تبعیض الگوریتمی، تبعیضی است که از تصمیم‌گیری‌های الگوریتمی ناشی می‌شود. الگوریتم‌ها که براساس محاسبات خاص، داده‌های ورودی را به خروجی‌های دلخواه هدایت می‌کنند، دو قسم‌اند: قاعده‌بنیان و یادگیری ماشین. الگوریتم‌های قاعده‌بنیان به سبب نقش فرعی و کمکی در تصمیم‌گیری، در حد کمتر و الگوریتم‌های یادگیری ماشین، به سبب نقش مؤثر و تعیین‌کننده در تصمیم‌گیری‌ها، در حد زیادی در بروز چنین تبعیض‌هایی نقش دارند.

سیستم‌های الگوریتمی به‌عنوان فناوری‌هایی که از داده‌ها تغذیه می‌کنند، در نتیجه دو عامل، ممکن است به تصمیمات تبعیض‌آمیز منتج شوند: اول، به دلیل داده‌های ورودی از این حیث که ممکن است با تبعیض‌های موجود در جامعه آلوده شده یا ناکافی و ناقص باشند و دوم، به دلیل عملکرد تبعیض‌آمیز سیستم‌ها که می‌تواند ناشی از طراحی یا آموزش آن‌ها باشد.

با توجه به اینکه مقابله با تبعیض از وظایف دولت‌هاست، ابتکارهای مختلفی در سطح بین‌المللی و ملی برای مقابله با این نوع تبعیض‌ها از سوی دولت‌ها ارائه شده است. کشورهای اروپایی در چارچوب کلی قوانین حمایت از داده‌های شخصی، استفاده از تصمیمات الگوریتمی خودکار را محدود و قاعده‌مند کرده‌اند و کشورهایی مانند کانادا و آمریکا نیز شفاف‌سازی چگونگی اتخاذ تصمیمات الگوریتمی برای نهادهای ناظر و عموم جامعه را در دستور کار قرار داده‌اند.

هر کدام از این رویکردها، نقاط قوت و ضعفی دارند. قوانین حمایت از داده‌های شخصی قابلیت‌های بیشتری برای مقابله با تبعیض علیه اشخاص موضوع داده دارند، زیرا افزون‌بر

پیش‌بینی احکام خاص در مورد این نوع تصمیم‌گیری، با شناسایی اصول پردازش داده‌های شخصی می‌توانند مانع از جمع‌آوری و پردازش داده‌های افراد به ضرر آنها شوند و با شناسایی حقوق اشخاص موضوع داده، از جمله حق دسترسی، حق اصلاح و حق حذف و فراموش شدن می‌توانند مانع از این شوند که داده‌های قدیمی، ناقص یا غلط مبنای تصمیمات الگوریتمی قرار گیرد. با وجود این، برای مقابله با تبعیض‌های نظام‌مند و گروهی ضعیف‌اند و قابلیت رویکرد آمریکایی و کانادایی برای مقابله با چنین تبعیض‌هایی بسیار بیشتر از رویکرد اروپایی است.

۱. در کشور ما نیز استفاده از سیستم‌های تصمیم‌گیری الگوریتمی به‌طور فزاینده‌ای در حال گسترش است. با وجود قوانین و مقرراتی در ممنوعیت تبعیض در سطح قانونگذاری و همین‌طور در نظام اداری کشور، نیاز جدی به مصوبه خاصی برای پیشگیری از تبعیض‌های الگوریتمی و مقابله با آنها احساس می‌شود. متأسفانه حمایت‌های حداقلی قانون تجارت الکترونیک (مصوب ۱۳۸۲) در مورد داده‌های شخصی نیز رعایت می‌شود تا چارچوبی برای کنترل داده‌های مورد استفاده این سیستم‌ها در دسترس باشد. از این‌رو به‌نظر می‌رسد ضروری است که شورای عالی فضای مجازی یا مجلس شورای اسلامی برای رفع خلأ قانونی در این زمینه اقدام کنند و در این مسیر استفاده از نقاط قوت مجموع راهکارهای معرفی شده در این مقاله، می‌تواند مؤثر باشد.

یادداشت‌ها

1. see: Panel for the Future of Science and Technology, 2019; Paterson et al. (2019, No 1.); Galajdová, 2019, Issue 1; Tene, Omer & Jules Polonetsky, 2017, Issue 1).
2. Data analytic systems.
3. Big data.
4. Artificial intelligence.
5. Machin learning.
6. Expert systems.
7. Natural language.
8. Computer vision.
9. Rule-based algorithms.
10. Machine-learning algorithms.
11. If this then that
12. Supervised learning.
13. Unsupervised learning.
14. Reinforcement learning.
15. Deep learning.
16. screening decisions.
17. Megvii .
18. Personalized price.
19. Personalized behavioral advertising.
20. Les biais cognitifs.
21. Les biais d'anticipation et de confirmation.
22. Le biais de corrélations illusoire.
23. le biais de stereotype.
24. target variable.
25. class labels.
26. a decision...which produces legal effects concerning him or her or similarly significantly affects him or her.
27. Federal Data Protection Act of 30 June 2017 (Federal Law Gazette I p. 2097), as last amended by Article 12 of the Act of 20 November 2019, https://www.gesetze-im-internet.de/englisch_bds/englisch_bds.html#p0310.
۲۸. شخصی است که هدف و روش‌های پردازش داده‌های شخصی را تعیین می کند.
29. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/>.
30. décision de justice.
31. Article 47, Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, Modifié par Ordonnance n°2018-1125 du 12 décembre 2018 - art. 1.
32. Des décisions administratives individuelles.
33. Aucune décision produisant des effets juridiques à l'égard d'une personne ou l'affectant de manière significative
34. Data Protection Act
- 2018, http://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpga_20180012_en.pdf.
35. <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592§ion=html>
36. Directive on Service and Digital.
37. <https://www.congress.gov/bill/116th-congress/senate-bill/1108/text?q=%7B%22search%22:%22algorithmic+accountability+act%22%7D&r=1&s=2>



►
منابع

الف) انگلیسی

1. Cofone, Ignacio N., (2019), “Algorithmic Discrimination Is an Information Problem”, **Hastings Law Journal**, Vol. 70, Issue 6 , Article 1, pp.1404-06.at:
https://repository.uchastings.edu/hastings_law_journal/vol70/iss6/1,
2. Committee of experts on internet intermediaries (2018), Study on the Human Rights Dimensions of Automated Data Processing Techniques (in particular algorithms) and possible regulatory implications, <https://edoc.coe.int/en/internet/7589-algorithms-and-human-rights-study-on-the-human-rights-dimensions-of-automated-data-processing-techniques-and-possible-regulatory-implications.html>.
3. European Commission, Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, Brussels, 21.4.2021, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206>
4. Executive Office of the President (2016) , Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights, https://obamawhitehouse.archives.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf.
5. Galajdová, Dominika (2019), “Artificial Intelligence as a New Challenge for Software Law”, **European Journal of Law and Technology**, Vol. 10, Issue 1.
6. Gerards, Janneke and Raphaële Xenidis (2021), Algorithmic discrimination in Europe: Challenges and opportunities for gender equality and non-discrimination law, A special report, European Union, <https://op.europa.eu/en/publication-detail/-/publication/082f1dbc-821d-11eb-9ac9-01aa75ed71a1>.
7. Guidelines on addressing the human rights impacts of algorithmic systems (2020), Appendix to Recommendation CM/Rec(2020)1, <https://rm.coe.int/09000016809e1154>.
8. International Working Group on Data Protection in Telecommunications, Working Paper on Privacy and Artificial Intelligence, 4th Meeting, 29-30 November 2018, Queenstown (New Zealand), at: <https://epic.org/IWG/WP-AI.pdf>.
9. ITU, Security, Infrastructure and Trust Working Group, Report of Trust Workstream, Big data, machine learning, consumer protection and privacy, 2018, https://www.itu.int/en/ITU-T/extcoop/figisymposium/Documents/FIGI_SIT_Technical%20report_Big%20data,%20Machine%20learning,%20Consumer%20protection%20and%20Privacy_f.pdf.



10. Jernigan, Carter, Mistree, Behram F.T. (2009), "Gaydar: Facebook friendships expose sexual orientation", **First Monday**, Vol. 14, No. 10 - 5 October 2009,
<https://firstmonday.org/ojs/index.php/fm/article/download/2611/2302>.
11. Kaye, David, Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), Report on Artificial Intelligence technologies and implications for freedom of expression and the information environment, 29 August 2018,
<https://www.undocs.org/A/73/348>. <https://www.ohchr.org/EN/Issues/FreedomOpinion/Pages/ReportGA73.aspx>
12. Kleinberg et al. (2018), "Discrimination In The Age Of Algorithms", **Journal of Legal Analysis**, Vol. 10.
13. Law Library of Congress (2019), Regulation of Artificial Intelligence in Selected Jurisdictions, January 2019, pp. 16-132.
<https://www.loc.gov/law/help/artificial-intelligence/regulation-artificial-intelligence.pdf>. Köchling, Alina, Marius Claus Wehner, Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development, <https://link.springer.com/article/10.1007/s40685-020-00134-w>.
14. Malgieri, ianclaudio (2109), "Automated decision-making in the EU Member States: The right to explanation and other "suitable safeguards" in the national legislations", **Computer Law & Security Review**, Vol. 35, Issue 5.
15. Orwat, Carsten (2020), Risks of Discrimination through the Use of Algorithms, <https://publikationen.bibliothek.kit.edu/1000123477>.
16. Panel for the Future of Science and Technology (2019), Understanding algorithmic decision-making: Opportunities and challenges, European Parliamentary Research Service, March 2019.
<https://www.statewatch.org/media/documents/news/2019/mar/ep-study-Understanding-algorithmic-decision-making.pdf>.
17. Paterson, Moira and McDonagh, Maeve (2019), "Data Protection in an Era of Big Data: The Challenges Posed by Big Personal Data", **Monash University Law Review**, Vol. 44, No. 1,
https://www.monash.edu/__data/assets/pdf_file/0009/1593630/Paterson-and-McDonagh.pdf.
18. Revell, T. (2018), Face-recognition software is perfect – if you're a white man, <https://www.newscientist.com/article/2161028-face-recognition-software-is-perfect-if-youre-a-white-man>.
19. Rovatsos, Michael, Brent Mittelstadt and Ansgar Koene (2019), Landscape Summary: Bias in Algorithmic Decision-Making, The Centre for Data Ethics and Innovation,

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/819055/Landscape_Summary_-_Bias_in_Algorithmic_Decision-Making.pdf.

20. Tene, Omer & Jules Polonetsky (2017), "Taming The Golem: Challenges of Ethical Algorithmic Decision-Making",

North Carolina Journal of Law and Technology , Vol.19, Issue 1.

21. The 80/20 data science dilemma,

https://www.infoworld.com/article/3228245/the-80-20-data-science-dilemma.html.

22. Xenidis, Raphaële and Linda Senden (2020), EU non-discrimination law in the era of artificial intelligence: Mapping the challenges of algorithmic discrimination in Ulf Bernitz et al (eds), General Principles of EU law and the EU Digital Order (Kluwer Law International),

https://cadmus.eui.eu/handle/1814/65845.

23. Zuiderveen Borgesius, Frederik (2018), **Discrimination**, Artificial Intelligence, and Algorithmic Decision-making. Council of Europe, https://rm.coe.int/discrimination-artificial-intelligence-and-algorithmic-decision-making/1680925d73.

ب) فرانسه

1. Bertail, Patrice, Bounie David, Clémenton Stephan, et Waelbroeck Patrick (2019), Algorithmes: biais, discrimination et équité, 14 février 2019, p.10. https://www.telecom-paris.fr/wp-content-

EvDsK19/uploads/2019/02/Algorithmes-Biais-discrimination-equite.pdf.

2. Institut Montaigne (2020), Algorithmes: contrôle des biais S.V.P., Rapport Mars 2020, pp. 17-18.

https://www.institutmontaigne.org/ressources/pdfs/publications/algorithmes-controlle-des-biais-svp.pdf .

3. Lucchesi, Laure, Simon Chignard (2019), Rapport-Ethique et responsabilité des algorithmes publics, Juin

2019. https://www.etalab.gouv.fr/wp-content/uploads/2020/01/Rapport-ENA-Ethique-et-responsabilit%C3%A9-des-algorithmes-publics.pdf.

ج) قوانین و پیش نویس مصوبات

1. Algorithmic Accountability Act of 2019, 116th Congress (2019-2020), https://www.congress.gov/bill/116th-congress/house-bill/2231.

2. Council of Europe Declaration on the manipulative capabilities of algorithmic processes, Adopted by the Committee of Ministers on 13 February 2019,

https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b.

3. Data Protection Act 2018, http://www.legislation.gov.uk/ukpga/2018/12/pdfs/ukpga_20180012_en.pdf
4. Directive on Automated Decision-Making, Effective Date on April 1, 2019, with compliance required by no later than April 1, 2020, <https://www.tbs-sct.gc.ca/pol/doc-eng.aspx?id=32592§ion=html>.
4. Federal Data Protection Act of 30 June 2017 (Federal Law Gazette I p. 2097), as last amended by Article 12 of the Act of 20 November 2019, https://www.gesetze-im-internet.de/englisch_bdsge/englisch_bdsge.html#p0310.
5. Loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000886460/>.

