بسم الله الرحمن الرحیم

فصلنامه‌ی علمی - تحقیقی علوم فنی
دانشگاه خاتم‌النبیین (ص)
ریاست تحقیقات
سال دوم، شماره‌ی سوم، تابستان ۱۴۰۳
صاحب امتیاز: دانشگاه خاتم‌النبیین (ص)
مدیرمسئول: سید زکریا حبیب
سردبیر: سید محمد احمدی

اعضای هیئت تحریر

سید محمد احمدی	عضوهیات علمی دانشکده کامپیوترساینس دانشگاه خاتم‌النبیین (ص)
محمد ناظم جعفری	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه کاتب
مرتضی نیکزاد	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه کاتب
محمد یحی اخلاقی	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه کابل
عباس محبی	عضوهیات علمی دانشکده کامپیوترساینس دانشگاه خاتم‌النبیین (ص)
محمد علی فهیمی	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه خاتم‌النبیین (ص)
سید محمد شریف شاکر	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه خاتم‌النبیین (ص)
محمد آصیف دولت نظر	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه تعلیم تربیه کابل
رجب علی احمدی	عضو هیات علمی دانشکده انجنیری دانشگاه خاتم‌النبیین (ص)
سید انورشاه موسوی	عضو هیات علمی دانشکده انجنیری دانشگاه خاتم‌النبیین (ص)
محمد نسیم زاهدی	عضو هیات علمی دانشکده انجنیری دانشگاه خاتم‌النبیین (ص)
نجیب الله خالقی	عضو هیات علمی دانشکده انجنیری دانشگاه خاتم‌النبیین (ص)
سید زکریا حبیب	عضوهیات علمی دانشکده کامپیوترساینس دانشگاه خاتم‌النبیین (ص)
محمدهادی فصیحی	عضو هیات علمی دانشکده انجنیری دانشگاه کاتب
محمد ضامن جعفری	دانشکده برق دانشگاه پل‌تخنیک کابل
قربان علی فروغ	عضو هیات علمی دانشکده کمپیوترساینس دانشگاه تعلیم تربیه کابل
محمد داود فروتن	عضو هیات علمی دانشکده کمپیوتر ساینس دانشگاه کاتب
محمد شاه امید	عضو هیات علمی دانشکده کامپیوترساینس، دانشگاه کاتب

ویراستار ادبی: سید محمد سجادی
صفحه آرا و طراح جلد: مجتبی احمدی

آدرس: کابل، سرک دارالامان، دانشگاه خاتم‌النبیین (ص)
شماره‌ی تماس: ۰۷۸۷۷۰۰۷۰۰
E-mail: info@knu.edu.af
Web: www.knu.edu.af

یادآوری: هرگونه استفاده از محتوای این فصلنامه تنها با ذکر منبع مجاز است.

راهنمای تنظیم مقالات

این رهنمود به منظور تدوین مقالات علمی و پژوهشی با استندردهای معین تهیه و تدوین شده است و از این رو، مقاله باید دارای مشخصات ذیل باشد:

۱. مقالات تحقیقی، تألیفی، مستند و تحلیلی باشند؛
۲. مقالات قبلاً در هیچ مجله‌ای چاپ نشده و یا همزمان به مجله‌ی دیگری ارسال نشده باشند.
۳. مقالات باید به همراه نام، نشانی دقیق، شماره تماس و ایمیل نویسنده به ایمیل مخصوص سردبیر مجله ارسال شود.
٤. عنوان دری، چکیده دری (حداقل ۲۰۰ حداکثر ۲۵۰ کلمه شامل هدف، روش و یافته-ها) و کلیدواژه‌ها (حداکثر ۷کلمه)؛
۵. عنوان انگلیسی، چکیده انگیسی (حداقل ۲۰۰ و حداکثر ۲۵۰ کلمه شامل هدف، روش و یافته‌ها) و کلیدواژه‌ها (حداکثر ۷کلمه)؛
٦. مقاله باید دارای مقدمه، متن (حداقل ٤۵۰۰ کلمه و حداکثر ۷۵۰۰ کلمه)، نتیجه‌گیری باشد؛
۷. مقدمه (حداکثر ۳ صفحه ۳۵۰ کلمه) باید شامل موارد ذیل باشد: بیان مسأله، سوال‌های اصلی و فرعی، فرضیه‌ها؛ ضرورت و اهمیت، اهداف، نوآوری، نوع، پیشینه و ساختار تحقیق؛
۸. رفرنس بر اساس شیوه‌نامه استناددهی درون متنی باشد؛ (تخلص، سال و شماره صفحه)
۹. عنوان و چکیده انگلیسی (حداکثر ۲۵۰ کلمه شامل هدف، روش و یافته‌ها) و کلیدواژه‌ها (حداکثر ۷کلمه)؛
۱۰. اسامی خاص و اصطلاحات لاتین در پاورقی ذکر شود؛
۱۱. عناوین اصلی و فرعی در بدنه (متن) مقاله با استفاده از حروف اعداد و به شکل زیر مشخص شوند:

عناوین اصلی مقاله ـ به جز مقدمه و نتیجه ـ با حروف ابجد (الف)، ب)، ج) و ...)؛

عناوین فرعی با اعداد تک رقمی (۱.، ۲. ۳.، ...)؛

عناوین جزئی‌تر با اعداد ترکیبی دو و سه رقمی (۱-۱.، ۲-۱. و ۳-۱. و ... ۱ -۱-۱، ۱-۱-۲، ۱-۱-۳ و ...).

۱۲. تمامی قواعد و استندردهای نگارشی رعایت گردد؛
نیم‌فاصله‌ها در مقاله رعایت شود. به عنوان مثال: می‌رود (صحیح) می رود (غلط)
۱۳. بین کلمات تنها یک فاصله باشد و فاصله‌های اضافه حذف شود. به عنوان مثال:
قانون.جزا (صحیح) قانون...جزا (غلط)

۱۴. فهرست منابع

فهرست منابع در عناوین کلی به کتاب، مقاله، پایان نامه و اسناد داخلی و بین الملل تقسسم بندی شود؛

فهرست کتاب: تخلص، نام کوچک، (سال نشر) نام کتاب، جلد (نام و تخلص مترجم، اگر ترجمه باشد) محل نشر، انتشارات، نوبت چاپ؛

فهرست مقاله: تخلص، نام نویسنده، نام مقاله، (نام و تخلص مترجم، اگر ترجمه باشد) شماره مسلسل، سال نشر، صفحات مقاله در مجله؛

در صورتی که سند داخلی و یا بین المللی باشد، نام سند و سال تصویب حتما ذکرگردد؛

یادآوری:

از آنجا که احترام به حقوق پدیدآورندگان آثار علمی یکی از اصول راهبردی مجله است، نقض این حقوق توسط ارسال‌کنندگان مقالات، موجب رد مقاله خواهد شد؛

سردبیر در تلخیص، اصلاح و ویرایش مقالات پذیرفته‌شده آزاد است.

در صورت لزوم به روزرسانی اطلاعات مندرج در مقاله هنگام چاپ، نویسنده(گان) متعهد به انجام آن است (هستند).

دفتر مجله مجاز خواهد بود مقالات چاپ شده را به صورت الکترونیکی نیز عرضه نماید.

مسئولیت آرا و دیدگاه‌های ارائه‌شده در مقاله‌ها به عهده نویسندگان است و چاپ آن‌ها به معنی تأیید مطالب نیست.

فهرست مطالب

تشخیص آی دی کانکور عموی افغانستان از روی ورقه امتحان..... ۳
محمد ناظم جعفری ، سید زکریا حبیب ، محمد منصور لطیفی

رویکرد انعطاف پذیری در طراحی مسکن تدریجی..... ۱۵
پوهنیار سید انورشاه موسوی

کاربردهای شبکه عصبی مصنوعی و یادگیری عمیق در شناسایی آسیب وپایش سلامت سترکچرها –مطالعه مروری..... ۳۱
عباس محبی ، سید مهدی موسوی

شبیه‌سازی ماشین‌های امن خودپرداز با استفاده از دیف..... ٤٥
سید زکریا حبیب

مروری بر اخلاق هوش مصنوعی..... ٦١
Changwu Huang, Zeqi Zhang, Bifei Mao, and Xin Yao

جایگاه (نقش) اخلاق حرفه‌ای معلمان در آموزش مجازی مکاتب..... ۱۱۹
سیروس قنبری و احمد عزیزی

سخن نخست

گرایش انسان به نوآوری، کار آفرینی و تولید دانش یکی از گرایش های ذاتی است که بخش قابل توجه از تاریخ مدنیت بر اساس آن شکل گرفته است، با توجه به سیرتکامل بشریت در جوامع مختلف، گرایش های ذاتی و تکامل انسان‌ها بر حسب نیاز صورت گرفته که همین اکنون نیز نیازمندی ها بزرگ ترین بهانه‌ی است که سبب تحول و دگرگونی در جوامع مختلف می‌گردد. سیر تکاملی انسان از مسیر دانش، یکی از امن ترین مسیرهایی بوده که انسان ها را به سوی خودکفایی و خود باوری سوق داده با آنهم این مسیر عاری از چالش ها و در برخی موارد موانع نبوده است.

فراگیری دانش و انتقال آن در مدت زمان طولانی محصور در چهارچوب دیدگاه‌ها و نظریه پردازی بوده که پر مخاطره ترین قرن‌ها را در گذشته تجربه کرده و پشت سر گذاشته است. دانشی که امروزه به دست ما رسیده در واقع محصول زحمات دانشمندانی است که بر مشکلات تمرکز کرده و نظریه ها و راهکارهایشان را به روش های مختلف به جوامع بشری عرضه نموده اند و از این طریق باعث توسعه دانش و تمدن بشریت شده اند، تا بجایی رسیده که باورها و مدنیت های مختلف در حال حاضر از آن بهره می‌برند. پس از تحول و دگرگونی بزرگ (به ویژه بعد از جنگ جهانی دوم) که جوامع بشری آنرا تجربه کرد گسترش دانش و پیاده سازی آن به شکل سریع و ساده‌تر است. پس از این جنگ زمینه گسترش دانش به روش های ساده و مفید بیشتر فراهم شده و دیگر این پدیده محصور از ذهنیت‌ها و اندیشه‌ها نیست.

از سوی دیگر تنها گرایش های ذهنی که سبب نو آوری و کار آفرینی می‌گردد به تنهایی نمی‌تواند به نیازمندی‌های بشریت پاسخ قانع کننده‌ی ارائه کند یا به واژه دیگر، دانش در چهارچوب نظری و دیدگاه‌ها حل کننده مشکلات و چالش های جوامع بشری به ویژه در حوزه علوم فنی و تجربی نیست بلکه علاوه بر آن باید به ابزاری دسترسی پیدا کنیم که بتوانیم با به کارگیری از آن روش های تجربی، تطبیقی و ساده سازی دانش را بهبود بخشیده و انکشاف جامعه را تسریع نمایم. تسریع یکی از مهمترین اموری است که در حوزه‌علوم فنی به آن دقت لازم را باید داشته باشیم، چون ممکن نیازمندی‌های جوامع بشری متکی به لمحه های زمانی خیلی کوتاه باشد که همین اکنون آنرا در حوزه تکنالوژی درحال تجربه هستیم. دراین حوزه یعنی تکنالوژی که در ماهیت خود یکی از قدرتمندترین ابزارها برای انکشاف جوامع بشری در حوزه های مختلف محسوب می‌شود سیر تحول بر حسب ثانیه صورت می‌گیرد.

پژوهش و تحقیقات در شروع جنگ سرد تبدیل به یک ابزار قدرتمند برای تولید، انتقال و پیاده سازی دانش شد که از روش‌های معیاری گوناگونی پیروی می‌کند. همچنان تحقیقات یکی از ساده‌ترین روش برای دسترسی به گزینه‌های فوق به شماره می‌رود که پیامدهای گوناگونی در

زمینه های مختلف مانند انکشاف، تکامل، توسعه و پیشرفت را درپی داشته است. در کشورهای توسعه یافته بزرگترین رسالت نهادهای علمی انجام تحقیقات بر حسب نیازمندی‌های جامعه است که نهادهای تحصیلات عالی را به سوی خودکفایی و شگوفایی سوق داده و به مصدر خدمت در جامعه تبدیل نموده است. به همین دلیل مجله علوم فنی دانشگاه خاتم النبیین (ص) تلاش می‌نماید تا زمینه تحقیق و پژوهش را برای خبره‌گان و پژوهشگران جامعه فراهم سازد، بر این اساس مجموعه مقاله های که دراین شماره به دست نشر سپرده شده نتیجه ماموریت و رسالت این نهاد و حاصل زحمات استادان و پژوهشگرانی است که برای دسترسی به نتایج قابل تامل به گونه شبانه روزی تلاش نموده اند.

به امید فردای بهتر از امروز

سید زکریا حبیب

مدیر مسئول مجله علوم فنی

تشخیص آی دی کانکور عموی افغانستان از روی ورقه امتحان

محمد ناظم جعفری[۱]، سید زکریا حبیب[۲]، محمد منصور لطیفی[۳]

چکیده

استفاده از هوش مصنوعی و ترویج آن در انکشاف صنایع یکی از دست آوردهای بشر در چند دهه اخیر است که در موارد از آن استفاده به عمل می‌آید. امروزه سازمان‌ها و نهادهای دولتی و غیر دولتی نیاز شدید به خودکار سازی اداری و رفع مشکلات شان با استفاده از هوش مصنوعی است. بر این أساس دراین پژوهش تلاش شده است تا بتوان با بکارگیری از هوش مصنوعی مشکل شناسایی شناسه کانکور را رفع کرد. در این پژوهش، تشخیص و بررسی شناسه یا آی دی نمبر[٤] داوطلب از روی ورقه امتحان کانکور در دو مرحله انجام می‌شود. در مرحله اول، تشخیص شناسه از روی اعدادی است که داوطلبان با دست‌خط خود در بخش شناسه در ورقه امتحان می‌نویسند. در بخش دوم، از کلیدهای پررنگ‌شده در بخش مربوط به شناسه استفاده می‌کنیم. این دو مرحله پس از اسکن ورقه امتحانی با کمک پردازش تصویر و عملیات برش ناحیه‌های تصویر به‌منظور کوچک‌سازی ناحیه مورد نظر صورت می‌گیرد. برای شناسایی اعداد

[۱]. دیپارتمنت مهندسی نرم افزار، دانشکده کامپیوترساینس، دانشگاه کاتب، m.n.jafari@kateb.edu.af

[۲]. دیپارتمنت مهندسی نرم افزار، دانشکده کامپیوترساینس، دانشگاه خاتم النبیین (ص) s.z.habib@knu.edu.af

[۳]. دیپارتمنت تکنالوژي معلوماتی، دانشکده کامپیوترساینس، دانشگاه کاتب، mohammadmansour.latifi@gmail.com

[٤]. Identification Number ID.

دست‌نویس، از دیتاست مینست[1] و کتابخانه‌های تینسرفلو[2] و کیراس[3] استفاده شده است تا مدلی با دقت بالا بدست آید. در بخش دوم نیز ناحیه کلیدها به ۹ ستون که هر کدام شامل یک کلید است برش می‌خورد و سپس سطرها با دقت پردازش می‌شوند. کلیدهای پررنگ شده با مقدار آستانه خاصی استخراج می‌شوند و در نهایت ای دی نمبر داوطلب به دو طریق شناسایی می شود تا اطمینان بیشتر حاصل شود.

واژگان کلیدی: شناسایی ارقام دست نویس، پردازش تصویر، امتحان کانکور، یادگیری عمیق.

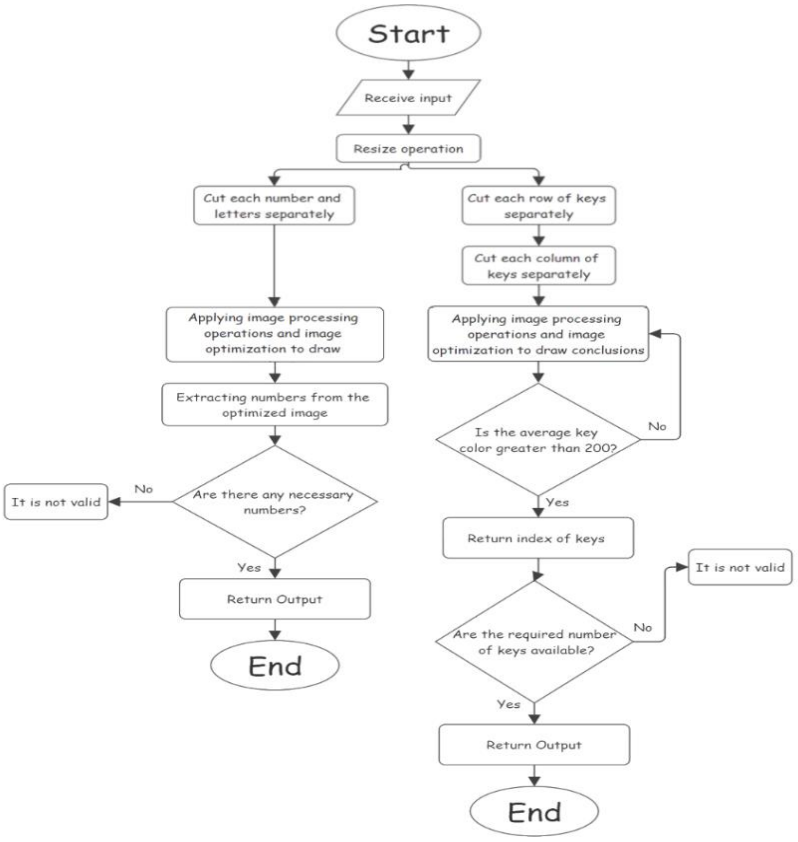
مقدمه

امتحان عمومی کانکور یکی از امتحاناتی است که سالیانه توسط دولت از تمام محصلین که از صنف دوازدهم فارغ التحصیل می‌شوند گرفته می‌شود، با توجه به روند ارزیابی پارچه‌های امتحان کانکور این روند نیاز مبرم به شفاف سازی و همین‌گونه تسریع در عملکرد دارد. از یک طرف تشخیص آی دی نمبر داوطلبان امتحان کانکور از روی ورقه امتحان کانکور همواره یک معضل جدی بوده است که در برخی موارد به ویژه هنگام نمره دهی سبب مشکلات ناخواسته از عدم تشخیص درست افراد و تامین شفافیت سازی در امتحان عمومی کانکور محسوب می‌شود. از سوی دیگر کمبود تحقیق که نیازهای حقیقی جامعه را در حوزه تکنالوژي شناسایی و حل نماید کمتر در افغانستان رویت می‌گردد و یا در برخی موارد تحقیقات انجام شده با دقت بالا کم‌رنگ تر به چشم می‌خورد که نادرستی و کم‌دقتی در فرآیند تشخیص سبب کاهش کیفیت و ایجاد نابرابری در نتایج داوطلبان می‌شود. روش‌های دستی که در حال حاضر برای بررسی این بخش استفاده می‌شوند، نه تنها زمان‌بر هستند بلکه دارای دقت کافی نیز نمی‌باشند. برای کاهش این نادرستی‌ها، استفاده از روش‌های الکترونیکی با کمک نرم‌افزارهای کمپیوتری الزامی پنداشته شده که می توانند نتایج دقیقی را در کمترین زمان ممکن ارائه دهند. در دنیای امروز،استفاده از پردازش تصویر و یادگیری عمیق در حال گسترش است و می‌توان از این فناوری‌ها در زمینه الکترونیکی‌سازی فرآیند بررسی آی دی نمبر داوطلب استفاده کرد. این سیستم می‌تواند به بهبود دقت و کاهش خطاها در نتایج آزمون کانکور کمک شایانی کرده و پروسه شفافیت سازی را در عرضه تشخیص بهبود بخشد.

[1] .MNIST.

[2] .TensorFlow.

[3] .Keras.



شکل ۱: جریان شناسایی آی دی نمبر کانکور

۱. بیان مسئله

پیشرفت جوامع بشری با نوآوری و تکنالوژی به ویژه در حوزه هوش مصنوعی و استفاده آن در صنعت سیر تکامل را توسعه داده و جوامع را بیشتر به سوی پیشرفت و ترقی به شمول ساده سازی سوق می‌دهد. استفاده از هوش مصنوعی امروزه در جوامع پیشرفته یکی از مهمترین و اساسی ترین پدیده ها است که برای ساده سازی و شفافیت در امور مختلف بکار برده می‌شود. اما درافغانستان از آنجایی که این کشور از جمله کشورهای در حال توسعه است استفاده از هوش مصنوعی کمرنگ‌تر بوده و صرف در دو دهه اخیر مورد توجه قرار گرفته است که استفاده و بکارگیری ازین تکنالوژی یکی از نیازهای مبرم می‌باشد به ویژه زیرشاخه آن یادگیری عمیق.

یکی از مهمترین ماموریت تکنالوژی از مرحله شروع و در چهارچوب آن استفاده از هوش مصنوعی در حوزه‌های مختلف به ارمغان آوری شفافیت در امور است. چون در بسیاری از موارد افغانستان کشوری است که از عدم وجود آن رنج می‌برد، یادگیری عمیق و استفاده از هوش

مصنوعی یکی از ابزارهای توانا و قدرتمند برای آوردن شفافیت و تسریع روند کاری بوده که همواره برای ارایه خدمات به کشورهای توسعه یافته ترویج یافته و مورد استفاده قرار می‌گیرد.
در حال حاضر تمام امور اداری به ویژه امتحانات کانکور در افغانستان و مشخصاً وزارت تحصیلات به روش سنتی و دستی انجام می‌شود که این روش ازیک سو از شفافیت لازم برخودار نبوده و از سوی دیگر زمانبر می‌باشد. برعلاوه استفاده از سیستم سنتی آسیب‌های مانند افزایش هزینه، درگیری بیش حد انسان‌ها که بخودی خود سبب بروز خطا و عدم شفافیت می‌گردد. باید اشاره کرد که استفاده از سیستم سنتی و دستی در امور به ویژه برگزاری امتحانات و ارزیابی آن پیچیدگی خلق می‌کند که تاثیر مستقیم به روند اداری دارد. باتوجه به مشکلات و چالش‌های فوق نیاز به یک‌سیستم متکی به فناوری‌های جدید از جمله و یادگیری عمیق احساس می‌گردد. دراین مقاله تلاش شده است تا در گام نخست بخشی از سیستمی را پیاده سازی نماید که در آن شفافیت، سرعت عمل، کاهش هزینه و ساده سازی در امور امتحان کانکور که یک امتحان عمومی و حیاتی در افغانستان است لحاظ شود.

۲. پیشینه تحقیق

لیکون و همکاران برای نخستین بار دراین زمینه برای شناسایی حروف با استفاده از پردازش تصویر کار کرده و به نتایج نسبی رسیدند[۱]، همچنان دراین زمینه بنرجی و همکاران در سال ۲۰۱۴ مقاله‌ی را نشر که در آن از هوش مصنوعی و یادگیری عمیق برای شناسایی حروف یا کاراکترها استفاده کرده اند[۲]. در رابطه پیرامون شناسایی کاراکترها کریژفسکی و همکاران به دست آوردهای قابل توجه رسیده بودند و در سال‌های اخیر باپیشرفت تکنالوژی و یادگیری عمیق براساس شبکه‌های عصبی نتایج نسبتاً خوبی بدست آمده است. برعلاوه در زمینه دیتاسیت‌ها و پایگاه‌های داده معیاری وجود دارد که از آنها در بخش‌های مختلف در سراسرجهان استفاده می‌شود ازجمله می‌توان به دیتاسیت امنست اشاره کرد[۳]. باتوجه به مشکلات و چالش‌های موجود که در افغانستان از یادگیری عمیق کمتر استفاده شده است دراین مقاله تلاش شده تا با توجه به استفاده از یادگیری عمیق تمرکز کرده و از آن برای تشخیص ای دی نمبر امتحان کانکور بکار برده شود. استفاده از هوش مصنوعی به ویژه یادگیری عمیق، دیتاسیت مربوطه از جمله پدیده‌های جدید است که برای نخستین بار از آن در شناسایی خودکار ورقه‌های امتحان کانکور افغانستان بکار برده می‌شود.

۳. روش تحقیق

با توجه به مقاله‌های نشر شده پیرامون یادگیری عمیق که در بخش قبلی به آن اشاره شد و با توجه به دیتاسیت‌های معیاری که در سراسر جهان از آن استفاده می‌شود دراین مقاله در گام نخست مقاله‌های و دیتاسیت‌های معیاری بررسی شده و در گام دوم با توجه به مشکلات و چالش‌های مطالعه شده و موجود در وزارت تحصیلات عالی و بخش امتحان کانکور تلاش شده تا با

استفاده از یادگیری عمیق و دیتاسیت‌های معیاری برای تشخیص آی دی نمبر ورقه امتحان کانکور عمومی بکار برده شود تا این مشکل مدیریت و حل شود. تشخیص ای دی نمبر داوطلب امتحان کانکور، در دو مرحله انجام می شود، در گام نخست تشخص آی دی نمر که در ورقه پاسخ‌ها، به گونه دستی توسط داوطلبان کانکور خانه پری می‌شوند و در بخش دیگر خانه های متناظر با این اعداد توسط داوطلب سیاه می شود، یعنی ای دی نمبر در دو جا ثبت می شود یکی با اعداد دست نویس، دیگری به وسیله سیاه کردن خانه های مستطیل شکل یا به عبارت دیگر مجموعه کلیدهای پر رنگ شده، این کار به تشخیص بهتر ای دی نمبر داوطلب کمک می نماید. در این پژوهش از هر دو روش جداگانه برای استخراج آی دی نمبر استفاده شده است که روش اول آن بر پایه تشخیص اعداد دست‌نویس و روش دوم بر اساس کلیدهای پررنگ‌شده است. هر دو روش به کمک الگوریتم‌های پردازش تصویر، تصویر ورقه امتحانی برش داده شده و نواحی مربوط به آی دی نمبر استخراج می‌شود، نتایج بدست آمده از روش نخست با دقت بین ۹۵٪ تا ۹۸ ٪ و در روش دوم دقت آن تقریبا ٪ ۱۰۰ است.

٤. تشخیص و استخراج اعداد دست‌نویس

برای پردازش ورقه‌های جوابات امتحان کانکور، ابتدا باید این ورقه‌ها به‌صورت دقیق اسکن شوند. این کار از اهمیت ویژه‌ای برخوردار است، زیرا صحت داده‌های ورودی نقش کلیدی در عملکرد صحیح مدل دارد. پس از اسکن، بخش‌هایی از تصویر که شامل آی‌دی کانکور هستند، جدا می‌شوند. این جداسازی به‌گونه‌ای انجام می‌شود که تنها نواحی مورد نیاز استخراج می گردند و بخش‌های غیرضروری حذف می شوند.

در مرحله بعد، تصاویر اعداد دست‌نویس که از این نواحی استخراج شده‌اند، به ابعاد استاندارد ۲۸x۲۸ پیکسل تبدیل می‌شوند. این مرحله، که به‌عنوان همسان‌سازی شناخته می‌شود، تنها بخشی از فرآیند آماده‌سازی داده‌ها است. علاوه بر همسان‌سازی، تصاویر باید به فرمت باینری تبدیل شوند، به این معنا که بیت‌های تصاویر به شکل دودویی تنظیم شوند. این فرآیند برای کاهش نویز و بهبود دقت تشخیص ضروری است.

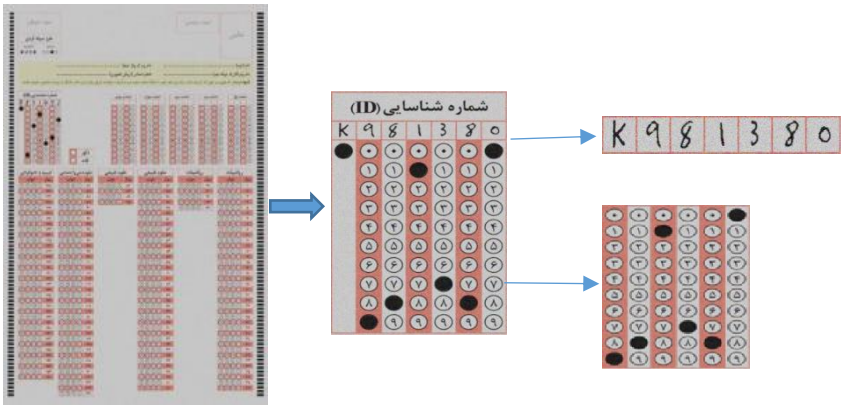
برای توسعه مدل پیشنهادی، از دیتاست مشهور ام‌نیست[1] استفاده شده است که شامل نمونه‌های متنوعی از اعداد دست‌نویس است. این مدل با بهره‌گیری از کتابخانه‌های قدرتمند تنسورفلو و کراس آموزش داده شده است. نتیجه نهایی، مدلی است که قادر است اعداد دست‌نویس را با دقت بسیار بالا شناسایی کند و بدین ترتیب، فرآیند تحلیل ورقه‌های امتحانی را به شکل چشم گیری بهبود بخشد.

---

۱. MNST .

٤-١. دریافت تصویر

مطابق با شکل ۲، پس از وارد کردن تصویر اسکن‌شده به سیستم، الگوریتم‌ها و فرآیندهای پردازشی طراحی‌شده وظیفه دارند تا بخش‌های مختلف ورقه را شناسایی و جداسازی کنند. ورقه امتحانی به دو بخش اصلی تقسیم می‌شود: بخش اعداد دست‌نویس و بخش کلیدها. این تقسیم‌بندی به سیستم کمک می‌کند تا هر بخش را به صورت مجزا و دقیق‌تر پردازش کند. بخش مربوط به اعداد دست‌نویس شامل اعدادی است که توسط داوطلب به صورت دستی نوشته شده‌اند و برای پردازش و تشخیص آن‌ها از روش‌های خاصی استفاده می‌شود. بخش کلیدها نیز خانه هایی است که متناظر با اعداد ای دی نمبر توسط داوطلب سیاه شده است به صورت جداگانه استخراج و پردازش می‌شود. این جداسازی و پردازش مستقل به سیستم این امکان را می‌دهد که بدون ایجاد تداخل بین بخش‌ها، هر قسمت را به بهترین شکل تحلیل کرده و نتایج دقیق‌تری ارائه دهد. به‌طور کلی، این رویکرد مرحله‌به‌مرحله و مبتنی بر جداسازی و پردازش مستقل بخش‌ها، دقت و کارایی سیستم را به شکل قابل‌توجهی افزایش می‌دهد.



شکل ۲: نحوه خوانش آی دی نمبر و تقسیم آن

٤-٢. برش و استخراج ناحیه آی دی نمبر

در این مرحله، آی‌دی نمبر بر اساس تعداد ثابت آن در ورقه امتحان کانکور که برابر با ۷ است، پردازش می‌شود. بخش ابتدایی آی‌دی که شامل یک حرف است، خارج از محدوده این مطالعه قرار دارد و در این فرآیند مورد بررسی قرار نمی‌گیرد. به همین دلیل، ابتدا بخش حروف از بخش اعداد تفکیک می‌شود تا تنها اعداد برای مراحل بعدی پردازش آماده شوند.

٤-٣. هم اندازه ساختن تصاویر برش شده

پس از برش تصاویر به ابعاد مورد نظر، لازم است اطمینان حاصل شود که اندازه این تصاویر با استانداردهای دیتاست امنیست مطابقت دارد. در این دیتاست، تمامی تصاویر به ابعاد ۲۸x۲۸

پیکسل تنظیم شده‌اند. بنابراین، تصاویر ورودی نیز باید به همین مقیاس تغییر اندازه داده شوند تا مدل بتواند آن‌ها را به‌درستی پردازش کرده و با داده‌های موجود در دیتاست سازگار کند. این همسان‌سازی ابعاد، بخش ضروری از فرآیند آماده‌سازی داده‌ها برای مدل است.

٤-٤. تبدیل تصاویر به فرمت باینری

در این مرحله، برای کاهش نویزهای موجود در تصاویر و افزایش سرعت پردازش مدل، تصاویر به‌دست‌آمده از مراحل قبلی به فرمت باینری تبدیل می‌شوند. این فرمت که تنها از دو مقدار صفر و یک برای نمایش پیکسل‌ها استفاده می‌کند، باعث ساده‌تر شدن ساختار داده‌ها می‌شود. تبدیل تصاویر به فرمت باینری نه تنها باعث می‌شود که دستگاه‌های دیجیتال بتوانند آن‌ها را به شکلی بهتر و آسان‌تر پردازش کنند، بلکه مزیت دیگری نیز به همراه دارد: سرعت پردازش تصاویر به‌طور قابل توجهی افزایش می‌یابد. این افزایش سرعت باعث می‌شود فاصله زمانی بین پردازش تصاویر ورودی و تولید نتایج خروجی به حداقل برسد. به این ترتیب، کارایی سیستم بهبود یافته و مدل می‌تواند با دقت و سرعت بیشتری به تحلیل داده‌ها بپردازد.

٥. مدل پیشنهادی و آموزش آن

این مدل پیشنهادی را با استفاده از کتابخانه های تینسرفلو وکراس در چهار لایه به صورت مسلسل، لایه اول مسطح سازی [1]، لایه دوم تمام متصل دینس [2] با ۱۲۸ نرون و تابع فعال سازی [3] و همین طور لایه سوم تمام متصل دینس با ۱۲۸ نرون و تابع فعال سازی ریلو و در لایه آخر با ۱۰ نرون و تابع فعال سازی سافت مکس [4] ایجاد شده است. سپس این مدل با استفاده از دیتاست ام‌نیست که دارای دو نوع داده، داده های آموزشی و داده‌های آزمایشی، همراه با جواب داده‌ها (کلیدها) که اعدادی (۰ تا ۹) می‌باشد. داده های آموزشی و آزمایشی عکس‌های سیاه و سفید که در ابعاد ۲۸×۲۸، با مقدار بین ۰ تا ۲۵۵ است را آموزش می دهیم. قبل از آموزش سیستم برای بهتر عمل کردن و کارایی بالا باید عملیات پیش پردازش روی داده ها انجام شود. در این جا چون عکسها به فرمت خاکستری یعنی بین ۰ تا ۲۵۵ است باید نرمالیزه شوند، برای این کار تمام مقادیر پیکسلهای عکسها را به ۲۵۵ تقسیم می نماییم، نتیجه حاصل یک عدد بین ۰ تا ۱ است. این عمل باعث عملکرد بهتر سیستم می شود. بعد از نرمالیزه کردن تصاویر را به سیستم می دهیم و مرحله آموزش سیستم شروع می شود. بعد از آموزش مدل با استفاده از دیتاست ام

---

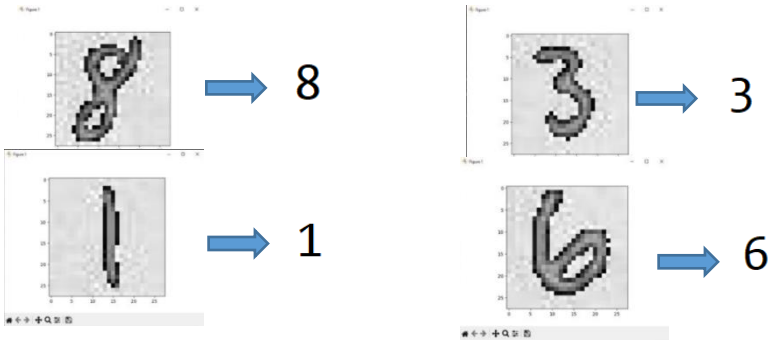
۱. Falatten.

۲. Dense .

۳. Relu .

۴. Softmax .

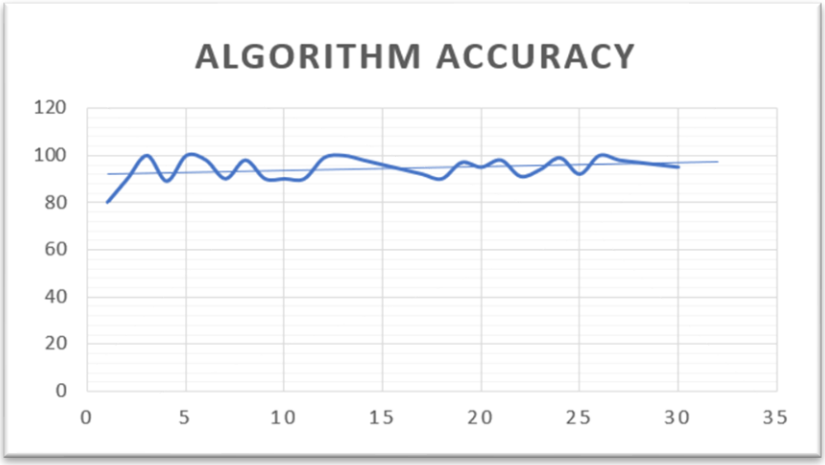
نیست[1]، داده های (عکس‌های ورق امتحان کانکور) خود را به مدل می دهیم که نتایج بدست آمده در صورت بالا بودن کیفیت تصاویر خیلی خوب (با۹۵٪ تا ۹۸٪ دقت در تشخیص درست اعداد دست نویس) می‌باشد که برای درک بهتر موضوع در زیر چند نمونه از تشخیص مدل آورده شده است:



شکل ۳: نمونه از تشخیص مدل

۵-۱. ویژگی‌های الگوریتم استخراج اعداد دست‌نویس

- دقت استخراج اعداد دست‌نویس بین ۹۵ تا ۹۸ درصد
- تشخیص اعداد در کمترین زمان ممکن
- بهینه سازی در زمان إتمام پروسس و تشخیص اعداد
- قابلیت تشخیص اعداد با دست‌خط‌های مختلف
- برش دقیق ناحیه‌های مختلف برای بهبود دقت
- تشخیص مناسب حتی در صورت وجود نویز



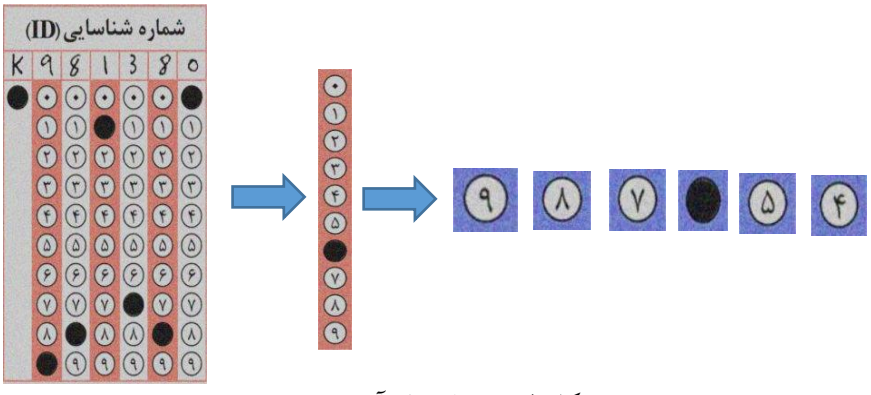
شکل ۴: نمودار آموزش و الگوریتم دقت سیستم

[1] . MNIST

با توجه به شکل ۴ با افزایش تعداد تکرارهای پروسه آموزش، دقت الگوریتم نیز افزایش می‌یابد که در شکل فوق به‌خوبی قابل مشاهده است.

۵-۲. برش ناحیه کلید ها در تصویر (ورق کانکور)

با توجه به تصویر زیر ابتداء برش به صورت عمودی انجام می شود تا ستون عدد مورد نظر جدا شود، سپس تک تک خانه ها جدا(برش) می شوند تا خانه ها از هم جدا شوند به تعداد ۱۰ تا تصویر بدست می آید، آنها را با سایزهای یکسان و به فرمت باینری درآورده تا بتوان به خوبی با سیستم‌های دیجیتل پروسس کرد.



شکل ۵: نمونه از برش آی دی نمبر

۵-۳. تشخیص عدد از روی کلیدها

در این مرحله اوسط مقدار پیکسل‌ها را بدست می آوریم اگر بیشتر از مقدار آستانه(۲۰۰) بود مقدار متناظر آن یعنی خانه چندم است را برمی گرداند و عدد مورد نظر تشخیص داده می شود و اگر کمتر بود یعنی خانه مورد نظر خالی است.

شکل ۶: تحلیل دیتا

۵-٤. ویژگی‌های الگوریتم تشخیص اعداد بر اساس کلیدها

- دقت تشخیص بالا حتی در صورت کم‌رنگ بودن کلیدها
- اعلام خطا در صورت عدم پررنگ‌سازی کامل کلیدها
- تشخیص دقیق در کمترین زمان ممکن (تقریبا٪۱۰۰)

نتیجه گیری و برنامه بعدی

در این پژوهش، پیرامون شناسایی آی دی نمبر از روی ورقه امتحان کانکور بحث شده و با استفاده از هوش مصنوعی و مشخصاً یادگیری عمیق برای تشخیص خودکار استفاده شده است. دو روش برای استخراج آی دی نمبر از روی ورق امتحان کانکور ارائه شده است که هر کدام دقت بالایی دارند که هر دو روش برای رسیدن به دقت بالا در تشخص مکمل هم‌دیگر هستند. دراین مطالعه نتیجه بدست آمده نشان می‌دهد که سیستم در مراحل اجراء از دقت بالای برخودار است که این یکی از مزیت‌های روش پیشنهادی می‌باشد. با این حال، الگوریتم‌های طراحی‌شده در برخی شرایط نیاز به بهبود دارند، از جمله زمانی‌که تصویر مورد استفاده از کیفیت کمتر برخودار و یا به شکل درست اسکن نشده باشد. به مشکلات فعلی می‌توان به عدم تشخیص آی دی در صورتی‌که اعداد خارج از کادر نوشته شوند اشاره کرد و یا خانه‌های متناظر با اعداد به‌صورت کامل پر نشده باشد. همچنین، بهبود دقت مدل در تصاویر با رزولوشن کم و محیط‌های نویزی از جمله چالش‌های آتی می‌باشد. طراحی یک محیط گرافیکی برای سیستم و افزایش دقت در استخراج اعداد دست‌نویس از دیگر کارهای پیشنهادی برای آینده است.

منابع

[1] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[2] Loizou, P. C., & Spanias, A. S. (1996). High-performance alphabet recognition. *IEEE Transactions on Speech and Audio Processing*, *4*(6), 430-445.

[3] https://www.kaggle.com/datasets/hojjatk/mnist-dataset .

[4] Kimura, F., & Shridhar, M. (1991). Handwritten numerical recognition based on multiple algorithms. *Pattern recognition*, *24*(10), 969-983.

[5] Shah, P., Karamchandani, S., Nadkar, T., Gulechha, N., Koli, K., & Lad, K. (2009, November). OCR-based chassis-number recognition using artificial neural networks. In *2009 IEEE International Conference on Vehicular Electronics and Safety (ICVES)* (pp. 31-34). IEEE.

[6] Qu, Y., Shi, H., & Xie, Y. (2010, October). ID numbers recognition by local similarity voting. In *2010 IEEE International Conference on Systems, Man and Cybernetics* (pp. 3881-3888). IEEE.

[7] Chakravorty, Pragnan (2018). "What is a Signal? [Lecture Notes]". *IEEE Signal Processing Magazine*. 35 (5): 175–177.

[8] Gonzalez, Rafael (2018). *Digital image processing*. New York, NY: Pearson. ISBN 978-0-13-335672.

[9] Banerjee R, Pavlides M, Tunnicliffe EM, Piechnik SK, Sarania N, Philips R, Collier JD, Booth JC, Schneider JE, Wang LM, Delaney DW, Fleming KA, Robson MD, Barnes E, Neubauer S (January 2014).

[10] A Brief, Early History of Computer Graphics in Film Archived 17 July 2012 at the Wayback Machine, Larry Yaeger, 16 August 2002 (last update), retrieved 24 March 2010.

[11] Optical Character Recognition (OCR) – How it works". Nicomsoft.com. Retrieved June 16, 2013.

[12] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *ImageNet classification* with *deep convolutional neural networks*. Advances in neural information processing systems, 25, 1097-1105

[13] Abdulbaqi, M. (2009). Higher education in Afghanistan. *Policy perspectives*, 99-117.

[14]Otsu, N. (1979). *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66.

رویکرد انعطاف پذیری در طراحی مسکن تدریجی

پوهنیار سید انورشاه موسوی[1]

چکیده

ازمسائل مهمی که بر سر راه بشر قرار دارد مسأله مسکنِ اقشار کم درآمد جامعه است و یکی از اشتباهاتی که تقریبا در تمامی طرح های تأمین مسکن حداقل تکرار شده است استفاده از طرح های پیچیده مهندسی در طراحی خانه هاست. اما جمعی از کارشناسان که درک بهتری از نیاز های سکونتی گروه های کم درآمد شهری دارند، بر این باورند که بهتر است مسئولان به جای طراحی و عرضه نمونه تکمیل شده مسکن، اقدام به ارائه نمونه هایی از تکمیل تدریجی واحد های مسکونی نمایند. به نظر می‌رسد این نوع مسکن نیاز به بودجه بالا ندارد و می‌تواند سریع ساخته شود، از این رو توسعه آن برای تضمین ثبات اجتماعی و اقتصادی در ترویج توسعه ملی بسیار حیاتی است. بنابرین مقاله حاضر در پی شناخت، ویژگی ها، ابعاد و شاخص های مسکن تدریجی و ارائه یک نمونه طراحی شده برای حل معضل مسکن گروه های کم‌درآمد در کشورهای در حال توسعه می باشد. در فرآیند طراحی و ساخت مسکن تدریجی، خانه‌ها انواع گوناگونی از نما و طرح بندی‌های مناسب مطابق خواسته‌های ساکنان وجود خواهد داشت. هزینه پایین، انعطاف‌پذیری و حق انتخاب آگاهانه در گسترش آینده و استفاده از مصالح ساختمانی در دسترس، از اصلی‌ترین ویژگی هایی است که باعث رضایت ساکنان در این روش ساخت مسکن می شود.

واژگان کلیدی: مسکن تدریجی، گروه های کم‌درآمد، انعطاف پذیری.

---

[1]. دانشکده انجنیری، دانشگاه خاتم النبیین (ص)، کابل، افغانستان، sa.mousawi@knu.edu.af، +۹۳۷۴۴۳۸۴۲۲۸

مقدمه

پیش بینی ها نشـان می دهد جمعیت جهان تا سـال ۲۰۵۰ میلادی به ۱۶ میلیارد نفر خواهد رسید و یکی ازمسائل مهم که بر سر راه بشر قرار دارد مسأله مسکن اقشار کم درآمد جامعه اسـت، که تعداد زیادی از این خانواده ها توانایی خرید و تامین مالی آن را ندارند و برای دولت ها نیز پرداخت این وجه دشـوار خواهد بود[۱]. میزان مشارکت بالقوه کاربر در مراحل طراحی و سـاخت مسکن رابطه مستقیم با سطح رضـایت آنها و موفقیت در محصـول نهایی دارد و در صـورت عدم اشـتراک کاربران در این فرایند محصـول نهایی با انتظارات سـاکنان همخوانی نخواهد داشـت. برای اشـتراک کاربران در فرایند طراحی و سـاخت مفهوم سـاده‌ای وجود دارد که می تواند به عنوان راه حل برای منافع متقابل جامعه آینده و دولت عمل کند. این راه حل خاص، مسکن تدریجی[۱] نامیده می شود.

مسکن تدریجی واژه جدیدی نیست اما از سال ۲۰۰۰ میلادی به طور گسترده مورد استفاده قرار گرفت. در این روش توسعه، مردم به استفاده از منابع محلی تشویق شده و کسب و کار را در جوامع ارتقاء می دهد[۲] و روش مقرون به صـرفه برای بسـیاری از خانواده ها با خدمات حداقلی مسکن است[۳]. این مفهوم سیستم مسکنی را ارائه می کند که ساکنان توانایی توسعه آن را در آینده داشـته باشـند[۴]. بناءً مسـکن تدریجی را می توان چنین تعریف کرد: مسـکن تدریجی یک فرایند مرحله به مرحله است، در این روش خانه با یک هسته اولیه آغاز شده وبا توجه به بُعد زمان و منابع کافی، آنها به خانه هایی با درآمد متوسط تبدیل می شوند[۳].

طبق نظر آدلر و وِرا[۲] مسکن تدریجی استراتیژی است که مبتنی بر یک سیستم مترقی می باشد، طوری که در آن ساخت و ساز ناقص اما قابل سکونت است. در این استراتیژی خانه ها به گونه‌ای طراحی می شوند که کاربران با استفاده از روش های شخصی مطابق نیاز و سلیقه خود آن را به پایان برسانند. بناءً مسکن تدریجی فرایندی است که طی آن افراد فضای قابل سکونت خود را در طول زمان تغییر دهند[۵].

بنابرین پژوهش حاضـر در پی شـناخت، ویژگی ها، ابعاد و شـاخص های مسـکن تدریجی و ارائه نمونه طراحی شـده برای حل معضـل مسـکن گروه‌های کم‌درآمد کشـور های جهان سـوم می باشد.

۱. روش تحقیق

هر تحقیق علمی نیاز به یک روش پژوهش متناسب با موضوع خود دارد. انتخاب روش تحقیقِ مناسب و تداوم آن در تمامی فرآیند و مسیر پژوهش از اصول راهبردی یک تحقیق علمی

۱. Incremental Housing

۲. Adler & Vera

است. بنابراین روش پژوهش در مقاله پیش‌رو با استفاده از منابع کتابخانه‌ای و مطالعه اسناد، کتب و مقالات، سعی شده برای مقابله با بی‌مسکنی و بدمسکنی و راه‌کارهای مقابله با حاشیه‌نشینی الگوها و طرح‌هایی پیشنهادی ارائه دهد.

۲. پیشینه تحقیق

رشد شهرنشینی از اواسط قرن بیستم منجر به یکی از چالش ها در تأمین مسکن مناسب برای جمعیت شهری شد[۶]. پس از جنگ جهانی اول در اروپا و اتحاد جماهیر شوروی و سپس در امریکا طرح مسکن تدریجی به عنوان پاسخ فوری به کمبود شدید مسکن ایجاد گردید[۷]. مسکن تدریجی یک روش توسعه مرحله به مرحله خانه است که با یک هسته اولیه آغاز شده و با توجه به نیاز و منابع مالی، تحت کنترول مالک، به تدریج اندازه و کیفیت آن گسترش می یابد[۸]. مفهوم مسکن تدریجی در دهه های ۱۹۷۰ تا ۱۹۸۰ بیشتر رایج شد تا برای دسترسی خانوار های کم درآمد به مسکن راه حل منطقی بیابند[۹]. متخصصین در این حوزه، مسکن تدریجی را برای بهبود کمبود مسکن موجود به ویژه برای خانواده های کم‌درآمد دارای اهمیت بالایی می دانند[۶].

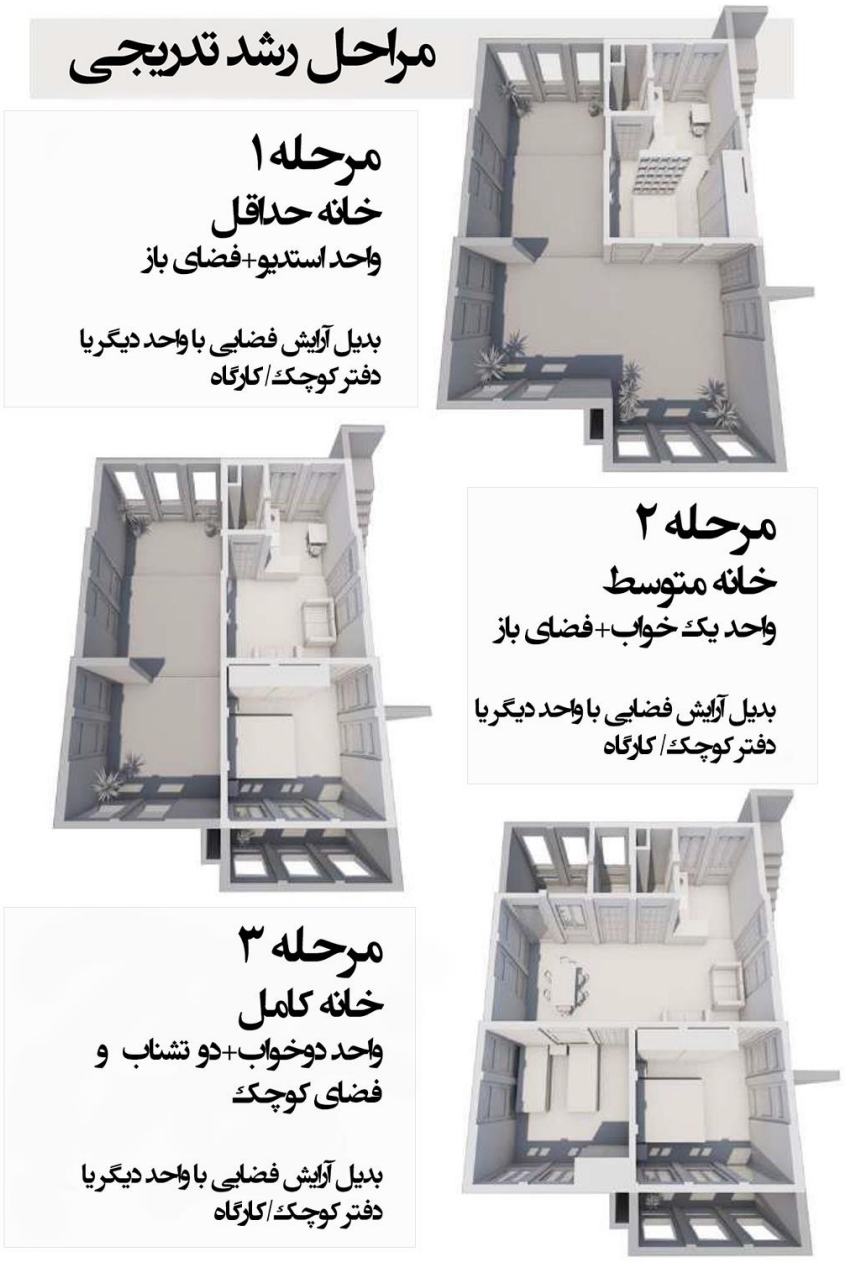
سازمان های بین‌المللی و برخی از دولت های دیگر که دیدگاه های حمایتی زیادی نسبت به مسکن تدریجی داشته‌اند؛ در حقیقت به رسمیت شناختن مسکن تدریجی را به عنوان یک استراتژی موثر مسکن که قدمت ۵۰ ساله دارد، تصدیق می کند[۵].

۳. مسکن تدریجی

مطالعات روز نشـان می دهد که حدود ۱٫۶ میلیارد نفر در سـطح جهان در مسـکن ناکافی به سر می برند و یک میلیارد نفر دارای اسکان غیررسمی در محله های فقیرنشین هستند. یعنی از هر چهار نفر یک نفر آن در شـرایط آسـیب پذیری زندگی می کنند [۱۰]. یکی از اشـتباهاتی که تقریبا در تمامی طرح های تامین مسکن حداقل[۱] تکرار شده است استفاده از طرح های پیچیده مهندسی در طراحی خانه هاست که خود از این تصور نادرست سرچشمه می گیرد که احداث مساکن خوش منظر و کامل راه حلی برای مسئله مسکن است [۱۱]. اما جمعی از کارشناسان که درک بهتری از نیاز های سـکونتی گروه های کم درآمد شـهری دارند، بر این باورند که بهتر است مسئولان به جای طراحی و عرضه نمونه تکمیل شده مسکن، اقدام به ارائه نمونه هایی از تکمیل تدریجی واحد های مسکونی نمایند. بدین ترتیب، نمونه اول می تواند شامل تشناب و یک دیوار مشـترک باشـد. نمونه دوم، تشـناب و یک اتاق نیمه کامل را ارائه دهد، نمونه سـوم، نشـاندهنده چگونگی تکمیل بیشـتر خانه و افزوده شـدن اتاق ها و ... باشـد. عرضـه این گونه نمونه ها، علاوه بر نشـان دادن اسـتندرد های مورد نظر، برای خانوار های سـهیم در پروژه، این

۱. Basic Housing

پیام را می رساند که تکمیل واحد های مسکونی به صورت تدریجی، مناسب و مطابق منابع مالی و مهارت های موجود بلامانع است [۱۱].

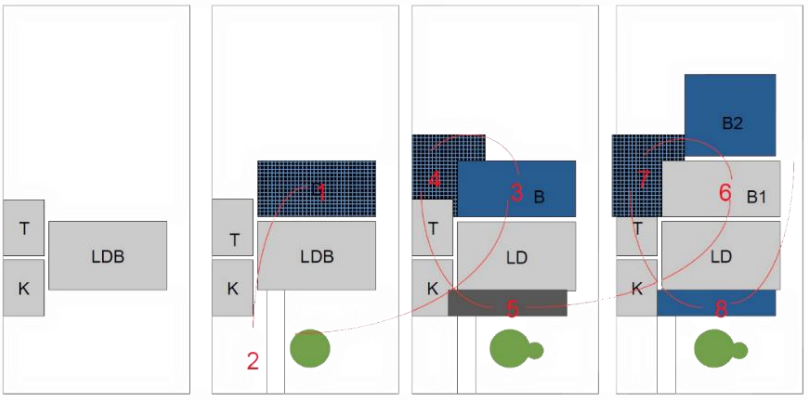


شکل ۱: مراحل گسترش تدریجی [۱]

۳-۱. هسته اولیه[۱]

تجارب بدست آمده از کشور های در حال توسعه در بخش مسکن نشان می دهد که برنامه ریزان مسکن برای گروه های کم درآمد دو روش اساسیِ ارتقاء کیفیت سکونت[۲] و تأمین زمین و خدمات[۳] [۱۱] و برخی هم با نقد تجارب گذشته آن را به چهار دستهِ ساخت مسکن، تأمین مالی مسکن، تأمین زمین و خدمات و راهبرد های توانمند سازی تقسیم بندی کرده اند[۱۲].

هسته اولیه مسکن بخشی از برنامه تأمین زمین و خدمات است که برای برآورده ساختن نیاز های مسکن گروه های کم‌درآمد در کشور های در حال توسعه به اجرا درآمد. امروزه اصطلاح هسته اولیه در سراسر جهان گسترش یافته است که بعد از دهه ۱۹۶۰ توسط تعدادی سازمان های بین المللی به ابزاری مهم برای بهبود مناطق کمتر توسعه یافته تبدیل شد. هسته اولیه، خانه‌ای است که حداقل فضای قابل سکونت را تامین می‌کند و سطح گسترش و رشد آن به شیوه‌ای تدریجی مطابق با نیاز و توانایی ساکنانش می‌باشد[۱۳]. مطالعات نشان می دهد بعد از انتخاب زمین، مراحل ساخت هسته اولیه (که با هدف حفاظت در برابر سرما، گرما، باد، باران و ... ساخته شده) و توسعه آن از مهمترین معیار ها در طراحی مسکن تدریجی به شمار می رود[۴]. مطابقت و سازگاری سرپناه با نیاز های متفاوت ساکنان همیشه بخشی از دغدغه های سکونت انسان را به همراه داشته است[۱۴] و یکی از مسائلی که در طرح های هسته اولیه مسکن با آن مواجه هستیم، این است که غالبا طرحی مورد پسند مردم است که همیشه با هسته مسکن اولیه منطبق و سازگار باشد. از این قبیل موارد دیده شده است که بعضی از خانوار ها قبل از توسعه هسته اولیه، اقدام به تخریب آن نموده و خود از اول همه چیز را می سازند. مسلم است که این عمل غیر اقتصادی بوده و هزینه ایجاد مسکن را افزایش می دهد [۱۱].



شکل ۲: رشد تدریجی هسته اولیه به مرور زمان

۱. Core House, Starter House, Starter Core, Phased-development house, Owner-driven house

۲. Community Upgrading

۳. Site and Services

۳-۲. مراحل گسترش هسته اولیه

استراتژی‌های توسعه فضا در بین خانوارها برای گسترش خانه متفاوت است. خانواده های متعددی از جمله اقشار متوسط و بالا هستند که دلایل اصلی توسعه فضای خانه هایشان الزمات فرهنگی و نیاز های کاربردی می باشد. اما برخی از خانواده های کم‌درآمد که نیاز های فوری آنها در اولویت است، متناسب با وسع مالی شان مسکن خود را توسعه و بهبود می بخشند[۹]. به نظر می‌رسد این نوع مسکن نیاز به بودجه بالا ندارد و می‌تواند سریع ساخته شود، از این رو توسعه آن برای تضمین ثبات اجتماعی و اقتصادی در ترویج توسعه ملی بسیار حیاتی است[۱]. این فرآیند گام به گام در پاسخ به تغییرات وضعیت اقتصادی است که به موجب آن ساکنان، هر زمان که بودجه یا زمان در دسترس داشته باشند به تدریج خانه های خود را بهبود بخشیده و یا گسترش دهند[۱۵]. تحقیقات انجام شده نشان می دهد که گسترش و رشد تدریجی هسته اولیه به شش نوع ذیل دسته بندی شده است:

الف) توسعه[۱]: افزودن در ملکیت خصوصی که محصور شده باشد.

ب) متصل[۲]: افزایش در قلمرو نیمه خصوصی مثل بالکن ها و حویلی ها.

ج) تعویض[۳]: ارتقاء ساخت و ساز از مواد کم دوام به مواد با دوام.

د) تقسیم[۴]: تقسیم بندی فضا های داخلی.

ه) پرکردن[۵]: محصور و یا دیوارکشی فریمی که از قبل ساخته شده است.

و) وصل کردن[۶]: ایجاد خدمات و تأسیسات در ساختمان

---

۱. Extend

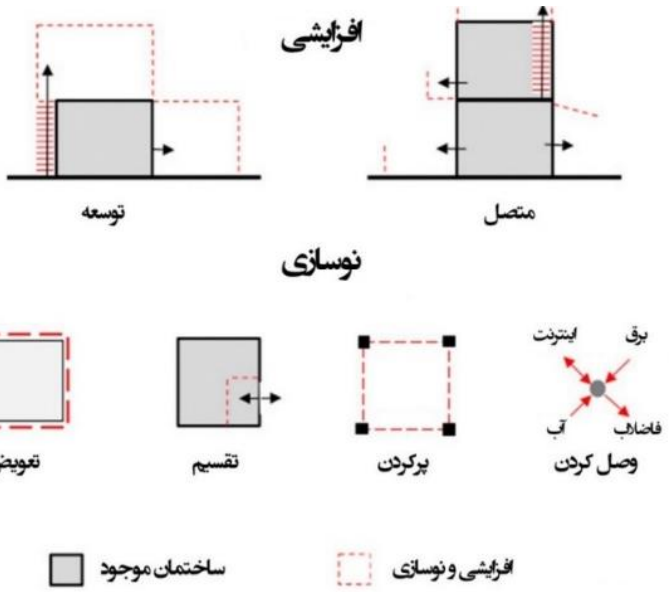
۲. Attach

۳. Replace

۴. Divide

۵. Infill

۶. Connect



شکل ۳: گونه هایی از گسترش در مسکن تدریجی

بنا بر آنچه در فوق بیان شد رشد و گسترش می تواند درونی یا بیرونی باشد. به عنوان مثال رشد درونی، یک فضای باز یا نیمه باز تبدیل به یک اتاق مجزا گردد و گسترش بیرونی نیز بام هایی است که در آینده می توان بر روی آن ها فضاهای جدیدی ساخت[۱۶].

در مرحله گسترش برای استفاده از انعطاف‌پذیری باید از راهبرد ها و روش های مختلف طراحی مسکن انعطاف پذیر پرداخته می شود. به عنوان مثال، طرح ساختمان باید انتخاب های مختلفی را برای کاربران دارای انعطاف‌پذیری اولیه جهت برآورده کردن انعطاف‌پذیری دائم روش های متفاوت زندگی ارائه دهد، و واحد مسکونی برای برآورده کردن باید مقدمات لازم را در طول زمان و با توجه به نیازهای در حال تغییر و خواسته های کاربران فراهم آورد[۱۷]. در رویکرد گسترش، جایگیری حداقل هسته اولیه در یک گوشه رایج به نظر می رسد. در بیشتر نمونه های برسی شده، فضای تر چسپیده به یک وجه در مرکز پلان یا در گوشه آن قرار گرفته است. دلیل اصلی این امر آن است که قرار گرفتن هسته اولیه در گوشه پلان و در مجاورت سطح توسعه پذیر، و به ویژه در صورت زیاد بودن مساحت آن، به علت سهولت دسترسی توسعه را آسان‌تر می کند[۱۶].

۴. انعطاف پذیری[۱]

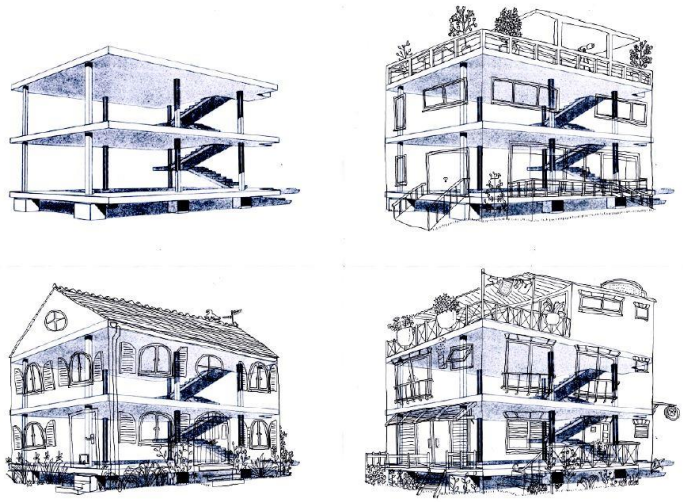
طوری که برای همگان مشهود است جوامع بشری به تغییرات دایمی و اجتناب‌ناپذیر خو گرفته‌اند و جامعه‌ای که دائما در حال تغییر و پیشرفت است، پیش بینی جمعیت، اقتصاد و الگوی زندگی مردم دشوار می باشد. پیش بینی چگونگی طراحی و اجرای خانه های آینده نیز از این امر مستثنی نیست. این تغییرات ابداع شیوه های طراحی معماری جدید را که در آن به لزوم انعطاف‌پذیری

۱ Flexibility

خانه‌ها و تطبیق آنها با تمایلات پویای اجتماع و زندگی ساکنان تاکید شده باشد الزامی می نماید. از این رو تلاش برای مطرح کردن انعطاف‌پذیری نتیجه سرعت تغییرات جمعیتی، فرهنگی، اجتماعی، اقتصادی، تکنولوژیکی و فرهنگ زندگی مردم در جامعه کنونی است[۱۴].

پژوهشگران اصطلاحاً تعریف از واژه انعطاف‌پذیری را در مهندسی و طراحی محیط و به طور خاص طراحی مسکن، در انعطاف‌پذیری فضایی و چیدمان فضای انسان‌ساخت و تغییر در آن برای دستیابی به شرایط، نیازها، و کاربری های جدید می دانند[۱۸]. حال اگر بتوان با مطرح کردن انعطاف‌پذیری، خانه‌ای طراحی نمود که در زمان اسکان پاسخگوی نیاز های خانواده بوده و در عین حال امکان تکمیل و گسترش خانه بدون این که به زندگی خللی وارد آید در طراحی آن پیش‌بینی شده باشد، آنگاه خانواده ها می توانند با وسع مالی کم، صاحب خانه شده و به مرور زمان و ارتقاء سطح مالی خود، کیفیت و کمیت خانه خود را بالا برده و به طور مناسبی از آن استفاده نمایند[۱۴]. قابلیت انتخاب در مرحله هسته اولیه خانه قبل از سکونت موسوم به انعطاف‌پذیری اولیه ایده طراحی بلاک های مسکونی مناسب برای کاربران گوناگون است و فضاها بصورت بالقوه انعطاف‌پذیر طراحی می گردد که نشان دهنده تفکر در طراحی می باشد [۱۷]. بدیهی است انعطاف‌پذیری نمی تواند تمامی این تغییرات را پوشش دهد، ولی با توجه به میزان تغییراتی که در سطوح مختلف اجتماعی، اقتصادی و فرهنگی خانواده رخ می دهد، می‌تواند تا حد قابل توجهی بخشی از نیاز های جدید و متغیر را تا حد ممکن بدست آورد.

بنابراین در فرآیند طراحی و ساخت مسکن، خانه‌ها انواع گوناگونی از نما و طرح بندی‌های مناسب برای خواسته‌های ساکنان وجود خواهد داشت. خانه هایی که در طراحی آن ها امکان بازسازی و ارتقاء کیفیت پیش‌بینی شده باشد، مطلوبیت بیشتری داشته و عمر مفید خانه ها نیز افزایش می یابد[۱۴].



شکل ٤: خانه دومینو لوکوبوزیه (بالا چپ) در طرح های متفاوت

هر خانه در یک دوره زمانی مشـخص و بر اسـاس سـبک مهندسـی آن دوره طراحی شـده، از تکنیک های خاص آن زمان استفاده می شـود و به وسـیله فرهنگ زیستی مردم هویت پیدا می کند. هنگامی که زمان می گذرد، عوامل موثر بر سـبک طراحی، سـاخت و هویت خانه‌ها به تاریخ می‌پیوندند و در نتیجه پاسـخگوی نیاز های سـاکنان معاصـر خود نمی‌باشـند. این عوامل که به صـورت اجتناب‌ناپذیری از خود سـاکنان شـروع شـده و باعث روند ایجاد تغییرات می شود، شامل موارد ذیل می‌گردد[۱۴]:

الف) دگرگونی در ساختار خانواده
ب) تغییر سطح اجتماعی، اقتصادی و فرهنگی خانواده
ج) نامشخص بودن مخاطبان خانه‌های مسکونی
د) روحیه تنوع طلبی انسان
ه) انطباق با تکنولوژی‌های جدید
و) ضرورت های اقلیمی

۵. نمونه طرح

در بخش طراحی هدف و تلاش اصلی ایجاد سکونتگاهی ارزان، با اجرای آسان و کارآمد و پربازده برای اسکان دائم قشر کم‌درآمد به عنوان راه‌حل اساسی می باشد. این نوع مسکن فرآیندی تدریجی در پاسخ به تغییرات جمعیتی و وضعیت اقتصادی است که به موجب آن منتفع‌شونده در مراحل بعدی، هر زمان که بودجه یا زمان در دسترس داشته باشند طبق سلیقه خویش فضای مدنظرش را به تدریج می‌تواند بهبود بخشیده و یا گسترش دهد. بناءً این طرح می‌تواند پایه‌ای برای گسترش‌های بعدی و الگویی برای طرح‌های مشابه باشد. در تطبیق این نوع طرح‌ها افزایش همبستگی و مشارکت شهروندان در فرآیند ساخت بسیار مهم می باشد، بنابراین فعالیت‌های ساکنین در ساخت این نوع مسکن می‌بایست نمود پیدا کند.

آشپزخانه و تشناب به عنوان فضاهای با قابلیت انعطاف کم و به علت سختی در تطبیق‌شان توسط خود ساکنین، به عنوان بخشی از هسته اولیه و اصلی خانه در مرحله ابتدایی ساخت طراحی می‌گردند تا بخش‌های باقی‌مانده خانه بدون هیچ‌گونه محدودیتی، شرایط لازم برای گسترش توسط ساکنین را پیدا کنند. بنابراین هسته اولیه خانه، شامل یک فضای عمومی برای فعالیت های مختلف، آشپزخانه و تشناب می‌باشد. به صورت کلی جهت تکمیل تدریجی یک واحد سه مرحله پیشنهاد می‌شود.

مرحله اول: ساخت هسته اولیه شامل تشناب، آشپزخانه و یک سالن چندمنظوره
مرحله دوم: توسعه فضاها با توجه به نیازهای خانوار
مرحله سوم: انعطاف‌پذیری در چگونگی تکمیل فضاها به صورت افقی و عمودی

۵-۱. مراحل توسعه تدریجی

روند طراحی این واحد با درنظر داشتن اهداف زیر مطابق نتایج و معلومات از قبل به دست آمده، در پیش‌برد طراحی تعقیب می شود.

الف) بهره‌گیری از روش تکمیل تدریجی در راستای کاهش هزینه‌ها با ایجاد هسته اصلی واحدها در مرحله ابتدایی و تکمیل تدریجی بنا به مرور زمان

ب) تلاش برای ایجاد انعطاف‌پذیری در مراحل گسترش فضا بر اساس نیاز خانوار

ج) توجه به افزایش همبستگی و تقویت حس تعلق شهروندان با مشارکت آنها در فرآیند ساخت

د) طراحی فضاهای داخلی منطبق و متناسب با نیازهای جامعه‌ی آماری مورد نظر و پیش‌بینی فضاهای مورد نیاز در مرحله توسعه آتی

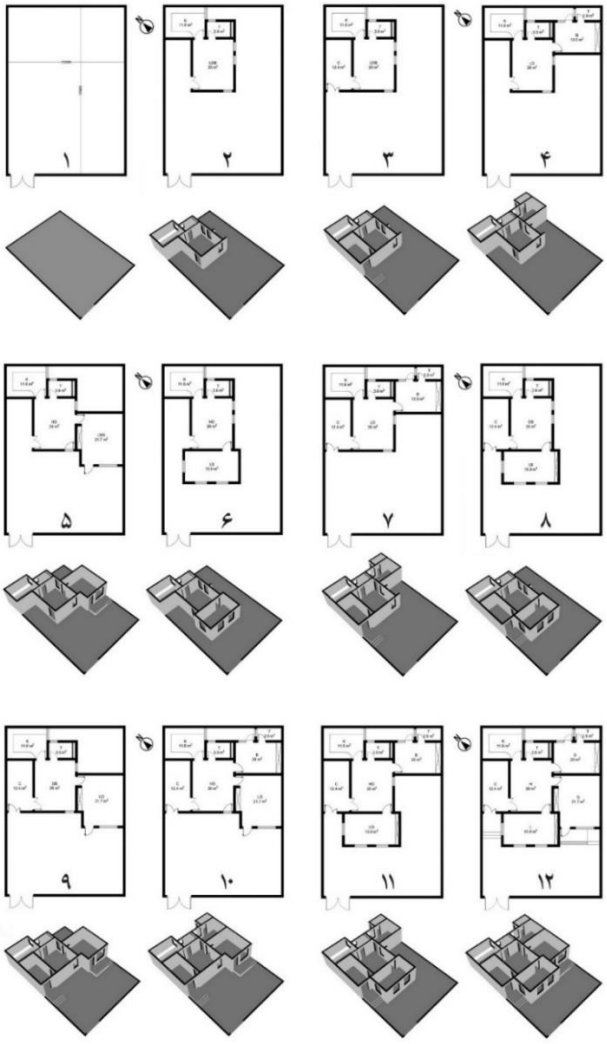
ه) توجه به فضاهای چند عملکردی عمومی

۵-۲. روند شکل گیری طرح

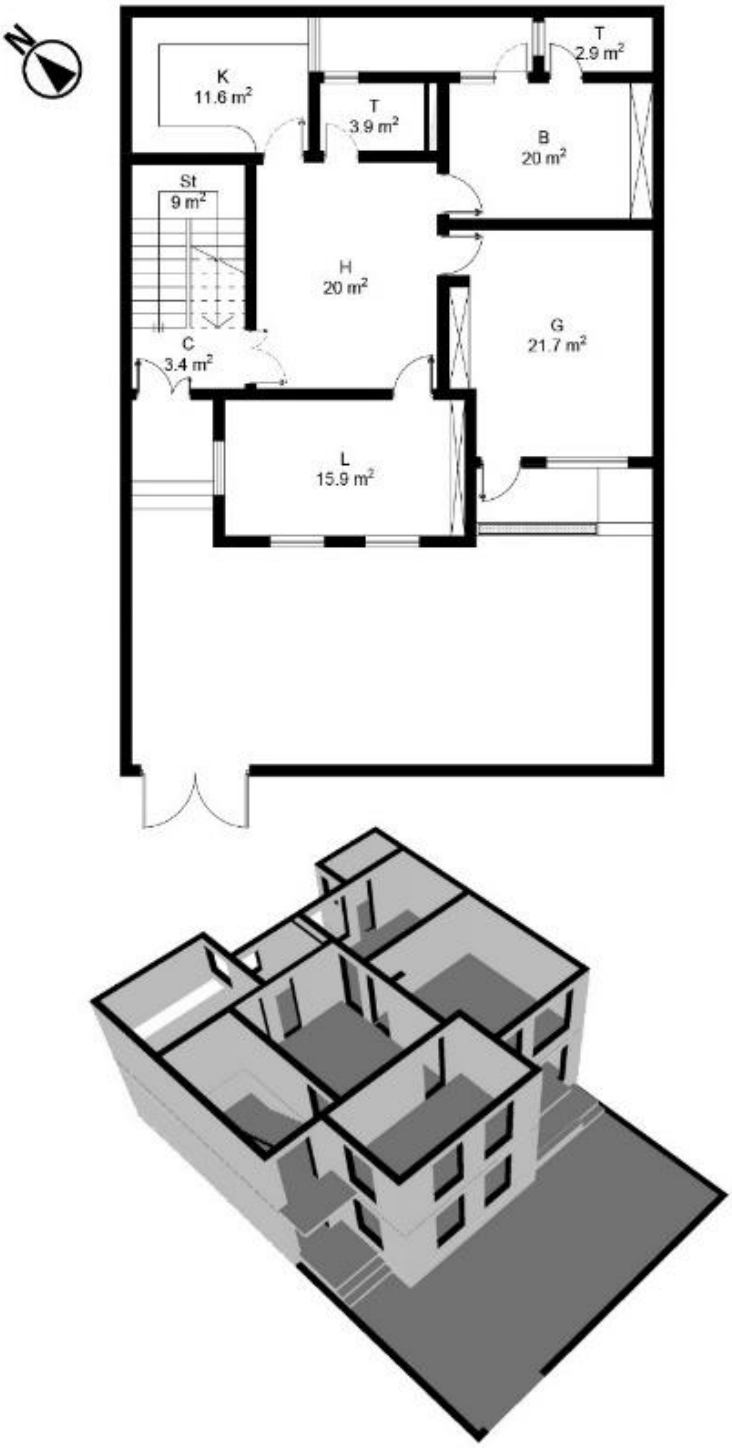
برای رسیدن به طرحی مطلوب، مراحلی در نظر گرفته شده است که روند شکل گیری واحد را نشان می‌دهد. هسته اولیه طرح با مساحت حداقلی ۴۳ مترمربع شروع شده که حداقل فضای مورد نیاز برای استراحت یک خانوار (۷ نفره) بوده و با تکمیل آن در دو طبقه، مساحت آن به ۲۵۶ مترمربع می رسد. طرح فضای هسته اولیه شامل فضای چندعملکردی (LDB)، آشپزخانه (K) و تشناب (T) می شود.

هسته اولیه این طرح طوری در نظر گرفته شده است که مراحل رشد و گسترش آن دارای انعطاف‌پذیری و انتخاب‌های مختلفی از فضا بر اساس نیاز، توانایی مالی و سلیقه خانوار باشد. موقعیت کلکین‌ها در فضای چندعملکردی هسته اولیه طوری جاگذاری شده است که در مراحل بعدی گسترش قابلیت استفاده مجدد و نصب دوباره در اتاق‌های خواب و نشیمن را دارند و از مصارف اضافی جلوگیری خواهد شد.

در مرحله دوم، خانوار گزینه های مختلف توسعه افقی را روی دست دارد تا فضای مورد نیازش را رشد دهد. در این مرحله رشد امکان انتخاب فضاهای اتاق خواب (B)، مهمانخانه (G)، نشیمن (L) و دهلیز (C) وجود دارد که در پلان های ۳ تا ۱۲ نشان داده شده است. این انعطاف‌پذیری در مراحل بعدی گسترشِ عمودی نیز وجود خواهد داشت.



شکل ۵: انعطاف‌پذیری در مراحل رشد مسکن تدریجی



شکل ٦: تکمیل رشد افقی و عمودی مسکن تدریجی

شکل ۷: مرحله هسته اولیه با مصالح ساختمانی خشت پخته، خشت خام و سنگ و گابیون

البته بر اساس معیارهای دسترسی به مصالح ساختمانی و هزینه آن، ممکن است در بعضی مناطق کشور بهترین گزینه ساخت مسکن تدریجی، استفاده از خشت خام و گِل، سنگ و گابیون و یا مصالح دیگر مناسب‌ترین گزینه باشند.

شکل ۸: مرحله تکمیل شده مسکن تدریجی با مصالح خشت پخته، خشت خام و سنگ و گابیون

نتیجه‌گیری

نتایج حاصله از بررسی پیشینه‌ی تحقیقِ موضوع نشان می دهد که مسکن تدریجی روشی مناسب برای تهیه مسکن گروه های کم‌درآمد جامعه در کشور های در حال توسعه است. مشارکت و همیاری ساکنان در مراحل طراحی و ساخت و ساز مسکن تدریجی باعث ایجاد حس تعلق در آنها شده و یکی از عوامل موفقیت در این روش به شمار می رود. همچنان محصول نهایی از نظر اقتصادی، عملکردی و زیبایی شناختی برای گروه های هدف مناسب‌تر خواهد بود.

در طرح نمونه پیشنهادی مشاهده شد که انعطاف‌پذیری و حق انتخاب آگاهانه در گسترش آینده و استفاده از مصالح ساختمانیِ در دسترس که از اصلی‌ترین ویژگی های استطاعت‌پذیری است، باعث رضایت‌مندی منتفع‌شونده در این روش ساخت مسکن می شود.

منابع

[1] Seo, K. W. (2016). Incremental SI (structure-infill) Housing for Low-income Population in Malaysia.

[2] Adhikari, S. (2019). Incremental Housing: Design Approach for Kathmandu.

[3] Goethert, R., & Director, S. (2010). Incremental housing. Monday developments, 9, 23-25.

[4] Wibowo, A. H., & Larasati, D. (2018, May). Incremental Housing Development6; an Approach in Meeting the Needs of Low Cost Housing in Indonesia. In IOP Conference Series: Earth and Environmental Science (Vol. 152, No. 1, p. 012006). IOP Publishing.

[5] Van Noorloos, F., Cirolia, L. R., Friendly, A., Jukur, S., Schramm, S., Steel, G., & Valenzuela, L. (2020). Incremental housing as a node for intersecting flows of city-making: rethinking the housing shortage in the global South. Environment and Urbanization, 32(1), 37-54.

[6] Croese, S., Cirolia, L. R., & Graham, N. (2016). Towards Habitat III: Confronting the disjuncture between global policy and local practice on Africa's 'challenge of slums'. Habitat International, 53, 237-242.

[7] Marinovic, G. I. (2020). The guideline for customizing incremental housing based on two Chilean case studies. Journal of Architecture and Urbanism, 44(2), 166-175.

[8] Richard Bahumwire,(2015) Innovative Financing for Housing the Urban Poor in Uganda The effect of incremental Housing Finance on access to a core-house and its upgrading: The Case of Stanbic Nama Project in Mukono Uganda.

[9] Rai, N. (2018). A study on self-help incremental housing: Lattakia case (Doctoral dissertation, Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü).

[10] Rosa, W. (2017). Goal 11. Make Cities and Human Settlements Inclusive, Safe, Resilient, and Sustainable. A New Era in Global Health: Nursing and the United Nations 2030 Agenda for Sustainable Development, 339.

[11] Laquian, A. A. (1983). Basic housing: policies for urban sites, services, and shelter in developing countries. IDRC, Ottawa, ON, CA.

[12] Pandelaki, E. E., & Shiozaki, Y. (2010). The core house concept and its implementation in Indonesia: past, present, future. International Journal for Housing Science and Its Applications, 34(4), 233.

[13] Park, J., Lim, Y., Kim, K., & Wang, H. (2019). Revisit to Incremental Housing Focusing on the Role of a Comprehensive Community Centre

[١٤]خدایی، زهرا، محمد، میره، ۱۳۹۰، مروری بر تجارت جهانی در زمینهٔ تأمین مسکن گروه‌های کم‌درآمد شهری، اولین کنفرانس اقتصاد شهری، دانشگاه فردوسی، مشهد.
[١٥]علی الحسابی،مهران، برهانی داریان،فرناز(١٣٨٦) «مسکن انعطاف پذیر، نشریه مسکن و محیط روستا» شماره ١١٧،صص ٢٠-٣٥
[١٦]رسولی ثانی آبادی, آیدا, فرهادی غفاری, علی. (٢٠٢٠). بررسی تطبیقی الگوهای جایگیری فضای خیس درانعطاف‌پذیری مسکن. معماری و شهرسازی ایران (JIAU), ١١(١), ١٣١-١٤٤.
[١٧]زندیه, اقبالی, سید رحمان, حصاری, پدرام. (٢٠١١). روش های طراحی مسکن انعطاف پذیر. نقش جهان-مطالعات نظری و فناوری های نوین معماری و شهرسازی, ١(١), ٩٥-١٠٦.
[١٨]عینی فر, علیرضا. (٢٠٠٣). الگویی برای تحلیل انعطاف‌پذیری در مسکن سنتی ایران. هنرهای زیبا, ١٣(١٣).

کاربردهای شبکه عصبی مصنوعی و یادگیری عمیق در شناسایی آسیب وپایش سلامت سترکچرها -مطالعه مروری

عباس محبی[۱]، سید مهدی موسوی[۲]

چکیده

در سال‌های اخیر، سیستم های پایش سلامت ستکچرها[۳] (SHM) بصورت گسترده در زیربناهای مختلف مدنی برای بررسی وضعیت صحت ستکچرها و تشخیص آسیب سترکچری یا غیرطبیعی بودن آنها از طریق نظارت طولانی مدت برروی سترکچرها نصب شده اند. شیوه-های مرسوم تحلیل معلومات با چالش‌های سروصدای محیطی، حجم معلومات اندازه گیری، پیچیدگی محاسبه، وغیره روبرو هستند و آنها کاربرد فراگیر تکنالوژی SHM را به شدت محدود می‌سازد. در سال‌های اخیر، با پیشرفت سریع سخت افزار محاسباتی و تجهیزات اکتساب تصویر، شیوه پروسس معلومات مبتنی بر یادگیری عمیق یک کانال جدیدی را برای حفر معلومات عظیم از یک سیستم SHM، به سمت پروسس خودمختار، دقیق و قوی معلومات نظارت فراهم می‌سازد. این مقاله یک مرور را در مورد SHM مبتنی بر یادگیری عمیق از جمله یک خلاصه مختصر از تاریخچه توسعه یادگیری عمیق، کاربردهای شیوه‌های پروسس داده‌های مبتنی بر یادگیری عمیق در SHM از انواع مختلف ارائه می‌دهد

واژگان کلیدی: شبکه عصبی مصنوعی، یادگیری عمیق، شناسایی آسیب، پایاش سلامت سترکچرها.

---

۱. استاد پوهنحی کامپیوترساینس دانشگاه خاتم النبیین؛ ab.mohebbi@knu.edu.af

۲. استاد پوهنحی انجنیری دانشگاه کاتب؛ s.m.mousavi@kateb.edu.af

۳. Structural Health Monitoring

مقدمه

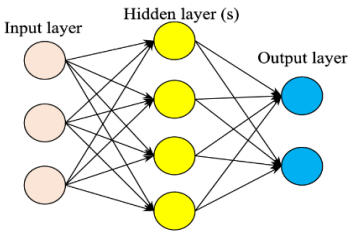
انجنیری سترکچر شامل تجزیه و تحلیل سترکچر و دیزاین آن در برابر بارهای وارده می‌باشد. برای سیستم‌های سترکچری پیچیده تحت بارهای شدید که در آن سترکچر رفتار بسیار غیرخطی از خود نشان می‌دهند، استفاده از روش‌های تحلیل و دیزاین سترکچری نیازمند یک فرآیند زمان‌بر است و به نوعی در عمل بسیار پیچیده و دوشوار خواهد بود. در این مورد، شبکه‌های عصبی و یادگیری عمیق می تواند یک جایگزین امیدوارکننده برای صرفه جویی در زمان انجام تحلیل، ارائه دهد. یکی از اولین کاربردهای شبکه عصبی مصنوعی در انجنیری سترکچر توسط عادلی و یه [۱] در سال ۱۹۸۹ برای دیزاین گادرهای فولادی انجام شد [۲]. از آن زمان، الگوریتم شبکه عصبی مصنوعی و یادگیری عمیق با موفقیت در بسیاری از کارهای پیشگام در انجنیری سترکچر از جمله تحلیل و دیزاین سترکچر [۳] ، شناسایی آسیب سترکچرها[٤]، پایش سلامت سترکچرها (SHM) [٥]، منحنی‌های شکنندگی [۶-۹] استفاده شده است. هدف از این مقاله، بررسی برخی از ویژگی‌های شبکه‌های عصبی و یادگیری عمیق و کاربرد آن در در شناسایی آسیب وپایش سلامت سترکچرها است.

۱. پیشینه تحقیق

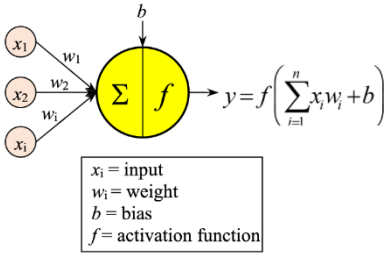
۱-۱. شبکه عصبی مصنوعی

یک ANN که ساختاری شبیه به مغز انسان دارد از نورون‌های مصنوعی که به عنوان واحدها یا گره ها نیز شناخته می‌شوند، تشکیل شده است. این گره‌ها به طور کامل به هم متصل شده و در سه لایه مختلف چیده شده‌اند که در شکل ۱ نشان داده شده است. لایه ورودی داده های ورودی x را دریافت می‌کند، در حالی که لایه خروجی نتایج پیش بینی شده y شبکه را نشان می‌دهد. ساندویچ بین لایه های ورودی و خروجی یک یا چند لایه پنهان از واحدهای محاسباتی است (شکل ۲)که محاسبات اصلی ریاضی را روی داده های ورودی انجام می‌دهد. هنگامی که یک ANN با دو یا چند لایه پنهان طراحی می‌شود، به آن پرسپترون چندلایه یا یادگیری عمیق (DL) می‌گویند [۱۰،۱۱]. رفتار هر واحد نورون با وزن‌هایی که به آن اختصاص داده می‌شود، تعریف می‌شود. هنگامی که داده‌های x_i به لایه ورودی وارد می‌شوند، در وزن‌های مربوطه w_i ضرب می‌شوند. سپس، از یک تابع انتقال برای محاسبه مجموع وزنی ورودی به اضافه یک بایاس b استفاده می‌شود که به تدریج تنظیم می‌شود تا اختلاف بین خروجی های پیش بینی شده و واقعی را به حداقل برساند. در نهایت مقدار تابع انتقال از طریق تابع فعال سازی f ارسال می‌شود تا بررسی شود که آیا گره باید داده‌ها را به لایه خروجی ارسال کند یا خیر. هنگامی که یک ANN آموزش داده می‌شود، مقادیر تصادفی وزن‌ها به همه گره‌ها اختصاص داده می‌شود. هنگامی که تابع فعال‌سازی مقدار پیش بینی شده y را به لایه خروجی منتقل می‌کند، از یک تابع خطا برای محاسبه تفاوت بین خروجی‌های پیش بینی شده و واقعی استفاده می‌شود. بر اساس نتیجه، مدل

ANN وزن تمام گره‌های خود را برای به حداقل رساندن خطا تنظیم می‌کند. چنین فرآیند آموزشی که به عنوان پس انتشار [۱۱] شناخته می‌شود تا زمانی که شرط همگرایی برآورده شود، تکرار می‌شود. تعدادی هایپر پارامتر وجود دارد که باید هنگام آموزش یک مدل ANN بهینه شوند زیرا فرآیند یادگیری و آموزش شبکه را کنترل می‌کنند. آن‌ها شامل انتخاب‌های معماری یک شبکه عصبی مصنوعی (یعنی تعداد لایه‌های پنهان استفاده‌شده، تعداد گره‌ها در هر لایه پنهان، و نوع توابع فعال‌سازی) و متغیرهای آموزشی (یعنی نرخ انحراف، تعداد دوره‌ها، تکانه، و اندازه دسته) می باشند. افزایش تعداد لایه‌های پنهان و گره‌های پنهان می‌توانند دقت شبکه را افزایش دهد، اما باعث هزینه محاسباتی می شود.تابع فعال‌سازی برای محاسبه غیر خطی بودن مدل ها استفاده می‌شود. نرخ یادگیری تعیین می‌کند که مدل ANN با چه سرعتی پارامترهای خود را به روز می‌کند. استفاده از نرخ بالای یادگیری باعث سرعت بخشیدن به یادگیری می‌شود، اما ممکن است نتواند همگرا شود. Epoch تعداد تکرارها را برای مجموعه داده آموزشی کنترل می‌کند، در حالی که حرکت برای انتخاب جهت مرحله بعدی از مرحله قبل استفاده می‌شود که می‌تواند به جلوگیری از نوسانات کمک کند. چنین پارامترهای فوق در طول مراحل آزمایش و اعتبار سنجی تنظیم می‌شوند.



شکل ۱: معماری شبکه عصبی



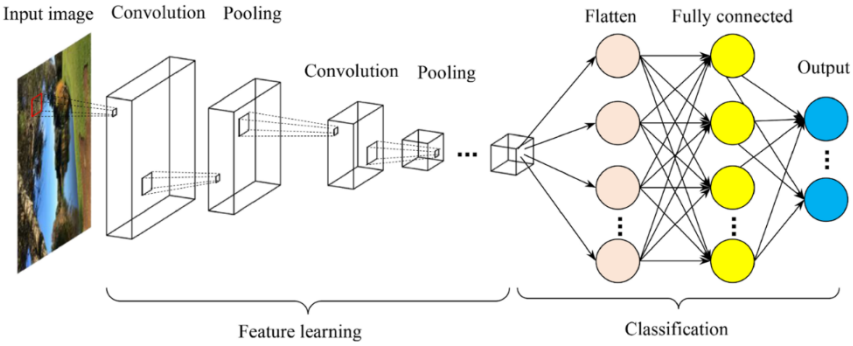
شکل ۲: واحد محاسبات شبکه عصبی

۲-۱. شبکه عصبی عمیق

الگوریتم‌های زیادی در شبکه عصبی عمیق وجود دارد که در اینجا به پر کاربرد ترین آن‌ها یعنی CNN1 یا همان شبکه عصبی کانولوشنی می‌باشد؛ پرداخته خواهد شد. این الگوریتم برای اولین بار در دهه ۱۹۸۰ توسط یان لیکون، یک محقق علوم کمپیوتر فرانسوی که یکی از نسخه های اولیه معماری های CNN را برای تشخیص دیجیتلی دست نویس به نام LeNet [۱۲] ساخت، معرفی گردید. نظریه اساسی CNN ایجاد یک شبکه است که در آن هر لایه می‌تواند معلومات را از لایه های قبلی به معلومات پیچیده‌تر تبدیل و به لایه‌های بعدی منتقل کند. طوری‌که در شکل ۳ نشان داده شده است، یک CNN بر اساس دو بلاک اساسی ساخته شده است: بلاک یادگیری ویژگی و بلاک طبقه بندی. یادگیری ویژگی که از تعدادی از لایه‌های

۲. Convolution Neural Network

متناوب پیچیدگی و استخراج تشکیل شده است برای استخراج و یادگیری ویژگی از تصویر ورودی استفاده می‌شود. ویژگی استخراج شده یا یاد گرفته شده بعد از طریق بلاک طبقه بندی طبقه‌بندی می‌شود. معماری‌های مختلف CNN همچنان برای بهبود عملکرد CNN در کاربردهای مختلف بوجود آمده‌اند. قابل توجه در میان آنها LeNet، AlexNet، VGG و ResNet می‌باشد [۱۳،۱٤]. در متن SHM، CNN بصورت گسترده برای تشخیص ترک در سترکچرها بر اساس شیوه‌های طبقه بندی تصویر یا تخنیک‌های بخش بندی استفاده شده است. میتود طبقه بندی تصویر به عوض سطح پیکسل که در میتود تقسیم‌بندی استفاده می‌شود، می‌تواند ترک را در سطح تصویر تشخیص دهد.



شکل ۳: معماری CNN معمولی

۲. کاربردهای شبکه عصبی و یادگیری عمیق در انجنیری سترکچر

۲-۱. شناسایی آسیب وپایش سلامت سترکچرها

بازرسی خسارت سترکچری برای ایمنی سترکچرهای در حال خدمت ضروری است و از این‌رو بسیاری از گروه های تحقیقاتی از روش های مبتنی بر یادگیری عمیق برای انجام تشخیص خسارت در سترکچرهای مختلف استفاده کرده‌اند .کاربردهای مطالعات مبتنی بر یادگیری عمیق جمع آوری و در جدول ۱ لیست شده است. مطالعات مبتنی بر تصویر و CNN متعدد وجود داشته است زیرا انواع زیادی از آسیب‌های سازه‌ای قابل مشاهده است. برای غلبه بر کمبود مجموعه معلومات تصویر حاشیه‌نویسی شده برای مقاصد مشخص بازرسی، یادگیری انتقال توسط آموزش قبلی با تعداد زیادی از مجموعه معلومات تصویر منبع باز و تنظیم دقیق با تعداد کمی از تصاویر جمع آوری شده تطبیق شده است. همچنان، تخنیک‌های مرسوم افزایش معلومات و همچنان شیوه‌های مبتنی بر یادگیری عمیق مانند GAN برای بزرگ کردن ست‌های معلومات استفاده شده است. برای تشخیص، بومی‌سازی و کمیت کردن آسیب‌های سترکچری مانند شکستن و ترک ها، شیوه های Faster R-CNN و FCN برای موقعیت‌یابی دقیق آسیب‌ها روی دست گرفته شدند و تخنیک‌های پروسس تصویر (IPTs) برای بدست آوردن پارامترهای

آسیب بکار برده شده‌اند. برعلاوه تصاویر، معلومات سلسله زمانی مانند شتاب و جابجایی برای تشخیص تخریب در آن مطالعات استفاده شده است. برای پروسس کردن معلومات سلسله زمانی، شبکه‌های رمزگذار خودکار و D-CNN۱ توسط چندین گروه تحقیقاتی بوجود آمدند. علاوه بر این، تبدیل کردن معلومات خام سلسله زمانی به طیف‌های فرکانس یا طیف‌های فرکانس زمانی فضایی برای پروسس بیشتر نیز در حال بررسی است.

جدول ۱: کاربرد یادگیری عمیق در پایش سلامت سترکچرها و شناسایی آسیب

Structure type	Application	Reference	Technology
Bridge	Crack detection	Alipour *et al.* (2019)	FCN
		Dung *et al.* (2019)	VGG-16+Transfer learning
		Kim *et al.* (2018)	UAV+R-CNN+Transfer learning+IPT
		Sajedi and Liang (2019)	SegNet
	Damage detection	Bao *et al.* (2019)	Auto-encoder+Unsupervised learning
		Duan *et al.* (2019)	CNN
		Liang (2019)	VGG-16+Faster R-CNN+SegNet
		Tang *et al.* (2019)	CNN
		Yeum *et al.* (2019)	CNN+UAV+Structure from motion
	Loosened bolt detection	Huynh *et al.* (2019)	R-CNN
	Damage state classification	Khodabandehlou *et al.* (2019)	CNN
Tunnel	Crack detection	Li *et al.* (2019)	Faster R-CNN
		Song *et al.* (2019)	ResNet+MobileNet+CrossNet
	Multiple damage detection	Huang *et al.* (2018)	FCN
		Gao *et al.* (2019)	Faster R-CNN+FCN
		Xue and Li (2018)	FCN+Faster R-CNN
Concrete building	Crack detection	Cha *et al.* (2017)	CNN
		Dorafshan *et al.* (2018)	AlexNet+Transfer learning
		Dung and Anh (2019)	FCN+Transfer learning
		Kang and Cha (2018)	UAV+CNN
		Kim and Cho (2018)	UAV+AlexNet+Transfer learning
		Kim and Cho (2019)	Mask R-CNN
		Ni *et al.* (2019)	GoogLeNet+ResNet
		Ni *et al.* (2019)	GoogleNet+Transfer learning
		Yang *et al.* (2018)	VGG-19+FCN
		Ye *et al.* (2019)	FCN
		Zhang *et al.* (2019)	SegNet
		Zhang *et al.* (2019)	ResNet+FCN
	Multiple damage detection	Gao and Mosalam (2018)	VGG+Transfer learning
		Li *et al.* (2018)	Faster R-CNN
		Li *et al.* (2019)	DenseNet+FCN
		Lin *et al.* (2017)	CNN
		Wang *et al.* (2018)	AlexNet+GoogLeNet
		Xu *et al.* (2019)	Faster R-CNN
		Yeum *et al.* (2018)	AlexNet
	Spalling detection	Beckman *et al.* (2019)	Faster R-CNN
	Damage dataset generation	Gao *et al.* (2019)	GAN
Steel building	Damage detection	Gulgec *et al.* (2019)	CNN
		Liu and Zhang (2019)	CNN
		Pathirage *et al.* (2018)	Auto-encoder
		Yu *et al.* (2019)	CNN
		Zhao *et al.* (2019)	VGG-16+MobileNet
	Multiple damage detection	Chen and Jahanshahi (2018)	CNN+Naive Bayes
		Wu *et al.* (2019)	VGG-16+ResNet-18
	Stiffness degradation detection	Zhou *et al.* (2019)	Auto-encoder
	Joint damage detection	Abdeljaber *et al.* (2017)	1D-CNN
	Corrosion detection	Atha and Jahanshahi (2018)	CNN
	Crack detection	Cha *et al.* (2018)	Faster R-CNN

۲-۲. پل‌ها

کیم و همکاران یک شیوه مبتنی بر UAV و R-CNN را برای تشخیص ترک‌ها در پل‌های کانکریتی قدیمی پیشنهاد کرد[15] . یک R-CNN از قبل آموزش دیده توسط تصاویر ترک برای تشخیص ترک تنظیم شده بود و IPTs برای سنجش ترک‌های تشخیص شده بکار برده شد. لیانگ یک میتود مبتنی بر یادگیری عمیق سه سطحی را برای بازرسی پل‌های پس از فاجعه پیشنهاد کرد. VGG-16 برای تشخیص خرابی سطح سیستم بکار برده شد، و Faster R-

CNN و SegNet برای تشخیص آسیب سطح جزء و سطح محلی به ترتیب بکار برده شدند [16]. یک CNN کم عمق که از ابتدا آموزش دیده بود، یک VGG-16 از قبل آموزش دیده با یک طبقه‌بندی کننده دقیق، و یک VGG-16 از قبل آموزش دیده با یک لایه پیچشی و طبقه‌بندی کننده دقیق با استفاده از میزان دقت، میزان
دقت و میزان یادآوری. تصاویر خام از آزمایش‌ها و بازرسی‌های روزانه برای ایجاد مجموعه معلومات جمع آوری شد و افزایش معلومات برای کاهش تناسب بیش از حد روی دست گرفته شد. هوین و همکاران یک شیوه مبتنی بر تبدیل خط R-CNN و Hough را برای تشخیص پیچ‌های شل شده اتصالات فولادی پیشنهاد کرد. یک R-CNN پانزده لایه بدون تصاویر پیچ از قبل آموزش دیده و با تصاویر پیچ تنظیم شده بود[17]. الگوریتم تبدیل خط هوف برای ارزیابی وضعیت شل شدن پیچ‌های تشخیص شده روی دست گرفته شد. علیپور و همکاران یک شیوه مبتنی بر FCN را برای تشخیص ترک‌ها برای ارزیابی ترک بازدید شده پیشنهاد کرد[18]. پنج مدل با میزان‌های مختلف نمونه‌گیری بر اساس حالت از قبل آموزش دیده آزمایش شدند. مجموعه معلومات تصویر توسط تصاویر شکاف جمع‌آوری شده در محل با حاشیه‌نویسی محتاطانه ایجاد شد و تاثیر اندازه مجموعه معلومات تحلیل شد.

۲-۳. تونل‌ها

ژو و لی یک چارچوب سه مرحله‌ای مبتنی بر یادگیری عمیق را برای طبقه‌بندی و محلی‌سازی آسیب‌های پوشش تونل توسعه دادند[19]. یک FCN برای استخراج نقشه‌های ویژگی تصاویر ورودی بوجود آمد، یک شبکه پیشنهاد سایت برای انتخاب ساحات مشکوک در نقشه های ویژگی بکار برده شد، و یک میتود استخراج حساس به موقعیت برای یافتن دقیق خسارات نیز بکار برده شد. هوانگ و همکاران یک شیوه دو جریان مبتنی بر FCN را برای تطبیق بخش‌بندی معنایی برای درزها و نشت‌ها در تونل‌ها بکار برد[20]. مقایسه عملکرد در بین شیوه پیشنهادی، یک الگوریتم رشد ساحه، و یک الگوریتم آستانه‌سازی تطبیقی انجام گردید. سونگ و دیگران عملکرد سه نوع مختلف DNNs را برای بخش‌بندی معنایی درزهای تونل مقایسه نمودند[21].

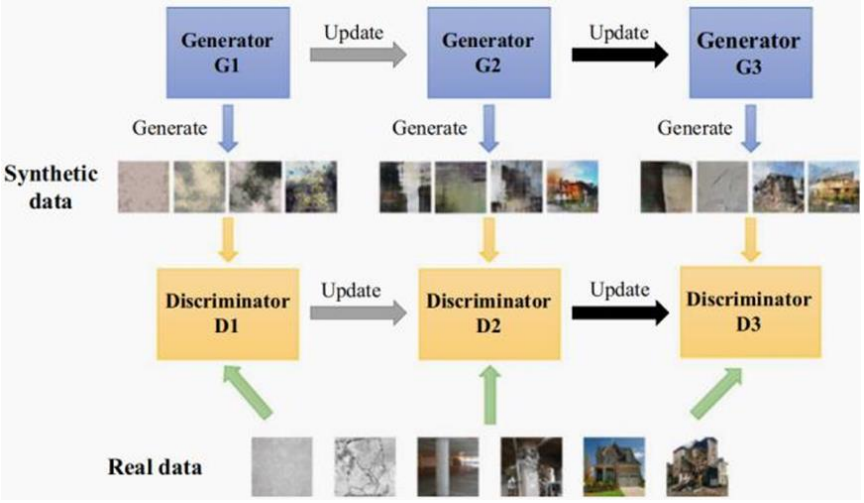
برای آموزش شبکه‌های آزمایش شده، تصاویر تونل از موقعیت‌های دنیای واقعی جمع آوری شد و یک مجموعه معلومات درز تونل با حاشیه نویسی بخش‌بندی معنایی ایجاد شد. گو و همکاران یک چارچوب مبتنی بر Faster R-CNN و FCN را برای تشخیص سریع و دقیق نقص‌های تونل متعدد ایجاد نمودند [22] . لی و همکاران (۲۰۱۹) یک پردازش تصویر و چارچوب مبتنی بر R-CNN را برای تشخیص ترک‌های تونل پیشنهاد نمودند که بر اساس آن یک مجموعه معلومات حاوی سه نوع ترک برای آموزش سریعتر R-CNN ساخته شد[23] .

۲-۴. ساختمان آهن‌کانکرتی

ژانگ و همکاران یک FCN مبتنی بر بلاک باقیمانده را با پیچیدگی اتساع یافته برای تشخیص ترک‌های کانکریت پیشنهاد کرد. بلاک‌های باقی مانده برای استخراج ویژگی‌ها استفاده شدند و پیچیدگی‌های اتساع شده با میزان‌های مختلف اتساع برای ساحات مختلف پذیرنده انجام شدند[24]. دونگ و ان یک میتود مبتنی بر FCN را برای تشخیص ترک‌ها در سطوح کانکریتی پیشنهاد کردند. یک مدل مبتنی بر VGG-16، یک مدل مبتنی بر InceptionV3 و یک مدل مبتنی بر ResNet برای عملکردهای استخراج ویژگی برای انتخاب بهترین رمزگذار برای FCN پیشنهادی مقایسه شدند[25]. نی و همکاران یک میتود مبتنی بر GoogLeNet و ResNet را برای تشخیص ترک‌ها پیشنهاد کرد[25]. اپریتور مومنت زرنیک برای پردازش کردن تصاویر ترک که توسط میتود پیشنهادی برای سنجش ترک‌های نازک تشخیص شده بود استفاده شد. لی و همکاران یک FCN مبتنی بر DenseNet-121 را برای تشخیص نواقص کانکریت از جمله پوسته شدن، ترک‌ها و سوراخ ها پیشنهاد کرد[27]. یادگیری انتقال مبتنی بر مدل برای اختصاص دادن پارامترهای اولیه FCN در پروسه آموزشی روی دست گرفته شد. ژانگ و همکاران یک مدل مبتنی بر SegNet را با آگاهی از متن برای تشخیص ترک‌ها در تصاویر با اندازه های دلخواه پیشنهاد کرد[28].نی و همکاران یک میتود دو مرحله یی مبتنی بر CNN را برای تشخیص ترک‌های سترکچری پیشنهاد کرد. گوگل نت از قبل آموزش دیده و خوب تنظیم شده برای تشخیص ترک‌ها استفاده شد و یک شبکه ترسیم ترک برای انجام ترکیب نقشه ویژگی برای ترسیم ترک‌های پیکسل وار روی دست گرفته شد[29].شو و همکاران یک مدل مبتنی بر R-CNN سریعتر را برای تشخیص و محلی سازی انواع مختلف آسیب‌های لرزه‌ای مانند ترک‌ها و شکاف‌ها پیشنهاد کرد. یک شبکه پیشنهادی منطقه‌ای با به اشتراک گذاشتن نقشه‌های ویژگی-های اولیه به یک Fast R-CNN ادغام شد.مجموعه معلومات تصویر توسط تصویر برداری در محل ایجاد شد و افزایش معلومات برای بزرگ کردن مجموعه معلومات روی دست گرفته شد[30].

کیم و چو یک چارچوب مبتنی بر Mask R-CNN را برای تشخیص و سنجش ترک‌های کانکریت پیشنهاد کردند. تصاویر آموزشی ترک‌های کانکریت از یک دیوار کانکریتی در محل جمع آوری شده بود و دارای ترک‌های با عرض‌های مختلف بود[31]. یی و همکاران یک FCN مبتنی بر U-Net را توسعه داد تا به طور خودکار ترک‌ها را در سطوح کانکریتی تشخیص دهد. یک مجموعه معلومات آنلاین تصاویر ترک با لیبل‌های پیکسل وار جمع آوری شد[32].گو و همکاران (۲۰۱۹) یک معماری مبتنی بر GAN را برای تولید تصاویر تخریب سترکچری کانکریتی برای ایجاد یک مجموعه معلومات آموزشی پیشنهاد کرد، طوریکه در شکل ۴ نشان داده شده است. یک میتود بوت استرپ کردن برای بهبود ظرفیت تولید مدل پیشنهادی روی دست گرفته شد. تصاویر مصنوعی تولید شده توسط یک نمره خود-آغاز و شاخص‌های توانایی

تعمیم ارزیابی شدند[33]. بیکمن و همکاران یک شیوه مبتنی بر کمره سریع R-CNN و عمق را برای تشخیص و سنجش پراکندگی اجزای سترکچری پیشنهاد کرد. R-CNN سریعتر توسط تصاویر پراکنده شدن در محل آموزش دیده و برای تشخیص ساحات پراکنده شدن در تصاویر بکار برده شد، و عمق پوشش توسط یک کمره عمق برای ارزیابی حجمی پوشش تشخیص‌شده اندازه‌گیری شد[34] .



شکل ۴: تولید مجموعه معلومات مبتنی بر GAN (گو و همکارانش ۲۰۱۹)

۲-۵. ساختمان‌های فولادی

عبدالجابر و همکاران یک CNN یک بعدی را برای تشخیص آسیب سترکچری مبتنی بر لرزش یک سترکچر فولادی با معلومات شتاب توسعه داد[35].اتا و جهانشاهی دو معماری مبتنی بر CNN به نام‌های خوردگی-۵ و خوردگی-۷ را برای تشخیص خوردگی در سطوح فلزی پیشنهاد کردند[36]. عملکرد معماری‌های پیشنهادی با ZF Net، VGG-15، و VGG-16 توسط میزان دقت، میزان یادآوری و نمره F1 مقایسه شد. چن و جهانشاهی یک شیوه مبتنی بر CNN را با یک میتود ادغام معلومات نایف بایس ترکیب کردند تا درزها را در چارچوب های ویدیویی نیروگاه‌های هسته‌ای تشخیص دهند[37]. یک CNN برای تشخیص ترک‌ها در هر فریم ویدیویی بکار برده شد، و یک طرح تصمیم گیری ساده بایس برای از بین بردن وصله‌های بدون ترک استفاده شد. چا و همکاران یک میتود سریعتر مبتنی بر R-CNN را برای بازرسی بصری نواقص سترکچری بشمول ترک‌های کانکریت، خوردگی پیچ، خوردگی فولاد، و لایه برداری فولاد پیشنهاد کرد[38]. پاتیراج و دیگران یک معماری مبتنی بر رمزگذار خودکار را برای شناسایی آسیب سترکچری توسط پاسخ‌های لرزش پیشنهاد کرد. مطالعات عددی و تجربوی برای تولید مجموعه معلومات برای آموزش، اعتبار سنجی و آزمایش معماری پیشنهادی انجام شد[39]. گلجک و همکاران یک شیوه مبتنی بر CNN را برای طبقه بندی

اجزای سترکچر فولادی آسیب دیده و آسیب ندیده تولید شده توسط شبیه‌سازی‌های عددی پیشنهاد کرد. برای انتخاب یک استخراج کننده ویژگی، ۵۰ CNN با میزان‌های یادگیری مختلف، لایه های پیچیده و کاملاً متصل آموزش دیده و مقایسه شدند. برای ساختن یک آشکارساز بومی‌سازی، یک مطالعه مقایسوی مشابه بر اساس ۷۰ تنظیمات انجام شد[40]. لیو و ژانگ یک میتود مبتنی بر CNN را برای ارزیابی شرایط تخریب برای ارزیابی پس از خطر اعضای فیوز فولادی سترکچری توسعه دادند. تصاویرکانتورهای فشار پلاستیکی تجمعی که توسط تحلیل عددی و مطالعه تجربوی تولید شده اند برای آموزش و اعتبار سنجی میتود پیشنهادی روی دست گرفته شده‌اند[41]. ژو و همکاران یک شبکه مبتنی بر رمزگذاری خودکار را توسط هیستوگرام سختی آموزش داد تا شناسایی آسیب را از طریق تخریب سختی تطبیق کند[42].

یو و همکاران یک چارچوب عمیق مبتنی بر CNN را برای تشخیص آسیب یک سترکچر فولادی هوشمند با جدا کننده‌ها پیشنهاد کرد. مجموعه معلومات آموزشی توسط شبیه‌سازی عددی مدل‌های سترکچری فولادی بوجود آمد[43]. وو و همکاران یک میتود مبتنی بر الگوریتم هرس کردن DNN و برای تشخیص آسیب‌های سترکچری پیشنهاد کرد. VGG-16 و ResNet-18 توسط یک سرور با عملکرد بالا آموزش دیده بودند و مجموعه معلومات آسیب حاوی تصاویر ترک و خوردگی بود که با دقت از زیرساخت‌های ساحوی جمع آوری شده بودند[44] . ژائو و همکاران (۲۰۱۹) یک میتود مبتنی بر VGG-16 را برای تشخیص وضعیت شل شدن پیچ برای سترکچرهای فولادی پیشنهاد کرد. بعد از آموزش، اعتبار سنجی و آزمایش، از یک MobileNet برای تطبیق پروسه تشخیص با یک موبایل هوشمند استفاده شد[45].

نتیجه‌گیری

این مقاله مروری بر تحقیقات اخیر و توسعه یادگیری عمیق برای SHM سترکچرها ارائه می‌کند. بر اساس بررسی جامع رویکردهای مبتنی بر یادگیری عمیق، موارد کاربرد، مسائل چالش برانگیز، می توان نتیجه گیری‌های زیر را انجام داد:

- توسعه یادگیری عمیق شامل معماری‌های جدید، آموزش کارآمد و الگوریتم‌های اعتبارسنجی، چارچوب‌های جدید و غیره، رویکردهای پردازش داده‌های آسان‌تر و قدرتمندتری را برای محققان وانجنیران فراهم می‌کند تا با مسائل حرفه‌ای مقابله کنند.
- کاربردهای اصلی رویکردهای مبتنی بر یادگیری عمیق برای SHM سترکچرها، تشخیص آسیب ستکچری و ارزیابی وضعیت سترکچر است. در میان آنها، کاربردهای مبتنی بر چشم انداز توجه زیادی را از سوی جامعه پژوهشی جلب می‌کند.
- غلبه بر چالش‌ها در کاربردهای رویکردهای مبتنی بر یادگیری عمیق برای SHM مستلزم مجموعه‌ای از مجموعه داده‌های خاص، توسعه معماری‌های جدید برای

عملکرد بهتر، و استراتژی‌های آموزشی جدید برای رفع مسائلی مانند برازش بیش از حد و ناپدید شدن گرادیان است.

ثابت شده است که رویکردهای مبتنی بر یادگیری عمیق ارزش قابل توجهی برای مقابله با انواع مختلف مشکلات SHM دارند. با توسعه الگوریتم‌ها و چارچوب‌های جدید، ایجاد مجموعه داده‌های کافی و بهبود قدرت محاسباتی، رویکردهای مبتنی بر یادگیری عمیق به طور قابل‌توجهی پیشرفت‌ها را در تحقیقات و کاربردهای SHM ارتقا می‌دهند.

منابع

[1] Adeli H, Yeh C. Perceptron learning in engineering design. Comput-Aided Civ Infrastruct Eng 1989;4:247–56.

[2] Adeli H. Neural networks in civil engineering: 1989–2000. Comput Aided Civ Infrastruct Eng 2001;16:126–42.

[3] Hajela P, Berke L. Neurobiological computational models in structural analysis and design. Comput Struct 1991;41:657–67.

[4] Wu X, Ghaboussi J, Garrett JH. Use of neural networks in detection of structural damage. Comput Struct 1992;42:649–59.

[5] Stephens JE, VanLuchene RD. Integrated assessment of seismic damage in structures. Comput Aided Civ Infrastruct Eng 1994;9:119–28.

[6] Berke L, Patnaik SN, Murthy PLN. Optimum design of aerospace structural components using neural networks. Comput Struct 1993;48:1001–10.

[7] Iranmanesh A, Kaveh A. Structural optimization by gradient-based neural networks. Int J Numer Meth Eng 1999;46:297–311.

[8] Kaveh A, Gholipour Y, Rahami H. Optimal design of transmission towers using genetic algorithm and neural networks. Int J Space Struct 2008;23:1–19.

[9] Goh ATC. Prediction of ultimate shear strength of deep beams using neural networks. ACI Struct J 1995;92:28–32.

[10] Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015;61:85–117.

[11] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[12] [12] Le Cun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:1–46.

[13] Khan A, Sohail A, Zahoora U, Qureshi AS. A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev 2020;53:5455–516.

[14] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data 2021;8:53.

[15] Kim, I.H., Jeon, H., Baek, S.C., Hong, W.H. and Jung, H.J. (2018), "Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle", Sensors, 18(6), 1881. DOI: 10.3390/S18061881.

[16] Liang, X. (2019), "Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization", Comput.-Aided Civil Infrastruct. Eng., 34(5), 415-430. DOI: 10.1111/mice.12425.

[17] Huynh, T.C., Park, J.H., Jung, H.J. and Kim, J.T. (2019), "Quasi autonomous bolt-loosening detection method using vision based deep learning and image processing", Automat. Constr., 105, UNSP 102844. DOI: 10.1016/J.Autcon.2019.102844.

[18] Alipour, M., Harris, D.K. and Miller, G.R. (2019), “Robust pixel level crack detection using deep fully convolutional neural networks”, J. Comput. Civil. Eng., 33(6), 04019040. DOI: 10.1061/(Asce)Cp.1943-5487.0000854.
[19] Xu, Y., Wei, S.Y., Bao, Y.Q. and Li, H. (2019), “Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network”, Struct. Control. Health Monit., 26(3), e2313. DOI: 10.1002/Stc.2313.
[20] Huang, H.W., Li, Q.T. and Zhang, D.M. (2018), “Deep learning based image recognition for crack and leakage defects of metro shield tunnel”, Tunn. Undergr. Sp. Tech., 77, 166-176. DOI: 10.1016/j.tust.2018.04.002.
[21] Song, Q., Wu, Y.Q., Xin, X.S., Yang, L., Yang, M., Chen, H.M., Liu, C., Hu, M.J., Chai, X.S. and Li, J.C. (2019), [21] “Real-time tunnel crack analysis system via deep learning”, IEEE Access, 7, 64186-64197. DOI: 10.1109/Access.2019.2916330.
[22] Gao, Y.Q., Kong, B.Y. and Mosalam, K.M. (2019), “Deep leaf bootstrapping generative adversarial network for structural image data augmentation”, Comput.-Aided Civil Infrastruct. Eng., 34(9), 755-773. DOI: 10.1111/mice.12458.
[23] Li, S.Y., Zhao, X.F. and Zhou, G.Y. (2019), “Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network”, Comput.-Aided Civil Infrastruct. Eng., 34(7), 616-634. DOI: 10.1111/mice.12433.
[24] Zhang, X.X., Rajan, D. and Story, B. (2019), “Concrete crack detection using context-aware deep semantic segmentation network”, Comput.-Aided Civil Infrastruct. Eng., DOI: 10.1111/mice.12477.
[25] Dung, C.V. and Anh, L.D. (2019), “Autonomous concrete crack A review on deep learning-based structural health monitoring of civil infrastructures detection using deep fully convolutional neural network”, Automat. Constr., 10.1016/j.autcon.2018.11.028. 99, 52-58. DOI: 10.1016/j.autcon.2018.11.028.
[26] Ni, F.T., Zhang, J. and Chen, Z.Q. (2019), “Zernike-moment measurement of thin-crack width in images enabled by dual scale deep learning”, Comput.-Aided Civil Infrastruct. Eng., 34(5), 367-384. DOI: 10.1111/mice.12421.
[27] Li, S.Y., Zhao, X.F. and Zhou, G.Y. (2019), “Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network”, Comput.-Aided Civil Infrastruct. Eng., 34(7), 616-634. DOI: 10.1111/mice.12433.
[28] Zhang, K.G., Cheng, H.D. and Zhang, B.Y. (2018), “Unified approach to pavement crack and sealed crack detection using preclassification based on transfer learning”, J. Comput. Civil. Eng., 32(2), 5487.0000736.
[29] Ni, F.T., Zhang, J. and Chen, Z.Q. (2019), “Pixel-level crack delineation in images with convolutional feature fusion”, Struct. Control. Health Monit., 26(1), e2286. DOI: 10.1002/Stc.2286.
[30] Xu, Y., Wei, S.Y., Bao, Y.Q. and Li, H. (2019), “Automatic seismic damage identification of reinforced concrete columns from

images by a region-based deep convolutional neural network", Struct. Control. Health Monit., 26(3), e2313. DOI: 10.1002/Stc.2313.
[31] Kim, B. and Cho, S. (2019), "Image-based concrete crack assessment using mask and region-based convolutional neural network", Struct. Control. Health Monit., 26(8), e2381. DOI: 10.1002/Stc.2381.
[32] Ye, X.W., Jin, T. and Chen, P.Y. (2019). "Structural crack detection using deep learning-based fully convolutional networks", Adv. Struct. Eng., DOI: 10.1177/1369433219836292.
[33] Gao, Y.Q., Kong, B.Y. and Mosalam, K.M. (2019), "Deep leaf bootstrapping generative adversarial network for structural image data augmentation", Comput.-Aided Civil Infrastruct. Eng., 34(9), 755-773. DOI: 10.1111/mice.12458.
[34] Beckman, G.H., Polyzois, D. and Cha, Y.J. (2019), "Deep learning-based automatic volumetric damage quantification using depth camera", Automat. Constr., 99, 114-124. DOI: 10.1016/j.autcon.2018.12.006.
[35] Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M. and Inman, D.J. (2017), "Real-time vibration-based structural damage detection using networks", J. one-dimensional convolutional neural Sound. Vib., 388, 154-170. DOI: 10.1016/j.jsv.2016.10.043.
[36] Atha, D.J. and Jahanshahi, M.R. (2018), "Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection", Struct. Health. Monit., 17(5), 1110-1128. DOI: 10.1177/1475921717737051.
[37] Chen, F.C. and Jahanshahi, M.R. (2018), "NB-CNN: deep learning-based crack detection using convolutional neural network and naive bayes data fusion", IEEE T. Ind. Electron., 65(5), 4392-4400. DOI: 10.1109/Tie.2017.2764844.
[38] Cha, Y.J., Choi, W., Suh, G., Mahmoudkhani, S. and Buyukozturk, O. (2018), "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types", Comput.-Aided Civil Infrastruct. Eng., 33, 731-747. DOI: 10.1111/mice.12334.
[39] Pathirage, C.S.N., Li, J., Li, L., Hao, H., Liu, W.Q. and Ni, P.H. (2018), "Structural damage identification based on autoencoder neural networks and deep learning", Eng. Struct., 172, 13-28. DOI: 10.1016/j.engstruct.2018.05.109.
[40] Gulgec, N.S., Takac, M. and Pakzad, S.N. (2019), "Convolutional neural network approach for robust structural damage detection and localization", J. Comput. Civil. Eng., 33(3), 04019005. DOI: 10.1061/(Asce)Cp.1943-5487.0000820.
[41] Liu, H. and Zhang, Y.F. (2019), "Image-driven structural steel damage condition assessment method using deep learning algorithm", Measurement, 133, 168-181. DOI: 10.1016/ j.measurement.2018.09.081.
[42] Zhou, C., Chase, J.G. and Rodgers, G.W. (2019), "Degradation evaluation of lateral story stiffness using HLA-based deep learning networks", Adv. Eng. Inform., 39, 259-268. DOI: 10.1016/j.aei.2019.01.007.

[43] Yu, Y., Wang, C.Y., Gu, X.Y. and Li, J.C. (2019), “A novel deep learning-based method for damage identification of smart building structures”, Struct. Health Monit., 18(1), 143-163. DOI: 10.1177/1475921718804132.
[44] Wu, R.T. and Jahanshahi, M.R. (2019), “Deep convolutional neural network for structural dynamic response estimation and system identification”, J. Eng. Mech., 145(1), 04018125. DOI: 10.1061/(Asce)Em.1943-7889.0001556.
[45] Zhao, X.F., Zhang, Y. and Wang, N.N. (2019), “Bolt loosening angle detection technology using deep learning”, Struct. Control. Health Monit., 26(1), e2292. DOI: 10.1002/Stc.2292.

شبیه‌سازی ماشین‌های امن خودپرداز با استفاده از دیف

سید زکریا حبیب[1]

چکیده

در این مقاله، فرمول دیف مشخصات سیستم‌های رویداد گسسته که توسط دکتر برنارد پی زایگلر در سال ۱۹۷٦ در کتاب با نام تئوری مدل‌سازی و شبیه‌سازی معرفی شد استفاده شده و برای مدلسازی و شبیه سازی عملکرد ماشین خود پرداز بحث شده است. امروزه سیستم‌های خود پرداز بدون شک به یک ضرورت در بانک‌ها تبدیل شده است. مردم دیگر مجبور نیستند برای انجام تراکنش‌های خود به بانک‌ها مراجعه کنند، اما در عوض ازین نوع سیستم‌ها استفاده می‌کنند، با این حال، همه کسانی که از سیستم خودپرداز استفاده می‌کنند از عملکرد واضح نحوه انجام فرآیندها در داخل سیستم آگاه نیستند که به چه رویه‌ی این سیستم‌ها امن است. با داشتن کارت خودپرداز و پین معتبر مشتریان بانک‌ها می توانند از دستگاه خودپرداز استفاده کنند که در آن اعتبار فنی به عنوان پین (شماره شناسایی شخصی) پذیرفته شده است. احراز هویت و مجوز تضمین می‌کند که حداکثر امنیت از طریق دستگاه‌های خودپرداز رعایت می‌شود. در واقع، بانک‌ها برای کاربران می‌توانند فراهم سازی عملکردهای مانند بررسی موجودی، سپرده نقدی و تراکنش‌های پولی خدمات ارائه کرده ولی بزرگترین چالش‌آن انتقال امن است که از آن به ویژه ماشین‌های سیار خودپرداز رنج می‌برند. با توجه به فرمول دیف دراین مقاله تلاش شده است تا با استفاده از فرمول فوق مدل کاربردی انتقال امن داده‌ها را فراهم ساخت.

واژگان کلیدی: ماشین خودپرداز ، احراز هویت، فرمول دیف، مدل سازی و شبیه سازی.

---

[1]. دیپارتمنت مهندسی نرم افزار، دانشکده کامپیوترساینس، دانشگاه خاتم النبیین(ص)، کابل، افغانستان ۰۷۸۵۷۲۷۲۱٦، s.z.habib@knu.edu.af

مقدمه

ماشین‌های خودپرداز[1] ماشین‌های است که به اهداف مشخص در سیستم بانک‌داری از آن بکار برده می‌شود یا بطور مشخص این ماشین‌ها بدون دخالت انسان خدماتی را فراهم می‌سازند که می‌توان در بانک انجام داد [۱]. همچنان این ماشین‌ها به کاربران اجازه می‌دهد که مجموعه از فعالیت‌های کلیدی که شامل پرداخت، انتقال پول، و پرداخت از طریق موبایل را انجام دهند[۲].

از آنجاییکه این ماشین‌ها جز مهم و اساسی برای ارایه خدمات بانکی بکاربرده شده و امروز یکی از مهمترین ابزارهای الکترونیکی برای ارایه خدمات بانکی محسوب می‌شود بناء براین انتقال امن داده‌ها و پول از جمله موارد قابل تامل دراین ماشین‌ها است. چالش‌های اصلی این ماشین‌ها در انتقال امن داده‌ها و پول نقد می‌باشد که با توجه به حساسیت داده‌های انتقالی باید به آن توجه جدی صورت گیرد. از طرف دیگر با توسعه خدمات بانکی و بکارگیری از ماشین‌های خود پرداز سیار این چالش‌های امنیتی بیشتر شده است. بدین معنی که بکارگیری ماشین‌ها انتقال داده‌ها و پول بدون چالش امنیتی بین چندین ابزار الکترونیکی صورت گرفته و سبب توسعه بهتر خدمات بانکی در سیستم بانکداری می‌گردد[۳]. بنابراین دراین مقاله تلاش شده تا در گام نخست انتقال داده‌ها را بطور امن در ماشین‌های خودپرداز سیار و غیر سیار را با استفاده از دیف که بواسطه برنارد زایگلر معرفی گردید انجام داده و سپس با بکار گیری مفهوم فوق پیاده سازی شده است.[۴]

درگام نخست تلاش شده تا شرح کوتا از دیف [2] و الگوریتم مربوط آن مانند خواص و فرمول های مرتبط به شی و وضعیت، وردی و خروجی به شمول مدل واحد[3] و زوجی[4] آن داده شده و سپس به بررسی شناسایی کاربران که مهم ترین بخش ماشین‌های خود پرداز بوده و از مدل زوجی که برای شناسایی کاربران استفاده می‌شود پرداخته شده است. با توجه به نتایج مطلوب ساختار کلی ماشین‌های امن شبه‌سازی و پیاده سازی گردیده است. خدمات که توسط ماشین‌های خود پرداز به مشتریان ارایه می‌گردد فرض براین شده، از آنجاییکه از ماشین‌های خودپرداز به اهداف مالی بکار برده می‌شود بناء این نوع خدمات بدون مداخله انسان صورت می‌گیرد.

---

۱. Automatic teller Machine ATM .

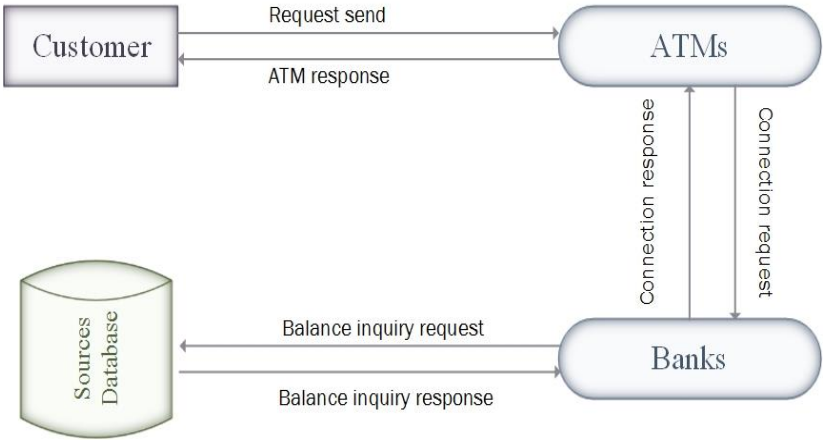
۲. DEVS . Discrete Event System Specification .

۳. Atomic Model .

۴. Coupled Model .

۱. جریان ارایه خدمات ماشین‌های خودپرداز

چنانچه جریان شناسایی کاربران دریک ماشین خود پرداز در شکل ۱ نمایش داده شده است بنابراین ماشین‌های خود پرداز بصورت کلی این قابلیت را داشته و از آن به عنوان یک منطق کلی می‌تواند نام برد.



شکل ۱: فرآیند کلی ماشین‌های خود پرداز

۲. عملیه‌های کلی ماشین‌های خودپرداز

- بطور عادی یک کاربر قانونی که برای دریافت خدمات ازیک ماشین خودپرداز ازیک رمز و کارت استفاده می‌نماید انجام عملیه بدون هیچ مشکلی صورت می‌گیرد. یعنی هرماشین خود پرداز که بصورت سیار یا غیر سیار بکار برده می‌شود باید بدون هیچ نقض فنی این کار را انجام دهد.
- هر کاربر قانونی بدون هیچ مشکلی بتواند از حساب خود برداشت نموده و نیز بتواند به حساب خویش پول اضافه کرده و یا هم به یک حساب بانکی معتبر دیگر ارز انتقال دهد.
- بررسی حساب موجودی و آگهی لازم در مورد موجودی حساب کاربران به شمول استخراج سابقه انتقال نظر به زمان و تاریخ که هر کاربر بدون هیچ چالشی بتواند ازین نوع خدمات با استفاده از ماشین خود پرداز بهره مند شود.
- هر ماشین خودپرداز بطور عادی به کاربران قانونی اجازه لغو اجرای عملیه‌ها را به جز موارد استثنایی داده تا کاربردان از آن مستفید شوند. در صورتی‌که به هر دلیلی ارایه خدمات توقف یا دچار خطا می‌شود، ماشین‌های خودپرداز باید به کاربران قانونی اگهی لازم را در زمینه ارایه دهد.

۲-۱. یوزکیس

راه‌اندازی سیستم: فرمول دیف برخی حالت‌های خاص را با توجه به حالت‌های سیستم توصیف می‌کند، که در آن دستگاه‌های خودپرداز، حالت راه‌اندازی یا «روشن» زمانی شروع می‌شود که اپراتور دکمه شروع را فشار می‌دهد. در این حالت، از کاربران معتبر برای وارد کردن مبلغی از تراکنش پرداخت در دستگاه مقدار نقدینه درخواست می‌گردد. با تأیید موفقیت آمیز کاربران، جلسه بین یک دستگاه خودپرداز و مرکز اصلی یا بانک به منظور خدمات رسانی به کاربران سازماندهی می شود.

احراز هویت: حالت شروع به منظور اعطاء فرآیندهای احراز هویت سیستم های خود پرداز مهم است که مدل دیف دراین زمینه کمک می‌نماید تا زمینه به روش بهتر این کار انجام شود. احراز هویت بخش اساسی فرآیند سیستم های خودپرداز است، به ویژه زمانی که مشتری یک پین نامعتبر را وارد می‌کند و سپس بازخورد یا پیام از بانک دریافت می‌کند و از او درخواست می شود که یک پین معتبر را به‌طور مداوم برای سه بار وارد کند. در حالتی تمام تراکنش‌های بعدی شروع می‌شود که بانک آنرا را تایید و دراین صورت جلسه امنی برای انتقال بین ماشین خودپرداز و بانک ایجاد می‌شود. در غیر این صورت، مشتریان با تکرار سه بار پین اشتباه بسته شده و کارت برای همیشه قفل می‌شود.

ایجادجلسه: جلسه تنها در مورد ذکر شده قابل ایجاد نیست، تا زمان‌که مشتری کارت معتبر را وارد کرده دستگاه خودپرداز مسئول خواندن ویژگی های کارت است که در این مرحله می توان چند مشکل را مطرح کرد؛ مانند اشکال در کارت یا کارت‌های آسیب دیده که وارد کارتخوان می‌شود یا حتا تعویض نامناسب کارت، که دراین صورت جلسه به طور خودکار با نمایش یک پیام خطا به مشتری خاتمه می‌یابد. در صورت درج صحیح کارت، کاربر معتبر ملزم به وارد کردن یک پین معتبر است، سپس برای انجام عملیه‌ها در حساب بانکی معتبر خود پذیرفته می شود. گزینه‌های متعددی از منوی ممکن به مشتری نشان داده شود که در هر مورد از کاربر برای معامله احتمالی دیگری درخواست می‌شود. در نتیجه، امکان سه عملکرد مهم را می توان در فرآیند احراز هویت برشمرد احراز هویت مشتری، مسترد کردن مشتری در صورت عدم وارد کردن پین درست و اجازه دادن برای انجام عملیه‌ها در صورت درست بودن پین و کارت معتبر.

- تراکنش های ارسالی: در صورت تایید تراکنش توسط بانک، تراکنش مورد تجزیه، تحلیل و ارزیابی قرار می گیرد و می توان تراکنش ها را همزمان با چاپ رسید برای مشتری انجام داد. یکی دیگر از تراکنش های احتمالی از مشتریان نیز پرسیده می‌شود که آیا تمایل به انجام آن را دارند یا خیر؟
- امکان ختم جلسه زمانی انجام می‌شود که به دلیل عدم وارد کردن پین درست و یا کارت اتفاق افتاده و به کاربران اجازه تلاش دوباره را می‌دهد. اگر مشتری خطا را

دوباره تکرار کند، خودپردازکارت مشتری را بسته کرده و از او درخواست دیگری نمی‌شود.

- در صورتی که مشتری یا کاربر تراکنش را لغو کرده باشد یا تراکنش به دلایل مشکلات فنی دیگر ناموفق شده باشد، با اطلاع رسانی مشکلات ناشی از خرابی سیستم یا لغو خودسرانه تراکنش، خطایی به مشتری نشان داده می شود.

برداشت وجه نقد: از کاربران قانونی درخواست می شود که گزینه های ممکن را از حساب خود برای برداشت پول انتخاب کنند. در این مرحله منوی احتمالی از قبیل نوع حساب، میزان برداشت وجه و انواع ارز به مشتری نمایش
داده می شود. در واقع سیستم‌ها با دخالت بانک‌ها تأیید می‌کنند که آیا مشتریان دارای مقادیر کافی و ارز درخواستی هستند یا خیر؟

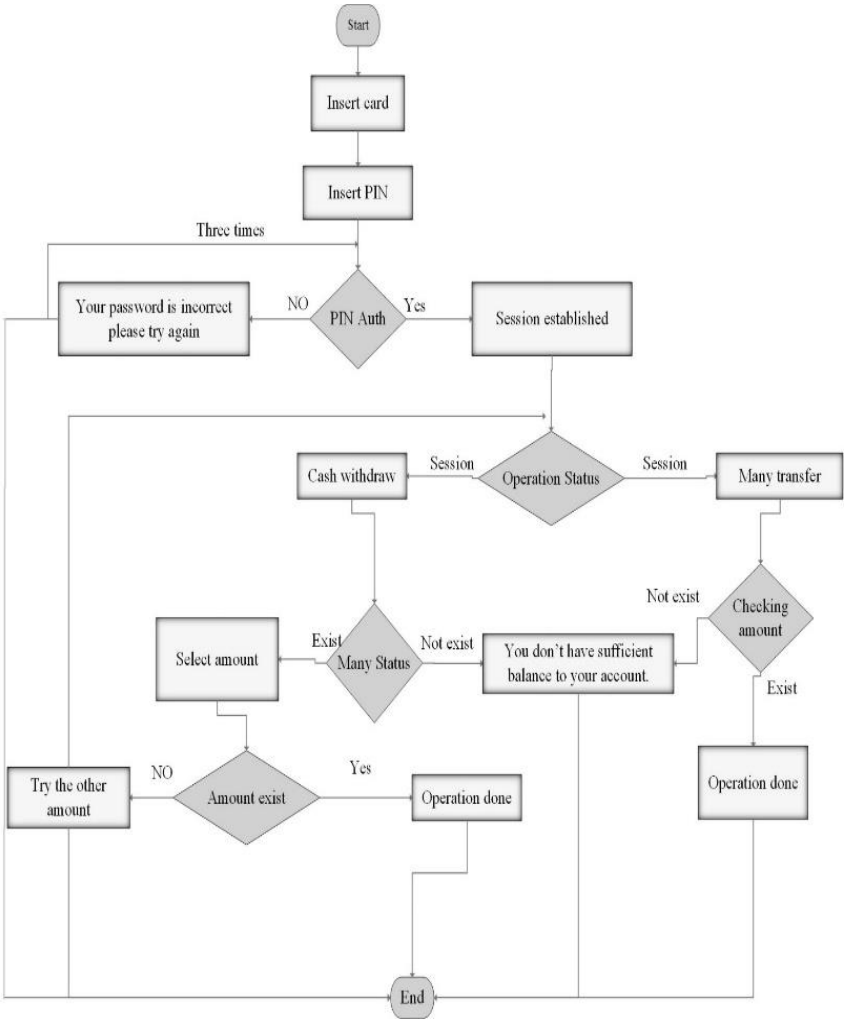
برداشت: با توجه به نوع حساب معتبر کاربر و با توجه به حساب احتمالی کاربر معتبر، از کاربران معتبر درخواست می‌شود تا برای انجام معامله برداشت موفق، نوع ارز معتبر را وارد کنند. پس از تایید اعتبار دستگاه های خودپرداز و بانک‌ها، دستگاه درخواست مشتری را می پذیرد. مشتریان می توانند پس از دریافت پیام از بانک، تراکنش دریافت یا برداشت را با صدور رسید انجام دهند، همچنین در مرحله دوم فرآیند تراکنش، پیام دوم برای تایید تراکنش از مشتری یا کاربر خواست می‌شود.

بررسی موجودی: انتخاب انواع حساب از منوی موجود به منظور استعلام ارز جاری از عملیات قابل دستیابی و حساب کاربری احتمالی، عملیات اساسی و مهمی است که در فرآیند بررسی موجودی توسط مشتری انجام می شود. این مرحله برای تراکنش بعدی پس از تایید درخواست توسط بانک ها الزامی نیست.

خاموش شدن سیستم: دستگاه خودپرداز زمانی که هیچ مشتری از دستگاه استفاده نمی کند خاموش می شود. و عملکرد باید به حالت خاموش باشد. [۵]

۲-۲. نمودار جریان سیستم

بر اساس حالت های دستگاه خودپرداز، وضعیت های خودپرداز، اجرای عملیه‌ها از طریق نمودار جریان و بیان فرآیند احراز هویت سیستم با شماره کارت کاربر معتبر و رمز عبور و حساب موجودی کافی مشتری در شکل ۲ نشان داده شده است.

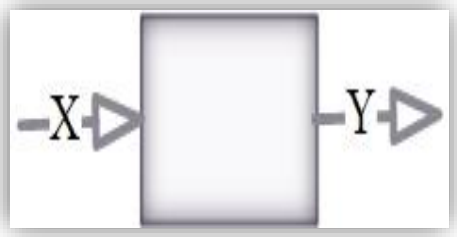


شکل ۲: ارایه جریان ماشین خودپرداز

۳. روش اجرای با استفاده از دیف

در واقع، رفتار و ساختار سیستم و موجودیت‌های کاربردان در سیستم توسط دیف تعریف می‌شود که در آن رفتار و ساختار آن در یک فرمول خاص بر اساس رویدادهای ورودی و خروجی خاص از جمله حالت‌هایی که برای ارزیابی سیستم‌ها تمرکز دارند، ارائه می‌شوند. برای پیاده سازی چنین مواردی دو مدل در دیف ارائه شده است، مدل کلاسیک فرمولیسم که بر روی دو مدل (اتمی و جفتی) شکل گرفته است. بر اساس مدل کلاسیک، رفتار سیستم ها توسط مدل اتمی مدیریت می شود، در حالی که مدل جفتی برای توضیح ساختار سیستم‌ها به کار برده می شود. توابع داخلی و بیرونی معروفترین ترین توابع حالت انتقالی هستند که از رفتار مدل اتمی دیف از جمله خروجی ها و عملکرد پیشروی زمانی آن پشتیبانی می‌کند. رفتار

مستقل مدل با تابع انتقال داخلی دیف مشخص می شود و نشان می‌دهد که چگونه یک انتقال داخلی بدون ورودی‌ای انجام شده که بنام کنترول رویدادهای داخلی نامیده می‌شود. دومین تابع مهم که تابع پیشروی زمانی نامیده می شود، وظیفه برنامه ریزی تغییرات همزمانی ذکر شده به ویژه حالت های سیستم را بر عهده دارد در حالی که رویدادهای داخلی رخ می دهند و سپس خروجی ها برای فراخوانی تابع خارجی سیستم تولید می شوند. به طور خاص، تابع انتقال خارجی بر تغییرات حالت های مدل در تأیید رویدادهای ورودی‌ها را تأکید می‌کند. در نهایت، تابع انتقال همزمان مسئول رسیدگی به ظواهر و شرایط همزمان رویدادهای داخلی و خارجی سیستم ها است که در شکل ۳ ساختار مدل اتمی ارائه شده است.



شکل ۳: نمای اولیه مدل اتمی دیف

در این بخش، مدولاسیون رفتاری دستگاه های خودپرداز با ویژگی کامل ساده شده است، در دستگاه های خودپرداز فرآیندهای درج کارت، پین، وجه موجود با یک عدد تصادفی تولید می شود، قبل از شروع تراکنش‌ها فرآیند احراز هویت باید به صورت خودکار انجام شود چون فرآیند احراز هویت مهمترین بلوک امنیتی جزئی ساختمان های خودپرداز است. جزئیات کامل احراز هویت و فرآیندهای تراکنش در شکل ٤ نشان داده شده است (احراز هویت خود پرداز یا مدل فرعی). این مدل پیش بینی مدت کل عملیات برداشت و تراکنش را نشان می دهد. با احراز هویت در این مدل، رفتارهای ساختاری خودپرداز به راحتی قابل تجزیه و تحلیل هستند. بنابراین، شکل معیاری دستگاه خودکار از سه قسمت تشکیل شده و برای شبیه‌سازی متشکل از سه مدل فرعی بر اساس آن ساخته شده‌اند:

کارتخوان: اعتبار سنجی منطقی اطلاعات کارت و اعتبار کاربر دو کارکرد مهمی است که توسط دستگاه های کارت خوان در داخل دستگاه های خودپرداز انجام می شود، همچنین در این مرحله اعتبارسنجی منطقی و فیزیکی اطلاعات مشتریان نیز انجام می‌شود.

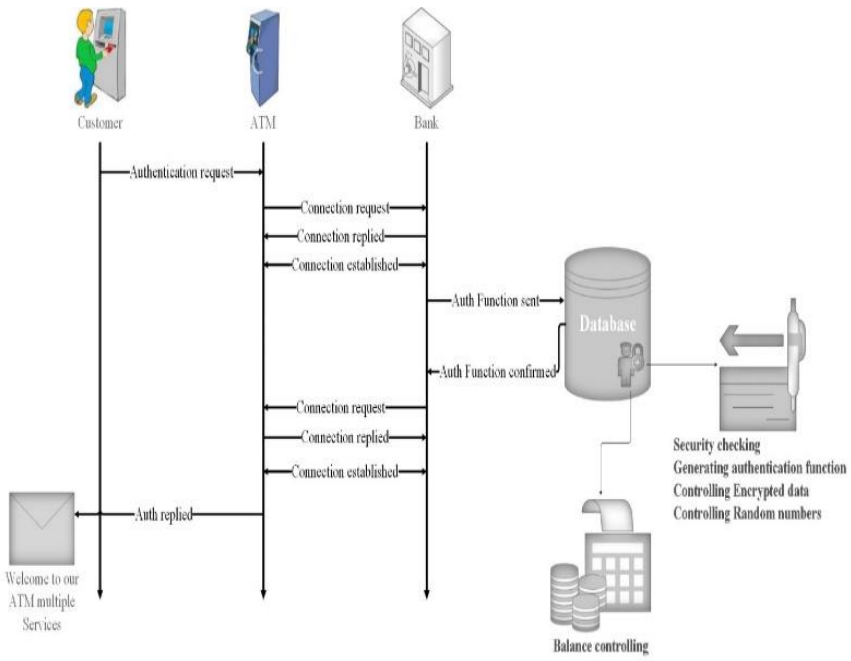
دستگاه پخش نقدی: مقدار پول یا وجه نقد اعمال شده را تایید و پس از آن توسط دستگاه پخش نقدی به منظور تضمین موفقیت آمیز تراکنش اعمال می‌کند.

مجوز طبق شبیه‌سازی دیف، صدور مجوز اصلی‌ترین مدل جفت شده را می‌سازد که اعتبار کاربر معتبر مانند پین را دریافت می‌کند و سپس شماره پین را اعتبار می‌دهد، اما کاربر

درخواست وجه کرده و تأیید آن نیز در این مرحله برای مشتریان انجام می‌شوذ که از موارد زیر ساخته شده است.

- واسط کاربری: رابطه بین مشتریان و دستگاه های خودپرداز است که به مشتری این امکان را می دهد تا پین و مقدار وجه نقد درخواستی را وارد کند تا برداشت با موفقیت انجام شود.
- تأیید کننده پین با تابع تصادفی و اعتبار دهی احتمالی به منظور تأیید شماره پین وارد شده توسط مشتری مجهز شده است. با این حال، تأیید کننده پین به طور کامل تسهیل می‌شود تا تصمیم مناسب را اتخاذ کند که آیا یک پین معتبر توسط کاربر وارد شود یا خیر. در صورت درست بودن پین وارد شده، با نمایش پیام تایید به کاربر، درست را برمی گرداند و سپس مشتری را به رابط کاربری متصل می کند. در این مرحله تایید کننده پین به طور کامل هر مرحله از درخواست کاربر را تجزیه و تحلیل می کند و منتظر می ماند تا مشتری مقدار پول برداشت را برای شروع عملیه‌های داخلی وارد کند. همچنین در صورت وارد کردن شماره پین نامعتبر و بازگشت نا درست کاربران را از طریق یک پیام خطا اگاه می‌سازد.

تأیید کننده موجودی: حساب فعلی مشتری و مقدار کافی اعتبار توسط تأیید کننده این بخش مدیریت می شود، به ویژه این بخش موجودی فعلی را تایید و میزان برداشت را به درستی بررسی می‌کند.

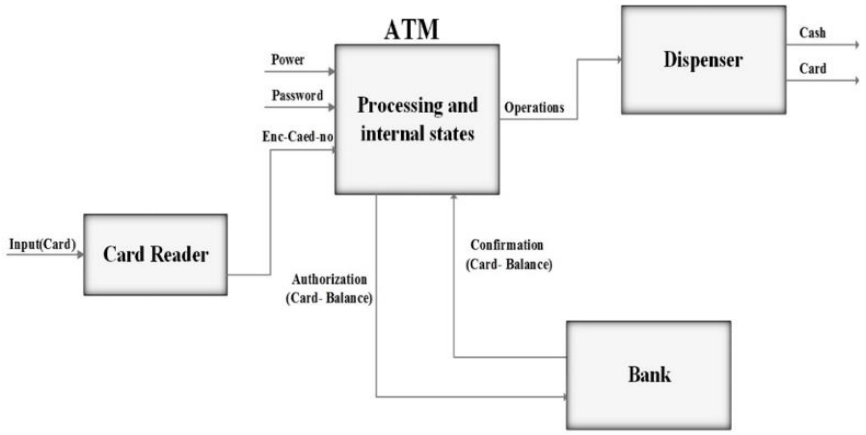


شکل ۴: مدل اصلی و فرعی احراز هویت ماشین‌های خودپرداز

در شکل فوق تعریف هر یک از مدل های اتمی در سلسله مراتب نشان داده شده به عنوان مثال، جزئیات مدل تأییدکننده پین و تأیید آن با تولید یک عدد تصادفی با توزیع یکنواخت شبیه سازی می‌شود که ۹۰ درصد موارد با پین ذخیره شده مطابقت دارد. [٦]

در ابتدا، رابط مدل همانطور که در فلوچارت سیستم نشان داده شد، با شرایط مناسب مربوط به مدل جفت شده و فرآیندهای احراز هویت خلاصه می‌شود که شامل عملکردهای احراز هویت دستگاه های خودپرداز (ورود کارت ها و پین) شروع می‌شود تعریف شده سپس خودپرداز با پارامترها عملیه شناسایی را تأیید می‌کند. اگر پارامترهای دسترسی به دستگاه خودپرداز(کارت و پین) معتبر باشد، دستگاه درخواست اتصالات را به بانک ارسال کرده چون بانک مسئول بررسی پارامترهای خودپرداز و دسترسی مشتری است .در صورت درست بودن پارامترهای وردی، بانک پارامترها را به منابع یا پایگاه های داده ارجاع می‌دهد. از این رو پایگاه داده به طور جامع مسئولیت اعتبارسنجی کلیه درخواست های مشتریان را بر عهده دارد و مشتریان مجاز یک پیام متنی دریافت می‌کنند که جزئیات مربوط به خواسته های مشتری را نشان می‌دهد.

براساس احراز هویت، هر مرحله یک به یک از هر بخش موارد مجاز پردازش می‌شود. یعنی اگر قرار باشد درخواست مشتری توسط خودپرداز احراز هویت شود، تکه دراین صورت خود پردازش به مشتری پاسخ می‌دهد، همینطور اگر بانک دستگاه‌های خودپرداز را بشناسد، بانک مجوز اتصال را می‌دهد و جلسه را بین خودپرداز، بانک‌ها و پایگاه‌های اطلاعاتی که توسط آن بررسی می‌شود برقرار می‌شود. جزئیات بیشتر دربخش فرمول دراین زمینه ارایه شده است.



شکل ۵: مدل جفتی ماشین خود پرداز

کارتخوان اطلاعات کارت را به عنوان ورودی و خروجی می گیرد و شماره کارت را رمزگذاری می‌کند. پارامترهای ورودی خودپرداز می تواند رمز عبور و شماره کارت رمزگذاری شده باشد. این فرآیند مدت زمانی طول می‌کشد که زمان احراز هویت نامیده می‌شود. خودپرداز برای بررسی جزئیات کلی با توجه به اطلاعات ذخیره شده در پایگاه داده با بانک ارتباط برقرار کرده و بررسی که شامل اطلاعات دقیق مدل زوج کارت مشخصات دستگاه های خودپرداز می‌شود

در آنجا صورت می‌گیرد. در شکل ۵، مقایسه مبلغ درخواستی همراه با مبلغ موجود در حساب مشتریان را نمایش می‌دهد، مشتریان پس از بررسی موفقیت آمیز کلیه اطلاعات معتبر یا نامعتبر، تایید نهایی را دریافت کرده و در نهایت خروجی خودپرداز که کارت و نقدینه است را دریافت می‌کند.

۳-۱. شبیه سازی خودپرداز با استفاده از فرمول دیف

مانطور که در بخش های قبلی اشاره شد در زیر شبیه سازی ماشین که توسط برنارد زایگلر[۱] در سال ۱۹۷۶ برای مدل سازی سیستم‌های رویداد گسسته[۲] به روش مدلاسیون تسلسلی[۳] در زبان سی شارب بیان شده است.

$$ATM=< X, S, Y, \delta_{int}, \delta_{out}, ta, \sigma>$$

X = (Card_No, Password, Power, Dispenser, Card_Reader)
Y = (Cash, Transaction, Card)
S = (Authentication, Checking_Amnt, Dispenser)
σ = (Active_Mode, Passive_Mode, Waiting_Mode)
$Ta = \ -\infty < ta < +\infty$
δ_{int} = (Processing (Auth_Process, Balance_Checking, Trasaction_Process, Controlling_Status), Rejected_Card)
δ_{out} = (Cash (Balance_Checking, Transaction_Controling), Card)

```
If (Phase (σ) = ="Passive")
  {Int X, Y;
  X = Y = R;
  Ta= ∞;
  δint=null;
  δout= Card;            }
Else if ( Phase (σ) ="Active")
  {Int PIN, Random_No, Card_No;
  Int No_Trails, Checking_Balances, Remains_Balances;
  Double Exsit_Amnt, Requst_Amnt;
  String Trans_status, Cash;
  X=PIN;
  Int PIN_Auth= (card_No, Random_No, Password)
    Case1 PIN_Auth: True
      {δint(PIN==1 && Rondom_No==1 && No_Trails<=3)
       While (Cash_Amnt< Exsit_Amnt)
        {Trans_status= Exsit_Amnt - Request_Amnt;
         Balance_Checking=Remains_Balances;
         Transaction=done;
         Cash=done; }
       δout= Cash, Card, Transaction;
       Ta= (X, e); }
    Case2 PIN_Auth: False
      { δint(PIN==0 || Rondom_No==0 || No_Trails>3)
         Phase="Passive";
         δout= Card;
         Ta="e"; }
       Break;   }
Else  {  σ = Waiting;
         δint= ta= X= Y= 0;
         δout= Card; }}
```

۱. P. Bernard Zeigler

۲. Discrete-Event systems

۳. Hierarchical and modular way

چنانچه مشخص است، دو حالت برای سیستم‌ها و زیرسیستم‌های[1] خودپرداز وجود دارد که شامل حالت‌های فعال[2] و غیرفعال[3] می‌شود که گرداننده اصلی سیستم است. هنگامی که سیستم ها درخواست های مشتری را به صورت منطقی یا فیزیکی پردازش، تجزیه و تحلیل می‌کنند، به آن حالت فعال می گویند. بالعکس اگر سیستم درخواست‌ها (منطقی یا فیزیکی) را بر اساس هر مشکلی (مشکل برق، مشکل اتصالات و برخی مشکلات متناقض) پردازش نکند، در این صورت سیستم در حالت غیرفعال است. در بخش بعدی حالت های سیستم از طریق شرایط با جزئیات بیشتر نشان داده شده است. اگر سیستم در حالت فعال باشد، تجزیه و تحلیل تمام پارامترهای ممکن از جمله احراز هویت کارت، قدرت، و درخواست های منطقی مانند برداشت وجه نقد، تراکنش ها، اعتبارسنجی رمز عبور و موجودی و غیره باشد.

برعلاوه در حالت فعال سیستم رفتارها و حالات داخلی سیستم‌ها توسط ورودی ها تحت تاثیر قرار می گیرند. به این معنی که فرآیندها و حالت های داخلی سلسله مراتبی سیستم مستقیماً با وارد کردن ورودی ها به سیستم ها و متناظر با ورودی ها تحت تأثیر است، سیستم در حال تولید خروجی های معادل است و بنابراین بر اساس بحث در اینجا سعی کردیم حالت های سیستم و فرآیندهای داخلی را از طریق فرمول دیف و اعمال الگوریتم بیان کنیم. در نهایت اگر سیستم در حالت غیرفعال باشد، یک ورودی و یک خروجی را تحت تاثیر قرار می دهد و دیگر شرایطی برای حالت غیرفعال سیستم وجود ندارد

۳-۲. نمودار فرمول دیف خودپرداز

در حالت فعال، سیستم ورودی ها را فراخوانی و توابع انتقال داخلی را راه اندازی می‌کند، توابع خروجی را پس از محاسبات و اجراهای متمایز نیز فراخوانده و در فرآیندهای احراز هویت، خروجی های استخراج شده شامل کارت و برداشت پول و حتی انتقال اعتبار می‌شود که حالت داخلی در شکل ٦ ارائه شده است (حالت فعال و غیر فعال). این فرآیند در صورتی اتفاق می‌افتد که همه ورودی‌ها (شماره تصادفی، شماره کارت و رمز عبور کاربر) صحیح باشند. در حالت غیرفعال سیستم، همانطور که در نمودار نشان داده شده است، هیچ ورودی برای فراخوانی حالت های داخلی خودپرداز وجود ندارد، تنها سیستم می تواند حالت غیرفعال را در صورتی که شماره تصادفی، شماره کارت و رمز عبور وارد شده نادرست را کنترل کند که یکی از مهترین دلیل استفاده از دیف و پیاده سازی آن در سیستم‌های خودپرداز برای دسترسی

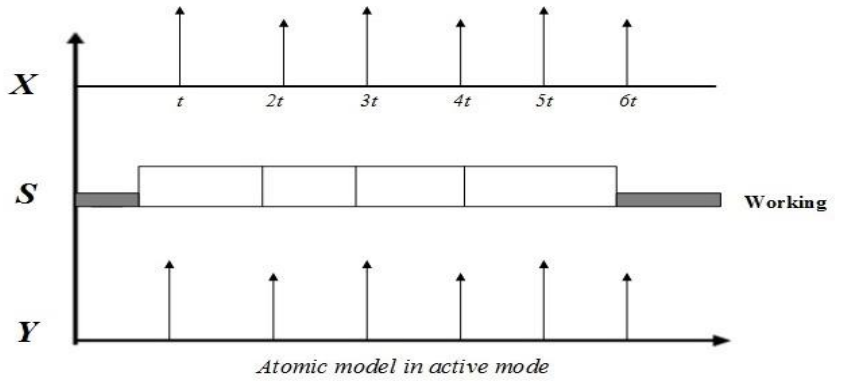
---

[1] . Sub-systems

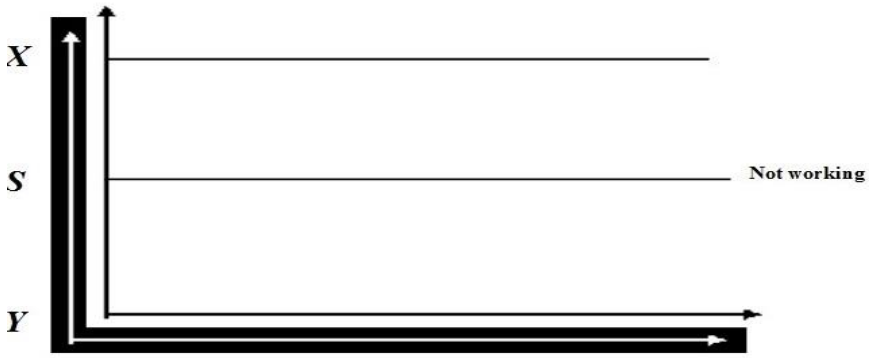
[2] . Active

[3] . Passive

و دستیابی به انجام تراکنش‌های امن به واسطه آن نمایش حالت‌ها یا وضعیت‌های مختلف سیستم است.



Atomic model in active mode



شکل ۶: حالت فعال و غیر فعال ماشین‌های خود پرداز

نتیجه‌گیری

با توجه به کاربرد ماشین‌های خودپرداز برای ارایه خدمات بانکی، استفاده ازاین نوع ماشین‌ها همزمانی‌که سهولت در روند ارایه خدمات بانکی به ارمغان می‌أورد، یک مجموعه چالش‌ها و آسیب‌ها نیز دارد به ویژه ماشین‌های خودپرداز که به گونه سیار خدمات ارایه می‌کند. یکی از چالش‌های مهم که باید به آن توجه داشت مشکلات انتقال ارز و تبادل اطلاعات که حاوی اطلاعات ارز است می‌باشد. بر أساس مباحث مطرح شده در بخش‌های قبلی به ویژه پیاده‌سازی و شبه‌سازی ماشین‌های خودپرداز می‌توان استنباط کرد که بکارگیری از فرمول دیف که توسط برنارد زایگلر مطرح شده نه تنها کمک بزرگی در پیاده‌سازی ماشین‌ها می‌نماید بلکه استفاده ازدیف کمک بزرگی می‌نماید تا تبادل اطلاعات حاوی ارزها، انتقال، احراز هویت و موارد مرتبط را که بخش مهم امنیت این ماشین را پوشش می‌دهد می‌شود. به این معنی که استفاده از دیف در شبه‌سازی ماشین‌های خودپرداز ازیک سو بکارگیری و پیاده‌سازی این نوع

ماشین‌ها را تسهیل کرده و از سوی دیگر امنیت این نوع ماشین‌ها به ویژه ماشین‌های سیار خودپرداز را تامین می‌کند.

نمودار جریان سیستم های خودپرداز رایج ترین الزاماتی است که دقیقاً در این مقاله پژوهش محور، بر اساس فرمالیسم دیف که توسط برنارد زیگلر در سال ۱۹۷٦ معرفی شد، مورد بحث و تمرکز قرار گرفته است. بدون شک، این تلاش در انجام مرحله بعدی که طراحی سیستم امن است کمک خواهد کرد. با این حال، دیف یک فرمول بسیار مهم برای شبیه‌سازی سیستم‌های واقعی پیچیده به ویژه امن است، ما همچنین با مشخص کردن اینکه چگونه و چه ورودی‌هایی به یک دستگاه خودپرداز مرتبط هستند، چه وضعیت‌هایی در داخل اتفاق می‌افتد و چه خروجی‌های ممکنی می‌توان به دست آورد، آن را در مدل خود اعمال کردیم تا بتوانیم آنرا در محیط عملی پیاده سازی نمایم. به طور کلی دیف برای نشان دادن رفتار سیستم یعنی خود پرداز که شکلی از سیستم رویداد گسسته است به صورت سلسله مراتبی و مدولار کمک می‌کند. بنابراین چنین سیستمی با استفاده از زبان‌های برنامه نویسی شی‌گرا نیز قابل تطبیق و پیاده سازی است با آنکه دراین مقاله پیاده سازی ماشین‌های خودپرداز با رویکرد ساختاری انجام شده است. این مدل نه تنها در روند تطبیق شی‌گرایی سیستم و نرم افزارها کمک می‌کند بلکه برای تامین امنیت ماشین‌های خودپراز سیار نیز کاربرد اساسی دارد تا بتوان هم انتقال تراکنش‌های امن را انجام داده و هم اطلاعات که حاوی ارزها می‌باشد را نیز تبادله نمود.

منابع

[1] Mohsin Karovaliya, Saifali Karedia, Sharad Oza and Dr.D.R.Kalbande "*Enhanced security for ATM machine with OTP and Facial recognition features*" Computer Engineering Department, Sardar Patel Institute of Technology, Mumbai 400058, India

[2] http://ezinearticles.com/?A-Brief-Introduction-to-the-Automated-ellerMachine&id=5397483

[3] http://www.thocp.net/hardware/atm.htm

[4] RICHA TULI, ABHIJEET KHATRI and ANITA YADAV "*A comparative study of customer attitude towards ATM of SBI and ICICI bank"* publisher IJMIT, Volume 2, Issue 8 and ISSN: 2249-1058

[5] http://www.math-cs.gordon.edu/courses/cs211/ATMExample/UseCases.html

[6] Gabrial A. Wainer, "*Discrete-Event Modeling and Simulation a Practitioner's Approach*"

[7] Bernard P. Zeigler, Herbert Praehofer, Tag Gon Kim "*Theory of modeling and simulation*"

[8] Hossein Farid Ghassem Nia and Huosheng Hu, "*Vision-based precise cash counting in ATM machines"*, IEEE International Conference 2014 on Mechatronics and Automation, INSPEC Accession Number: 14547172, Tianjin, China

[9] Z.F. Kamran, "*ATM - AAL1 machine"*, IEEE conference 27-28 Dec. 2002 Published in: Multi Topic Conference, 2002, INMIC 2002. International, DOI: 10.1109/INMIC.2002.1310142, Karachi.

[10] "*Biometric skimmers threaten the security of biometric ATM authentication Publication, Biometric Technology",* Today publisher, Elsevier

[11] Dr. Bernard P Zeigler, "*Theory of Modeling and Simulation, Integrating Discrete Event and Continuous Complex Dynamic Systems*".

[12] Rishigesh Murugesh, "*Advanced biometric ATM machine with AES 256 and steganography implementation*", Department of Informatics, Technical University of Munich, Munich, Germany, published on IEEE conference 13-15 Dec. 2012, Print ISSN: 2377-6927, Chennai, India

[13] Rupesh Mandal and Nupur Choudhury "*Automatic video surveillance for theft detection in ATM machines: An enhanced approach"*, IEEE Conference: 16-18 March 2016, INSPEC Accession Number: 16426601, New Delhi, India

[14] Sweta Singh, Akhilesh Singh and Rakesh Kumar "*A constraint-based biometric scheme on ATM and swiping machine"*, Department of Computer Science and Engineering, Madan Mohan Malaviya University of Technology, Gorakhpur, India, presented on IEEE

Conference: 11-13 March 2016, INSPEC Accession Number: 16156324, New Delhi, India

[15] Nischal Bansal and Nepali Singla, "*Cash withdrawal from ATM machine using Mobile banking"*, IEEE Conference: 11-13 March 2016, INSPEC Accession Number: 16143274, New Delhi, India

[16] Iwasokun Gabriel Babatunde, Akinyokun Oluwole Charles and Munda Josiah Lange "*Experimental study of thumbprint-based authentication framework for ATM machines"*, IEEE, Science and Information Conference (SAI), 2014, INSPEC Accession Number: 14650918, London, UK

مروری بر اخلاق هوش مصنوعی

Changwu Huang, Zeqi Zhang, Bifei Mao, and Xin Yao
مترجم: Ryan Heida [1]

چکیده

هوش مصنوعی (AI) به طور عمیق زندگی ما را تغییر داده و همچنان به تغییر آن ادامه خواهد داد. هوش مصنوعی در حوزه‌ها و سناریوهای متعددی مانند رانندگی خودران، مراقبت‌های پزشکی، رسانه، امور مالی، ربات‌های صنعتی و خدمات اینترنتی به کار گرفته می‌شود. کاربرد گسترده هوش مصنوعی و ادغام عمیق آن با اقتصاد و جامعه باعث افزایش بهره‌وری و تولید مزایا شده است. در عین حال، این فناوری به ناچار بر نظم اجتماعی موجود تأثیر می‌گذارد و نگرانی‌های اخلاقی را مطرح می‌کند. مسائلی مانند نشت حریم خصوصی، تبعیض، بیکاری، و خطرات امنیتی که توسط سیستم‌های هوش مصنوعی ایجاد می‌شوند، مشکلات زیادی برای افراد به وجود آورده‌اند. بنابراین، اخلاق هوش مصنوعی، که حوزه‌ای مرتبط با مطالعه مسائل اخلاقی در هوش مصنوعی است، نه تنها به یک موضوع مهم تحقیقاتی در میان محققان دانشگاهی تبدیل شده است، بلکه یک موضوع مشترک مورد توجه افراد، سازمان‌ها، کشورها، و جامعه نیز می‌باشد. این مقاله یک مرور جامع از این حوزه ارائه می‌دهد، که شامل خلاصه و تحلیل ریسک‌ها و مسائل اخلاقی ناشی از هوش مصنوعی، دستورالعمل‌ها و اصول اخلاقی صادر شده توسط سازمان‌های مختلف، رویکردهایی برای مقابله با مسائل اخلاقی در هوش مصنوعی، و روش‌هایی برای ارزیابی اخلاق در هوش مصنوعی است. علاوه بر این، چالش‌های اجرای اخلاق در هوش مصنوعی و دیدگاه‌های آینده نیز مورد بحث قرار گرفته‌اند. امیدواریم این پژوهش دیدگاهی سیستماتیک و جامع از اخلاق هوش مصنوعی برای پژوهشگران و متخصصان این حوزه، به‌ویژه تازه‌کاران این رشته تحقیقاتی، ارائه دهد.

واژگان کلیدی: هوش مصنوعی (AI)، اخلاق هوش مصنوعی، مسائل اخلاقی، نظریه اخلاقی، اصل اخلاقی.

[1] (links.ryanheida.com).

مقدمه

هوش مصنوعی (AI) در دهه گذشته پیشرفت‌های سریع و چشمگیری داشته است. فناوری‌های هوش مصنوعی مانند یادگیری ماشین(ML) ، پردازش زبان طبیعی و بینایی کامپیوتری به طور فزاینده‌ای در حوزه‌ها و جنبه‌های مختلف جامعه ما نفوذ کرده و گسترش یافته‌اند. هوش مصنوعی به تدریج در حال جایگزینی وظایف انسانی و تصمیم‌گیری‌های انسانی است و در بخش‌های متعددی مانند تجارت، لجستیک، تولید، حمل‌ونقل، مراقبت‌های بهداشتی، آموزش و مدیریت دولتی به کار گرفته شده است.

کاربرد هوش مصنوعی باعث بهبود بهره‌وری و کاهش هزینه‌ها شده است که این امر برای رشد اقتصادی، توسعه اجتماعی و رفاه انسانی مفید است. برای مثال، چت‌بات‌های هوش مصنوعی می‌توانند در هر زمان به سوالات مشتریان پاسخ دهند و رضایت مشتریان و فروش شرکت را افزایش دهند. همچنین، هوش مصنوعی این امکان را برای پزشکان فراهم کرده است که از طریق خدمات تله‌مدیسین به بیماران در مناطق دورافتاده خدمت کنند. بدون شک، توسعه سریع و کاربرد گسترده هوش مصنوعی در حال حاضر زندگی روزمره، انسانیت و جامعه را تحت تأثیر قرار داده است.

با این حال، هوش مصنوعی هم‌زمان ریسک‌ها و مسائل اخلاقی قابل توجهی را برای کاربران، توسعه‌دهندگان، انسان‌ها و جامعه ایجاد می‌کند. در سال‌های اخیر، موارد بسیاری از نتایج نامطلوب ناشی از هوش مصنوعی مشاهده شده است. به عنوان مثال، در سال ۲۰۱۶، راننده یک خودروی تسلا در تصادف جاده‌ای کشته شد، زیرا حالت Autopilot این خودرو نتوانست یک کامیون در حال عبور را تشخیص دهد. چت‌بات هوش مصنوعی شرکت مایکروسافت، **Tay.ai**، به دلیل نژادپرست و جنسیت‌گرا شدن در کمتر از یک روز پس از ورود به توییتر از دسترس خارج شد. نمونه‌های بسیاری دیگر نیز وجود دارند که به مسائل مربوط به شکست، انصاف، تعصب، حریم خصوصی و دیگر مشکلات اخلاقی سیستم‌های هوش مصنوعی مربوط می‌شوند. حتی جدی‌تر از این، فناوری هوش مصنوعی توسط مجرمان برای آسیب رساندن به دیگران یا جامعه مورد استفاده قرار گرفته است؛ به عنوان مثال، مجرمان با استفاده از نرم‌افزار مبتنی بر هوش مصنوعی صدای یک مدیر اجرایی را جعل کردند و درخواست انتقال جعلی ۲۴۳٬۰۰۰ دلار کردند. بنابراین، ضروری و حیاتی است که مسائل و ریسک‌های اخلاقی هوش مصنوعی مورد بررسی قرار گیرد تا این فناوری به شکلی اخلاقی ساخته، اعمال و توسعه یابد.

اخلاق هوش مصنوعی یا اخلاق ماشین یک حوزه نوظهور و میان‌رشته‌ای است که به بررسی مسائل اخلاقی هوش مصنوعی می‌پردازد. اخلاق هوش مصنوعی شامل دو جنبه است: اخلاق هوش مصنوعی که نظریه‌های اخلاقی، دستورالعمل‌ها، سیاست‌ها، اصول، قوانین و مقررات مرتبط با هوش مصنوعی را مطالعه می‌کند؛ و هوش مصنوعی اخلاقی که به هوش مصنوعی‌ای

اشاره دارد که می‌تواند هنجارهای اخلاقی را رعایت کند و رفتار اخلاقی داشته باشد. اخلاق هوش مصنوعی پیش‌نیازی برای ساخت یا رفتار اخلاقی هوش مصنوعی است. این حوزه ارزش‌ها و اصول اخلاقی را بررسی می‌کند که تعیین می‌کنند چه چیزی از نظر اخلاقی درست یا غلط است. با داشتن اخلاق مناسب برای هوش مصنوعی، می‌توان از طریق برخی روش‌ها و فناوری‌ها هوش مصنوعی اخلاقی ساخت یا پیاده‌سازی کرد.

با وجود اینکه اخلاق هوش مصنوعی طی چندین سال گذشته به طور گسترده توسط پژوهشگران میان‌رشته‌ای مورد بحث قرار گرفته است، این حوزه همچنان در مراحل ابتدایی خود قرار دارد. اخلاق هوش مصنوعی یک حوزه پژوهشی بسیار گسترده و به سرعت در حال توسعه است که در سال‌های اخیر توجه بیشتری از سوی پژوهشگران جلب کرده است. اگرچه در سال‌های گذشته چندین مقاله مروری منتشر شده است، هر یک از آن‌ها بر جنبه یا جنبه‌هایی خاص از اخلاق هوش مصنوعی تمرکز داشته‌اند و هنوز کمبود بررسی‌های جامع برای ارائه یک تصویر کامل از این حوزه وجود دارد.

اهداف و مشارکت‌های اصلی مقاله:

این مقاله با هدف ارائه یک مرور سیستماتیک و جامع از اخلاق هوش مصنوعی از جنبه‌های مختلف، برای ارائه راهنمایی‌های کاربردی به جامعه برای تحقق هوش مصنوعی اخلاقی در آینده نوشته شده است. این مقاله با جمع‌بندی و تحلیل مسائل اخلاقی هوش مصنوعی، دستورالعمل‌ها و اصول اخلاقی، رویکردها برای مقابله با مسائل اخلاقی و روش‌هایی برای ارزیابی اخلاقی هوش مصنوعی به جامعه علمی و متخصصان کمک می‌کند.

مشارکت‌های اصلی مقاله به شرح زیر است:

۱. ارائه یک مرور جامع درباره اخلاق هوش مصنوعی، شامل مسائل اخلاقی و ریسک‌های هوش مصنوعی، دستورالعمل‌ها و اصول اخلاقی، رویکردهای مقابله با مسائل اخلاقی و روش‌های ارزیابی هوش مصنوعی اخلاقی.

۲. دسته‌بندی جدیدی از مسائل اخلاقی هوش مصنوعی ارائه شده که به شناسایی، درک و تحلیل مشکلات اخلاقی در هوش مصنوعی و توسعه راه‌حل‌های مربوطه کمک می‌کند.

۳. مرور دستورالعمل‌ها و اصول اخلاقی جهانی مربوط به هوش مصنوعی که توسط شرکت‌ها، سازمان‌ها و دولت‌ها منتشر شده است.

۴. بررسی رویکردهای میان‌رشته‌ای برای مقابله با مشکلات اخلاقی هوش مصنوعی، از جمله رویکردهای اخلاقی، فناورانه و قانونی.

۵. مرور روش‌های ارزیابی اخلاق هوش مصنوعی که جنبه‌ای کمتر مورد توجه در ادبیات موجود است.

۶. شناسایی چالش‌های موجود در اخلاق هوش مصنوعی و ارائه دیدگاه‌های آینده برای طراحی هوش مصنوعی اخلاقی.

این مقاله راهنمایی کامل و جامعی برای پژوهشگران و متخصصان این حوزه، به ویژه مبتدیان، ارائه می‌دهد تا آن‌ها بتوانند تحقیقات و مطالعات بیشتری در این زمینه انجام دهند.

۱. دامنه و روش‌شناسی

در این بخش، ابتدا جنبه‌ها و موضوعات تحت پوشش این بررسی و ارتباط میان این موضوعات را روشن می‌کنیم. سپس روش‌شناسی استفاده‌شده در انجام این بررسی، از جمله استراتژی جستجوی منابع و معیارهای انتخاب آن‌ها، را شرح می‌دهیم.

۲. دامنه

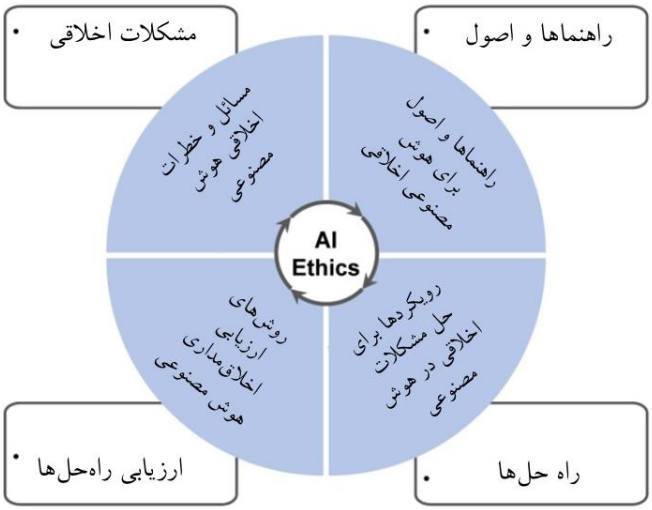
دامنه و موضوعات این مقاله به شرح زیر است: بررسی مسائل و ریسک‌های اخلاقی هوش مصنوعی نقطه شروع این مطالعه است، زیرا وجود مسائل اخلاقی در هوش مصنوعی زمینه‌ساز شکل‌گیری حوزه تحقیقاتی اخلاق هوش مصنوعی است. بنابراین، ضروری و مهم است که مشکلات اخلاقی موجود در هوش مصنوعی شفاف‌سازی و درک شود.

سپس دستورالعمل‌ها و اصول اخلاقی که توسعه و استفاده از هوش مصنوعی را هدایت می‌کنند، بررسی می‌شوند. با توجه به اینکه مسائل اخلاقی هوش مصنوعی توجه روزافزون بخش‌های مختلف جامعه را به خود جلب کرده است، بسیاری از سازمان‌ها (از جمله نهادهای آکادمیک، صنعت و دولت‌ها) به بحث و جستجوی چارچوب‌ها، دستورالعمل‌ها و اصول ممکن برای حل مسائل اخلاقی هوش مصنوعی پرداخته‌اند. این دستورالعمل‌ها و اصول جهت‌گیری‌های ارزشمندی برای پیاده‌سازی هوش مصنوعی اخلاقی ارائه می‌دهند.

پس از شفاف‌سازی مسائل اخلاقی موجود و دستورالعمل‌ها، رویکردهای حل مسائل اخلاقی در هوش مصنوعی بررسی می‌شوند. این مقاله رویکردهای اخلاقی، فناورانه و قانونی را پوشش می‌دهد، اما تمرکز بیشتری بر دو دسته اول (رویکردهای اخلاقی و فناورانه) دارد، زیرا پژوهشگران جامعه هوش مصنوعی ممکن است بیشتر به این دو دسته علاقه‌مند باشند.

در نهایت، نحوه ارزیابی هوش مصنوعی اخلاقی که شامل بررسی اخلاقی یا اخلاق‌مداری هوش مصنوعی است، خلاصه می‌شود؛ به عبارت دیگر، بررسی اینکه مشکلات اخلاقی تا چه

حد برطرف شده یا اینکه آیا یک سیستم هوش مصنوعی الزامات اخلاقی را برآورده می‌کند یا خیر.



شکل ۱: موضوعات مطرح‌شده در این مقاله و ارتباط بین آن‌ها

به وضوح، این چهار جنبه برای حل مسائل اخلاقی در هوش مصنوعی ضروری هستند. بنابراین، این چهار جنبه محتوای اصلی این مقاله را تشکیل می‌دهند و یک مرور سیستماتیک از اخلاق هوش مصنوعی ارائه می‌دهند.

موضوعات یا جنبه‌های مورد بررسی در این مقاله و ارتباط میان آن‌ها در **شکل ۱** نشان داده شده است.

۱-۲. روش‌شناسی

این مرور طیف گسترده‌ای از اسناد را پوشش می‌دهد، از جمله منابع علمی، سازمانی، منابع خاکستری دولتی، و گزارش‌های خبری. جستجوی ادبیات مرتبط در دو مرحله انجام شد. در مرحله اول، ورودی‌ها یا کلمات کلیدی که منعکس‌کننده اصطلاحات مختلف مرتبط با اخلاق هوش مصنوعی هستند، برای جستجو در پایگاه‌های Google Scholar، Web of Science، IEEE Xplore، ACM Digital Library، Science Direct، Springer Link، arXiv و Google به کار گرفته شدند. کلمات کلیدی استفاده‌شده شامل موارد زیر بودند: (اخلاق، اخلاقی، مسئولیت، مسئولانه، قابل اعتماد، شفاف، توضیح‌پذیر، عادلانه، مفید، قوی، ایمن، خصوصی، پایدار) و/یا (مسائل، ریسک‌ها، دستورالعمل، اصل، رویکرد، روش، ارزیابی، سنجش، چالش) و) هوش مصنوعی، AI، یادگیری ماشین، ML، سیستم هوشمند، عامل هوشمند .(ما عمدتاً به ادبیاتی که از سال ۲۰۱۰ به بعد منتشر یا ارائه شده باشد توجه کردیم و تا حد امکان این کلمات کلیدی را در عناوین جستجو کردیم.

در مرحله دوم، به بررسی کارهای مرتبط با ادبیات یافت‌شده در مرحله اول پرداختیم، از جمله مقالات ارجاع‌شده و سایر آثار نویسندگان همان مقالات. در مورد دستورالعمل‌های اخلاقی هوش مصنوعی، فقط اسنادی را جمع‌آوری کردیم که به زبان انگلیسی (یا با ترجمه رسمی به انگلیسی) بودند و امکان مشاهده یا دانلود آن‌ها از اینترنت وجود داشت. فهرست کاملی از این دستورالعمل‌های اخلاقی هوش مصنوعی همراه با لینک‌های URL در مواد تکمیلی این مقاله ارائه شده است.

۲. مسائل اخلاقی و ریسک‌های هوش مصنوعی

برای پرداختن به مشکلات اخلاقی هوش مصنوعی، ابتدا باید مسائل اخلاقی یا ریسک‌های بالقوه‌ای را که هوش مصنوعی ممکن است به همراه داشته باشد، شناسایی و درک کنیم. سپس می‌توان دستورالعمل‌ها، سیاست‌ها، اصول، و قواعد اخلاقی لازم (یعنی اخلاق هوش مصنوعی) را به طور مناسب تدوین کرد. با داشتن اخلاق مناسب برای هوش مصنوعی، می‌توان سیستم‌های هوش مصنوعی‌ای طراحی و ایجاد کرد که به طور اخلاقی رفتار کنند (یعنی هوش مصنوعی اخلاقی). مسئله اخلاقی در هوش مصنوعی معمولاً به چیزهای غیراخلاقی یا نتایج مشکل‌زای مرتبط با هوش مصنوعی اشاره دارد (یعنی مسائل و ریسک‌هایی که از توسعه، استقرار، و استفاده از هوش مصنوعی ناشی می‌شوند) که باید به آن‌ها پرداخته شود. بسیاری از مسائل اخلاقی، مانند کمبود شفافیت، حریم خصوصی و مسئولیت‌پذیری، تبعیض و جانبداری، مشکلات ایمنی و امنیت، و همچنین استفاده‌های جنایی و مخرب شناسایی شده‌اند. این بخش بر مسائل و ریسک‌های اخلاقی هوش مصنوعی تمرکز دارد. ابتدا، در بخش III-A ، چهار دسته‌بندی مختلف از مسائل اخلاقی هوش مصنوعی در ادبیات موجود بررسی می‌شود. از آنجا که این چهار دسته‌بندی یا برخی مسائل اخلاقی را نادیده می‌گیرند یا بیش از حد پیچیده هستند، در بخش III-B یک دسته‌بندی جدید پیشنهاد شده است که مسائل اخلاقی هوش مصنوعی را در سه سطح فردی، اجتماعی، و زیست‌محیطی طبقه‌بندی می‌کند. این دسته‌بندی پیشنهادی تمام مسائل اخلاقی موجود را به صورت جامع پوشش می‌دهد و درک و تحلیل مشکلات اخلاقی ناشی از هوش مصنوعی را تسهیل می‌کند. علاوه بر این، در بخش III-C تلاش می‌کنیم مسائل اخلاقی مرتبط با مراحل چرخه عمر سیستم‌های هوش مصنوعی را ترسیم کنیم، که این امر می‌تواند در شناسایی این مسائل در طی فرایند توسعه سیستم‌های هوش مصنوعی مفید باشد.

هدف اصلی این بخش بحث و روشن‌سازی مسائل اخلاقی هوش مصنوعی است تا متخصصان بتوانند این مسائل را شناسایی و درک کنند و سپس به آن‌ها در مطالعه بیشتر برای حل مشکلات اخلاقی هوش مصنوعی کمک شود. مشارکت اصلی این بخش پیشنهاد یک دسته‌بندی جدید برای مسائل اخلاقی هوش مصنوعی است که مسائل اخلاقی موجود را به شیوه‌ای روشن و آسان

برای درک پوشش می‌دهد. علاوه بر این، مسائل اخلاقی مرتبط با مراحل چرخه عمر سیستم‌های هوش مصنوعی مورد بحث قرار گرفته است.

۲-۱. مرور دسته‌بندی‌های مسائل اخلاقی هوش مصنوعی

این بخش به بررسی نگرانی‌ها یا مسائل اخلاقی هوش مصنوعی از دیدگاه‌های مختلف می‌پردازد و چهار دسته‌بندی متفاوت که در ادبیات جمع‌آوری‌شده یافت شده‌اند، مرور می‌کند. دو مورد از این دسته‌بندی‌ها از گزارش‌های دولتی و دو مورد دیگر از انتشارات علمی گرفته شده‌اند. از دیدگاه‌ها و دسته‌بندی‌های مختلف، مسائل اخلاقی مرتبط نیز تا حدی متفاوت هستند. در ادامه، چهار دسته‌بندی مختلف از مسائل اخلاقی هوش مصنوعی به ترتیب مرور می‌شوند. چهار دسته‌بندی مورد بررسی و دسته‌بندی پیشنهادی ما در جدول I ذکر شده‌اند.

۲-۱-۱. دسته‌بندی بر اساس ویژگی‌های هوش مصنوعی، عوامل انسانی و تأثیر اجتماعی

در مرجع [۱۱]، مسائل اخلاقی هوش مصنوعی به سه دسته اصلی تقسیم شده‌اند: مسائل اخلاقی ناشی از ویژگی‌های هوش مصنوعی، ریسک‌های اخلاقی ناشی از عوامل انسانی، و تأثیر اجتماعی مسائل اخلاقی هوش مصنوعی.

۲-۱-۱-۱. مسائل اخلاقی ناشی از ویژگی‌های هوش مصنوعی

شفافیت: یادگیری ماشین (ML) فناوری اصلی هوش مصنوعی کنونی، به ویژه شبکه‌های عصبی عمیق است. با این حال، فرایند استنباط یادگیری ماشین، که معمولاً به عنوان "جعبه سیاه" شناخته می‌شود، به سختی قابل توضیح و درک است. این عدم شفافیت باعث می‌شود الگوریتم‌ها یا مدل‌ها برای کاربران و حتی توسعه‌دهندگان مرموز به نظر برسند. این مسئله به طور مستقیم به مشکل شفافیت منجر می‌شود. کمبود شفافیت نه تنها مشکلات توضیحی ایجاد می‌کند، بلکه مانع نظارت و هدایت انسانی بر یادگیری ماشین یا هوش مصنوعی نیز می‌شود. بنابراین، شفافیت یا توضیح‌پذیری یکی از معایب به‌شدت مورد بحث هوش مصنوعی است.

امنیت داده و حریم خصوصی: عملکرد هوش مصنوعی کنونی به شدت به داده‌های آموزشی وابسته است. معمولاً مقدار زیادی داده، که احتمالاً شامل داده‌های شخصی و خصوصی است، برای آموزش یک مدل هوش مصنوعی (به ویژه مدل‌های یادگیری عمیق) مورد نیاز است. سوءاستفاده و استفاده مخرب از داده‌ها، مانند نشت اطلاعات شخصی یا دستکاری، از مسائل اخلاقی جدی است که به شدت به افراد، مؤسسات، سازمان‌ها و حتی کشورها مربوط می‌شود. امنیت داده و حریم خصوصی از مسائل کلیدی در توسعه و کاربرد فناوری هوش مصنوعی هستند.

استقلال، قصد و مسئولیت‌پذیری: با پیشرفت هوش مصنوعی، سیستم‌ها یا عامل‌های هوش مصنوعی کنونی، مانند ربات‌های مراقبت بهداشتی، دارای درجه‌ای از استقلال، قصد، و

مسئولیت‌پذیری هستند. استقلال هوش مصنوعی به توانایی یک سیستم هوش مصنوعی برای عمل بدون دخالت یا کنترل مستقیم انسانی اشاره دارد. قصد به توانایی یک سیستم هوش مصنوعی برای انجام اعمالی که می‌توانند اخلاقاً مضر یا مفید باشند اشاره دارد. مسئولیت‌پذیری نشان می‌دهد که سیستم هوش مصنوعی برخی از قوانین اجتماعی و مسئولیت‌های فرضی را برآورده می‌کند. اما اینکه یک سیستم هوش مصنوعی تا چه حد باید استقلال، قصد، و مسئولیت‌پذیری داشته باشد، یک سؤال و مسئله چالش‌برانگیز است.

۲-۱-۱-۲. مسائل اخلاقی ناشی از عوامل انسانی

مسئولیت‌پذیری: وقتی یک سیستم یا عامل هوش مصنوعی در انجام یک وظیفه مشخص شکست بخورد و نتایج نامطلوبی به بار آورد، چه کسی باید پاسخگو باشد؟ نتیجه نامطلوب ممکن است به دلایل مختلفی، از جمله کدهای برنامه‌نویسی، داده‌های ورودی، عملیات نادرست یا عوامل دیگر ایجاد شده باشد. این امر به مشکل معروف به "مسئله دست‌های متعدد" منجر می‌شود. بنابراین، مسئولیت‌پذیری یکی از مسائل اخلاقی مربوط به عوامل انسانی در طراحی، پیاده‌سازی، استقرار، و استفاده از هوش مصنوعی است.

استانداردهای اخلاقی: از آنجا که هدف نهایی اخلاق هوش مصنوعی ایجاد هوش مصنوعی‌ای است که بتواند از اصول اخلاقی پیروی کند و به طور اخلاقی رفتار نماید، ضروری است که استانداردهای جامع و بی‌طرف اخلاقی برای آموزش یا تنظیم هوش مصنوعی ایجاد شود. برای تدوین استانداردهای اخلاقی برای هوش مصنوعی، پژوهشگران و متخصصان باید نظریه‌ها و اصول اخلاقی موجود را به خوبی درک کنند.

قوانین حقوق بشر: طراحان، مهندسان نرم‌افزار و سایر افرادی که در طراحی و کاربرد سیستم‌های هوش مصنوعی مشارکت دارند، باید قوانین حقوق بشر را فرا گیرند. بدون آموزش در زمینه قوانین حقوق بشر، ممکن است به صورت ناآگاهانه حقوق اساسی بشر را نقض کنند. قوانین یا اسناد حقوق بشر که در کشورهای مختلف یا مناطق مختلف دنبال می‌شوند، اغلب با یکدیگر متفاوت هستند. قوانین حقوق بشر متعددی، مانند حقوق بین‌المللی بشر، میثاق بین‌المللی حقوق مدنی و سیاسی، میثاق بین‌المللی حقوق اقتصادی، اجتماعی و فرهنگی، اعلامیه جهانی حقوق بشر، منشور سازمان ملل متحد و کنوانسیون اروپایی حفاظت از حقوق بشر و آزادی‌های بنیادین، توسط دولت‌های مختلف منتشر شده‌اند.

۲-۱-۱-۳. تأثیر اجتماعی مسائل اخلاقی هوش مصنوعی

اتوماسیون و جایگزینی شغل

با جایگزینی تعداد بیشتری از کارگران کارخانه با سیستم‌های خودکار و ربات‌ها، هوش مصنوعی بازار کار را دچار اختلال و تحول می‌کند. بنابراین، بسیاری از افراد نگران اتوماسیون و جایگزینی شغلی هستند.

دسترس‌پذیری

دسترس‌پذیری یا قابلیت استفاده از فناوری‌های نوظهور، مانند هوش مصنوعی، تأثیر مستقیمی بر رفاه انسان دارد. با این حال، اگر تنها بخشی از جمعیت از مزایای هوش مصنوعی بهره‌مند شوند، این امر غیراخلاقی و ناعادلانه خواهد بود. باید توجه شود که محصولات و خدمات هوش مصنوعی به گونه‌ای توسعه یابند که برای همه قابل دسترس باشند، به طوری که مزایای هوش مصنوعی به طور مساوی در میان همه گسترش یابد.

دموکراسی و حقوق مدنی

هوش مصنوعی غیراخلاقی حقیقت را تحریف می‌کند و در نهایت منجر به از دست دادن اعتماد و حمایت عمومی از فناوری هوش مصنوعی می‌شود. قدرت دموکراسی‌ها با از دست دادن جوامع آگاه و اعتمادکننده آسیب می‌بیند. با آسیب‌دیدن دموکراسی‌ها و تشدید تعصب‌های ساختاری، بهره‌مندی آزادانه از حقوق مدنی دیگر به طور مداوم برای همه قابل دسترس نخواهد بود. بنابراین، دموکراسی و حقوق مدنی باید در اخلاق هوش مصنوعی مدنظر قرار گیرند.

۲-۱-۲. دسته‌بندی بر اساس آسیب‌پذیری‌های هوش مصنوعی و انسان

در مرجع [29]، **لیائو** مسائل اخلاقی هوش مصنوعی را به دو دسته تقسیم می‌کند:

۳. مسائل اخلاقی که به دلیل محدودیت‌های سیستم‌های یادگیری ماشین (ML) کنونی به وجود می‌آیند و به عنوان "آسیب‌پذیری‌های هوش مصنوعی (به ویژه یادگیری ماشین)" شناخته می‌شوند.

۴. مسائل اخلاقی که به دلیل عملکرد بیش از حد خوب سیستم‌های یادگیری ماشین کنونی به وجود می‌آیند و انسان‌ها در حضور یا تعامل با این سیستم‌های هوشمند آسیب‌پذیر می‌شوند، که به عنوان "آسیب‌پذیری‌های انسانی" مطرح می‌شوند.

۲-۱-۲-۱. مسائل اخلاقی ناشی از آسیب‌پذیری‌های هوش مصنوعی

یادگیری ماشین تشنه‌ی داده است: معمولاً یادگیری ماشین به مقدار زیادی داده نیاز دارد تا عملکرد خوبی داشته باشد. این امر شرکت‌ها و سازمان‌ها را به جمع‌آوری یا خرید داده، از جمله داده‌های حساس شخصی، حتی اگر این کار ممکن است حق حریم خصوصی فرد را نقض کند، ترغیب می‌کند.

زباله وارد شود، زباله خارج می‌شود: عملکرد الگوریتم یادگیری ماشین به شدت به داده‌هایی که از آن‌ها یاد می‌گیرد وابسته است. اگر یک الگوریتم یادگیری ماشین با داده‌های ناکافی یا نادرست آموزش داده شود، حتی اگر طراحی خوبی داشته باشد، نتایج نامطلوبی ارائه خواهد کرد.

الگوریتم‌های معیوب: حتی اگر یک الگوریتم یادگیری ماشین با داده‌های کافی و دقیق وارد شود، اگر خود الگوریتم بد باشد، پیش‌بینی‌های نادرستی ارائه می‌دهد. برای مثال، یک الگوریتم

بد ممکن است نتواند یک الگو را تشخیص دهد حتی اگر وجود داشته باشد، یا ممکن است یک الگو را شناسایی کند حتی اگر وجود نداشته باشد، که به ترتیب به عنوان "کم‌برازش" و "بیش‌برازش" شناخته می‌شوند.

یادگیری عمیق یک جعبه‌ی سیاه است: یادگیری عمیق یک جعبه سیاه است که مسائلی مانند توضیح‌پذیری، تفسیرپذیری و اعتماد را ایجاد می‌کند. حتی برای طراحان و توسعه‌دهندگان یادگیری عمیق، این مدل غیرقابل درک است، زیرا معمولاً شامل هزاران یا میلیون‌ها ارتباط بین نورون‌های مختلف است. بنابراین، توضیح چگونگی تعامل این ارتباطات و دلیل پیش‌بینی‌های خاص مدل دشوار است.

۲-۱-۲-۲. مسائل اخلاقی ناشی از آسیب‌پذیری‌های انسانی

سوءاستفاده از هوش مصنوعی: فناوری‌های هوش مصنوعی، مانند تشخیص چهره و تولید تصویر، می‌توانند بهتر از انسان عمل کنند. با این حال، مسائل اخلاقی وجود دارند زیرا مردم ممکن است به استفاده نادرست از این فناوری‌ها وسوسه شوند. برای مثال، یک دولت می‌تواند از فناوری تشخیص چهره برای نظارت بر شهروندان خود استفاده کند یا یادگیری ماشین می‌تواند برای ساخت عکس‌ها یا ویدیوهای جعلی که انسان‌ها نمی‌توانند واقعی یا جعلی بودن آن‌ها را تشخیص دهند، استفاده شود.

جایگزینی شغل: از آنجا که ربات‌های هوشمند می‌توانند وظایف خاصی را سریع‌تر و بهتر از انسان‌ها انجام دهند، بسیاری از افراد نگران هستند که ربات‌ها و سایر فناوری‌های هوش مصنوعی بخش بزرگی از نیروی کار انسانی را در آینده‌ای نزدیک جایگزین کنند.

مسائل مربوط به همدم‌های رباتیک: با پیچیده‌تر شدن ربات‌های هوش مصنوعی، آن‌ها به عنوان همدم‌های انسان در نظر گرفته شده‌اند. این امر مسائل اخلاقی‌ای را در مورد رابطه بین انسان و همدم‌های رباتیک ایجاد می‌کند.

۲-۱-۳. دسته‌بندی بر اساس الگوریتم، داده‌ها، کاربرد و ریسک‌های بلندمدت و غیرمستقیم اخلاقی

در گزارش تحلیلی ریسک‌های اخلاقی هوش مصنوعی که توسط **گروه کاری استانداردسازی ملی هوش مصنوعی چین** منتشر شده است، مسائل اخلاقی هوش مصنوعی به چهار جنبه زیر تقسیم می‌شود:

- مسائل اخلاقی مرتبط با الگوریتم‌های هوش مصنوعی؛
- مسائل اخلاقی مرتبط با داده‌ها؛
- مسائل اخلاقی مرتبط با کاربرد هوش مصنوعی؛
- ریسک‌های بلندمدت و غیرمستقیم اخلاقی.

۲-۱-۳-۱. مسائل اخلاقی مرتبط با الگوریتم‌ها

امنیت الگوریتم: الگوریتم‌های هوش مصنوعی چندین مشکل امنیتی ایجاد می‌کنند. نخست، خطر نشت الگوریتم یا مدل وجود دارد. به طور معمول، مدل از طریق آموزش با داده‌های آموزشی و بهینه‌سازی پارامترهایش به دست می‌آید. اگر پارامترهای مدل یک الگوریتم افشا شوند، یک طرف ثالث ممکن است بتواند مدل را کپی کند. این امر به مالک مدل خسارات اقتصادی وارد می‌کند، زیرا طرف ثالث بدون پرداخت هزینه داده‌های آموزشی، مدل مشابهی به دست می‌آورد. دوم، پارامترهای مدل الگوریتم هوش مصنوعی ممکن است به صورت غیرقانونی توسط مهاجمان تغییر کنند، که این امر باعث کاهش عملکرد مدل هوش مصنوعی و ایجاد نتایج نامطلوب می‌شود. علاوه بر این، در بسیاری از سناریوها، خروجی مدل به شدت با امنیت شخصی مرتبط است، مانند حوزه‌های پزشکی و رانندگی خودکار. در صورت وجود حفره‌ها یا اشتباهات در الگوریتم‌ها در این حوزه‌ها، آسیب‌های مستقیم به انسان وارد می‌شود و عواقب جدی به دنبال خواهد داشت.

توضیح‌پذیری الگوریتم: به دلیل ویژگی "جعبه سیاه" بسیاری از الگوریتم‌های یادگیری ماشین، به ویژه یادگیری عمیق یا شبکه‌های عصبی، فرایند تصمیم‌گیری الگوریتم‌های هوش مصنوعی دشوار است. تفسیرپذیری یا توضیح‌پذیری الگوریتم‌ها یک مسئله اخلاقی اساسی در هوش مصنوعی است، زیرا به حق انسان برای دانستن مربوط می‌شود.

معضل تصمیم‌گیری الگوریتمی: پس از به دست آوردن مدل هوش مصنوعی، نتایج الگوریتم معمولاً برای ما غیرقابل پیش‌بینی است. به عبارت دیگر، حتی اگر یک مدل هوش مصنوعی به خوبی طراحی شود، نمی‌توان تصمیمات الگوریتم و نتایج آن را پیش‌بینی کرد. این امر به معضل یا خطر تصمیم‌گیری الگوریتمی هوش مصنوعی منجر می‌شود. به عنوان مثال، خودروهای خودران باید تصادفات را کاهش دهند، اما گاهی باید بین دو گزینه بد انتخاب کنند، مانند برخورد با عابران پیاده یا فدا کردن خود و سرنشینان برای نجات عابران.

۲-۱-۳-۲. مسائل اخلاقی مرتبط با داده‌ها

حفظ حریم خصوصی: با توسعه داده‌های بزرگ و هوش مصنوعی، تنش بین فناوری هوش مصنوعی و حفاظت از حریم خصوصی کاربران به طور فزاینده‌ای جدی‌تر شده است. مجرمان راه‌های بیشتری برای دسترسی به داده‌های خصوصی شخصی با هزینه کمتر و سود بیشتر دارند. حوادث امنیتی داده‌ها در سال‌های اخیر به طور مکرر رخ داده است. حفاظت از حریم خصوصی به یک مسئله اخلاقی جدی و شناخته‌شده در استفاده از هوش مصنوعی تبدیل شده است.

شناسایی و پردازش اطلاعات شخص و حساس: قوانین و مقررات سنتی تنها بر حفاظت از اطلاعات شخصی و حساس تمرکز دارند. اگر اطلاعات شخصی یا حساس از طریق روش‌هایی مانند تصادفی‌سازی یا سنتز داده‌ها ناشناس شوند، دیگر به عنوان اطلاعات شخصی یا حساس

در نظر گرفته نمی‌شوند و تحت حفاظت قوانین سنتی قرار نمی‌گیرند. استفاده، اشتراک‌گذاری و انتقال چنین اطلاعاتی، مسائل اخلاقی جدیدی ایجاد می‌کند.

۲-۱-۳-۳. مسائل اخلاقی مرتبط با کاربرد

تبعیض الگوریتمی: نتایج اجرای الگوریتم‌ها مستقیماً بر تصمیم‌گیری سیستم‌های هوش مصنوعی تأثیر می‌گذارد. با این حال، تبعیض یا تعصب الگوریتمی در بسیاری از کاربردهای هوش مصنوعی مشاهده شده است. برای مثال، تعصب نژادی در سیستم‌های عدالت کیفری یا تبعیض جنسیتی در استخدام.

سوءاستفاده از الگوریتم‌ها: سوءاستفاده از الگوریتم‌ها به وضعیتی اشاره دارد که افراد از الگوریتم‌ها برای تحلیل، تصمیم‌گیری، هماهنگی و فعالیت‌های دیگر استفاده می‌کنند، اما هدف استفاده، روش استفاده، یا دامنه استفاده نادرست بوده و اثرات منفی ایجاد می‌کند. برای مثال، الگوریتم‌های تشخیص چهره می‌توانند برای ارتقای امنیت عمومی و تسریع در کشف مظنونان جنایی استفاده شوند، اما اگر برای شناسایی مجرمان بالقوه یا تعیین احتمال ارتکاب جرم بر اساس چهره فرد به کار گرفته شوند، این امر سوءاستفاده از الگوریتم محسوب می‌شود.

۲-۱-۳-٤. ریسک‌های بلندمدت و غیرمستقیم اخلاقی

اشتغال: با پیشرفت سریع و کاربرد گسترده هوش مصنوعی، کارهای بیشتری توسط برخی محصولات هوش مصنوعی انجام می‌شود. این امر تأثیر قابل‌توجهی بر مسئله اشتغال خواهد داشت.

مالکیت: با پیشرفت مداوم هوش مصنوعی، تفاوت‌های فکری بین عامل‌های هوش مصنوعی و انسان‌ها به تدریج کاهش می‌یابد. در نتیجه، بحث‌های متعددی درباره مالکیت مطرح می‌شود، مانند اینکه آیا عامل‌های هوش مصنوعی باید به عنوان "موضوع حقوقی" در نظر گرفته شوند و آیا محصولات هوش مصنوعی دارای حقوق مالکیت (کپی‌رایت یا حقوق ثبت اختراع) هستند یا خیر.

رقابت: رقابت ناعادلانه، رقابت مخرب و رفتارهای انحصارگرایانه با مزایای فناوری تأثیرات منفی بر ثبات اجتماعی، آزادی بازار، عدالت و ارزش برابر دارند و به شدت منافع مصرف‌کنندگان را آسیب می‌رسانند و بهبود رفاه اجتماعی را مانع می‌شوند.

مسئولیت‌پذیری: با کاربرد گسترده هوش مصنوعی، موارد بسیاری مشاهده شده است که محصولات هوش مصنوعی قوانین یا اخلاق را نقض کرده‌اند، مانند آسیب شخصی یا تبعیض الگوریتمی. در چنین مواردی، مسئله اساسی این است که چه کسی مسئول این عواقب بد است. برای مثال، در رانندگی خودکار که موضوعات مختلفی مانند مالک خودرو، راننده، سرنشینان، تولیدکنندگان خودرو و ارائه‌دهندگان سیستم خودران را درگیر می‌کند، مسئولیت‌ها در صورت وقوع تصادف چگونه تقسیم می‌شود.

۲-۱-٤. دسته‌بندی بر اساس استقرار هوش مصنوعی

در جدیدترین مطالعه **خدمات پژوهشی پارلمان اروپا** درباره پیامدهای اخلاقی و سؤالات اخلاقی مرتبط با هوش مصنوعی، مسائل اخلاقی بر اساس تأثیرات هوش مصنوعی بر جامعه انسانی، روان‌شناسی انسانی، سیستم مالی، سیستم قانونی، محیط زیست و سیاره، و اعتماد طبقه‌بندی شده‌اند.

۲-۱-٤-۱. تأثیر بر جامعه

بازار کار: هوش مصنوعی در حال حاضر در بخش‌هایی مانند امور مالی، تولید پیشرفته، حمل‌ونقل، توسعه انرژی، مراقبت‌های بهداشتی و بسیاری حوزه‌های دیگر به کار گرفته شده است. اثرات اتوماسیون بر مشاغل کارگری یا "یقه آبی" به وضوح قابل مشاهده است. با پیشرفته‌تر شدن عامل‌های هوش مصنوعی یا ربات‌ها، تعداد بیشتری از شغل‌ها تحت تأثیر فناوری‌های هوش مصنوعی قرار می‌گیرند و بسیاری از موقعیت‌های شغلی از بین خواهند رفت. این امر می‌تواند خطر بیکاری گسترده را در بسیاری از بخش‌های شغلی به همراه داشته باشد.

نابرابری: فناوری‌های هوش مصنوعی انتظار می‌رود که عملیات تجاری شرکت‌ها را ساده‌تر و بهره‌وری را افزایش دهند. با این حال، برخی افراد معتقدند این کار به هزینه نیروی کار انسانی انجام خواهد شد. در نتیجه، درآمدها در میان افراد کمتری توزیع می‌شود و صاحبان شرکت‌های مبتنی بر هوش مصنوعی از مزایای نامتناسبی بهره‌مند خواهند شد، که به افزایش نابرابری‌های اجتماعی منجر می‌شود.

حریم خصوصی، حقوق بشر و کرامت انسانی: دستیارهای شخصی هوشمند مانند **سیری اپل**، **اکوی آمازون** و **هوم گوگل** می‌توانند علایق و رفتار کاربران را بیاموزند، اما در عین حال نگرانی‌هایی درباره اینکه این دستگاه‌ها دائماً فعال و در حال گوش دادن هستند، مطرح می‌شود. همچنین، هوش مصنوعی می‌تواند برای تعیین باورهای سیاسی افراد استفاده شود، که ممکن است آن‌ها را در برابر دست‌کاری آسیب‌پذیر کند.

تعصب: تعصبات انسانی مانند تعصب جنسیتی یا نژادی می‌توانند به هوش مصنوعی منتقل شوند. تعصب هوش مصنوعی ممکن است ناشی از داده‌های آموزشی، ارزش‌های توسعه‌دهندگان یا کاربران، یا فرایند یادگیری خود هوش مصنوعی باشد.

دموکراسی: پیاده‌سازی و پذیرش هوش مصنوعی می‌تواند به چندین روش به دموکراسی آسیب برساند. از جمله تمرکز قدرت در دستان چند شرکت بزرگ، تأثیرگذاری بر انتخابات سیاسی، و قطبی‌شدن اجتماعی از طریق سیستم‌های توصیه خبر مبتنی بر هوش مصنوعی.

۲-۱-٤-۲. تأثیر بر روان‌شناسی انسان

روابط: هوش مصنوعی در حال پیشرفت در تقلید از تفکر، تجربه، رفتار، و روابط انسانی است. این امر ممکن است بر روابط واقعی انسانی تأثیر بگذارد و مسائل اخلاقی جدیدی ایجاد کند.

شخصیت: با انجام وظایف و تصمیماتی که به طور سنتی توسط انسان انجام می‌شوند، این سؤال اخلاقی مطرح می‌شود که آیا سیستم‌های هوش مصنوعی باید دارای حقوق و شخصیت حقوقی شوند.

۲-۱-٤-۳. تأثیر بر سیستم مالی

استفاده از هوش مصنوعی در بازارهای مالی به طور قابل توجهی کارایی تراکنش و حجم معاملات را بهبود بخشیده است. بازارها برای اتوماسیون، بسیار مناسب هستند، زیرا در حال حاضر تقریباً به طور کامل به صورت الکترونیکی کار می‌کنند و حجم عظیمی از داده‌ها با سرعت بالایی تولید می‌شود که نیاز به استفاده از الگوریتم‌هایی برای هضم و تجزیه و تحلیل آن دارد. علاوه بر این، به دلیل پویایی بازارها، واکنش سریع به اطلاعات بسیار مهم است [۵۹]، که انگیزه‌های قابل توجهی را برای جایگزینی فرایند تصمیم گیری کُند افراد با تصمیم گیری الگوریتمی فراهم می‌کند. علاوه بر این، جوایز برای تصمیمات تجاری موثر قابل توجه است، که توضیح می‌دهد که چرا شرکت‌ها در فناوری هوش مصنوعی سرمایه گذاری زیادی کرده اند. با این حال، عوامل معاملاتی خودکار مبتنی بر هوش مصنوعی نیز ممکن است به طور مخرب برای بی ثبات کردن بازارها یا آسیب رساندن به طرف‌های بی گناه از راه‌های دیگر استفاده شوند. حتی اگر قصد اصلی، مخرب بودن آنها نباشد. استقلال و انعطاف‌پذیری استراتژی‌های معاملاتی الگوریتمی، از جمله استفاده روزافزون از تکنیک‌های ML، پیش‌بینی عملکرد آنها در موقعیت‌های غیرمنتظره را برای افراد دشوار می‌کند.

۲-۱-٤-٤. تأثیر بر سیستم قانونی

حقوق جزا: بر اساس قوانین جزایی فعلی، جرم از دو عنصر تشکیل شده است، یعنی فعل (یا ترک فعل) اختیاری و قصد ارتکاب جرم. اگر نشان داده شود که محصولات یا ربات‌های هوش مصنوعی از هشیاری یا آگاهی کافی برخوردارند، ممکن است آنها مرتکب مستقیم جرایم جنایی یا مسئول جنایات یا سهل‌انگاری‌ها باشند. اگر بپذیریم که محصولات هوش مصنوعی دارای ذهن، اراده آزاد مانند انسان، استقلال یا حس اخلاقی خاص خود هستند، در این صورت قوانین جزایی ما و حتی کل سیستم حقوقی ما باید مورد بازنگری قرار گیرند [٦٠].

قانون شکنجه: قانون شکنجه شرایطی مانند آسیب رفتاری یک فرد، رنج، ضرر ناعادلانه یا آسیب رساندن به شخص دیگر را پوشش می‌دهد. هنگامی که تصادفی با خودرو(های) خودران رخ می‌دهد، دو حوزه قانونی مرتبط وجود دارد: سهل انگاری و مسئولیت محصول. در حالی که امروزه بیشتر تصادفات، ناشی از خطای راننده است که نشان می‌دهد مسئولیت تصادفات بر اساس اصل سهل انگاری تنظیم می‌شود. بنابراین، در آینده، قانون جرم، که شامل انواع مختلفی از دعاوی آسیب شخصی است، به طور قابل توجهی تحت تأثیر قرار خواهد گرفت [٦١] زیرا محصولات هوش مصنوعی (مانند اتومبیل‌های خودران یا سایر ربات‌های هوشمند) در دعاوی

صدمات شخصی دخیل خواهند بود. به عنوان تصادف بین اتومبیل‌های خودران یا ادعای جراحت که در آن ربات به انسان آسیب می‌رساند.

۲-۱-٤-٥. تأثیر بر محیط زیست و سیاره

استفاده از منابع طبیعی: توسعه و کاربرد هوش مصنوعی تقاضای بسیاری از منابع طبیعی مانند فلزات خاکی کمیاب مانند نیکل، کبالت، گرافیت و غیره را افزایش می‌دهد. با کاهش عرضه‌ی موجود، اپراتورها ممکن است مجبور شوند در محیط‌های جدید و پیچیده‌تر برای استخراج کار کنند. این امر باعث افزایش میزان تولید و مصرف فلزات کمیاب خاکی و آسیب بیشتر به محیط زیست می‌شود [٦٢].

آلودگی و ضایعات: افزایش تولید و مصرف دستگاه‌های فناوری هوش مصنوعی مانند ربات‌ها باعث تشدید آلودگی و ضایعات مانند تجمع فلزات سنگین و مواد سمی در محیط می‌شود [٦٣].

نگرانی‌های انرژی: استفاده از فناوری هوش مصنوعی، به‌ویژه یادگیری عمیق، عموماً شامل آموزش مدل‌های ML بر روی حجم عظیمی از داده است که معمولاً مقادیر زیادی انرژی مصرف می‌کند. با توجه به داده‌های فهرست شده در [٦٤]، اثر کربن آموزش یک مدل پردازش زبان طبیعی (مدل ترانسفورماتور) تقریباً ٥ برابر اثر کربن یک ماشین متوسط در کل طول عمر آن است.

۲-۱-٤-٦. تأثیر بر اعتماد

هوش مصنوعی نوید تغییرات و مزایای متعددی را برای زندگی افراد و جامعه می‌دهد. این مسئله در حال تغییر زندگی روزمره ما در بسیاری از حوزه ها، مانند حمل و نقل، صنعت خدمات، مراقبت‌های بهداشتی، آموزش، ایمنی و امنیت عمومی، و سرگرمی است. با این وجود، این سیستم‌های هوش مصنوعی باید به گونه‌ای معرفی شوند که اعتماد و درک را تقویت کند و به حقوق بشر و مدنی احترام بگذارد [٦٥]. اتفاق نظر در میان جامعه‌ی تحقیقاتی این است که اعتماد به هوش مصنوعی تنها از طریق انصاف، شفافیت، مسئولیت پذیری و مقررات (یا کنترل) حاصل می‌شود.

انصاف: اعتماد به هوش مصنوعی، باید منصفانه و بی طرفانه باشد. همانطور که تصمیمات بیشتر و بیشتری به هوش مصنوعی واگذار می‌شود، ما باید اطمینان حاصل کنیم که این تصمیمات عاری از تعصب و تبعیض هستند [٦٦]. چه فیلتر کردن رزومه‌ها برای مصاحبه‌های شغلی باشد یا تصمیم گیری در مورد پذیرش در دانشگاه یا انجام رتبه بندی اعتباری برای شرکت‌های وام، اساساً ضروری است که تصمیمات اتخاذ شده توسط هوش مصنوعی منصفانه باشد.

شفافیت: شفافیت برای ایجاد اعتماد در هوش مصنوعی مهم است، زیرا باید دانست که چرا یک سیستم هوش مصنوعی تصمیم خاصی گرفته است، به خصوص اگر آن تصمیم باعث

عواقب نامطلوب یا آسیب شود. با توجه به اینکه اتوپایلوت یک خودروی هوشمند منجر به تصادفات مرگبار متعددی شده است، واضح است که برای کشف چگونگی و چرایی وقوع این تصادفات و رفع نقص فنی یا عملیاتی، شفاف سازی ضروری است. مشخص نبودن هسته‌ی ML و ایهام در ML، که به جعبه سیاه معروف است، یکی از موانع اصلی شفافیت هوش مصنوعی است [۵۱].

مسئولیت پذیری: مسئولیت پذیری [٦٧] تضمین می‌کند که اگر یک سیستم هوش مصنوعی مرتکب اشتباهی شود یا به کسی آسیب برساند، می‌توان مسئولیت آن را بر عهده گرفت، خواه طراح باشد، توسعه دهنده یا شرکتی باشد که هوش مصنوعی را می‌فروشد. در صورت بروز خسارت، پاسخگویی برای ایجاد یک مکانیسم اصلاحی ضروری است تا قربانیان بتوانند غرامت کافی را دریافت کنند. بنابراین، پاسخگویی برای اطمینان از اعتماد هوش مصنوعی بسیار مهم است.

کنترل: موضوع دیگری که بر اعتماد عمومی به هوش مصنوعی تأثیر می‌گذارد، کنترل پذیری هوش مصنوعی است [٦٨]. این مسئله تا حد زیادی به ترس مردم از ایده‌ی "اَبَر هوش (super-intelligence)" مربوط می‌شود که به معنی افزایش هوش AI به حدی است که از توانایی‌های انسانی پیشی می‌گیرد، ممکن است هوش مصنوعی کنترل منابع ما را به دست گرفته و از گونه‌های ما پیشی بگیرد. حتی می‌تواند منجر به انقراض انسان شود. یک نگرانی مرتبط به این موضوع این است که حتی اگر یک عامل هوش مصنوعی به دقت طراحی شده باشد تا اهداف خود را با نیازهای انسان هماهنگ کند، ممکن است به تنهایی اهداف فرعی غیرقابل پیش بینی ایجاد کند. بنابراین، برای حفظ اعتماد به هوش مصنوعی، مهم است که انسان‌ها نظارت یا کنترل نهایی بر فناوری هوش مصنوعی داشته باشند.

جدول ۱: مسائل اخلاقی هوش مصنوعی

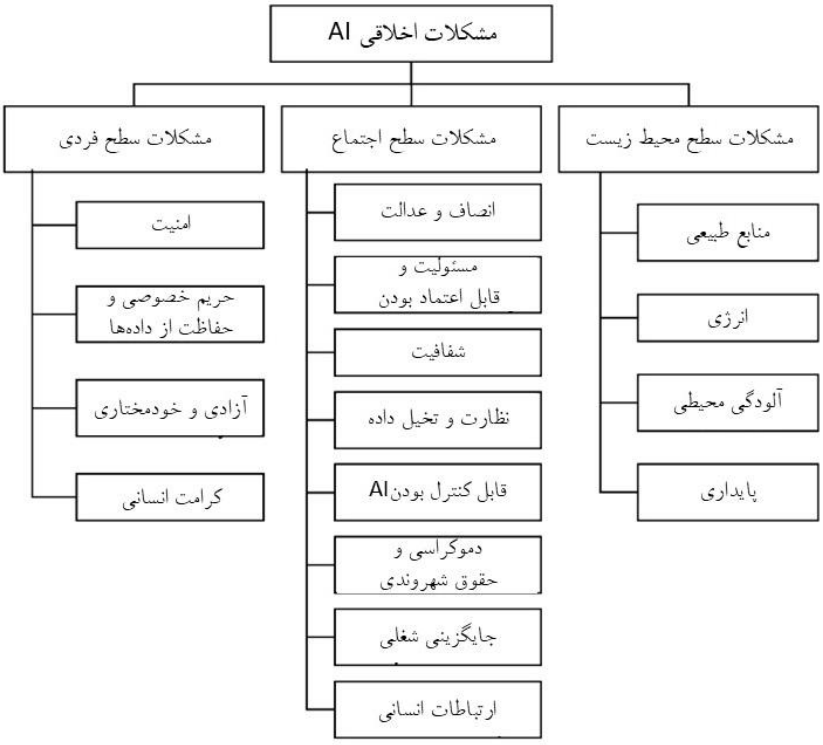
طبقه‌بندی	کلاس	مسائل اخلاقی	بحث
طبقه‌بندی مسائل اخلاقی هوش مصنوعی بر اساس ویژگی‌های هوش مصنوعی، عوامل انسانی و تأثیر اجتماعی [۱۱]	مسائل اخلاقی ناشی از ویژگی‌های هوش مصنوعی، مسائل اخلاقی ناشی از عوامل انسانی، تأثیر اجتماعی مسائل اخلاقی هوش مصنوعی	شفافیت، امنیت داده‌ها و حریم خصوصی، خودمختاری، قصد و مسئولیت‌پذیری؛ پاسخگویی، استانداردهای اخلاقی، قوانین حقوق بشر؛ اتوماسیون و جایگزینی شغل، دسترسی، دموکراسی و حقوق مدنی	تأثیرات هوش مصنوعی بر محیط زیست، مانند مصرف منابع طبیعی و آلودگی محیط زیست، نادیده گرفته شده است.

طبقه‌بندی مسائل اخلاقی هوش مصنوعی بر اساس آسیب‌پذیری‌های هوش مصنوعی و انسان [۲۹]	مسائل اخلاقی ناشی از آسیب‌پذیری‌های هوش مصنوعی، مسائل اخلاقی ناشی از آسیب‌پذیری‌های انسانی	هوش مصنوعی نیازمند داده‌های زیاد است، ورود داده‌های نامعتبر/خروج داده‌های نامعتبر، الگوریتم‌های معیوب، یادگیری عمیق به عنوان جعبه سیاه؛ سوءاستفاده از هوش مصنوعی، جایگزینی شغل، مسائل مربوط به همراهان روباتیک	چندین مسئله مهم، مانند مسئولیت‌پذیری، ایمنی، آزادی، و مشکلات زیست‌محیطی، نادیده گرفته شده‌اند.
طبقه‌بندی مسائل اخلاقی هوش مصنوعی بر اساس مسائل مربوط به الگوریتم، داده، کاربرد، و ریسک‌های بلندمدت و غیرمستقیم [۳۸]	مسائل اخلاقی مربوط به الگوریتم، مسائل اخلاقی مربوط به داده، مسائل اخلاقی مربوط به کاربرد، ریسک‌های بلندمدت و غیرمستقیم	امنیت الگوریتم، قابلیت توضیح الگوریتم، معضلات تصمیم‌گیری الگوریتمی؛ حفاظت از حریم خصوصی، شناسایی و پردازش اطلاعات حساس شخصی؛ تبعیض الگوریتمی، سوءاستفاده از الگوریتم؛ اشتغال، مالکیت، رقابت، مسئولیت‌پذیری	مسائل مرتبط با پاسخگویی، عدالت، خودمختاری و آزادی، کرامت انسانی، مشکلات زیست‌محیطی شامل نمی‌شوند.
طبقه‌بندی مسائل اخلاقی هوش مصنوعی بر اساس استقرار هوش مصنوعی [۵۱]	تأثیر بر جامعه، تأثیر بر روانشناسی انسانی، تأثیر بر سیستم مالی، تأثیر بر سیستم حقوقی، تأثیر بر محیط زیست و سیاره، تأثیر بر اعتماد	بازار کار، نابرابری، حریم خصوصی، حقوق بشر و کرامت انسانی، تعصب، دموکراسی؛ روابط، شخصیت انسانی؛ قوانین کیفری، قوانین حقوقی؛ استفاده از منابع طبیعی، آلودگی و زباله، نگرانی‌های انرژی؛ عدالت، شفافیت، پاسخگویی، کنترل	برخی مسائل، از جمله مسئولیت‌پذیری، ایمنی، و پایداری، حذف شده‌اند و این طبقه‌بندی پیچیده و دشوار درک است.
طبقه‌بندی ما: طبقه‌بندی مسائل	مسائل اخلاقی در سطح فردی، مسائل	ایمنی، حریم خصوصی و حفاظت از داده‌ها، آزادی	طبقه‌بندی ما مسائل اخلاقی هوش

اخلاقی هوش مصنوعی در سطح فردی، اجتماعی و زیست‌محیطی	اخلاقی در سطح اجتماعی، مسائل اخلاقی در سطح زیست‌محیطی	و خودمختاری، کرامت انسانی؛ عدالت و انصاف، مسئولیت‌پذیری و پاسخگویی، شفافیت، نظارت و داده‌سازی، کنترل‌پذیری هوش مصنوعی، دموکراسی و حقوق مدنی، جایگزینی شغل، روابط انسانی؛ منابع طبیعی، انرژی، آلودگی زیست‌محیطی، پایداری	مصنوعی را از سطح فردی، اجتماعی و زیست‌محیطی دسته‌بندی می‌کند. این طبقه‌بندی نه‌تنها واضح و قابل فهم است، بلکه به طور جامع مسائل اخلاقی مورد بحث را پوشش می‌دهد.

۲-۲. دسته بندی پیشنهادی ما: مسائل اخلاقی در سطوح فردی، اجتماعی و محیطی

در بخش قبل، ما مسائل اخلاقی هوش مصنوعی را که در بخش ادبیات تحقیق، شرح داده شده و طبقه بندی شده اند، مرور کرده ایم (جدول ۱ را ببینید). با این حال، دسته بندی‌های ارائه شده در بالا دارای نقص‌های آشکار هستند. به طور خاص، طبقه‌بندی بر اساس ویژگی‌های هوش مصنوعی، عوامل انسانی و تأثیرات اجتماعی [۱۱]، به وضوح تأثیر هوش مصنوعی بر محیط‌زیست، مانند مصرف منابع طبیعی و آلودگی محیط‌زیست را نادیده می‌گیرد. طبقه بندی بر اساس آسیب پذیری‌های هوش مصنوعی و انسان [۲۹] چندین موضوع مهم مانند مسئولیت، ایمنی و مشکلات زیست محیطی را حذف می‌کند. طبقه‌بندی بر اساس الگوریتم، داده‌ها، کاربردها و ریسک‌های اخلاقی بلندمدت و غیرمستقیم [۳۸] ملاحظات انصاف، استقلال و آزادی، کرامت انسانی، مشکلات زیست‌محیطی و غیره را نادیده می‌گیرد. اگرچه طبقه بندی بر اساس استقرار هوش مصنوعی [۵۱] مسائل اخلاقی را به طور جامع پوشش می‌دهد، این طبقه بندی بسیار دست و پا گیر است و برخی مسائل از جمله مسئولیت، ایمنی و پایداری حذف شده اند. این موضوع به ما انگیزه می‌دهد تا مسائل اخلاقی هوش مصنوعی را بیشتر تحلیل و مرتب کنیم.



شکل ۲: دسته‌بندی ارائه‌شده برای مشکلات اخلاقی AI

بدون شک سیستم‌های هوش مصنوعی عمدتاً به افراد یا عموم جامعه خدمت ارائه می‌دهند. از این رو، می‌توانیم مسائل اخلاقی هوش مصنوعی را از منظر فردی و اجتماعی تحلیل و روشن کنیم. در عین حال، محصولات هوش مصنوعی به‌عنوان نهادهای روی کره‌ی زمین، ناگزیر بر محیط‌زیست تأثیر خواهند داشت. بنابراین، مسائل اخلاقی مربوط به جنبه‌های زیست محیطی نیز باید مورد توجه قرار گیرند. بنابراین، در این بخش، ما پیشنهاد کردیم که موضوعات اخلاقی هوش مصنوعی را در سه سطح مختلف، یعنی مسائل اخلاقی در سطوح فردی، اجتماعی و محیطی طبقه بندی کنیم. مسائل اخلاقی در سطح فردی عمدتاً شامل موضوعاتی می‌شود که پیامدهای نامطلوبی برای افراد انسانی، حقوق و رفاه آنها دارد [۶۹]. مسائل اخلاقی هوش مصنوعی در سطح اجتماعی، پیامدهای اجتماعی را در نظر می‌گیرد که هوش مصنوعی برای گروه‌ها یا جامعه‌ای از افراد به عنوان یک کل، به ارمغان آورده یا ممکن است به همراه داشته باشد [۶۹]. مسائل اخلاقی هوش مصنوعی در سطح زیست محیطی بر تأثیرات هوش مصنوعی بر محیط طبیعی متمرکز است. دسته بندی پیشنهادی ما در شکل ۲ نشان داده شده است.

۲-۲-۱. مسائل اخلاقی در سطح فردی

در سطح فردی، هوش مصنوعی بر ایمنی، حریم خصوصی، استقلال و کرامت انسانی افراد تأثیر گذاشته است. استفاده از هوش مصنوعی خطراتی را برای ایمنی افراد به همراه داشته است. به عنوان مثال، در چند سال گذشته تصادفات آسیب دیدگی با خودروهای خودران و ربات‌ها رخ داده و گزارش شده است. مسئله حریم خصوصی یکی از خطرات جدی است که هوش مصنوعی برای ما به ارمغان می‌آورد. برای دستیابی به عملکرد خوب، سیستم‌های هوش مصنوعی معمولاً به مقدار زیادی داده نیاز دارند که اغلب شامل داده‌های خصوصی کاربران می‌شود. با این حال، خطرات جدی مرتبط با این مجموعه داده وجود دارد. یکی از مسائل اصلی حفظ حریم خصوصی و اطلاعات است. علاوه بر این، همانطور که در بخش قبل توضیح داده شد، استفاده از هوش مصنوعی ممکن است چالش‌هایی را برای حقوق بشر مانند استقلال و کرامت به همراه داشته باشد. خودمختاری به ظرفیت تفکر، تصمیم گیری و عمل مستقل، آزادانه و بدون تأثیر دیگران اشاره دارد [۷۰]. کرامت انسانی که یکی از حقوق اصلی بشر است، مربوط به حق احترام و رفتار اخلاقی انسان است [۷۱]. حفاظت از کرامت در زمینه هوش مصنوعی بسیار مهم است. کرامت انسانی باید یکی از مفاهیم اساسی برای محافظت از انسان در برابر آسیب باشد و هنگام توسعه فناوری‌های هوش مصنوعی باید مورد احترام قرار گیرد. به عنوان مثال، یک سیستم تسلیحاتی خودمختار مرگبار [۷۲] ممکن است اصل کرامت انسانی را نقض کند.

۲-۲-۲. مسائل اخلاقی در سطح اجتماعی

هنگام در نظر گرفتن مسائل اخلاقی هوش مصنوعی در سطح اجتماعی، ما عمدتاً بر پیامدها و تأثیرات گسترده‌ای که هوش مصنوعی برای جامعه و رفاه جوامع و ملل در سراسر جهان به ارمغان می‌آورد تمرکز می‌کنیم. در طبقه بندی مسائل اخلاقی در سطح اجتماعی، ما در مورد انصاف و عدالت، مسئولیت و پاسخگویی، شفافیت، نظارت و اطلاعات، قابلیت کنترل هوش مصنوعی، دموکراسی و حقوق مدنی، جایگزینی شغل و روابط انسانی بحث می‌کنیم.

وجود تعصب و تبعیض در هوش مصنوعی چالش‌هایی را برای انصاف و عدالت ایجاد کرده است. سوگیری‌ها و تبعیض‌های تعبیه شده در هوش مصنوعی ممکن است شکاف‌های اجتماعی را افزایش داده و به گروه‌های اجتماعی خاصی آسیب برساند [۷۰]. به عنوان مثال، در سیستم عدالت کیفری ایالات متحده، الگوریتم‌های هوش مصنوعی که برای ارزیابی خطر ارتکاب جرم استفاده می‌شوند، مورد توجه قرار گرفته‌اند که سوگیری نژادی را نشان می‌دهند [۷۳]. مسئولیت به معنای مسئول چیزی بودن یا در چیزی مسئول بودن است. بر اساس این مفهوم، مسئولیت پذیری در اصل یعنی کسی که از نظر حقوقی یا سیاسی مسئول خسارت است، باید نوعی توجیه یا جبران خسارت ارائه کند. با مسئولیت ارائه راه حل‌های حقوقی منعکس می‌شود [۷۰]. بنابراین، مکانیسم‌هایی باید ایجاد شود تا از مسئولیت و پاسخگویی سیستم‌های هوش مصنوعی و نتایج آن‌ها قبل و بعد از اجرای آنها اطمینان حاصل شود. به دلیل ماهیت

جعبه سیاه الگوریتم‌های هوش مصنوعی، عدم شفافیت به یکی از موضوعاتی تبدیل شده است که به طور گسترده مورد بحث قرار گرفته است. شفافیت، یعنی درک نحوه عملکرد سیستم‌های هوش مصنوعی، برای پاسخگویی نیز بسیار مهم است. در عصری که ما زندگی می‌کنیم که به آن عصر دیجیتال و هوشمند نیز می‌گویند، نظارت و اطلاعات [۷۴]، از دغدغه‌های رایج هستند. داده‌ها از زندگی روزمره کاربران از طریق دستگاه‌های هوشمند جمع آوری می‌شود و ما تحت نظارت انبوه، زندگی می‌کنیم. از آنجایی که قدرت هوش مصنوعی به سرعت افزایش یافته است، توسعه سیستم‌های هوش مصنوعی باید دارای پادمان‌هایی باشد تا از کنترل‌پذیری سیستم‌های هوش مصنوعی توسط انسان اطمینان حاصل شود. سایر موضوعاتی که قبلاً مورد بحث قرار گرفت، از جمله دموکراسی و حقوق مدنی، جایگزینی شغل و روابط انسانی نیز در این دسته قرار می‌گیرند.

۲-۲-۳. مسائل اخلاقی در سطح محیطی

مسائل اخلاقی هوش مصنوعی در سطح زیست محیطی بر تأثیرات هوش مصنوعی بر محیط زیست و سیاره‌ی زمین تمرکز دارد. هوش مصنوعی می‌تواند راحتی زیادی را برای زندگی ما به ارمغان بیاورد و می‌تواند به ما در مقابله با برخی چالش‌ها کمک کند، اما برای سیاره زمین نیز هزینه دارد. کاربرد گسترده هوش مصنوعی اغلب مستلزم استقرار تعداد زیادی دستگاه پایانه سخت افزاری از جمله تراشه ها، حسگرها، دستگاه‌های ذخیره سازی و غیره است. تولید این سخت افزارها منابع طبیعی زیادی به خصوص برخی عناصر کمیاب را مصرف می‌کند. علاوه بر این، در پایان چرخه عمر، این سخت افزارها معمولاً دور ریخته می‌شوند که باعث آلودگی جدی زیست محیطی خواهند شد. جنبه مهم دیگر این است که سیستم‌های هوش مصنوعی معمولاً به قدرت محاسباتی قابل توجهی نیاز دارند که با مصرف انرژی بالا همراه است. علاوه بر این، از دیدگاه بلندمدت و جهانی، توسعه هوش مصنوعی باید پایدار باشد، یعنی فناوری هوش مصنوعی باید اهداف توسعه انسانی را برآورده کند و در عین حال توانایی سیستم‌های طبیعی را برای ارائه منابع طبیعی و خدمات اکوسیستمی که اقتصاد و جامعه به آن وابسته است، حفظ کند [۲]. به طور خلاصه، مصرف منابع طبیعی، آلودگی محیط زیست، هزینه‌های مصرف انرژی و پایداری درگیر در توسعه هوش مصنوعی مسائل و نگرانی‌های اصلی در سطح زیست محیطی هستند.

دسته بندی پیشنهادی ما مسائل اخلاقی را از سه سطح اصلی، یعنی تأثیر هوش مصنوعی بر فرد، جامعه و محیط، روشن می‌کند. مهم نیست که هوش مصنوعی در چه زمینه یا بخشی استفاده می‌شود، می‌توانیم مسائل اخلاقی مربوطه را از این سه سطح در نظر بگیریم. بدیهی است که این روش طبقه بندی ساده و واضح است و به طور جامع مسائل اخلاقی هوش مصنوعی را پوشش می‌دهد.

۳-۲. مسائل اخلاقی کلیدی مرتبط با هر مرحله از چرخه حیات سیستم هوش مصنوعی

پس از بررسی مسائل اخلاقی و خطرات مورد بحث در ادبیات، ما در مورد مسائل اخلاقی مرتبط با مراحل مختلف چرخه حیات یک سیستم هوش مصنوعی بحث می‌کنیم. اگر بدانیم مشکلات اخلاقی موجود در کدام مراحل یا مراحل چرخه حیات سیستم هوش مصنوعی ایجاد می‌شوند یا مطرح می‌شوند، در رفع این مشکلات بسیار مفید خواهد بود. این موضوع انگیزه‌ای برای بحث در مورد مسائل اخلاقی بالقوه در هر مرحله از چرخه حیات یک سیستم هوش مصنوعی است.

چرخه عمر کلی یا فرایند توسعه یک سیستم هوش مصنوعی مبتنی بر ML [۷۵] یا محصول [۷۶]، اغلب شامل مراحل زیر است: تجزیه و تحلیل کسب و کار، مهندسی داده، مدل سازی ML، استقرار مدل، و بهره برداری و نظارت. معمولاً چرخه عمر محصولات هوش مصنوعی از تجزیه و تحلیل کسب و کار شروع می‌شود که عمدتاً شامل شناسایی و درک مشکل تجاری‌ای که باید حل شود و معیارهای تجاری (یا معیارهای موفقیت) است. این معیارها باید شامل معیارهای عملکرد مدل و همچنین شاخص‌های عملکرد کلیدی کسب و کار باشد تا با استفاده از مدل‌های هوش مصنوعی بهبود یابد. گام بعدی در مورد مهندسی داده است که به جمع آوری داده ها، برچسب گذاری داده ها، پاکسازی داده ها، ساختار داده ها، مهندسی ویژگی‌ها و سایر عملیات مربوط به داده‌ها می‌پردازد. پس از این، ادامه‌ی فرایند، وارد مرحله‌ی به اصطلاح مدل سازی ML می‌شود. این مرحله به طور کلی شامل فرایند تکراری طراحی یا انتخاب الگوریتم، آموزش مدل و ارزیابی مدل است. اگر ساخت مدل رضایت‌بخش باشد، ادامه‌ی فرایند، به مرحله استقرار مدل می‌رود که مدل ML را در دسترس سایر سیستم‌های درون سازمان یا وب قرار می‌دهد تا مدل بتواند داده‌ها را دریافت کند و پاسخ آنها را برگرداند. مرحله عملیات و نظارت شامل عملیات سیستم هوش مصنوعی و ارزیابی مداوم عملکرد و تأثیرات آن است. این مرحله مشکلات را شناسایی کرده و سیستم هوش مصنوعی را با بازگشت به مراحل دیگر یا در صورت لزوم کنار گذاشتن سیستم هوش مصنوعی از تولید، یا تنظیم می‌کند یا تکامل می‌دهد.

ما سعی می‌کنیم نقشه‌ای ایجاد کنیم که مسائل اخلاقی را با مراحل چرخه‌ی عمر هوش مصنوعی مرتبط می‌کند، جایی که این ارتباط به این معنی است که موضوع اخلاقی در مرحله‌ی خاصی از چرخه عمر هوش مصنوعی بیشتر رخ می‌دهد، یا اغلب به دلایلی در این مرحله ایجاد می‌شود. این نقشه برداری در جدول ۲ ارائه شده است، جایی که چندین مشکل اخلاقی حیاتی با پنج مرحله چرخه حیات هوش مصنوعی مرتبط است. این نقشه برداری برای پرداختن به مشکل اخلاقی به روشی فعال در طول فرایند طراحی یک سیستم هوش مصنوعی مفید خواهد بود.

جدول ۲: مراحل ارتباط مسائل اخلاقی با چرخه عمر هوش مصنوعی

مرحله چرخه حیات هوش مصنوعی	ملاحظات اخلاقی در طول مرحله
تحلیل کسب‌وکار	شفافیت، عدالت (آیا معماری محصول هوش مصنوعی طراحی‌شده شامل متغیرها، ویژگی‌ها یا فرایندهایی است که غیرمنطقی، غیراخلاقی یا غیرقابل توجیه هستند؟)، مسئولیت‌پذیری و پاسخگویی، دموکراسی و حقوق مدنی، پایداری
مهندسی داده	حریم خصوصی (چگونه امنیت داده‌ها و حفظ اطلاعات خصوصی و حساس موجود در مجموعه داده‌ها تضمین شود؟)، شفافیت (چگونه می‌توان روش‌های جمع‌آوری داده‌ها را برای مصرف‌کنندگان شفاف کرد؟)، عدالت (آیا داده‌ها به درستی نماینده، مرتبط، دقیق و قابل تعمیم هستند؟)، دموکراسی و حقوق مدنی (چگونه به کاربران نهایی امکان کنترل استفاده از داده‌هایشان را می‌دهید؟)
مدل‌سازی یادگیری ماشین	شفافیت (آیا فرایند تصمیم‌گیری یا استنتاج مدل قابل درک است؟)، ایمنی (دقت، قابلیت اطمینان، امنیت و استحکام مدل)، عدالت (آیا خروجی‌های مدل نتایج متفاوتی برای گروه‌های مختلف افراد نشان می‌دهند؟)
استقرار مدل	حریم خصوصی (اطمینان حاصل شود که اطلاعات خصوصی از طریق مدل مستقر قابل شناسایی مجدد نیستند)، ایمنی (چگونه می‌توان از ایمنی مدل مستقر در برابر تغییرات مخرب و حملات محافظت کرد؟)
عملیات و نظارت	حریم خصوصی (حریم خصوصی باید در طول فرایند عملیات و نظارت تضمین شود)، عدالت (آیا محصول هوش مصنوعی تأثیرات تبعیض‌آمیز یا ناعادلانه بر افرادی که تحت تأثیر قرار می‌گیرند دارد؟)، دموکراسی و حقوق مدنی (حقوق مدنی یا حقوق کاربران را نقض نکنید)

۵. رهنمودها و اصول اخلاقی برای هوش مصنوعی

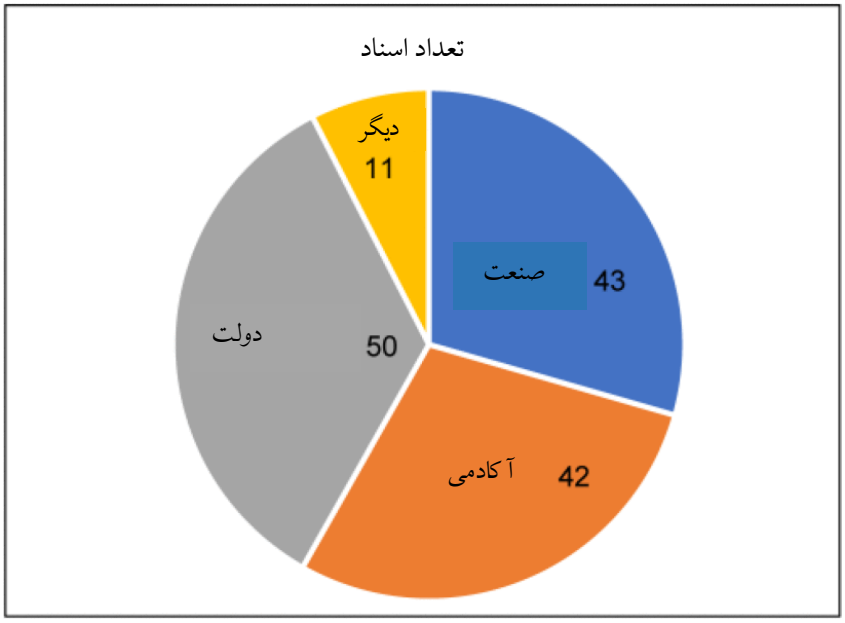
از آنجایی که مسائل اخلاقی هوش مصنوعی بیشتر و بیشتر مورد توجه و بحث‌های بخش‌های مختلف جامعه قرار گرفته است، بسیاری از سازمان‌ها (از جمله دانشگاه، صنعت و دولت) شروع به بحث و جست‌وجوی چارچوب‌ها، دستورالعمل‌ها و اصول ممکن برای حل مسائل اخلاقی هوش مصنوعی کرده‌اند [۷۸]. این اصول و دستورالعمل ها، دستورالعمل‌های مفیدی را برای تمرین هوش مصنوعی اخلاقی ارائه می‌دهند. این بخش به ارائه‌ی یک چشم انداز جهانی به روز از دستورالعمل‌ها و اصول اخلاق هوش مصنوعی اختصاص دارد که از طریق بررسی ۱۴۶ گزارش دستورالعمل و توصیه مربوط به اخلاق هوش مصنوعی منتشر شده توسط شرکت ها، سازمان‌ها و دولت‌ها از سال ۲۰۱۵ در سراسر جهان به دست می‌آید. این اصول و دستورالعمل ها، راهنمایی‌های سطح بالایی را برای برنامه ریزی، توسعه، تولید، و استفاده از هوش مصنوعی و نیز دستورالعمل‌هایی برای پرداختن به مسائل اخلاقی هوش مصنوعی ارائه می‌دهند.

۳-۱. رهنمودهایی برای اخلاق هوش مصنوعی

یک بررسی و تجزیه و تحلیل عالی از اصول و دستورالعمل‌های فعلی در مورد هوش مصنوعی اخلاقی در سال ۲۰۱۹ توسط Jobin و همکاران ارائه شده است [۱۲]، که بررسی ۸۴ دستورالعمل اخلاقی منتشر شده توسط سازمان‌های ملی یا بین المللی از کشورهای مختلف را انجام داده‌اند. جوبین و همکاران [۱۲] توافق گسترده و قوی در مورد پنج اصل کلیدی، یعنی شفافیت، عدالت و انصاف، عدم سوء استفاده، مسئولیت و حریم خصوصی در میان بسیاری از آنها یافتند. با این حال، بسیاری از دستورالعمل‌ها و توصیه‌های جدید برای اخلاق هوش مصنوعی در دو سال گذشته منتشر شده‌اند که باعث شده است مقاله Jobin منسوخ شود زیرا بسیاری از اسناد مهم گنجانده نشده‌اند. به عنوان مثال، در ۲۴ نوامبر ۲۰۲۱، یونسکو (سازمان آموزشی، علمی و فرهنگی ملل متحد) توصیه‌ای را در مورد اخلاق هوش مصنوعی تصویب کرد که اولین توافق جهانی در مورد اخلاق هوش مصنوعی است [۷۹]. برای به‌روزرسانی و غنی‌سازی تحقیقات در مورد دستورالعمل‌ها و اصول هوش مصنوعی اخلاقی، بر اساس جدول دستورالعمل‌های اخلاقی هوش مصنوعی که در مقاله Jobin [۱۲] ارائه شده است (فقط شامل ۸۴ سند)، ما بسیاری از دستورالعمل‌های اخلاقی هوش مصنوعی جدید منتشر شده را جمع‌آوری کرده‌ایم که در مقاله بررسی جوبین گنجانده نشده‌اند. در نهایت، در مجموع ۱۴۶ دستورالعمل اخلاقی AI جمع آوری شده است. فهرستی از تمام دستورالعمل‌ها یا اسناد جمع‌آوری‌شده در شکل ۵ از مواد تکمیلی آورده شده است. تعداد دستورالعمل‌های صادر شده در هر سال از سال ۲۰۱۵ تا ۲۰۲۱ در جدول ۳ شمارش و فهرست شده است. واضح است که اکثر دستورالعمل‌ها در پنج سال گذشته منتشر شده اند، یعنی از سال ۲۰۱۶ تا ۲۰۲۰. تعداد راهنماهای منتشر شده در سال ۲۰۱۸ با ۵۳ مورد، که ۳۶٫۳ درصد از کل تعداد را به خود اختصاص داده است، بیشترین تعداد راهنما را به خود اختصاص داده است. علاوه بر این، تعداد

دستورالعمل‌های AI صادر شده توسط هر کشور در جدول ۴ فهرست شده است. علاوه بر این، درصد دستورالعمل‌های منتشر شده توسط انواع مختلف صادرکنندگان (از جمله دولت، صنعت، دانشگاه و سایر سازمان ها) در شکل ۳ نشان داده شده است. از شکل ۳ می‌توان دریافت که دولت‌ها، شرکت‌ها و دانشگاه‌ها همه، نگرانی‌های شدیدی در مورد اخلاق هوش مصنوعی نشان داده‌اند.

شکل ۳: درصد دستورالعمل‌های منتشرشده براساس صادرکننده



جدول ۳: دستورالعمل‌های صادرشده در هر سال

سال	تعداد اسناد
2015	2
2016	7
2017	25
2018	53
2019	31
2020	24
2021	4

جدول ۴: تمام دستورالعمل‌های صادرشده به تفکیک کشورها

کشور	تعداد
Australia	3
Canada	4
China	5
Denmark	4

15	EU
4	Finland
3	France
7	Germany
1	Iceland
1	India
12	International
3	Ireland
6	Japan
3	N/A
4	Netherlands
1	Norway
1	Russia
3	Singapore
3	South Korea
2	Spain
1	Sweden
1	Switzerland
1	Turkey
2	UAE
16	UK
39	USA
1	Vatican

جدول ۵: تمام دستورالعمل‌های صادرشده

نظریه اخلاقی	توضیحات	تمرکز در تأمل	معیار تصمیم‌گیری	استدلال عملی
اخلاق فضیلت	یک عمل درست است اگر کاری باشد که یک فرد با فضیلت در آن موقعیت انجام دهد.	انگیزه‌ها (آیا عمل با فضیلت انگیزه یافته است؟)	فضایل	تجسم فضایل/ویژگی‌های انسانی
اخلاق وظیفه‌گرا	یک عمل درست است اگر با یک قاعده یا اصل اخلاقی مطابقت داشته باشد.	عمل (آیا عمل با یک الزام سازگار است؟)	وظایف/قوانین	پیروی از قوانین

اخلاق پیامدگرا	یک عمل درست است اگر بهترین پیامدها را ترویج کند، یعنی خوشبختی را به حداکثر برساند.	پیامدها (نتیجه عمل چیست؟)	رفاه نسبی	بهینه‌سازی سودمندی یا خوشبختی

۳-۲. اصول اخلاق هوش مصنوعی

اصول اخلاقی که در ۱۴۶ دستورالعمل جمع آوری شده، در جدول ۱ از مواد تکمیلی ذکر شده است. طبق جدول، یک همگرایی آشکار حول پنج اصل اخلاقی مهم وجود دارد: شفافیت، انصاف و عدالت، مسئولیت، عدم سوء استفاده و حریم خصوصی. ۱۱ اصل اخلاقی مشخص شده در دستورالعمل‌های هوش مصنوعی موجود در ادامه شرح و توضیح داده شده است.

۱) شفافیت: شفافیت یکی از اصولی است که به طور گسترده در بحث اخلاق هوش مصنوعی مورد بحث قرار گرفته است. شفافیت هوش مصنوعی عمدتاً شامل شفافیت خود فناوری هوش مصنوعی و شفافیت توسعه و پذیرش هوش مصنوعی است [۱۳]. از یک طرف، شفافیت هوش مصنوعی شامل تفسیرپذیری یک سیستم هوش مصنوعی معین است، یعنی توانایی دانستن اینکه چگونه و چرا یک مدل به روشی که در یک زمینه خاص در پیش می‌گیرد و بنابراین درک منطقی که در پس تصمیم یا رفتار آن وجود دارد. این جنبه از شفافیت معمولاً به عنوان استعاره "باز کردن جعبه سیاه هوش مصنوعی" ذکر می‌شود. این مبحث به تفسیر پذیری، توضیح پذیری یا قابل فهم بودن مربوط می‌شود. از سوی دیگر، شفافیت هوش مصنوعی شامل توجیه پذیری یا منطقی بودن فرایند طراحی و اجرای سیستم هوش مصنوعی و نتیجه آن است. به عبارت دیگر فرایند طراحی و پیاده سازی سیستم هوش مصنوعی و تصمیم یا رفتار آن باید قابل توجیه و قابل مشاهده باشد.

۲) انصاف و عدالت: اصل عدالت و انصاف بیان می‌کند که توسعه، استقرار و استفاده از هوش مصنوعی باید عادلانه و منصفانه باشد تا سیستم هوش مصنوعی نباید منجر به تبعیض یا سوگیری علیه افراد، جوامع یا گروه‌ها شود [۸۰]. تبعیض و نتایج ناعادلانه ناشی از الگوریتم‌های هوش مصنوعی به موضوع داغ رسانه‌ها و دانشگاه‌ها تبدیل شده است. در نتیجه اصل انصاف و عدالت در چند سال اخیر توجه قابل توجهی را به خود جلب کرده است.

۳) مسئولیت و پاسخگویی: اصل مسئولیت و پاسخگویی ایجاب می‌کند که هوش مصنوعی باید قابل ممیزی باشد؛ یعنی طراحان، توسعه دهندگان، مالکان و اپراتورهای هوش مصنوعی در قبال رفتارها یا تصمیمات یک سیستم هوش مصنوعی مسئول و پاسخگو هستند و بنابراین برای مضرات یا پیامدهای بدی که ممکن است ایجاد کند مسئول تلقی می‌شوند [۵۱]. طراحان، سازندگان و کاربران سیستم‌های هوش مصنوعی ذینفعان پیامدهای اخلاقی استفاده، سوء

استفاده و رفتارشان هستند و مسئولیت و فرصت شکل‌دهی به این مفاهیم را دارند. این موضوع مستلزم آن است که مکانیسم‌های مناسبی برای اطمینان از مسئولیت و پاسخگویی سیستم‌های هوش مصنوعی و نتایج آن‌ها، قبل و بعد از توسعه، استقرار و استفاده از آن‌ها ایجاد شود.

۴) NonMeficence (بدخواهی نکردن): اساساً به معنای عدم آسیب رساندن یا اجتناب از تحمیل خطرات آسیب به دیگران است [۸۱]،[۸۲]. بنابراین، اصل عدم سوء استفاده از هوش مصنوعی به طور کلی به این اشاره دارد که سیستم‌های هوش مصنوعی نباید باعث آسیب یا تشدید آسیب به انسان یا تأثیر نامطلوب بر انسان‌ها شوند. این امر مستلزم حفظ کرامت انسانی و نیز تمامیت روحی و جسمی است. اصل عدم سوء استفاده مستلزم آن است که سیستم‌های هوش مصنوعی و محیط‌هایی که در آن کار می‌کنند باید ایمن و مطمئن باشند تا برای استفاده مخرب باز نباشند. با توجه به برخی از تصادفات مرگبار که از خودروهای خودران و ربات‌ها رخ می‌دهد، اجتناب از آسیب به انسان یکی از بزرگترین نگرانی‌ها در اخلاق هوش مصنوعی است. از این رو، بیشتر دستورالعمل‌های اخلاقی تاکید زیادی بر اطمینان از عدم آسیب به انسان از طریق ایمنی و امنیت هوش مصنوعی دارند.

۵) حریم خصوصی: هدف اصل حریم خصوصی، اطمینان از احترام به حریم خصوصی و حفاظت از داده‌ها هنگام استفاده از سیستم‌های هوش مصنوعی است. سیستم‌های هوش مصنوعی باید حقوق حریم خصوصی و حفاظت از داده‌ها را حفظ کرده و به آنها احترام بگذارند و همچنین امنیت داده‌ها را حفظ کنند. این موضوع شامل ارائه حاکمیت و مدیریت داده موثر برای تمام داده‌های مورد استفاده و تولید شده توسط سیستم هوش مصنوعی در کل چرخه عمر آن سیستم است [۸۳]. به طور خاص، جمع آوری، استفاده و ذخیره داده‌ها باید با قوانین و مقررات مربوط به حریم خصوصی و حفاظت از داده‌ها مطابقت داشته باشد. داده‌ها و الگوریتم‌ها باید در برابر سرقت محافظت شوند. هنگامی که نشت اطلاعات رخ می‌دهد، کارفرمایان یا ارائه دهندگان هوش مصنوعی باید در اسرع وقت به کارمندان، مشتریان، شرکا و سایر افراد مرتبط اطلاع دهند تا ضرر یا تأثیر ناشی از نشت به حداقل برسد.

۶) سودمندی: اصل سودمندی بیان می‌کند که هوش مصنوعی باید به مردم کمک کند و به نفع بشریت باشد [۸۲]. این اصل نشان می‌دهد که فناوری هوش مصنوعی باید برای به ارمغان آوردن نتایج و تأثیرات مفید برای افراد، جامعه و محیط استفاده شود [۸۴]. هنگام توسعه یک سیستم هوش مصنوعی، اهداف آن باید به وضوح تعریف و توجیه شوند. استفاده از فناوری هوش مصنوعی برای کمک به رسیدگی به نگرانی‌های جهانی باید تشویق شود، مانند استفاده از هوش مصنوعی برای کمک به ما در مدیریت امنیت غذایی، آلودگی و مسری‌هایی مانند ایدز و کووید ۱۹.

۷) آزادی و خودمختاری: آزادی و خودمختاری که عموماً به توانایی فرد در تصمیم گیری با توجه به اهداف و خواسته‌های خود اشاره دارد، ارزش اصلی شهروندان در جوامع دموکراتیک است. بنابراین، مهم است که استفاده از هوش مصنوعی به آزادی و خودمختاری ما آسیب نرساند یا آن را

محدود نکند. هنگامی که ما از عوامل هوش مصنوعی استفاده می‌کنیم، مایلیم بخشی از اختیارات تصمیم گیری خود را به ماشین‌های هوش مصنوعی واگذار کنیم. بنابراین، حفظ اصل آزادی و خودمختاری در زمینه هوش مصنوعی به معنای ایجاد تعادل بین قدرت تصمیم‌گیری که برای خود حفظ می‌کنیم و قدرتی که به هوش مصنوعی واگذار می‌کنیم [۸۴] است.

۸) همبستگی: اصل همبستگی مستلزم این است که توسعه و کاربرد یک سیستم هوش مصنوعی باید با حفظ مرزهای همبستگی بین مردم و نسل‌ها سازگار باشد. به عبارت دیگر، هوش مصنوعی باید امنیت اجتماعی و انسجام را ارتقا دهد و پیوندها و روابط اجتماعی را به خطر اندازد [۱۳].

۹) پایداری: با توجه به تغییرات اقلیمی و آسیب‌های زیست محیطی مداوم، اهمیت پایداری بیش از پیش مورد توجه قرار گرفته است. مانند سایر زمینه‌ها و رشته ها، هوش مصنوعی تحت تأثیر قرار گرفته و باید در دستور کار توسعه پایدار گنجانده شود. اصل پایداری بیانگر این است که تولید، مدیریت و اجرای هوش مصنوعی باید پایدار باشد و از آسیب‌های زیست محیطی جلوگیری کند. به عبارت دیگر، فناوری هوش مصنوعی باید الزامات تضمین تداوم رفاه بشر و حفظ محیطی خوب برای نسل‌های آینده را برآورده کند [۸۵]. سیستم‌های هوش مصنوعی قول می‌دهند که به رفع برخی از مهم‌ترین نگرانی‌های اجتماعی کمک کنند، اما باید اطمینان حاصل شود که این امر به سازگارترین شکل ممکن به محیط‌زیست اتفاق می‌افتد.

۱۰) اعتماد: قابل اعتماد بودن، پیش نیاز افراد و جوامع برای پذیرش هوش مصنوعی است، زیرا اعتماد یک اصل اساسی برای تعاملات بین فردی و عملکرد اجتماعی است. اعتماد در توسعه، استقرار و استفاده از سیستم‌های هوش مصنوعی نه تنها به ویژگی‌های ذاتی فناوری مربوط می‌شود، بلکه به کیفیت سیستم اجتماعی-فنی مربوط به برنامه‌های کاربردی هوش مصنوعی نیز مرتبط است. بنابراین، حرکت به سمت هوش مصنوعی قابل اعتماد نه تنها به قابلیت اعتماد خود سیستم هوش مصنوعی مربوط می‌شود، بلکه نیازمند رویکردی جامع و سیستماتیک است که قابلیت اطمینان همه شرکت‌کنندگان و فرایندهایی را که کل چرخه حیات سیستم را تشکیل می‌دهند، پوشش می‌دهد [۸۶].

۱۱) کرامت: کرامت انسانی شامل این باور است که همه مردم دارای یک ارزش ذاتی هستند که صرفاً به انسانیت آنها گره خورده است، یعنی هیچ ربطی به طبقه، نژاد، جنسیت، مذهب، توانایی‌ها یا هر عامل دیگری غیر از آنها ندارد. انسان بودن و این ارزش ذاتی هرگز نباید توسط افراد دیگر یا توسط فناوری‌هایی مانند هوش مصنوعی کاهش یابد، به خطر بیفتد، یا سرکوب شود. این مهم است که هوش مصنوعی نباید به حیثیت کاربران یا سایر اعضای جامعه آسیب برساند. در نتیجه، احترام به کرامت انسانی یک اصل مهم است که باید در اخلاق هوش مصنوعی مورد توجه قرار گیرد. بنابراین سیستم هوش مصنوعی باید به گونه‌ای توسعه یابد که به تمامیت جسمی و روانی افراد، احساس هویت فردی و فرهنگی و ارضای نیازهای اساسی آنها احترام بگذارد، حمایت کند و از آنها محافظت کند [۱۳].

۴. رویکردهایی برای پرداختن به مسائل اخلاقی در هوش مصنوعی

این بخش رویکردهای مربوط به رسیدگی یا کاهش مسائل اخلاقی هوش مصنوعی را بررسی می‌کند. از آنجایی که اخلاق هوش مصنوعی یک زمینه گسترده و چند رشته‌ای است، ما سعی می‌کنیم به جای تمرکز صرف بر روی رویکردهای فناورانه‌ای که مورد علاقه جامعه‌ی AI/ML هستند، یک مرور کلی از رویکردهای موجود و بالقوه برای پرداختن به مسائل اخلاقی هوش مصنوعی، از جمله رویکردهای اخلاقی، تکنولوژیکی و قانونی ارائه کنیم. این بررسی از رویکردهای چند رشته‌ای برای پرداختن به مشکلات اخلاقی هوش مصنوعی نه تنها خلاصه‌ای آموزنده در مورد رویکردهای هوش مصنوعی اخلاقی ارائه می‌کند، بلکه به محققان جامعه هوش مصنوعی پیشنهاد می‌کند تا به جای تکیه بر رویکردهای فناوری، راه‌حل‌هایی برای مسائل اخلاقی هوش مصنوعی از دیدگاه‌های مختلف جستجو کنند. از آنجایی که مسائل اخلاقی هوش مصنوعی با مشکلات چند رشته‌ای در هم آمیخته است، ممکن است تنها از طریق همکاری روش‌های مختلف بتوان این مشکلات را به طور موثر حل کرد.

رویکردهای اخلاقی به توسعه‌ی سیستم‌ها یا عوامل هوش مصنوعی اخلاقی اختصاص دارد که قادر به استدلال و عمل اخلاقی بر اساس نظریه‌های اخلاقی [۸۷] با پیاده‌سازی یا تعبیه اخلاق در هوش مصنوعی هستند. رویکردهای فناوری برای توسعه فناوری‌های جدید (به ویژه فناوری‌های ML) برای حذف یا کاهش کاستی‌های هوش مصنوعی فعلی طراحی شده‌اند. به عنوان مثال، تحقیق در مورد ML قابل توضیح قصد دارد رویکردهای جدیدی را برای توضیح دلیل و مکانیسم کار الگوریتم‌های ML ایجاد کند. ML منصفانه تکنیک‌هایی را مطالعه می‌کند که ML را قادر می‌سازد تا تصمیمات یا پیش‌بینی‌های منصفانه بگیرد، یعنی تعصب یا تبعیض ML را کاهش دهد. رویکردهای حقوقی در نظر دارند تحقیقات، استقرار، کاربرد و سایر جنبه‌های هوش مصنوعی را از طریق قانون‌گذاری و مقررات، با هدف اجتناب از موضوعات اخلاقی مورد بحث قبلی، تنظیم یا کنترل کنند.

۴-۱. رویکردهای اخلاقی: اجرای اخلاق در هوش مصنوعی

طراحی سیستم‌های هوش مصنوعی اخلاقی، که می‌توانند استدلال کنند و اخلاقی عمل کنند، نیاز به درک درستی از رفتار اخلاقی دارد. این مسئله شامل قضاوت درست و نادرست، خوب و بد، و همچنین مسائل مربوط به عدالت، انصاف، فضیلت و سایر اصول اخلاقی است. بنابراین، نظریه‌های اخلاقی، که با مفاهیم رفتار درست و نادرست مرتبط هستند، ارتباط نزدیکی با اخلاق هوش مصنوعی دارند. این بخش به رویکردهای پیاده سازی اخلاق در سیستم‌های هوش مصنوعی بر اساس تئوری‌های اخلاقی موجود اختصاص دارد. ابتدا، نظریه‌های اخلاقی، به ویژه اخلاق هنجاری که با اخلاق هوش مصنوعی مرتبط است، بررسی می‌شود. سپس، سه نوع رویکرد اصلی برای طراحی سیستم‌های هوش مصنوعی اخلاقی خلاصه می‌شود.

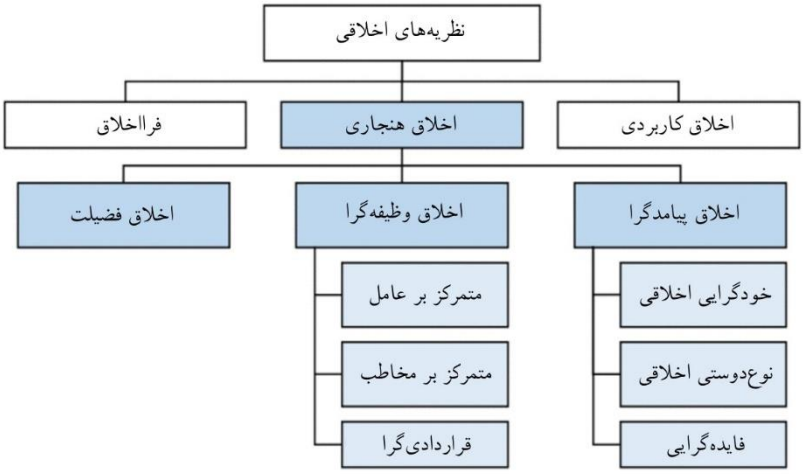
۶. نظریه‌های اخلاقی

حوزه اخلاق (همچنین به عنوان فلسفه اخلاق شناخته می‌شود) به نظام‌مند کردن، دفاع و توصیه مفاهیم رفتار درست و نادرست می‌پردازد. اخلاق بر قضاوت و تعیین اینکه کدام عمل در شرایط معین خوب یا اخلاقی است تمرکز دارد [۸۸]. مطالعه فلسفی اخلاق معمولاً شامل سه حوزه‌ی موضوعی اصلی است: فرا اخلاق، اخلاق هنجاری و اخلاق کاربردی [۸۹]. شاخه‌های نظریه‌های اخلاقی در شکل ۴ نشان داده شده است.

فرا اخلاق، ماهیت، دامنه و معنای اصول اخلاقی یا قضاوت اخلاقی را بررسی می‌کند. تلاش برای درک معنا و منشأ اصطلاحات اخلاقی، نقش عقل در قضاوت‌های اخلاقی و مسائل مربوط به حقایق جهانی یا ارزش‌های انسانی است [۹۰].

اخلاق هنجاری به دنبال رسیدن به معیارها و قوانین اخلاقی است که رفتار درست و نادرست را تنظیم می‌کند. به این معنا که هدف آن ایجاد مجموعه‌ای از قواعد است که بر رفتار انسان و یا اینکه چگونه چیزها باید باشند، با بررسی اینکه چگونه انسان‌ها برای چیزها ارزش قائل هستند و تشخیص درست از غلط یا خوب از بد قضاوت می‌کنند.

اخلاق کاربردی، اخلاق حوزه‌های کاربردی خاص است که شامل تحلیل مسائل اخلاقی خاص و بحث‌برانگیز مانند سقط جنین، مجازات اعدام، حقوق حیوانات، نگرانی‌های زیست‌محیطی، جنگ هسته‌ای و غیره است.



شکل ۴: شاخه‌های نظریه‌های اخلاقی

الف) اخلاق هنجاری

اخلاق هنجاری مخصوصاً به درک و به کارگیری اصول اخلاقی در طراحی، استقرار و استفاده از سیستم‌های هوش مصنوعی [۸۹] مربوط می‌شود، زیرا یک رشته‌ی فلسفی عملی-هنجاری است که به نحوه‌ی رفتار انسان‌ها یا عوامل با دیگران مربوط می‌شود. سه شاخه‌ی اخلاقی هنجاری، یعنی فضیلت، اخلاق دین شناختی و پیامدگرایی در زیر ارائه و خلاصه می‌شود.

اخلاق فضیلت: اخلاق فضیلت بر فضایل یا شخصیت اخلاقی تأکید می‌کند و بر اهمیت پرورش عادات خوبِ مَنِش مانند خیرخواهی تأکید می‌کند [۹۲]. از این رو، اخلاق فضیلت، بر شخصیت ذاتی عامل، تمرکز دارد تا پیامدهای اعمالی که توسط عامل انجام می‌شود. در اخلاق فضیلتی، اگر فاعل بر اساس برخی ارزش‌های اخلاقی عمل کرده و بیندیشد، عمل آن فاعل را از نظر اخلاقی خوب تعریف می‌کند [۹۳]. به عبارت دیگر، بر اساس نظریات فضیلت، فاعل در صورتی اخلاقی است که برخی از فضایل اخلاقی را از طریق اَعمال خود بروز دهد [۹۴]، [۹۵].

اخلاق دئونتولوژیک: نظریه‌های دین‌شناختی، که گاهی اوقات نظریه‌های وظیفه نامیده می‌شوند، با استفاده از قواعد اخلاقی خاصی که به‌عنوان اصول بنیادین تعهد عمل می‌کنند، درباره اخلاقی بودن یک عمل قضاوت می‌کنند. دئونتولوژی نوعی نظریه اخلاق هنجاری است که در مورد آن، انتخاب‌ها یا اعمالی از نظر اخلاقی میتوانند مورد نیاز، ممنوع یا مجاز باشند. به عبارت دیگر، deontology یک نظریه اخلاقی است که تصمیمات ما را در مورد آنچه که باید انجام دهیم، هدایت و ارزیابی می‌کند [۹۶]. دئونتولوژیست‌ها یک عمل اخلاقی خوب را به عنوان عملی تعریف می‌کنند که به برخی تعهدات پایبند باشد، که ممکن است قوانین یا وظایف اخلاقی، مقررات و هنجارهای قابل اجرا باشد.

سه مکتب اصلی از نظریه‌های deontological وجود دارد، یعنی نظریه‌های عامل محور، بیمار محور (که قربانی محور نیز نامیده می‌شود) و نظریه‌ی قراردادی. نظریه‌ی عامل محور، عامل را در مرکز قرار می‌دهد و بر وظایف عامل نِسبی تمرکز می‌کند. تئوری بیمارمحور، همانطور که از دیونتولوژی عامل‌محور متمایز می‌شوند، به جای اینکه مبتنی بر وظیفه باشند، مبتنی بر حقوق هستند. بر حقوق بیماران یا قربانیان بالقوه تمرکز دارد، مانند حق استفاده نکردن به عنوان وسیله‌ای برای رسیدن به هدف توسط شخص دیگری. تئوری‌های قراردادگرا با هر دو نظریه عامل محور و بیمار محور متفاوت است. در نظریه‌ی قراردادگرا، اعمال نادرست اخلاقی آن دسته از اعمالی هستند که توسط اصولی ممنوع می‌شوند که افراد در یک قرارداد اجتماعی مناسب و توصیف شده آن را می‌پذیرند یا با اصولی ممنوع می‌شوند که چنین افرادی نمی‌توانند «معقولانه رد کنند» [۹۶].

اخلاق نتیجه گرایانه: اخلاق نتیجه گرایانه، همانطور که از نامش پیداست، بر نتایج سودگرایانه‌ی اَعمال، تأکید دارد [۹۷]. اخلاق پیامدگرایانه، اخلاقی بودن یک عمل را صرفاً بر اساس نتیجه

یا پیامدهای آن ارزیابی می‌کند. به عبارت دیگر، در نظریه‌های نتیجه گرا، صحت اخلاقی یک عمل با توجه به نتیجه یا نتایج آن عمل تعیین می‌شود. به عقیده نتیجه گرایان، اگر پیامد آن عمل سودمند، یعنی مطلوب‌تر از نامطلوب تلقی شود، از نظر اخلاقی درست است. فرض کنید یک مورد ساده که در آن فرد با انتخاب بین چندین عمل ممکن مواجه می‌شود، نتیجه‌گرایی مشخص می‌کند که عمل اخلاقاً درست همان کاری است که بهترین پیامدهای کلی را دارد.

اخلاق پیامدگرا از نظر تاریخی مهم و هنوز هم محبوب است، زیرا این شهود اساسی را در بر می‌گیرد که آنچه خوب یا درست است هر چیزی است که جهان را در آینده، بهترین می‌کند، زیرا ما نمی توانیم گذشته را تغییر دهیم. نظریه‌های نتیجه گرایی را می‌توان به موارد زیر تقسیم کرد [۹۸]، [۹۹].

- اگویسم اخلاقی: بیان می‌کند که یک عمل، زمانی از نظر اخلاقی خوب است که پیامدها یا آثار آن عمل فقط برای عاملی که آن عمل را انجام می‌دهد، مطلوب‌تر از نامطلوب باشد.
- نوع دوستی اخلاقی: بیان می‌کند که یک عمل، زمانی از نظر اخلاقی خوب است که پیامدها یا آثار آن عمل برای همه به جز فاعل، مطلوب‌تر از نامطلوب باشد.
- فایده گرایی: بیان می‌کند که یک عمل، زمانی از نظر اخلاقی خوب است که پیامدها یا آثار آن عمل برای همه مطلوب‌تر از نامطلوب باشد.

هر سه‌ی این نظریه‌ها بر پیامدهای اَعمال برای گروه‌های مختلف مردم تمرکز دارند. اما مانند همه نظریه‌های هنجاری، سه نظریه فوق رقیب یکدیگر هستند. آنها همچنین نتایج متفاوتی را ارائه می‌دهند.

ب) خلاصه‌ای از اخلاق هنجاری

از توصیفات بالا مشخص می‌شود که نظریه‌های اخلاقی هنجاری مختلف، قضاوت متفاوتی را برای یک اقدام یا تصمیم به همراه خواهند داشت. به تصویر زیر توجه کنید [۱۰۰]: یک آقای مسن توسط گروهی از نوجوانان متکبر در مترو عذاب می‌کشد و زنی مصمم به کمک او می‌آید. اخلاق مدار فضیلت، عمل او را از نظر اخلاقی مناسب می‌داند، زیرا فضیلت‌های خیرخواهی و شجاعت را نشان می‌دهد. متخصص دئونتولوژی اقدام او را ستودنی می‌داند زیرا با قانون کمک به نیازمندان مطابقت دارد. نتیجه گرا از عمل او به خوبی دفاع می‌کند، زیرا او رفاه کلی همه طرف‌های درگیر را به حداکثر رسانده - نجیب زاده‌ی سالخورده از رنج و رسوایی در امان است، که از سرگرمی نوجوانان پیشی می‌گیرد-. مقایسه‌ی مختصری بین سه نظریه اخلاقی هنجاری در جدول ۵ ارائه شده است.

۲) رویکردهای پیاده سازی اخلاق در هوش مصنوعی

در بخش قبل، نظریه‌های اخلاقی مرتبط با اخلاق هوش مصنوعی را مورد بحث قرار دادیم. این بخش به طور خلاصه روش‌ها و رویکردهای پیاده‌سازی اخلاق در سیستم‌های هوش مصنوعی، یعنی طراحی سیستم‌های هوش مصنوعی اخلاقی را بررسی می‌کند. روش‌ها یا رویکردهای موجود برای کاشت اخلاق در هوش مصنوعی را می‌توان به سه نوع اصلی تقسیم کرد: رویکردهای بالا به پایین، رویکردهای پایین به بالا و رویکردهای ترکیبی [۱۰۱].

الف) رویکردهای بالا به پایین

رویکرد بالا به پایین به رویکردی اشاره دارد که یک نظریه اخلاقی خاص را اتخاذ می‌کند و الزامات محاسباتی آن را برای هدایت طراحی الگوریتم‌ها و زیرسیستم‌هایی که می‌توانند آن نظریه را تحقق بخشند، تجزیه و تحلیل کند [۱۰۲]. رویکردهای بالا به پایین، استدلال اخلاقی را بر اساس نظریه‌های اخلاقی یا اصول اخلاقی معین انجام می‌دهند. در رویکردهای بالا به پایین، اصول اخلاقی و نظریه‌های اخلاقی به عنوان قوانینی برای انتخاب اقدامات اخلاقی مناسب [۱۰۱] استفاده می‌شود یا برای توصیف آنچه عامل هوش مصنوعی باید در یک موقعیت خاص انجام دهد استفاده می‌شود. بنابراین، یک رویکرد از بالا به پایین مستلزم قوانین، تعهدات و حقوق تعریف شده‌ی رسمی برای هدایت عامل هوش مصنوعی در فرایند تصمیم‌گیری است. به عنوان مثال، سه قانون رباتیک آسیموف [۱۰۳] که بر رفتار ربات‌ها حاکم است، می‌تواند یک سیستم اخلاقی از بالا به پایین برای روبات‌ها در نظر بگیرد [۱۰۱]. بسیاری از پیاده سازی‌های دیگر با استفاده از رویکردهای بالا به پایین را می‌توان در [۱۰۴]-[۱۱۱] و غیره یافت.

رویکردهای بالا به پایین معمولاً به عنوان داشتن مجموعه‌ای از قوانین درک می‌شوند که می‌توانند به یک الگوریتم تبدیل شوند. این قوانین وظایف یک نماینده یا نیاز نماینده را برای ارزیابی پیامدهای اقدامات احتمالی مختلفی که ممکن است انجام دهد مشخص می‌کند. رویکردهای بالا به پایین در نظریه‌های اخلاقی متفاوت مورد استفاده هستند. به عنوان مثال، زمانی که نظریه‌ی نتیجه گرایی در رویکرد از بالا به پایین استفاده می‌شود، مدل استدلال باید نتیجه یا پیامد اَعمال را به عنوان مبنای تصمیم ارزیابی کند، یعنی عملی که منجر به نتیجه‌ی خوب می‌شود اخلاقی است و در غیر این صورت غیراخلاقی است. در حالی که اگر تئوری دئونتولوژیک به کار رود، مدل استدلال رضایت، یک ارزش معین را برای تصمیم گیری در نظر می‌گیرد، یعنی عملی که از وظایف تبعیت می‌کند اخلاقی است و عمل زیر پا گذاشتن وظایف غیراخلاقی است.

ب) رویکردهای پایین به بالا

رویکردهای پایین به بالا فرض می‌کنند که رفتار اخلاقی از مشاهدات رفتارهای دیگران آموخته می‌شود. در رویکرد پایین به بالا، تأکید بر ایجاد محیطی است که در آن یک عامل هوش

مصنوعی مسیر عمل را بررسی می‌کند و عمل اخلاقی قابل ستایش، با پاداش یا انتخاب مشخص می‌شود [۱۰۱]. برخلاف رویکردهای بالا به پایین، که برای تعریف اَعمال اخلاقی و غیر اخلاقی، به نظریه‌ها یا اصول اخلاقی نیاز دارند، اصول اخلاقی از مشاهدات یا تجربه در رویکردهای پایین به بالا کشف یا آموخته می‌شوند. این رویکرد نشان می‌دهد که عامل هوش مصنوعی باید مانند بچه‌های کوچک هنجارها و اخلاقیات را بیاموزد تا از نظر اخلاقی شایسته باشد. برای مثال، آقایان هنرور و آقایی، یک عامل BDI کاسوئیستی [۱۱۲] را پیشنهاد کردند که روش استدلال مبتنی بر مورد در هوش مصنوعی و رویکرد موردی از پایین به بالا را در اخلاق ترکیب می‌کند تا قابلیت استدلال اخلاقی را به عامل باور-میل-نیت (belief-desire-intention) اضافه کند. [۱۱۳]. سایر پیاده سازی‌های رویکردهای پایین به بالا را می‌توان در [۱۱۴]-[۱۱۸] و غیره یافت.

رویکردهای پایین به بالا می‌توانند از خرد جمعی به عنوان وسیله‌ای برای آگاهی دادن به قضاوت اخلاقی عامل، استفاده کنند و سپس آن عامل می‌تواند یاد بگیرد که چگونه اخلاقی بودن عمل خود را قضاوت کند و بنابراین اخلاقی رفتار کند. ظاهراً، رویکردهای پایین به بالا فرض می‌کنند که حجم کافی از داده‌ها یا مشاهدات در مورد تصمیمات اخلاقی و نتایج آنها را می‌توان از مجموعه‌ای از موضوعات یا سناریوها جمع‌آوری کرد. این موضوع لازمه‌ی استفاده از رویکردهای پایین به بالا برای پیاده سازی سیستم‌های هوش مصنوعی اخلاقی است. با این حال، در عمل، این نیاز به راحتی برآورده نمی شود.

ج) رویکردهای ترکیبی

رویکرد ترکیبی تلاش می‌کند تا مزایای رویکردهای بالا به پایین و پایین و همچنین پایین به بالا را ترکیب کند. رویکردهای بالا به پایین از نظریه‌ها و اصول اخلاقی استفاده می‌کنند و بر اهمیت نگرانی‌های اخلاقی صریح که از خارج از نهاد (موضوع اخلاقی) ناشی می‌شوند، تأکید می‌کنند. در حالی که رویکردهای پایین به بالا بیشتر بر پرورش اخلاقی که از درون موجودیت، از طریق تکامل و یادگیری ناشی می‌شود، تمرکز می‌کنند. هر دو رویکرد از بالا به پایین و پایین به بالا جنبه‌های مختلفی از حساسیت اخلاقی را در بر می‌گیرند. با ترکیب این رویکردها، ممکن است بتوانیم عامل هوش مصنوعی ایجاد کنیم که بتواند اخلاق پویا و انعطاف پذیر رویکرد پایین به بالا را در عین رعایت اصول از بالا به پایین حفظ کند. رویکردهای ترکیبی متفاوتی در [۱۱۹] - [۱۲۴] پیاده‌سازی شده‌اند.

همانطور که Gigerenzer [۱۲۵] بیان کرد، ماهیت رفتار اخلاقی، ناشی از تعامل بین ذهن و محیط است. بر اساس این دیدگاه، هم طبیعت و هم تربیت، در شکل گیری رفتار اخلاقی مهم هستند. رویکرد ترکیبی با این مفهوم سازگار است. در رویکرد ترکیبی، رویکرد بالا به پایین از قوانین برنامه ریزی شده و رویکرد پایین به بالا از قوانین آموخته شده از مشاهدات یا تجربیات

زمینه استفاده می‌کند که به ترتیب شبیه به ماهیت و پرورش جنبه‌های اخلاقی هستند. بنابراین، از این منظر، هم طبیعت و هم پرورش در رویکردهای ترکیبی مورد توجه قرار می‌گیرند.

د) نکاتی در مورد رویکردهای اخلاقی

رویکرد از بالا به پایین، نظریه‌ها و اصول اخلاقی مشخص شده را به تصمیم گیری اخلاقی تبدیل می‌کند یا نظریه‌ها و اصول اخلاقی داده شده را به الگوریتم تبدیل می‌کند. این رویکرد برای طراحی و تحقق عوامل هوش مصنوعی اخلاقی با اصول اخلاقی شناخته شده و کدهای اخلاقی مناسب است. مزیت این رویکرد این است که بر اساس تئوری‌ها و قواعد اخلاقی از پیش تعیین شده، تصمیمات و اقدامات کارگزاران اخلاقی قابل پیش بینی است و هنجارها یا قوانین اخلاقی اجرا شده از طریق کدهای برنامه یا ابزارهای دیگر را می‌توان در طول تصمیم گیری اخلاقی درک کرد. بنابراین، اعتبار عامل هوش مصنوعی اخلاقی ایجاد شده توسط رویکرد بالا به پایین را می‌توان بهتر تضمین کرد زیرا فرایند تصمیم گیری آن دارای قابلیت تفسیر و شفافیت قوی است. نقطه ضعف رویکرد از بالا به پایین این است که عامل اخلاقی نظریه‌های اخلاقی یا قواعد اخلاقی از پیش تعیین شده را اتخاذ می‌کند، هنگام تصمیم گیری در یک محیط پیچیده و متغیر، این روش فاقد انعطاف پذیری و سازگاری است.

رویکرد پایین به بالا تأکید می‌کند که کارگزاران اخلاقی اخلاق را به طور مستقل از محیط اجتماعی می‌آموزند، به تدریج دارای استدلال اخلاقی و توانایی‌های اخلاقی هستند و می‌توانند با تغییرات محیطی سازگار شوند. رویکرد پایین به پایین برای طراحی و اجرای عوامل هوش مصنوعی اخلاقی بدون تئوری‌ها و دستورالعمل‌های اخلاقی روشن، مناسب است. مزیت این رویکرد این است که عامل می‌تواند از طریق یادگیری مستمر توسعه یافته و تکامل یابد تا با تغییرات محیطی سازگار شود. این دسته از رویکردها، سازگاری و انعطاف پذیری خوبی دارند و می‌توان نظریه‌ها یا دستورالعمل‌های اخلاقی متفاوت و جدیدی را برای سناریوهای کاربردی مختلف ساخت. نقطه ضعف این رویکرد این است که به دلیل عدم هدایت نظریه‌ها یا قوانین اخلاقی، فرایند تصمیم گیری عوامل هوش مصنوعی اخلاقی دارای درجه خاصی از اطاعت کورکورانه است و تکمیل آموزش در کوتاه مدت دشوار است. پس در این رویکرد باید زمان و تصمیمات اخلاقی مناسب اتخاذ کنیم. در عین حال، تضمین تفسیرپذیری و شفافیت فرایند تصمیم‌گیری عوامل AI اخلاقی طراحی‌شده دشوار است.

رویکرد ترکیبی، مزایای رویکردهای بالا به پایین و پایین به بالا را با هم ترکیب می‌کند و تا حدودی بر کاستی‌های این دو روش غلبه می‌کند. اگر یک رویکرد واحد (از بالا به پایین یا پایین به بالا) الزامات را پوشش ندهد، یک رویکرد ترکیبی ضروری و امیدوارکننده در نظر گرفته می‌شود. با این حال، چالش اصلی ترکیب مناسب ویژگی‌های رویکردهای بالا به پایین و پایین به بالا است. ویژگی‌های سه رویکرد برای پیاده سازی اخلاق در هوش مصنوعی خلاصه شده و در جدول ۶ فهرست شده است.

جدول ۶: ویژگی های سه رویکرد برای پیاده‌سازی اخلاق در هوش مصنوعی

رویکرد	توضیحات	آیا نیاز به قوانین اخلاقی دارد؟	توانایی یادگیری	توانایی تطبیق	قابلیت تفسیر
بالا به پایین	برنامه‌ریزی نظریه و اصول اخلاقی داده‌شده	بله	خیر	ضعیف	زیاد
پایین به بالا	یادگیری قوانین کلی از موارد فردی	خیر	قوی	قوی	کم
ترکیبی	ترکیب رویکردهای پایین به بالا و بالا به پایین	بله	قوی	قوی	متوسط

۴-۲. رویکردهای فناورانه

در این بخش، به طور خلاصه وضعیت تحقیق در مورد رویکردهای فناورانه برای پرداختن به مسائل اخلاقی هوش مصنوعی را در راستای اصول مورد بحث در بخش ۵-ب خلاصه می‌کنیم. در حال حاضر، رویکردهای فناورانه برای کاهش مسائل مرتبط هنوز در مرحله رشد اولیه هستند. در سال‌های اخیر، جوامع تحقیقاتی هوش مصنوعی تلاش‌های خاصی را برای پرداختن به مسائل اخلاقی هوش مصنوعی انجام داده‌اند. به عنوان مثال، ACM (انجمن ماشین‌های محاسباتی) کنفرانس سالانه ACM FaccT (که محققان و متخصصان علاقه‌مند به عدالت، پاسخگویی و شفافیت در سیستم‌های فنی-اجتماعی را گرد هم می‌آورد) را از سال ۲۰۱۸ تا به حال برگزار کرده است. AAAI (انجمن پیشرفت هوش مصنوعی) و ACM کنفرانس AAAI/ACM را در زمینه هوش مصنوعی، اخلاق و جامعه (AIES) از سال ۲۰۱۸ تأسیس کرده اند و سی و یکمین کنفرانس بین المللی مشترک هوش مصنوعی و بیست و سومین کنفرانس اروپایی در زمینه هوش مصنوعی (IJCAI-ECAI 2022) مسیر ویژه‌ای را در مورد "AI برای خوب" ارائه می‌دهد.

کار موجود، تا جایی که ما می‌دانیم، عمدتاً بر روی چند موضوع و اصول اصلی و کلیدی تمرکز دارد و سایر مسائل و اصول به ندرت مطرح می‌شوند. بنابراین، ما فقط خلاصه‌ای کوتاه از

رویکردهای فناورانه ارائه می‌دهیم که شامل پنج اصل کلیدی اخلاقی است. به ویژه، برای پنج اصل کلیدی (به عنوان مثال، شفافیت، انصاف و عدالت، عدم سوء استفاده، مسئولیت و پاسخگویی و حریم خصوصی)، برخی از موضوعات پژوهشی نماینده و مراجع مربوطه در جدول ۲ مواد تکمیلی فهرست شده‌اند.

هوش مصنوعی قابل توضیح (XAI)، که به عنوان هوش مصنوعی قابل تفسیر نیز شناخته می‌شود، در حال حاضر مسیر اصلی تحقیق و روش فنی برای رسیدگی به مسائل عدم شفافیت در هوش مصنوعی است. هدف XAI این است که به کاربران انسانی اجازه دهد تا نتایج و خروجی‌های ارائه شده توسط یک سیستم هوش مصنوعی، به ویژه توسط الگوریتم‌های ML را درک کنند. کریستوف و همکارانش [۱۲۸] تاریخچه مختصری از حوزه XAI ارائه کرد و با یک مرور کلی از روش‌های تفسیری پیشرفته، برخی از چالش‌های تحقیق را مورد بحث قرار داد. علاوه بر این، کریستوف کتابی در مورد ML قابل تفسیر [۱۲۹] نوشته است که یک کتاب محبوب در زمینه XAI است.

در مورد اصل انصاف، آثار زیادی نیز وجود دارد که به حذف یا کاهش تعصب یا تبعیض نشان داده شده توسط سیستم‌های هوش مصنوعی، به ویژه در ML اختصاص داده شده است. هوش مصنوعی منصفانه [۱۳۰] با هدف جلوگیری از آسیب (یا منفعت) متفاوت برای زیرگروه‌های مختلف، یک موضوع تحقیقاتی بسیار فعال است که به پرداختن به مسائل عدم انصاف در هوش مصنوعی اختصاص دارد. در بررسی انصاف در ML توسط سایمون و کریستین [۱۳۱]، مکاتب فکری و رویکردهای مختلف برای کاهش سوگیری‌ها و افزایش انصاف در ML بررسی شد.

اصل Nonmaleficence شامل چندین کد مانند ایمنی، امنیت و استحکام است. از این رو، کارهایی برای هر یک از کدهای مرتبط با اصل عدم سوء استفاده وجود دارد. در حال حاضر، هوش مصنوعی ایمن، هوش مصنوعی مطمئن و هوش مصنوعی قوی سه جهت اصلی تحقیقاتی برای تحقق اصل عدم سوء استفاده در هوش مصنوعی هستند. خوانندگان علاقه‌مند می‌توانند جزئیات بیشتر را از طریق مراجع مربوطه فهرست شده در جدول ۲ مواد تکمیلی دریافت کنند.

از آنجایی که هوش مصنوعی به طور گسترده در زندگی ما استفاده می‌شود، هوش مصنوعی مسئول، در حال تبدیل شدن به یک موضوع حیاتی است. مسئولیت مفهومی نسبتاً انتزاعی و گسترده است. در حال حاضر، هیچ تعریف یا مفهوم جهانی و یکپارچه‌ای برای هوش مصنوعی مسئول وجود ندارد که عمدتاً شامل پاسخگویی، مسئولیت پذیری، انصاف، استحکام و توضیح پذیری است [۱۳۲]. دوریان و همکاران [۱۳۳] دو چارچوب برای هوش مصنوعی مسئول با ادغام تجزیه و تحلیل اخلاقی در عملکرد مهندسی در هوش مصنوعی پیشنهاد کردند. علاوه بر این، مقاله‌ی [۱۳۴] مقدمه‌ای سیستماتیک در مورد هوش مصنوعی مسئول ارائه می‌دهد.

به منظور رسیدگی به مسائل حریم خصوصی در هوش مصنوعی، محققان تلاش‌های زیادی انجام داده اند. حریم خصوصی متفاوت [۱۳۵]، یکی از رویکردهای اصلی برای حفظ حریم خصوصی و تجزیه و تحلیل داده‌ها است. اخیراً یک پارادایم جدید ML، یعنی یادگیری فدرال [۱۳۶]، [۱۳۷] (که ML توزیع شده نیز نامیده می‌شود)، برای کاهش خطر نشت حریم خصوصی در ML پیشنهاد شده است. علاوه بر این، برخی دیگر از تکنیک‌های حفظ حریم خصوصی برای ML [138]، [۱۳۹] پیشنهاد شده است.

در مورد سایر اصول مانند خیرخواهی، آزادی و خودمختاری، کرامت و غیره، رویکردهای تکنولوژیکی مرتبطی در ادبیات پیدا نکرده‌ایم. این امر، ممکن است به دلیل دشواری یا نامناسب بودن استفاده از روش‌های فنی برای رسیدگی به مسائل مربوط به این اصول باشد. به طور کلی، اخلاق هوش مصنوعی یک حوزه نسبتاً جدید است و رویکردهای تحقق این اصول هنوز نیاز به مطالعه در آینده دارد.

۴-۳. رویکردهای حقوقی: قانونگذاری و مقررات

با توجه به استفاده روزافزون از فناوری‌های هوش مصنوعی در بسیاری از بخش‌ها و نمایش مسائل اخلاقی و خطرات در کاربردهای هوش مصنوعی، قوانین و مقررات بسیاری توسط دولت‌ها و سازمان‌ها برای کنترل توسعه و کاربرد هوش مصنوعی وضع شده است. رویکردهای حقوقی به یکی از ابزارهای پرداختن به مسائل اخلاقی در هوش مصنوعی تبدیل شده است. در ادامه، چندین قانون و مقررات مرتبط با هوش مصنوعی را که در چند سال گذشته پیشنهاد شده‌اند، فهرست می‌کنیم.

- در سال ۲۰۱۶، پارلمان اروپا و شورای اتحادیه اروپا (EU) مقررات عمومی حفاظت از داده‌ها [۱۴۰] را منتشر کردند که مقرراتی در قانون اتحادیه اروپا در مورد حفاظت از داده‌ها و حریم خصوصی در اتحادیه اروپا و منطقه اقتصادی اروپا است.
- در سال ۲۰۱۷، ایالات متحده، لایحه‌ای را برای تضمین ایمنی وسایل نقلیه‌ی خودکار با تشویق آزمایش و استقرار چنین وسایل نقلیه‌ای تصویب کرد به نام"قانون استقرار و تحقیقات ایمن در آینده - استقرار و تحقیقات در خودروها در آینده" [۱۴۱].
- در سال ۲۰۱۸، برزیل قانون شماره ۱۳ ۷۰۹، قانون کلی حفاظت از داده‌ها (Lei Geral de Proteção de Dados) [۱۴۲] را برای حفاظت از داده‌های شخصی در این کشور به تصویب رساند.
- در سال ۲۰۲۱، کمیسیون اروپا قانون هوش مصنوعی [۱۴۳] را منتشر کرد که یک رویکرد نظارتی بین بخشی را برای استفاده از سیستم‌های هوش مصنوعی در سراسر اتحادیه اروپا و بازار آن تعیین می‌کند.

۵. روش‌های ارزیابی هوش مصنوعی اخلاقی

هدف رشته‌ی اخلاق هوش مصنوعی، طراحی سیستم‌های هوش مصنوعی اخلاقی برای رفتار اخلاقی یا پایبندی به اصول و قواعد اخلاقی و معنوی است. نحوه ارزیابی یا محاسبه‌ی اخلاقیات یا معنویات (صلاحیت معنوی) هوش مصنوعی اخلاقی طراحی شده بسیار مهم و ضروری است، زیرا سیستم‌های هوش مصنوعی طراحی شده باید قبل از استقرار، آزمایش یا ارزیابی شوند که آیا سیستم هوش مصنوعی الزامات اخلاقی را برآورده می‌کند یا خیر. با این حال، این جنبه اغلب در ادبیات موجود نادیده گرفته شده یا نادیده گرفته می‌شود. این بخش سه نوع رویکرد آزمایش، تأیید و استانداردها را برای ارزیابی اخلاق هوش مصنوعی بررسی می‌کند.

الف. آزمایش

تست، یک روش معمولی است که برای ارزیابی قابلیت‌های اخلاقی یک سیستم هوش مصنوعی استفاده می‌شود. معمولاً هنگام آزمایش یک سیستم، خروجی سیستم باید با یک حقیقت زمینی یا خروجی مورد انتظار مقایسه شود [۱۰۰]. این بخش بر روی رویکردهای آزمایشی برای ارزیابی هوش مصنوعی اخلاقی تمرکز دارد.

۷. آزمون تورینگ اخلاقی

هم در نظریه‌های اخلاقی و هم در بحث‌های روزمره درباره اخلاق، مردم معمولاً نظرات متفاوتی در مورد اخلاقی بودن اَعمال مختلف دارند. به عنوان مثال، کانت ادعا کرد که دروغ، صرف نظر از عواقب آن، همیشه غیراخلاقی است. اخلاق‌گرایان فایده‌گرا این را انکار می‌کنند و معتقدند که دروغ تا زمانی که پیامدهای آن در مجموع به اندازه کافی خوب باشد موجه است. از آنجایی که نظریه‌های اخلاقی مختلف معیارهای ارزیابی متفاوتی برای رفتار اخلاقی دارند، آلن و همکاران [۱۴۴] پیشنهاد کردند که از آزمون تورینگ اخلاقی (MTT) برای ارزیابی عوامل اخلاقی مصنوعی استفاده شود.

در نسخه استاندارد آزمون تورینگ [۱۴۵]، یک بازجوی انسانی از راه دور وظیفه دارد بین یک ماشین (کامپیوتر) و یک موضوع انسانی بر اساس پاسخ آنها به سؤالات مختلف مطرح شده توسط بازجو تمایز قائل شود. اگر ماشینی با شانس کافی به عنوان سوژه انسانی اشتباه شناسایی شود، موفق به گذراندن تست می‌شود و ماشین به عنوان موجودی باهوش و متفکر در نظر گرفته می‌شود. Turing Test مستقیماً آزمون رفتاری را انجام می‌دهد تا اختلاف نظر در مورد معیارهای تعریف هوش یا کسب موفقیت در زبان طبیعی را دور بزند. آزمون تورینگ اخلاقی (MTT) به طور مشابه برای دور زدن اختلاف نظرها در مورد استانداردهای اخلاقی با محدود کردن مکالمات در آزمون عطف استاندارد به سؤالات مربوط به اخلاق پیشنهاد شد. اگر انسان بازجو نتواند ماشین را از سوژه انسانی در سطحی بالاتر از شانس تشخیص دهد، ماشین یک عامل اخلاقی است.

با این حال، آلن و همکاران [۱۴۴] اعتراف کردند که یکی از محدودیت‌های MTT این است که بر توانایی ماشین‌ها برای بیان واضح قضاوت‌های اخلاقی تأکید می‌کند. ریشه شناسان یا کانتیان ممکن است با این تاکید راضی باشند، اما نتیجه گرایان استدلال می‌کنند که MTT تاکید زیادی بر توانایی بیان دلیل اَعمال فرد دارد. به منظور تغییر تمرکز از توانایی مکالمه به عمل، آلن و همکاران [۱۴۴] همچنین یک MTT جایگزین را پیشنهاد کرد که "MTT مقایسه ای" (cMTT) نامیده شد. در cMTT، به انسانی بازجوی دو جفت توصیف از اَعمال واقعی و اخلاقی مهم یک سوژه انسانی و یک ماشین (یا عامل هوش مصنوعی) داده می‌شود، که از تمام مراجعی که بازیگر را شناسایی می‌کنند پاک می‌شود. اگر بازجو دستگاه را در درصد معینی به درستی شناسایی کند، آنگاه دستگاه نمی تواند آزمون را پشت سر بگذارد. یک مشکل این نسخه از MTT این است که نحوه رفتار ماشین راحت‌تر از انسان تشخیص داده می‌شود، زیرا ماشین به طور مداوم در شرایط یکسان رفتار می‌کند. بنابراین، باید از بازجو خواسته شود تا ارزیابی کند که آیا یک بازیگر نسبت به دیگری اخلاقی‌تر است یا نه؟ اگر دستگاه به عنوان دستگاهی که دارای اخلاق کمتری است، بیشتر از انسان شناسایی نشود، دستگاه آزمایش را به خوبی گذرانده است.

اگرچه cMTT مشکلات متعددی دارد، برای مثال، ممکن است کسی استدلال کند که این استاندارد بسیار پایین است، والاک و آلن [۱۴۶] معتقدند که cMTT یک روش عملی و قابل قبول برای ارزیابی اخلاقیات عوامل هوش مصنوعی است، زیرا هیچ معیار ارزیابی دیگری وجود ندارد که به طور مشترک پذیرفته شده و مورد توافق باشد.

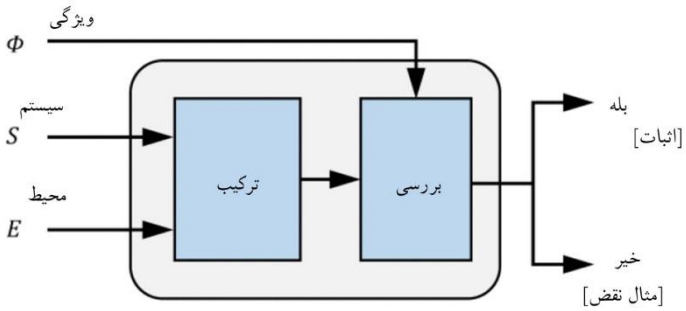
۲) تست‌های تخصصی و غیر کارشناسی

علاوه بر MTT، محققان سعی کرده‌اند صلاحیت اخلاقی سیستم‌های هوش مصنوعی را از طریق آزمون‌های خبره یا غیرمتخصص ارزیابی کنند، که در آن نتیجه سیستم با حقیقت اصلی ارائه شده توسط افراد غیرمتخصص یا متخصص مقایسه می‌شود. آزمون کارشناسی استانداردهای متخصصان در اخلاق هنجاری را برای ارزیابی اخلاقیات عوامل هوش مصنوعی اتخاذ می‌کند. آزمون‌های غیرکارشناسی، اخلاق عامیانه را به عنوان معیار در نظر می‌گیرند و توانایی اخلاقی عامل یا سیستم هوش مصنوعی را در آزمون معیار مربوطه ارزیابی می‌کنند. در آزمون‌های غیرکارشناسی، شهروندان می‌توانند نقش خود را در ارزیابی و ارزیابی قابلیت‌های اخلاقی یک سیستم هوش مصنوعی بر اساس مواضع اخلاقی و بررسی دقیق خود ایفا کنند.

ب. تأیید

دسته دیگری از رویکردها برای ارزیابی اخلاقی بودن هوش مصنوعی شامل اثبات این است که سیستم هوش مصنوعی طبق برخی مشخصات شناخته شده به درستی رفتار می‌کند. سشیا و همکاران [۱۴۷] این نوع رویکرد را مورد بحث قرار دادند. یک فرایند تأیید رسمی معمولی در شکل ۵ نشان داده شده است که در آن S مدلی از سیستمی است که باید تأیید شود، E مدلی از

محیط است و Φ خاصیتی است که باید تأیید شود. برنامه راستی آزمایی، یک پاسخ بله یا خیر به عنوان خروجی می‌دهد که نشان دهنده‌ی آن است که آیا S ویژگی Φ را در محیط E برآورده می‌کند یا نه. و یک مدرک صحت شامل یک پاسخ بله در برخی از ابزارهای تأیید رسمی است.



شکل ۵: فرایند تأیید رسمی معمولی سشیا و همکاران

آرنولد و شوتز [۱۴۸] نقایص MTT را بررسی کردند و خاطرنشان کردند که ارزیابی‌های مبتنی بر MTT در برابر فریب، استدلال ناکافی و عملکرد اخلاقی ضعیف آسیب‌پذیر هستند و آنها مفهوم "تأیید طراحی" را برای ارزیابی شایستگی اخلاقی سیستم هوش مصنوعی پیشنهاد کردند. برای ارزیابی طراحی اخلاقی هوش مصنوعی، می‌توان از معیارهای ارزیابی متنوع استفاده کرد. صرف نظر از روشی که هوش مصنوعی استدلال اخلاقی را انجام می‌دهد، بسیار مهم است که فعالیت‌های اخلاقی آن با اهداف طراحی اخلاقی مطابقت داشته باشد.

ج. استانداردها

بسیاری از استانداردهای صنعتی برای هدایت توسعه و کاربرد هوش مصنوعی و ارزیابی محصولات هوش مصنوعی پیشنهاد شده است. در این بخش برخی از استانداردهای مرتبط با هوش مصنوعی معرفی می‌شوند.

- در سال ۲۰۱۴، انجمن کامپیوتر استرالیا کد رفتار حرفه‌ای ASC را برای پیروی از همه متخصصان فناوری ارتباطات اطلاعات ایجاد کرد که شش ارزش اصلی اخلاقی و الزامات مرتبط با رفتار حرفه‌ای را مشخص می‌کند.
- در سال ۲۰۱۸، ACM کد اخلاقی و رفتار حرفه‌ای ACM را برای پاسخ به تغییرات در حرفه رایانه از سال ۱۹۹۲ به روز کرد. این کد بیانگر وجدان حرفه است و برای الهام بخشیدن و هدایت رفتار اخلاقی همه متخصصان رایانه، از جمله متخصصان فعلی طراحی شده است و پزشکان مشتاق، مدرسان، دانش‌آموزان، تأثیرگذاران و هر کسی که از فناوری محاسباتی به روشی تأثیرگذار استفاده می‌کند. علاوه بر این، این

کد به عنوان مبنایی برای اصلاح در صورت وقوع تخلف عمل می‌کند. این آیین نامه شامل اصولی است که به عنوان بیانیه‌های مسئولیت، بر اساس این درک که خیر عمومی همیشه ملاحظات اولیه است، فرموله شده است. هر اصل با دستورالعمل‌هایی تکمیل می‌شود که توضیحاتی را برای کمک به متخصصان محاسبات در درک و به کارگیری اصل ارائه می‌دهد [۱۴۹].

- پروژه‌ی IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems [۱۵۰] توسعه‌ی سری استانداردهای IEEE P7000™ [۱۵۱] را تایید کرد (فهرست شده در جدول ۳ مواد تکمیلی) که موضوعاتی از جمع آوری داده‌ها تا حریم خصوصی و سوگیری الگوریتمی و فراتر از آن را پوشش می‌دهد.
- ISO/IEC JTC 1/SC 42 [۱۵۲] که کمیته مشترک بین ISO و IEC مسئول استانداردسازی در حوزه هوش مصنوعی است، به توسعه مجموعه‌ی بزرگی از استانداردها شامل حوزه‌های استانداردهای اساسی هوش مصنوعی، داده‌های بزرگ، قابلیت اعتماد هوش مصنوعی، موارد استفاده، برنامه‌های کاربردی، مفاهیم حاکمیتی هوش مصنوعی، رویکردهای محاسباتی هوش مصنوعی، نگرانی‌های اخلاقی و اجتماعی هستند. استانداردهای منتشر شده و در دست توسعه توسط ISO/IEC JTC 1/SC 42 در جدول ۴ از مواد تکمیلی فهرست شده است.

با نگرانی در مورد مسائل اخلاقی هوش مصنوعی، علاقه به استانداردهای هوش مصنوعی برای شکل دادن به طراحی، استقرار و ارزیابی هوش مصنوعی به سرعت در حال رشد است. اگرچه استانداردهای زیادی پیشنهاد شده است، شکاف بین استانداردها (یا اصول) و عمل هنوز زیاد است. در حال حاضر، تنها برخی از شرکت‌های بزرگ، مانند IBM [۱۵۳] و مایکروسافت [۱۵۴]، استانداردها، چارچوب‌ها و دستورالعمل‌های صنعتی خود را برای ایجاد فرهنگ هوش مصنوعی پیاده‌سازی کرده‌اند. اما برای کسب و کارهای کوچکتر با منابع کمتر، اصول تمرین شکاف یک مشکل بزرگ است. بنابراین، هنوز تلاش‌های زیادی لازم است. از یک سو، ارائه استانداردهای توسعه یافته ضروری است و از سوی دیگر، لازم است به شدت عمل به استانداردها ترویج شوند.

۸. چالش‌ها و چشم‌اندازهای آینده

از آنجایی که اخلاق هوش مصنوعی یک رشته نوظهور است و هنوز چالش‌ها و مشکلات زیادی وجود دارد که باید در این زمینه مورد توجه قرار گیرد. در این بخش، برخی از چالش‌ها در اخلاق هوش مصنوعی را مورد بحث قرار می‌دهیم و دیدگاه‌های آینده را از دیدگاه خود ارائه می‌کنیم. هدف از این بخش ارائه برخی سوالات احتمالی تحقیقاتی و رهنمودها برای تحقیقات بیشتر در آینده است و از این طریق پیشرفت تحقیقات در زمینه اخلاق هوش مصنوعی را تسهیل می‌کند.

الف. چالش‌ها در دستورالعمل‌ها و اصول اخلاقی هوش مصنوعی

همانطور که در بخش چهارم بررسی شد، تعداد زیادی دستورالعمل توسط سازمان‌ها، شرکت‌ها و دولت‌های مختلف پیشنهاد و منتشر شده‌اند و اصول مختلفی را می‌توان در این دستورالعمل‌ها شناسایی کرد. اما در حال حاضر هنوز دستورالعملی وجود ندارد که توسط سازمان ها، بخش‌ها و دولت‌های مختلف تصویب شده باشد. به عبارت دیگر، سازمان‌های مختلف، شرکت‌های حوزه‌های مختلف و حتی شرکت‌های مختلف در یک حوزه، نظرات متفاوتی در مورد اخلاق هوش مصنوعی دارند. اجماع در مورد اخلاق هوش مصنوعی هنوز حاصل نشده است و مشخص نیست که هوش مصنوعی باید از چه اصول و ارزش‌های مشترکی پیروی کند. علاوه بر این، زمانی که هوش مصنوعی در حوزه‌های مختلف به کار می‌رود، ممکن است اصول اخلاقی متفاوتی مورد نیاز باشد. در حال حاضر، مطالعه و بحث در مورد اخلاق هوش مصنوعی در زمینه‌های مختلف کاربردی خاص به ندرت در طول مطالعه ادبیات ما دیده می‌شود.

بنابراین، بسیار مهم و ضروری است که اصول اخلاقی اساسی و مشترک هوش مصنوعی از طریق بحث و همکاری بین سازمان‌ها، حوزه‌ها و دولت‌های مختلف به دست آید و به خوبی تثبیت شود. سپس بر اساس اصول اولیه و رایج، هر رشته می‌تواند این اصول را بیشتر ارتقا دهد تا به طور کلی در این زمینه خاص قابل اجرا باشد. شفاف سازی اصول و ارزش‌های اخلاقی که یک سیستم هوش مصنوعی باید با آنها مطابقت داشته باشد، پیش نیاز و اساس طراحی چنین سیستمی است که این الزامات را برآورده کند.

ب. چالش‌های پیاده سازی اخلاق در هوش مصنوعی

در پیاده سازی اخلاق در هوش مصنوعی، چالش‌های زیادی وجود دارد. این بخش به تحلیل چالش‌هایی می‌پردازد که ممکن است در عمل، با اتخاذ انواع مختلف نظریه‌های اخلاقی با آن مواجه شوند.

۹. چالش‌های اخلاق فضیلت در عمل

بر اساس اخلاق فضیلتی، اگر فاعل، فضیلتی را مصداق دهد، یعنی بر اساس برخی ارزش‌های اخلاقی عمل کند و بیندیشد، از نظر اخلاقی خوب است [۹۳]. نمی توان فقط با مشاهده یک عمل یا یک سری اقدامات که به نظر می‌رسد دلالت بر آن فضیلت دارد، قضاوت کرد که آیا یک سیستم یا عامل هوش مصنوعی فضیلت دارد یا نه، دلایل پشت این اعمال باید روشن شود، یعنی انگیزه‌های پشت این کارها روشن شود. اقدامات باید واضح باشد با این حال، انگیزه‌های پشت اَعمال سیستم‌های هوش مصنوعی، معمولاً برای ما نامشخص و ناشناخته هستند و کشف آن دشوار است. این مورد، چالش اصلی برای اجرای اخلاق فضیلت است. به‌علاوه، وقتی طراحی اخلاقی را بر اساس اخلاق فضیلت انجام می‌دهیم، اینکه سیستم هوش مصنوعی با

کدام ویژگی‌ها یا ویژگی‌های فضیلت هماهنگ می‌شود، سؤال دشواری است. حتی اگر صفات فضیلت به دقت انتخاب شده باشد، نحوه توصیف و اندازه گیری فضیلت همچنان یک کار چالش برانگیز است.

۲) چالش‌های اخلاق دئونتولوژیک در عمل

دئونتولوژیست‌ها یک عمل را در صورتی از نظر اخلاقی خوب می‌دانند که به برخی از قوانین یا وظایف، مقررات و هنجارهای اخلاقی پایبند باشد. اگرچه ماهیت مبتنی بر قاعده اخلاق deontological برای عمل، مناسب به نظر می‌رسد، چالش‌هایی در طول فرایند اجرا بوجود می‌آیند. اول اینکه کدام قواعد اخلاقی باید در طراحی اخلاقی اجرا شوند. دوم، ممکن است در برخی شرایط بین قوانین تضاد وجود داشته باشد. اگرچه دستور دادن یا سنجیدن قوانین اخلاقی ممکن است این مشکل را حل کند، تعیین ترتیب اهمیت قواعد اخلاقی مختلف اغلب دشوار است.

۳) چالش‌های اخلاق نتیجه گرایی در عمل

اخلاق نتیجه گرایانه اخلاقی بودن یک عمل را صرفاً بر اساس نتیجه آن ارزیابی می‌کند. دو چالش اصلی در طول اجرای اخلاق نتیجه گرا درگیر است. اولاً، تعیین پیامدهای یک اقدام یا تصمیم، دشوار است. برای سیستم هوش مصنوعی فعلی، عواقب احتمالی اقدامات سیستم، معمولاً از قبل با توجه به عدم شفافیت یا تفسیرپذیری مدل‌های هوش مصنوعی فعلی، به‌ویژه شبکه‌های عصبی مصنوعی، مشخص نیست. چالش دوم مربوط به کمّی کردن پیامدها است. از آنجایی که در اخلاق نتیجه گرا، هدف، به حداکثر رساندن مطلوبیت است، چگونگی تعریف و محاسبه‌ی مطلوبیت یک مشکل اساسی است.

۴) چالش‌های هماهنگی در بین استانداردهای اخلاقی مختلف

به دلیل تفاوت در فرهنگ، مذهب و سازمان، معیارهای اخلاقی نیز متفاوت است، حتی اگر در یک زمینه باشند. دستیابی به پیشنهاد استاندارد اخلاقی یکپارچه نه تنها دشوار است، بلکه غیر ضروری است. بنابراین، چگونگی دستیابی به هماهنگی بین استانداردهای اخلاقی کشورها و سازمان‌های مختلف مهم و به‌ویژه چالش‌برانگیز است.

ج. چالش در توسعه‌ی رویکردهای فناورانه برای کاهش مسائل اخلاقی هوش مصنوعی

در حال حاضر، بهبود توضیح پذیری، انصاف، حفاظت از حریم خصوصی، امنیت، استحکام و سایر شایستگی‌های مرتبط با الزامات هوش مصنوعی اخلاقی از موضوعات داغ تحقیقاتی در جوامع هوش مصنوعی هستند. با این حال، بیشتر کارهای تحقیقاتی کنونی از یک بعد واحد از اصول اخلاقی انجام می‌شوند، برای مثال، XAI بر افزایش تفسیرپذیری هوش مصنوعی تمرکز دارد و ML منصفانه به کاهش بی‌عدالتی یا سوگیری ML اختصاص دارد. هنوز عدم ادغام اصول یا الزامات اخلاقی متعدد در کار تحقیقاتی جاری وجود دارد. بدیهی است که ادغام چند بعد اخلاقی که تعادل هم افزایی بین چندین اصل اخلاقی مختلف را ممکن می‌سازد، برای

ساختن سیستم‌های هوش مصنوعی اخلاقی که می‌توانند اصول اخلاقی متعدد را برآورده کنند، ضروری و حیاتی است. اما ادغام چند بعد اخلاقی در یک سیستم هوش مصنوعی از طریق رویکردهای فناورانه به دلیل تضاد یا ناسازگاری بین الزامات اخلاقی مختلف بسیار چالش برانگیز است.

د. چالش در ارزیابی اخلاق در هوش مصنوعی

اخلاق ذاتاً یک مفهوم کیفی است که به بسیاری از ویژگی‌هایی که کمّی کردن آنها دشوار است، مانند ویژگی‌های فرهنگی یا نژادی وابسته است. از این رو، تعریف دقیق اخلاق، اگر نگوییم غیرممکن، بسیار دشوار است. در نتیجه، ارزیابی اخلاق هوش مصنوعی بسته به افرادی که هوش مصنوعی را ارزیابی می‌کنند، همیشه دارای عناصر ذهنی است. این امر تحقیقات و کاربردهای اخلاق هوش مصنوعی را با چالش‌هایی مواجه می‌کند.

ه. چشم‌اندازهای آینده

در این بخش به برخی دیدگاه‌های آینده اشاره می‌شود که ممکن است برای تحقیقات آتی ارزشمند باشند. ابتدا برای پیاده‌سازی اخلاق در هوش مصنوعی باید به این نکته اشاره کرد که انسان‌ها هرگز از یک نظریه اخلاقی استفاده نمی‌کنند، بلکه با توجه به موقعیت یا زمینه‌ای که با آن مواجه هستند، بین نظریه‌های مختلف جابه‌جا می‌شوند [۱۳۴]. این مسئله نه تنها به این دلیل است که انسان‌ها کارگزاران عقلانی محض نیستند که نظریه اقتصادی می‌خواهد آن‌ها را باور کنیم، بلکه به این دلیل است که پیروی دقیق از هر نظریه اخلاقی می‌تواند منجر به نتایج نامطلوب شود. این بدان معناست که سیستم‌های هوش مصنوعی باید دارای بازنمایی از نظریه‌های اخلاقی مختلف و توانایی انتخاب بین این نظریه‌های اخلاقی باشند. در اینجا ما این رویکرد را رویکرد چند نظریه‌ای می‌نامیم. در رویکرد چند تئوری، سیستم‌های هوش مصنوعی بسته به نوع موقعیت می‌توانند به جای یکدیگر نظریه‌های مختلفی را اعمال کنند. علاوه بر این، ترکیبی از تئوری‌های اخلاقی هنجاری و اخلاق دامنه‌ی خاص که توسط متخصصان این حوزه پذیرفته شده است، شایسته‌ی پیاده‌سازی است زیرا یک سیستم هوش مصنوعی اخلاقی باید توسط کاربرانش پذیرفته شود.

از نظر رویکردهای فن‌آوری برای پرداختن به مسائل اخلاقی در هوش مصنوعی، توسعه ML جدید و سایر فناوری‌های هوش مصنوعی تحت هدایت دستورالعمل‌های اخلاقی و اصول بررسی‌شده در بخش ۴، مطلوب است. اگرچه در نظر گرفتن چندین اصل اخلاقی مختلف به طور همزمان هنگام طراحی عوامل جدید هوش مصنوعی چالش برانگیز است، اما یک گام بسیار مهم و ضروری در توسعه هوش مصنوعی اخلاقی در آینده خواهد بود.

از بررسی رویکردهای ارزشیابی اخلاقی، می‌توان دریافت که روش‌های ارزیابی موثر به فوریت مورد نیاز است زیرا ما باید سیستم هوش مصنوعی طراحی شده را قبل از استقرار ارزیابی کنیم.

در حال حاضر، پیشنهاد یک روش ارزیابی کلی دشوار است. بنابراین، محققان اغلب بر حوزه‌های خاصی تمرکز می‌کردند و به وظایف ارزیابی شایستگی اخلاقی در این حوزه‌ها می‌پرداختند. معیارهای خاص دامنه، به عنوان مثال، مجموعه داده‌های جامع، برای آزمایش اخلاقی سیستم‌های هوش مصنوعی نیز برای برخی از زمینه‌های کاربردی حیاتی، مانند اتومبیل‌های خودران و مراقبت‌های بهداشتی مهم به نظر می‌رسند.

به عنوان آخرین مورد و نه کم اهمیت‌ترین آن، از آنجایی که هم طبیعت و هم تربیت، در شکل‌دهی رفتارهای اخلاقی مهم هستند، ما ترکیب اخلاق هنجاری و اخلاق تکاملی [۱۵۵] را برای طراحی سیستم‌های هوش مصنوعی اخلاقی پیشنهاد می‌کنیم. اخلاق هنجاری مانند توانایی‌های اخلاقی فطری است، در حالی که رویکرد اخلاق تکاملی می‌تواند با یادگیری و تکامل مستمر، شایستگی اخلاقی جدیدی کسب کند. این مسئله ممکن است یک مسیر امیدوارکننده برای توسعه سیستم هوش مصنوعی اخلاقی در آینده باشد.

نتیجه‌گیری

بر اساس بررسی ما از اخلاق هوش مصنوعی و پیچیدگی‌ها و چالش‌های فراوانی که در این مقاله توضیح داده شد، واضح است که تلاش برای پرداختن به مسائل اخلاقی در هوش مصنوعی و طراحی سیستم‌های هوش مصنوعی اخلاقی که قادر به رفتار اخلاقی باشند، کاری دشوار و پیچیده است. با این حال، اینکه آیا هوش مصنوعی می‌تواند نقش مهمی را در جامعه‌ی آینده ما ایفا کند تا حد زیادی به موفقیت سیستم‌های هوش مصنوعی اخلاقی بستگی دارد. نظم و انضباط اخلاق هوش مصنوعی مستلزم تلاش مشترک دانشمندان، مهندسان، فیلسوفان، کاربران و سیاستگذاران دولتی است.

این مقاله با خلاصه و تجزیه و تحلیل خطرات اخلاقی و مسائل مطرح شده توسط هوش مصنوعی، دستورالعمل‌ها و اصول اخلاقی صادر شده توسط سازمان‌های مختلف، رویکردهایی برای پرداختن به مسائل اخلاقی در هوش مصنوعی یا رعایت اصول اخلاقی هوش مصنوعی، و روش‌هایی برای ارزیابی اخلاق (یا معنویت) هوش مصنوعی و علاوه بر این موارد، به برخی از چالش‌ها در عمل اخلاق هوش مصنوعی و برخی از جهت‌گیری‌های تحقیقاتی آینده اشاره می‌شود.

با این حال، اخلاق هوش مصنوعی یک حوزه‌ی تحقیقاتی بسیار گسترده و چند رشته‌ای است. پوشش همه موضوعات ممکن در این زمینه با یک مقاله مروری غیرممکن است. امیدواریم این مقاله بتواند نقطه شروعی برای افرادی باشد که به اخلاق هوش مصنوعی علاقه‌مند هستند تا پیشینه‌ی کافی و دید پرنده‌ای به دست آورند تا تحقیقات بیشتر توسط آنها انجام شود.

منابع

[1] M. Haenlein and A. Kaplan, “A brief history of artificial intelligence: On the past, present, and future of artificial intelligence,” California Manage. Rev., vol. 61, no. 4, pp. 5–14, 2019.

[2] R. Vinuesa et al., “The role of artificial intelligence in achieving the sustainable development goals,” Nature Commun., vol. 11, no. 1, 2020, Art. No. 233.HUANG et al.: OVERVIEW OF ARTIFICIAL INTELLIGENCE ETHICS 817.

[3] Gartner, “Chatbots will appeal to modern workers,” 2019. Accessed: Feb. 10, 2022. [Online]. Available: https://www.gartner.com/smarterwithgartner/chatbots-will-appeal-to-modern-workers

[4] M. J. Haleem, R. P. Singh, and R. Suman, “Telemedicine for healthcare: Capabilities, features, barriers, and applications,” Sensors Int., vol. 2, 2021, Art. No. 100117.

[5] A. Morby, “Tesla driver killed in first fatal crash using autopilot,” 2016. Accessed: Feb. 10, 2022. [Online]. Available: https://www.dezeen.com/2016/07/01/tesla-driver-killed-car-crashnews-driverless-car-autopilot/

[6] S. McGregor, Ed., “Incident number 6,” in AI Incident Database, 2016. [Online]. Available: https://incidentdatabase.ai/cite/6

[7] R. V. Yampolskiy, “Predicting future AI failures from historic examples,” Foresight, vol. 21, no. 1, pp. 138–152, 2019.

[8] C. Stupp, “Fraudsters used AI to mimic CEO’s voice in unusual cybercrime case: Scams using artificial intelligence are a new challenge for companies,” 2019. Accessed: Feb. 10, 2022. [Online]. Available: https://www.wsj.com/articles/fraudsters-use-ai-tomimic-ceos-voice-in-unusual-cybercrime-case-11567157402

[9] C. Allen, W. Wallach, and I. Smit, “Why machine ethics?,” IEEE Intell. Syst., vol. 21, no. 4, pp. 12–17, Jul./Aug. 2006.

[10] M. Anderson and S. L. Anderson, “Machine ethics: Creating an ethical intelligent agent,” AI Mag., vol. 28, no. 4, pp. 15–26, 2007.

[11] K. Siau and W. Wang, “Artificial intelligence (AI) ethics,” J. Database Manage., vol. 31, no. 2, pp. 74–87, 2020.

[12] A. Jobin, M. Ienca, and E. Vayena, “The global landscape of AI ethics guidelines,” Nature Mach. Intell., vol. 1, no. 9, pp. 389–399, 2019.

[13] M. Ryan and B. C. Stahl, “Artificial intelligence ethics guidelines for developers and users: Clarifying their content and normative implications,” JICES, vol. 19, no. 1, pp. 61–86, 2021.

[14] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, “A survey on bias and fairness in machine learning,” ACM Comput. Surv., vol. 54, no. 6, pp. 1–35, 2021.

[15] J. García and F. Fernández, “A comprehensive survey on safe reinforcement learning,” J. Mach. Learn. Res., vol. 16, no. 42, pp. 1437–1480, 2015.

[16] V. Mothukuri, R. M. Parizi, S. Pouriyeh, Y. Huang, A. Dehghantanha, and G. Srivastava, "A survey on security and privacy of federated learning," Future Gener. Comput. Syst., vol. 115, pp. 619–640, 2021.
[17] X. Liu et al., "Privacy and security issues in deep learning: A survey," IEEE Access, vol. 9, pp. 4566–4593, 2021.
[18] B. Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, pp. 82–115, 2020.
[19] Y. Zhang, M. Wu, G. Y. Tian, G. Zhang, and J. Lu, "Ethics and privacy of artificial intelligence: Understandings from bibliometrics," Knowl. – Based Syst., vol. 222, 2021, Art. No. 106994.
[20] D. Castelvecchi, "Can we open the black box of AI?," Nature, vol. 538, no. 7623, pp. 20–23, 2016.
[21] S. Dilmaghani, M. R. Brust, G. Danoy, N. Cassagnes, J. Pecero, and P. Bouvry, "Privacy and security of big data in AI systems: A research and standards perspective," in Proc. IEEE Int. Conf. Big Data, 2019, pp. 5737–5743.
[22] J. P. Sullins, "When is a robot a moral agent?," in Machine Ethics, M. Anderson and S. L. Anderson, Eds., Cambridge, U.K.: Cambridge Univ. Press, 2011, pp. 151–161.
[23] J. Timmermans, B. C. Stahl, V. Ikonen, and E. Bozdag, "The ethics of cloud computing: A conceptual review," in Proc. IEEE 2nd Int. Conf. Cloud Comput. Technol. Sci., 2010, pp. 614–620.
[24] W. Wang and K. Siau, "Ethical and moral issues with AI: A case study on healthcare robots," in Proc. 24th Americas Conf. Inf. Syst., 2018, pp. 1–5.
[25] I. Bantekas and L. Oette, International Human Rights Law and Practice. Cambridge U. K.: Cambridge Univ. Press, 2018.
[26] R. Rodrigues, "Legal and human rights issues of AI: Gaps, challenges and vulnerabilities," J. Responsible Technol., vol. 4, 2020, Art. No. 100005.
[27] W. Wang and K. Siau, "Artificial intelligence, machine learning, automation, robotics, future of work and future of humanity: A review and research agenda," J. Database Manage., vol. 30, no. 1, pp. 61–79, 2019.
[28] W. Wang and K. Siau, "Industry 4.0: Ethical and moral predicaments," Cutter Bus. Technol. J., vol. 32, no. 6, pp. 36–45, 2019.
[29] S. M. Liao, Ed., Ethics of Artificial Intelligence. New York, NY, USA: Oxford Univ. Press, 2020.
[30] A. Adadi, "A survey on data-efficient algorithms in big data era," J. Big Data, vol. 8, no. 1, pp. 1–54, 2021.
[31] R. S. Geiger et al., "Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled

training data comes from?," in Proc. Conf. Fairness, Accountability, Transparency, 2020, pp. 325–336.
[32] W. M. P. van der Aalst, V. Rubin, H. M. W. Verbeek, B. F. van Dongen, E. Kindler, and C. W. Günther, "Process mining: A two-step approach to balance between underfitting and overfitting," Softw. Syst. Model., vol. 9, no. 1, pp. 87–111, 2010.
[33] Z. C. Lipton, "The mythos of model interpretability," Queue, vol. 16, no. 3, pp. 31–57, 2018.
[34] Y. Wang and M. Kosinski, "Deep neural networks are more accurate than humans at detecting sexual orientation from facial images," J. Pers. Social Psychol., vol. 114, no. 2, pp. 246–257, 2018.
[35] D. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in Proc. IEEE Int. Conf. Adv. Video Signal-based Surveill., 2018, pp. 1–6.
[36] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to 110icrosoft110ation?," Technological Forecasting Social Change, vol. 114, pp. 254–280, 2017.
[37] R. Maines, "Love + sex with robots: The evolution of human-robot relationships (Levy, D.; 2007) [Book review]," IEEE Technol. Soc. Mag., vol. 27, no. 4, pp. 10–12, Dec. 2008.
[38] National AI Standardization General, "Artificial intelligence ethical risk analysis report," 2019. Accessed: Apr. 19, 2022. [Online]. Available: http://www.cesi.cn/201904/5036.html
[39] A. Hannun, C. Guo, and L. van der Maaten, "Measuring data leakage in machine-learning models with fisher information," in Proc. 37th Conf. Uncertainty Artif. Intell., 2021, pp. 760–770.
[40] A. Salem, M. Backes, and Y. Zhang, "Get a model! Model hijacking attack against machine learning models," Nov. 2021. [Online]. Available: https://arxiv.org/pdf/2111.04394
[41] A. Pereira and C. Thomas, "Challenges of machine learning applied to safety-critical cyber-physical systems," MAKE, vol. 2, no. 4, pp. 579–602, 2020.
[42] J. A. McDermid, Y. Jia, Z. Porter, and I. Habli, "Artificial intelligence explainability: The technical and ethical dimensions," Philos. Trans.. Ser. A, Math. Phys. Eng. Sci., vol. 379, no. 2207, 2021, Art. No. 20200363.
[43] J.-F. Bonnefon, A. Shariff, and I. Rahwan, "The social dilemma of autonomous vehicles," Science, vol. 352, no. 6293, pp. 1573–1576, 2016.
[44] B. C. Stahl and D. Wright, "Ethics and privacy in AI and big data: Implementing responsible research and innovation," IEEE Secur. Privacy, vol. 16, no. 3, pp. 26–33, May/Jun. 2018.
[45] S. Ribaric, A. Ariyaeeinia, and N. Pavesic, "De-identification for privacy protection in multimedia content: A survey," Signal Process., Image Commun., vol. 47, pp. 131–151, 2016.

[46] A. Julia, L. Jeff, M. Surya, and K. Lauren, “Machine bias: There’s software used across the country to predict future criminals. And it’s biased against blacks,” 2016. Accessed: Apr. 19, 2022. [Online]. Available: https://www.propublica.org/article/machine-bias-riskassessments-in-criminal-sentencing
[47] J. Dastin, “Amazon scraps secret AI recruiting tool that showed bias against women,” 2018. Accessed: Apr. 19, 2022. [Online]. Available: https://www.reuters.com/article/us-amazon-com-jobs-automationinsight/amazon-scraps-secret-ai-recruiting-tool-that-showed-biasagainst-women-idUSKCN1MK08G
[48] D. Castelvecchi, “AI pioneer: ‘The dangers of abuse are very real’,” Nature, Apr. 4, 2019, [Online]. Available: https://www.nature.com/articles/ d41586-019-00505-2
[49] K. Hristov, “Artificial intelligence and the copyright dilemma,” IDEA, IP Law Rev., vol. 57, 2017, Art. No. 3. [Online]. Available: https://ssrn. Com/abstract=2976428
[50] C. Bartneck, C. Lütge, A. Wagner, and S. Welsh, “Responsibility and liability in the case of AI systems,” in An Introduction to Ethics in Robotics and AI (SpringerBriefs in Ethics), C. Bartneck, C. Lütge, A. Wagner, and S. Welsh, Eds., Cham, Switzerland: Springer, 2021, pp. 39–44.
[51] E. Bird, J. Fox-Skelly, N. Jenner, R. Larbey, E. Weitkamp, and A. Winfield, “The ethics of artificial intelligence: Issues and initiatives,” European Parliamentary Research Service, Brussels, Belgium, 2020. Accessed: Apr. 19, 2022. [Online]. Available: https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)634452
[52] C. Lutz, “Digital inequalities in the age of artificial intelligence and big data,” Hum. Behav. Emerg. Technol., vol. 1, no. 2, pp. 141–148, 2019.
[53] L. Manikonda, A. Deotale, and S. Kambhampati, “What’s up with Privacy? User preferences and privacy concerns in intelligent personal assistants,” in Proc. AAAI/ACM Conf. AI, Ethics Soc., 2018, pp. 229–235.818 IEEE TRANSACTIONS ON ARTIFICIAL INTELLIGENCE, VOL. 4, NO. 4, AUGUST 2023
[54] D. Roselli, J. Matthews, and N. Talagala, “Managing bias in AI,” in Proc. World Wide Web Conf., 2019, pp. 539–544.
[55] Y. Gorodnichenko, T. Pham, and O. Talavera, “Social media, sentiment and public opinions:Evidencefrom #Brexit and #USElection,” Eur.Econ. Rev., vol. 136, Jul. 2021, Art. No. 103772.
[56] N. Thurman, “Making ‘The daily me’: Technology, economics and habit in the mainstream assimilation of personalized news,” Journalism, vol. 12, no. 4, pp. 395–415, 2011.

[57] J. Donath, “Ethical issues in our relationship with artificial entities,” in The Oxford Handbook of Ethics of AI. M. D. Dubber, F. Pasquale, and S. Das, Eds., Oxford, U.K.: Oxford Univ. Press, 2020, pp. 51–73.
[58] E. Magrani, “New perspectives on ethics and the laws of artificial intelligence,” Internet Policy Rev., vol. 8, 2019, Art. No. 3.
[59] M. P. Wellman and U. Rajan, “Ethical issues for autonomous trading agents,” Minds Mach., vol. 27, no. 4, pp. 609–624, 2017.
[60] U. Pagallo, “The impact of AI on criminal law, and its two fold procedures,” in Research Handbook on the Law of Artificial Intelligence, W. Barfield and U. Pagallo, Eds., Cheltenham U.K.: Edward Elgar Publishing, 2018, pp. 385–409.
[61] E. Dacoronia, “Tort law and new technologies,” in Legal Challenges in the New Digital Age, A. M. López Rodríguez, M. D. Green, and M. L. Kubica, Eds., Leiden, The Netherlands: Koninklijke Brill NV, 2021, pp. 3–12.
[62] J. Khakurel, B. Penzenstadler, J. Porras, A. Knutas, and W. Zhang, “The rise of artificial intelligence under the lens of sustainability,” Technologies, vol. 6, no. 4, 2018, Art. No. 100.
[63] S. Herat, “Sustainable management of electronic waste (e-Waste),” Clean Soil Air Water, vol. 35, no. 4, pp. 305–310, 2007.
[64] E. Strubell, A. Ganesh, and A. McCallum, “Energy and policy considerations for deep learning in NLP,” in Proc. 57th Annu. Meeting Assoc. Comput. Linguistics, 2019, pp. 3645–3650.
[65] V. Dignum, “Ethics in artificial intelligence: Introduction to the special issue,” Ethics Inf. Technol., vol. 20, no. 1, pp. 1–3, 2018.
[66] S. Corbett-Davies, E. Pierson, A. Feller, S. Goel, and A. Huq, “Algorithmic decision making and the cost of fairness,” in Proc. 23rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining, 2017, pp. 797–806.
[67] R. Caplan, J. Donovan, L. Hanson, and J. Matthews, “Algorithmic accountability: A primer,” Data Soc., vol. 18, pp. 1–13, 2018.
[68] R. V. Yampolskiy, “On controllability of AI,” Jul. 2020. [Online]. Available: https://arxiv.org/pdf/2008.04071
[69] B. C. Stahl, J. Timmermans, and C. Flick, “Ethics of emerging information and communication technologies,” Sci. Public Policy, vol. 44, no. 3, pp. 369–381, 2017.
[70] L. Vesnic-Alujevic, S. Nascimento, and A. Pólvora, “Societal and ethical impacts of artificial intelligence: Critical notes on European policy frameworks,” Telecommun. Policy, vol. 44, no. 6, 2020, Art. No. 101961.
[71] U. G. Assembly, “Universal declaration of human rights,” UN Gen. Assem., vol. 302, no. 2, pp. 14–25, 1948.
[72] S. Russell, S. Hauert, R. Altman, and M. Veloso, “Robotics: Ethics of artificial intelligence,” Nature, vol. 521, no. 7553, pp. 415–418, 2015.

[73] A. Chouldechova, “Fair prediction with disparate impact: A study of bias in recidivism prediction instruments,” Big Data, vol. 5, no. 2, pp. 153–163, 2017.
[74] J. van Dijck, “Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology,” Surveill. Soc., vol. 12, no. 2, pp. 197–208, 2014.
[75] E. de Souza Nascimento, I. Ahmed, E. Oliveira, M. P. Palheta, I. Steinmacher, and T. Conte, “Understanding development process of machine learning systems: Challenges and solutions,” in Proc. ACM/IEEE Int. Symp. Empirical Softw. Eng. Meas., 2019, pp. 1–6.
[76] K. A. Crockett, L. Gerber, A. Latham, and E. Colyer, “Building trustworthy AI solutions: A case for practical solutions for small businesses,” IEEE Trans. Artif. Intell., early access, 2021, doi: 10.1109/TAI.2021.3137091.
[77] D. Leslie, “Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector,” 2019. Accessed: Apr. 19, 2022. [Online]. Available: https://www.turing.ac.uk/research/publications/understanding-artificial-intelligence-ethics-and-safety
[78] B. Buruk, P. E. Ekmekci, and B. Arda, “A critical perspective on guidelines for responsible and trustworthy artificial intelligence,” Med. Health Care Philosophy, vol. 23, no. 3, pp. 387–399, 2020.
[79] UNESCO, “Recommendation on the ethics of artificial intelligence,” 2021. Accessed: Feb. 15 2022. [Online]. Available: https://en.unesco.org/artificial-intelligence/ethics
[80] B. C. Stahl, Ed., Artificial Intelligence for a Better Future: An Ecosystem Perspective on the Ethics of AI and Emerging Digital Technologies. Cham, Switzerland: Springer, 2021.
[81] P. D. Motloba, “Non-maleficence – A disremembered moral obligation,” South Afr. Dent. J., vol. 74, 2019, Art. No. 1.
[82] L. Floridi and J. Cowls, “A unified framework of five principles for AI in society,” in Ethics, Governance, and Policies in Artificial Intelligence (Philosophical Studies Series), vol. 144, L. Floridi, Ed. Cham, Switzerland: Springer, 2021, pp. 5–17.
[83] S. Jain, M. Luthra, S. Sharma, and M. Fatima, “Trustworthiness of artificial intelligence,” in Proc. 6th Int. Conf. Adv. Comput. Commun. Syst., 2020, pp. 907–912.
[84] L. Floridi et al., “AI4People-An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations,” Minds Mach., vol. 28, no. 4, pp. 689–707, 2018.
[85] R. Nishant, M. Kennedy, and J. Corbett, “Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda,” Int. J. Inf. Manage., vol. 53, 2020, Art. No. 102104.
[86] C. S. Wickramasinghe, D. L. Marino, J. Grandio, and M. Manic, “Trustworthy AI development guidelines for human system

interaction," in Proc. 13th Int. Conf. Hum. Syst. Interaction, 2020, pp. 130–136.
[87] V. Dignum, "Can AI systems be ethical?," in Artificial Intelligence: Foundations Theory and Algorithms, Responsible Artificial Intelligence. V. Dignum, Ed., Cham, Switzerland: Springer, 2019, pp. 71–92.
[88] S. L. Anderson and M. Anderson, "AI and ethics," AI Ethics, vol. 1, no. 1, pp. 27–31, 2021.
[89] V. Dignum, "Ethical decision-making," in Artificial Intelligence: Foundations, Theory, and Algorithms, Responsible Artificial Intelligence. V. Dignum, Ed., Cham, Switzerland: Springer, 2019, pp. 35–46.
[90] G. Sayre-McCord, "Metaethics," in The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed., Stanford, CA, USA: Metaphys. Res. Lab, Stanford Univ., 2014. [Online]. Available: https://plato.stanford.edu/entries/metaethics/#:~:text=Metaethics%20is%20the%20attempt%20to,matter%20of%20taste%20than%20truth%3F
[91] Ethics | Internet Encyclopedia of Philosophy, 1995. Accessed: Aug. 2, 2021. [Online]. Available: https://iep.utm.edu/ethics/#SH2c
[92] R. Hursthouse and G. Pettigrove, "Virtue ethics," in The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed. Stanford, CA, USA: Metaphys. Res. Lab, Stanford Univ., 2018. [Online]. Available: https://plato. 114icrosof.edu/entries/ethics-virtue/
[93] N. Cointe, G. Bonnet, and O. Boissier, "Ethical judgment of agents' behaviors in multi-agent systems," in Proc. Int. Conf. Auton. Agents Multiagent Syst., 2016, pp. 1106–1114.
[94] H. Yu, Z. Shen, C. Miao, C. Leung, V. R. Lesser, and Q. Yang, "Building ethics into artificial intelligence," in Proc. 27th Int. Joint Conf. Artif. Intell., 2018, pp. 5527–5533.
[95] [95] H. J. Curzer, Aristotle and the Virtues. New York, NY, USA: Oxford Univ. Press, 2012.
[96] L. Alexander and M. Moore, "Deontological ethics," in The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed., 2020, Stanford, CA, USA: Metaphys. Res. Lab, Stanford Univ., 2020. [Online]. Available: https://plato.stanford.edu/entries/ethics-deontological/
[97] W. Sinnott-Armstrong, "Consequentialism," in The Stanford Encyclopedia of Philosophy, E. N. Zalta, Ed., 2019. Stanford, CA, USA: Metaphys. Res. Lab, Stanford Univ., [Online]. Available: https://plato.stanford.edu/entries/consequentialism/
[98] D. O. Brink, "Some forms and limits of consequentialism," in Oxford Handbooks in Philosophy, The Oxford Handbook of Ethical Theory. D. Copp, Ed., New York, NY, USA: Oxford Univ. Press, 2006, pp. 380–423.

[99] H. A. M. J. ten Have, Ed., Encyclopedia of Global Bioethics. Cham, Switzerland: Springer, 2016.
[100] S. Tolmeijer, M. Kneer, C. Sarasua, M. Christen, and A. Bernstein, "Implementations in machine ethics: A survey," ACM Comput. Surv., vol. 53, no. 6, pp. 1–38, 2021.
[101] C. Allen, I. Smit, and W. Wallach, "Artificial morality: Top-down, bottom-up, and hybrid approaches," Ethics Inf. Technol., vol. 7, no. 3, pp. 149–155, 2005.
[102] W. Wallach and C. Allen, "Top-down morality," in Moral Machines, W. Wallach and C. Allen, Eds., Oxford, U.K: Oxford Univ. Press, 2009, pp. 83–98.
[103] I. Asimov, "Runaround," Astounding Sci. Fiction, vol. 29, no. 1, pp. 94–103, 1942.
[104] J.-G. Ganascia, "Ethical system formalization using nonmonotonic ogics," in Proc. Annu. Meeting Cogn. Sci. Soc., 2007, pp. 1013–1018.HUANG et al.: OVERVIEW OF ARTIFICIAL INTELLIGENCE ETHICS 819
[105] K. Arkoudas, S. Bringsjord, and P. Bello, "Toward ethical robots via mechanized deontic logic," in Proc. AAAI Fall Symp. Mach. Ethics, 2005, pp. 17–23.
[106] S. Bringsjord and J. Taylor, "Introducing divine-command robot ethics," in Robot Ethics: The Ethical and Social Implication of Robotics. 2012, pp. 85–108.
[107] N. S. Govindarajulu and S. Bringsjord, "On automating the doctrine of double effect," in Proc. 26th Int. Joint Conf. Artif. Intell., 2017, pp. 4722–4730.
[108] F. Berreby, G. Bourgne, and J.-G. Ganascia, "A declarative modular framework for representing and applying ethical principles," in Proc. 16th Conf. Auton. Agents MultiAgent Syst., 2017, pp. 96–104.
[109] V. Bonnemains, C. Saurel, and C. Tessier, "Embedded ethics: Some echnical and ethical challenges," Ethics Inf. Technol., vol. 20, no. 1, pp. 41–58, 2018.
[110] G. S. Reed, M. D. Petty, N. J. Jones, A. W. Morris, J. P. Ballenger, and H. S. Delugach, "A principles-based model of ethical considerations in military decision making," J. Defense Model. Simul., vol. 13, no. 2, pp. 195–211, 2016.
[111] L. Dennis, M. Fisher, M. Slavkovik, and M.Webster, "Formal verification of ethical choices in autonomous systems," Robot. Auton. Syst., vol. 77, pp. 1–14, 2016.
[112] A. R. Honarvar and N. Ghasem-Aghaee, "Casuist BDI-Agent: A new extended BDI architecture with the capability of ethical reasoning," in Proc. Int. Conf. Artif. Intell. Comput. Intell., 2009, pp. 86–95.

[113] A. S. Rao and M. P. Georgeff, “BDI agents: From theory to practice,” in Proc. 1st Int. Conf. Multiagent Syst., 1995, pp. 312–319.
[114] S. Armstrong, “Motivated value selection for artificial agents,” in Proc. AAAI Workshop Artif. Intell. Ethics, Jan. 2015, pp. 12–20.
[115] U. Furbach, C. Schon, and F. Stolzenburg, “Automated reasoning in deontic logic,” in Proc. 8th Int. Workshop Multi-Disciplinary Trends Artif. Intell., 2014, pp. 57–68.
[116] D. Howard and I. Muntean, “Artificial moral cognition: Moral functionalism and autonomous moral agency,” in Philosophical Studies Series, Philosophy and Computing, T. M. Powers, Ed. Cham, Switzerland: Springer, 2017, pp. 121–159.
[117] Y.-H. Wu and S.-D. Lin, “A low-cost ethics shaping approach for designing reinforcement learning agents,” in Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 1687–1694.
[118] R. Noothigattu et al., “A voting-based system for ethical decision making,” in Proc. 32nd AAAI Conf. Artif. Intell., 2018, pp. 1587–1594.
[119] M. Guarini, “Particularism and the classification and reclassification of moral cases,” IEEE Intell. Syst., vol. 21, no. 4, pp. 22–28, Jul./Aug. 2006.
[120] M. Anderson and S. L. Anderson, “GenEth: A general ethical dilemma analyzer,” in Proc. 28th AAAI Conf. Artif. Intell., 2014, pp. 253–261.
[121] M. Azad-Manjiri, “A new architecture for making moral agents based on C4.5 decision tree algorithm,” Int. J. Inf. Technol. Comput. Sci., vol. 6, no. 5, pp. 50–57, 2014.
[122] L. Yilmaz, A. Franco-Watkins, and T. S. Kroecker, “Computational models of ethical decision-making: A coherence-driven reflective equilibrium model,” Cogn. Syst. Res., vol. 46, pp. 61–74, 2017.
[123] T. A. Han, A. Saptawijaya, and L. M. Pereira, “Moral reasoning under uncertainty,” in Logic For Programming, Artificial Intelligence, and Reasoning. Berlin, Germany: Springer, 2012, pp. 212–227.
[124] M. Anderson, S. Anderson, and C. Armen, “Towards machine ethics implementing two action-based ethical theories,” in Proc. AAAI Fall Symp. Mach. Ethics, 2005, pp. 1–7.
[125] G. Gigerenzer, “Moral satisficing: Rethinking moral behavior as bounded rationality,” Topics Cogn. Sci., vol. 2, no. 3, pp. 528–554, 2010.
[126] J. Skorin-Kapov, “Ethical positions and decision-making,” in Professional and Business Ethics Through Film, J. Skorin-Kapov, Ed., New York, NY, USA: Springer, 2018, pp. 19–54.
[127] T.-L. Gu and L. Li, “Artificialmoral agents and their designmethodology: Retrospect and prospect,” Chin. J. Comput., vol. 44, pp. 632–651, 2021.

[128] C.Molnar, G. Casalicchio, and B. Bischl, “Interpretablemachine learning – A brief history, state-of-the-art and challenges,” in Proc. Joint Eur. Conf. Mach. Learn. Knowl. Discov. Databases, 2020, pp. 417–431.
[129] C. Molnar, Interpretable Machine Learning: A Guide For Making Black Box Models Interpretable. Morisville, NC, USA: Lulu, 2019.
[130] S. Feuerriegel, M. Dolata, and G. Schwabe, “Fair AI,” Bus. Inf. Syst. Eng., vol. 62, no. 4, pp. 379–384, 2020.
[131] S. Caton and C. Haas, “Fairness in machine learning: A survey,” Oct. 2020. [Online]. Available: https://arxiv.org/pdf/2010.04053
[132] S. E. Whang, K. H. Tae, Y. Roh, and G. Heo, “Responsible AI challenges n end-to-end machine learning,” Jan. 2021. [Online]. Available: https://arxiv.org/pdf/2101.05967
[133] D. Peters, K. Vold, D. Robinson, and R. A. Calvo, “Responsible AI—Two frameworks for ethical design practice,” IEEE Trans. Technol. Soc., vol. 1, no. 1, pp. 34–47, Mar. 2020.
[134] V. Dignum, ed., Responsible Artificial Intelligence. Cham, Switzerland: Springer, 2019.
[135] C. Dwork, “Differential privacy: A survey of results,” in Proc. Int. Conf. Theory Appl. Models Computation, 2008, pp. 1–19.
[136] Q. Yang, Y. Liu, Y. Cheng, Y. Kang, T. Chen, and H. Yu, “Federated learning,” Synth. Lectures Artif. Intell. Mach. Learn., vol. 13, no. 3, pp. 1–207, 2019.
[137] M. Kirienko et al., “Distributed learning: A reliable privacy-preserving strategy to change multicenter collaborations using AI,” Eur. J. Nucl. Med. Mol. Imag., vol. 48, no. 12, pp. 3791–3804.2021.
[138] R. Shokri and V. Shmatikov, “Privacy-preserving deep learning,” in Proc. 22nd ACM SIGSAC Conf. Comput. Commun. Secur., 2015, pp. 1310–1321.
[139] C. Meurisch, B. Bayrak, and M. Mühlhäuser, “Privacy-preserving AI services through data decentralization,” in Proc. Web Conf., 2020, pp. 190–200.
[140] UR-Lex – 02016R0679-20160504 – EN – EUR-Lex, 2016. Accessed: Jun. 28, 2021. [Online]. Available: https://eur-lex.europa.eu/legalcontent/EN/TXT/?uri=CELEX%3A02016R0679-20160504&qid=1532348683434
[141] R. E. Latta, H.R.3388 – 115th Congress (2017-2018): SELF DRIVE Act, 2017. Accessed: Jun. 28, 2021. [Online]. Available: https://www.congress.gov/bill/115th-congress/house-bill/3388
[142] 7. Lei No. 13, de 14 de Agosto de 2018, 2018. Accessed: Jun. 25, 2021. [Online]. Available: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.html
[143] EUR-Lex – 52021PC0206 – EN – EUR-Lex, 2021. Accessed: Jun. 28, 2021. [Online]. Available: https://eur-lex.europa.eu/legal-

content/EN/TXT/?qid=1623335154975&uri=CELEX%3A52021PC0206
[144] C. Allen, G. Varner, and J. Zinser, “Prolegomena to any future artificial moral agent,” J. Exp. Theor. Artif. Intell., vol. 12, no. 3, pp. 251–261, 2000.
[145] A. M. Turing, “Computing machinery and intelligence,” Mind, vol. LIX, no. 236, pp. 433–460, 1950.
[146] W. Wallach and C. Allen, Moral Machines: Teaching Robots Right From Wrong. Oxford, U.K.: Oxford Univ. Press, 2009.
[147] S. A. Seshia, D. Sadigh, and S. S. Sastry, “Towards verified artificial intelligence,” Jun. 2016. [Online]. Available: http://arxiv.org/pdf/1606.08514v4
[148] T. Arnold and M. Scheutz, “Against the moral turing test: Accountable design and the moral reasoning of autonomous systems,” Ethics Inf. Technol., vol. 18, no. 2, pp. 103–115, 2016.
[149] ACM Code of Ethics and Professional Conduct, 2018. Accessed: Jun. 25, 2021. [Online]. Available: https://www.acm.org/code-of-ethics
[150] IEEE SA – The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2019. Accessed: Jun. 28 2021. [Online]. Available: https://standards.ieee.org/industry-connections/ec/autonomous-systems.html
[151] IEEE Ethics In Action | Ethically Aligned Design, IEEE 7000TM Projects, 2020. Accessed: Jun. 28, 2021. [Online]. Available: https://ethicsinaction.ieee.org/p7000/
[152] ISO, ISO/IEC JTC 1/SC 42 – Artificial intelligence, 2017. Accessed: Jun. 28, 2021. [Online]. Available: https://www.iso.org/committee/6794475.html
[153] B. Goehring, F. Rossi, and D. Zaharchuk, “Advancing AI ethics beyond compliance: From principles to practice,” IBM Corporation, Apr. 2020. Accessed: Apr. 19, 2022. [Online]. Available: https://www.ibm.com/thought-leadership/institute-business-value/report/ai-ethics
[154] Responsible AI, 2017. Accessed: Apr. 19, 2022. [Online]. Available: https://www.microsoft.com/en-us/ai/responsible-ai?activetab=pivot1:primaryr6
[155] F. Allhoff, “Evolutionary ethics from Darwin to Moore,” Hist. Philosophy Life Sci., vol. 25, no. 1, pp. 51–79, 2003.

جایگاه (نقش) اخلاق حرفه‌ای معلمان در آموزش مجازی مکاتب

سیروس قنبری و احمد عزیزی
ترجمه: سید محمدشریف شاکر[1]

چکیده

افزایش کیفیت آموزش مجازی یکی از اساسی‌ترین دغدغه‌ی سیستم آموزشی در اکثر کشورهای جهان است. از آنجایی‌که آموزش مجازی در مراکز آموزشی از جایگاه ویژه‌ای برخوردار است، برای رشد و ارتقای آن باید اخلاق حرفه‌ای معلمان مورد توجه قرار گیرد. با درنظرداشت اهمیت این موضوع، هدف این تحقیق، بررسی جایگاه اخلاق حرفه‌ای معلمان در آموزش مجازی مکاتب است.

روش تحقیق به‌کار رفته در این تحقیق کتابخانه‌ای (مروری) بوده و برای دستیابی به هدف تحقیق، علاوه بر کتاب‌های آموزش‌های الکترونیکی و مجازی در این زمینه، مقالات مرتبط با کلیدواژه‌های تحقیق طی سال‌های 2004 تا 2022 از پایگاه‌های معتبر سیویلیکا (Civilica)، مگ‌ایران (Magiran)، سید(Sid)، ریسرچ‌گیت (ResearchGate) و ساینس‌دایرکت (Science Direct) نیز مورد مطالعه قرار گرفته است.

نتایج نشان‌داد که بهبود آموزش مجازی در مکاتب، علاوه بر زیرساخت‌های فناوری اطلاعات و ارتباطات، توانایی معلمان و دانش‌آموزان در استفاده از فناوری‌های نوین و تعامل معلمان و دانش‌آموزان، مستلزم اخلاق حرفه‌ای معلمان و رعایت اخلاق تدریس نیز می‌باشد. بنابراین، یکی از عواملی که می‌تواند در آموزش مجازی مکاتب تأثیرگذار باشد، اخلاق حرفه‌ای معلمان است. بر این اساس مباحث اخلاقی تدریس و مفاهیم آن باید جزء موضوعات و عناوین آموزشی دانشجویان در دانشگاه‌های علوم ترتیبی و آموزش‌های داخل‌خدمت معلمان در نظر گرفته شود. فناوری‌های نوین در آموزش و پرورش در زمان استفاده باید مورد توجه و تأکید جدی قرار گیرد و تلاش شود تا اخلاق حرفه‌ای تدریس مجازی در بین معلمان نهادینه شود.

کلید واژه‌ها: اخلاق حرفه‌ای معلمان، آموزش مجازی، معلمان.

[1] s.shakir”knu.edu.af
+93 780233275

مقدمه

کشورها و جوامع پیشرفته در همه ابعاد زندگي سياسي، اقتصادي و اجتماعي خود به مؤسسات و انجمن‌های آموزشي به عنوان راه‌گشاي اصلي رفاه، پيشرفت و موفقيت در كليه مقاطع تحصيلي (به ويژه در مكاتب)، خط‌مشی آموزشي، آماده‌سازی معلمان و برنامه‌های آموزشی توجه خاصی دارند. زیرا؛ نصاب درسی، تکنالوژی، محیط درونی و بیرونی، نظام ارزشی، اخلاق، مدیران و رهبران آموزش‌دیده بر کیفیت آموزش و دستیابی به اهداف تعلیمی تأثیرگذار است. از سوی دیگر، با پیشرفت چشمگیر تکنالوژی، تمامی جنبه‌های زندگی بشر توسط فناوری اطلاعات و ارتباطات متحول شده است[۱].

در نتیجهٔ این تغییرات و پیشرفت، مراکز آموزشی با پدیده‌ی جدیدی به نام فناوری آموزشی مواجه شده اند که آموزش را از حالت سنتی خارج کرده و به آن جریان خاصی بخشیده است[۲]. استفاده از فناوری در حوزه‌ی آموزش با شروع همه‌گیری کرونا و بسته‌شدن مراکز آموزشی در سال ۲۰۲۰ اهمیتی دوچندان یافت و نقش معلمان و خانواده‌ها در آموزش کودکان و نوجوانان دچار دگرگونی بی‌سابقه گردید. در چنین شرایطی، والدین و مربیان کودک بیش از هرزمان دیگر درگیرِ فراگیری آموزش فرزندانشان بودند و پرسونل آموزشی و معلمان برای ارائه خدمات آموزشی و تدریس، به فناوری‌های دیجیتال و آنلاین روی آوردند.

برای پایان‌دادن به چنین وضعیت ناامیدکننده، معلمان در همکاری با مدیران و خانواده‌ها از فناوری دیجیتال و آموزش مجازی به عنوان ابزار اساسی تدریس استفاده نمودند[۳]. هویدا است که اخلاق و رفتار مناسب یکی از اساسی‌ترین نیازهای بشر و از مهم‌ترین عوامل فرهنگی رشد یک جامعه به شمار می‌رود[۴]. از سوی دیگر؛ در کنار مزیت‌های فراوان آموزش مجازی، در عین‌زمان، شانس رفتارهای نامناسب و غیراخلاقی در الگوهای یاددهی- یادگیری افزایش یافته و رفتارهای محیط آموزشی را با چالش‌های جدید و جدی مواجه کرده است[۵]. به دلیل این‌که معلمان یکی از عوامل کلیدی برای اجرای مؤثر آموزش از راه دور هستند[۶]، جریان و محیط آزاد اطلاعات و ارتباطات در آموزش مجازی و مشاهده‌ی رفتارهای غیرحرفه‌ای و ناپسندِ معلمان، نیازمند توجه جدی و فوری برای اتخاذ رفتارهای اخلاقی است[۷]. بنابراین؛ معلمان ضمن مهارت در تدریس، باید در زمینه‌ی اصول و اخلاق حرفه‌ای در آموزش مجازی نیز آشنا باشند[۸].

۱۰. مواد و روش‌های تحقیق

این تحقیق به روش کتابخانه‌ای(مروری) بوده و برای دستیابی به هدف تحقیق، علاوه بر کتاب‌های آموزش‌های الکترونیکی و مجازی در این زمینه، مقالات مرتبط با کلیدواژه‌های تحقیق از سال 2004 تا 2022 از پایگاه‌های معتبر سیویلیکا(Civilica)، مگ‌ایران(Magiran)، سید(Sid)، ریسرچ گیت(ResearchGate) و ساینس دایرکت(Science Direct) نیز مورد مطالعه قرار گرفته است.

۲. موضوع

کلمات و اصطلاحات مختلفی مانند؛ آموزش مبتنی بر کامپیوتر، آموزش مبتنی بر سی‌دی، آموزش مبتنی بر وِب، آموزش انترنتی، آموزش آنلاین، آموزش مجازی، آموزش الکترونیکی و مفاهیم دیگر برای تعریف آموزشی که از طریق کامپیوتر، رسانه‌ها و عوامل انترنتی به دانش‌آموز ارائه می شود، به عنوان آموزش مجازی در نظر گرفته می‌شود[۹]. برای افزایش کیفیت آموزش مجازی، اهداف و محتوای برنامه‌ها، استراتژی‌های آموزشی، وظایف و مسئولیت‌های دانش‌آموزان و فراگیران در طول فرآیند یادگیری، فناوری مناسب برای انتقال دانش و برنامه‌های اساسی، پاسخ به سوالاتی مانند «به چه کسی، چه چیزی، چرا و چه زمانی برای تهیه و ارائه محتوای درسی؟» نیز ضروری پنداشته می‌شود. این امر مستلزم همکاری و هماهنگی بین معلمان، دانش‌آموزان و مدیران مؤسسات و مراکز آموزشی بوده [۱۰،۱۱]، هم‌چنان، چگونگی نحوه‌ی تعامل معلمان و دانش‌آموزان باید متناسب با موضوع، محتوا، گروه سنی[۱۲]، انطباق آموزش مجازی با برنامه درسی، ارزشیابی منسجم، باثبات و شفاف، یادگیری مؤثر، بهره‌گیری و استفاده آسان از ابزارهای آموزش مجازی، مقرون به صرفه‌بودن آموزش مجازی نیز از عوامل مؤثر در این نوع آموزش است[۱۳].

آموزش مجازی به عناصر و مهارت‌های بیشتری نسبت به آموزش حضوری نیاز دارد. این مهارت‌ها چارچوبی برای درک عمیق از شبکه پیچیده‌ای روابط بین محتوا، پیداگوژی، فناوری و شرایط هستند که در آن فعالیت می‌کنند. بنابراین، معلمان برای تدریس در محیط مجازی علاوه بر مهارت لازم در محیط حضوری، به مهارت‌ها و قابلیت‌های دیگری نیز نیاز دارند.

معلمان در آموزش مجازی باید با بازنگری در ماهیت حرفه‌ای خود، فراگیران را به برقراری ارتباط و مشارکت با دیگران، جستجوی منابع و مشارکت در فعالیت‌ها تشویق کنند[۱۴]. یک معلم تأثیرگذار در آموزش مجازی، مسئولیت اجرای درست و کامل پنج وظیفه مهم و اساسی ترغیب دانش‌آموزان به استفاده از امکانات تکنالوژیکی و ارتباطی، سازماندهی، ایجاد جو و محیط دوستانه، تجزیه و تحلیل موضوع یادگیری و راهنمایی را به عهده دارد[۱۵].

۲-۱. مزایا و معایب آموزش مجازی

آموزش مجازی نه تنها جایگزین آموزش حضوری در شرایط ناگوار مانند همه‌گیری کرونا، آلودگی هوا، تعطیلی طولانی مدت مکاتب و سایر مراکز آموزشی شده است، بلکه در شرایط عادی نیز آموزش مجازی [به عنوان] جزء جداناپذیر آموزش مورد توجه قرار گرفته است[۱۶].

آموزش مجازی دارای پنج هدف اصلی شامل غلبه برمحدودیت‌های جغرافیایی، فرهنگی، اقتصادی، فردی و محدودیت سیستم آموزشی رایج است که به فراگیران و معلمان اجازه می‌دهد تا به طور همزمان یا غیر همزمان، مشترک یا انفرادی با یکدیگر در محیط انترنت ایجاد شده تعامل داشته باشند[۱۷]. سرعت یادگیری، آموزش بِه‌روز، تکرار پذیری، جبران اشکالات درسی، مشارکت به جای معلم‌محوری، کاهش نابرابری جنسیتی، یادگیری بدون ترس و گسترش فرهنگ یادگیری مستقل[۱۸]، کاهش هزینه تدریس، ذخیره سازی مؤثر دانش، افزایش مزایای یادگیری، آموزش و یادگیری همزمان تکنالوژی و فناوری‌های جدید، یادگیری تعاملی و شبکه‌ای غنی از منابع و اطلاعات دوره‌ای تعلیمی برای دانش‌آموزان و معلمان از مهم‌ترین مزایا و ویژگی‌های آموزش مجازی است[۱۹].

آموزش مجازی علاوه بر مزایای فراوان، موانع و مشکلاتی نیز برای دانش‌آموزان به همراه دارد. سیاست‌های مکاتب و مراکز آموزشی، نداشتن مهارت لازم در استفاده از امکانات دیجیتالی و تکنالوژیکی، تنوع منابع دیجیتالی، در دسترس نبودن دیجیتال، مشکلات شبکه و مقاومت در برابر تغییرات شیوه یادگیری، فقدان دستورالعمل‌های روشن، فقدان خط‌مشی منسجم مکاتب و مراکز آموزشی از جمله مواردی هستند که نیاز به توجه دارند[۲۰،۲۱]. آسیب آموزشی، آسیب مدیریتی، آسیب تربیتی آسیب فنی و تکنالوژیکی، خسارت مالی و آسیب سلامتی از جمله آسیب‌های تأثیرگذار بر آموزش مجازی اند [۲۲].

۳. اخلاق حرفه‌ای تدریس

امروزه اخلاق جزء لاینفک زندگی بشر به شمار می‌رود، اخلاق ساختار پیچیده‌ای دارد که تعریف واحدی برای آن وجود ندارد [۲۳]. اخلاق علمی است که از اصول و ارزش‌های اخلاقی، خوب و بد، شایستگی و ناشایستگی، حسن و قبح رفتار انسان و صفات اختیاری انسان بحث می کند[۲۴]. حفظ چارچوب اخلاقی و انسانی در انجام کاری، لازمه انجام صحیح یک عمل و اثربخشی آن است[۲۵]. افرادی که ارزش‌ها و اصول اخلاقی را در درون خود نهادینه می‌کنند و به آن اصول پایبند هستند، کار را مایهٔ استقلال و خودشکوفایی می‌دانند و این ویژگی‌ها بر انگیزه درونی فرد تأثیر می‌گذارد و میل و رغبت زیادی برای انجام کار ایجاد می‌کند[۲۶].

اخلاق حرفه‌ای یک فرایند تفکر عقلانی است و هدف آن این است که بدانیم چه ارزش‌هایی باید در سازمان حفظ و منتشر گردد. اخلاق حرفه‌ای بر فعالیت‌ها و عملکردهای فردی و گروهی و نتایج سازمان تأثیر بسزایی دارد[۲۷]. اخلاق حرفه‌ای مانند راهنمایی برای معلم در چگونگی ارائه آموزش با کیفیت و تلقین ارزش‌های خوب در بین فراگیران است[۲۸]. در این رابطه،

اخلاق در تدریس، شاخه‌ای از اخلاق در تعلیم و تربیت است که مسئولیت‌های حرفه‌ای معلم را در تدریس و آموزش مشخص می‌کند. در واقع، تدریس به عنوان بخش قوی و مهم فرآیند آموزش، مستلزم مهارت در اجرای روش‌های مناسب و ملاحظات اخلاقی ویژه شغل معلمی است[۲۹]. معلمی، مانند سایر مشاغل، دارای مفاهیم اخلاقی و ارزش‌های حرفه‌ای است که رفتار اخلاقی مسئولانه معلمان را به طور دسته‌جمعی تعیین می‌کند و اغلب در قالب طیف گسترده هنجارهای قابل قبول رفتار حرفه‌ای ظاهر می‌شود[۳۰]. آنچه ضامن سلامت فرایند یاددهی - یادگیری در مکاتب و افزایش تعهد معلمان به نیازهای دانش‌آموز می‌گردد، رعایت اخلاق تربیتی است[۳۱].

اخلاق حرفه‌ای و توجه به شاخص‌های اخلاقی ابزاری مؤثر و مزیت راهبردی در شناسایی و ایجاد فرصت‌های جدید و دستیابی به اهداف سازمان است[۳۲]. فرهنگ سازمانی در توسعه اخلاق حرفه‌ای معلمان نقش دارد، و بیشتر تغییرات اخلاق حرفه‌ای معلمان در مدارس با فرهنگ سازمانی تبیین می‌شود[۳۳]. اخلاق حرفه‌ای برای بهبود عملکرد کاری کارکنان نیاز است[۳۴]. اگر اخلاق حرفه‌ای معلمان تقویت شود، عملکرد آموزشی نیز بهبود می‌یابد[۳۵]. موفقیت در فعالیت حرفه‌ای معلم مستلزم رعایت اصول، راهبردها و استانداردهای اخلاق حرفه‌ای است[۳۶]. اخلاق حرفه‌ای مدیران باعث می‌شود معلمان نسبت به شغل خود مشتاق شوند[۳۷]. هم‌چنین اخلاق حرفه‌ای به توانمندسازی روانی معلمان می‌انجامد[۳۸].

۴. رابطه میان اخلاق حرفه‌ای تدریس و آموزش مجازی

با روی آوردن به دنیای دیجیتالی، یافته‌ها بینش‌هایی را در مورد طراحی و توسعه برنامه درسی ارائه می‌دهند که در موفقیت معلمان در محیط آنلاین کمک می‌کند. از میان عمده‌ترین عوامل مؤثر در تغییر به سمت دیجیتال شدن آموزشی که توسط نویسنده شناسایی گردیده است، پشتیبانی و بازخورد از فن‌آوران یادگیری، پایداری مالی، اهداف و رویه‌های برنامه، تغییر در مهارت‌ها و نگرش‌های معلمان، ارتباطات و ارزیابی و حمایت از توسعه حرفه‌ای معلمان می‌باشد. نتایج این تحلیل منجر به شناسایی یافته‌های ترکیبی در مورد ارزشمندترین عناصر توسعه اخلاق حرفه‌ای برای یادگیری آنلاین و ترکیبی شد[۳۹]. هم‌چنین با توجه به این‌که نظارت کافی بر کار معلمان در آموزش مجازی وجود ندارد و روند یادگیری دانش‌آموزان رعایت نمی شود، اخلاق حرفه‌ای معلمان جایگاه ویژه‌ای پیدا می‌کند.

نتیجه‌گیری

در مجموع نتایج نشان داد که می‌توان کیفیت آموزش مجازی را با به‌کارگیری اخلاق حرفه‌ای تدریس معلمان و مؤلفه‌های آن که شامل اشراف بر محتوا، صلاحیت تدریس، توانمندی در برخورد با مسائل حساس، کیفیت تدریس، رشد دانش‌آموز، توانایی در برقراری روابط دو طرفه با دانش‌آموزان، رازداری، احترام به همکاران و ارزشیابی صحیح افزایش داد. با توجه به این‌که

امروزه آموزش مجازی به عنوان بخشی جداناپذیر از آموزش در سیستم‌های آموزشی جهان پذیرفته شده است، نیاز به رویکرد جدیدی در فرایند آموزش وجود دارد. این تغییر نگرش، سیستم آموزشی را ایجاب می‌کند که با استفاده از روش‌های نوین تدریس، در بهبود فرآیندهای تدریس و یادگیری تلاش کند. در این میان نقش معلمان و رعایت اخلاق حرفه‌ای می‌تواند راه را برای اثربخشی و کیفیت آموزش مجازی باز کند. بنابراین، یکی از موضوعاتی که سیستم‌های آموزشی مجازی با آن سروکار دارند، لزوم توسعه‌ی رفتارهای مبتنی بر اخلاق و اخلاق حرفه‌ای توسط اعضای سیستم، به‌ویژه معلمان است. با توجه به نقش نظام آموزشی در مواجهه با مشکلات جامعه، توجه به اخلاق حرفه‌ای معلمان ضروری به نظر می‌رسد. برای این‌که بتوانیم سناریوهای ایجاد شغل را مطابق با آموزش حرفه‌ای شهروندی ارائه دهیم و ارتباط مؤثری بین سیستم آموزشی و جامعه برقرار کنیم.

منابع

1. *Aljaloudi M A, Battah A. Degree of availability and practice of level 5 leadership by the principals of Palestinian private high schools based on Jim Collins concepts. Modern Applied Science, 2019, 13(1): 172-182. DOI:*10.5539/MAS.V13n1p172
2. *Verkuy M, Hughes M. Virtual gaming simulation in bridging nursing education: A mixed methods study. Clinical Simulation in Nursing, 2019, 29(3): 9-14. DOI:* 101016/j.ecns.2019.02.00
3. *Leonard B, Rosienne F. Teachers' response to the sudden shift to online learning during COVID-19 pandamic: implications for policy and practice, Faculty of Education, 2020, 14(2):211-241.*
4. *Chanbari S, Aziz A. The relationship between moral behaviour and academic procrastination and cheating attitude. Teaching Research, 2022, 10(1): 274-296. (In Persian). DOI:* https://www.doi.ogr/10.34785/J012.2022.011
5. *Talib A A, Mahasneh o m k. Awareness of ethical issues when using a e-learning system. International Journal of Advanced Computer Science and Applications, 2020, 11(1):1-10. DOI:*10,14569/IJACSA, 2020.0110116
6. *Alikhani F, Samadi M, Saidipour B, Masoumi- Fard M, Gudarzi M. Identifying the components of professional ethics in the context of electronic learning and distance education. Research in Virtual Education Learning, 2021, 9(2):87-99. (In Persian). DOI:*https://doi.org/1030473/etl.2021.60605.3604
7. *Miglani A, Awadhiya A K. Mobile learning: reading and perceptions of teachers of Open Universities of Commonwealth Asia. Journal of learning for Development-JL4D, 2017, 4(1):1-10*
8. *Hsanpour A, Nazari K, Akbari P. Design the model of effective factors on the professional ethics of teaching in Payam University the light of the west of the county. Journal of Sociology of Education, 2018, 6:1-10. (In Persian). DOI:* https://1022034/ijes.2017.34678.
9. *Mirzaei K, Saadi H, Movahed Mohamadi SH, Movahedi R. Checking the e-lrearning acceptability of Iranian agricultural higher education centres from the viewpoints of faculty staff and graduate students. Journal of Technology and education, 2018; 12(4):259-272. (In Persian). DOI:* https://sid.ir/paper/155434/en
10. *Orhan G, Beyahan O. Teachers' perceptions and teaching experiences on distance education through synchronous video conferencing during Covid – 19 pandemic. Social Science and Education Research Review, 2020; 7(1): 8-44*
11. *Marsh B, Mitchell N, Adamczyk P. Interactive video technology: Enhancing professional learning in initial teacher education. Comptures & Education, 2010; 54(3):742-784. DOI:* 10.1016/j.compedu.2009.09.011
12. *Moore M G, Kearsley G. Distance education: Systems view of online learning. Cengage Learning. 2011.*
13. *Mohammad G, Sri S, Irma-Yulia, B. Online learning quality control in the pandemic Covid-19 Era in Indonesia. Journal of Nonformal Education, 2020; 6)2): 168-175. DOI:* https://doi.org/10.15294/jne.v6i2.25594.

14. *Miguel B, Mcpherson M. Developing innovation in online learning. London: Rouledgfalmer. 2004.*
15. *Mandernach B J, Gonzales R M, Garrett A L. An examination of online instructor presence via threaded distance participation. MERLOT Journal of Online Learning and Teaching, 2006; 2(4):248-261.*
16. *Novak E. Toward a mathematical model of motivation, volition, and performance. Computers & Education, 2014; 74: 73-80. DOD:*https://doi.ogr/10.1016/j.compedu.2014.01.009
17. *Pourghfar L, Jafarzadeh H. Nazargiglu S. Virtual education-parent/student communication. New Achievements in Humanities Studies 2022; 4: (37). (In Persian).*
18. *Eskandari H, Vahadani Asadi M. Barriers to intelligentization and the effect of in-service virtual training on its use and the quality of the teaching and learning process among primary school teachers. Teaching and learning Technology, 2016; 3(12):71-93.(In Persian). DOI:* https://doi.ogr/10.22054/jti.2020.47256.1285.
19. *Miyoshi M, Tsuboyama-Kasaoka N, Nishi N School-based" Shokuiku "program in Japan: application to nutrition education in Asian countries. Asia Pacific Journal of Clinical Nutrition, 2012; 21(1):159-162.*
20. *Ferdig RE, Baumgartner E, Hartshome R, Kaplan-Rakowski R, Mouza C. Teaching, technology, and teachger education during the COVID-19 pandemic: stories from the field. Association for the advancement of Computing in Education. Retrieved June 15, 2020 from* https://www.learntechlib.org/p/216903.
21. *Onyema E M, Educheria N C, Obafemi F A, Sen S, Atonye F G, Sharma A, Alsayed A O. Impact of Coronavirus pandemic on education. Journal of Education and Paractice,2020; 41:108-121.DOI:*10.7176/JEP/11-13-12
22. *Khalili M Z, Feridoni F. Pathology of Virtual education from the perspective of elementary school teachers; a qualitative case study. Quarterly Journal of New Developments in Educational Management, 2020; 1(2):43-53. (In Persian).*
23. *Nasirian F, Jafarian H, Seifi Selseleh Y. The role of emotional and cognitive empathy, moral metacognition and quality of life in predicting moral behaviour of nurses. Int. J. Ethics Soc. 2021; 3(2), 44-51. DOI:10.25547/ijethics.3.2.44.*
24. *Kazemi Malek Mhamodi A, Jabbari N, Niyaz Azari K. Identification and ranking of antecedents affecting ethical-emotional acting in education.Int. J. Ethics Soc. 2022; 4(2), 40-40. DOI:*10.25547/ijethics.4.2.240.
25. *Arabshahi Krizi A, Gholaminejad A. Investigating the effect of Islamic human resource management on organizational virtue with the mediating role of work ethics. Marine Science Education Quarterly, 2021; 8(1): 45-59. (In Persian).*
26. *Jahanshahi F, Ghasemi M, Ahmadi M. Investigating the effect of Islamic work ethics on job motivation with the mediating role of*

organizational justice. Nursing Management Quarterly, 2021; 10 (4): 1-12. (In Persian).
27. *Ghasemi M, Shariatmadari M. The relationship between professional ethics and trust building among faculty members. Ethics in Science and Technology, 2017; 12(4): 39-47. (In Persian). DIO:*20.1001.1.22517634.1396.12.4.5.6.
28. *Sherpa K. Importance of professional ethics for teachers. International Education and Research Journal, 2018; 4(3): 16- 18.*
29. *Moradi-Doliskani M, Mohammadi – Mehr M. investigating the relationship between professional ethics and teaching quality among the professors of the Army University of Medical Science. Journal of Teaching Strategies in Medical Science, 2020; 13(3): 143-150. (In Persian).*
30. *Maxwell B, Schwimer M. Seeking the elusive ethical base of teacher professionalism in Canadian codes of ethics. Teaching and Teacher Education, 2016; (59): 468-480. DOI:ttps*://doi.org/10.1016.j.tate.2016.07.015
31. *Salmanzadeh P, Farhadi A. Professional ethics of teaching 1st Scientific Research Conference of Psychology, Educational Science and Pathology of Society, Kerman. 2015. (In Persian).*
32. *Khayatmoghadam S. The role of organizational culture. Int.J Ethics Soc., 2020; 2(1):21-28.* https://ijethecs. Com/article-1-74-en.html.
33. *Ali A, Sajjad B. The role of organizational culture in the emergence of teachers' professional ethics. Integration of Education. 2022; 26(1): 10-26. DOI:9468.106.026.202201.010-026*
34. *Gholampour M, Pourshafei H, Ferastkhah M, Ayati M. The components of teachers' professional ethics: a systematic review based on Wright's model. Iranian Curriculum Studies Quarterly, 2021; 15(58): 145-174. (In Persian). DOR:* 20.1001.1.17354986.1399.15.58.5.5
35. *Ayeni A J. Teachers' professional ethics and instructional performance as correlates of students' academic performance in secondary schools in Owo local government Ondo State, Nigeria. Advances in Social Sciences Research Journal, 2018, 5(8): 611-622. DOI:*10.14738/ASSRJ.584973.
36. *Faiz S A, Elahi Z. History of teacher professional ethics in education. Professional Ethics in Education, 2021; 1(1): 154-175. (In Persian).*
37. *Zahed Bablan A, Gharibzadeh R, Nasrabadi H, , Gharibzadeh S. The role of managers' professional ethics in teachers' job enthusiasm; with the mediation of organizational identity. Ethics in Science and Technology, 2018; 13(4): 110-116. (In Persian). DOR:*10.1001.1.22517634.1397.13.4.15.3
38. *Asadian S, Rabiei M, Gharibzadeh A. The relationship between professional ethics and organizational trust with teachers' psychological empowerment. Ethics Quarterly, 2016; 6(22): 165-190. (In Persian).*
39. *Eom- Sean B, Ashill – Nichlas J. A systems' view of e-learning success model. Decision Science Journal of Innovative Education, 2018; 16, (1): 1-36. DOI:*10.10.1111/dsji12144

40. *Philipsen B, Tondeur J, Pareja – Roblin N, Vanslambrouck S, Chang Z. Improving teachers professional development for online and blended learning: A systematic meta-aggregative review. Educational Technology and Research Development, 2019; 67, 1145-1174. DOI:*https://doi.org/10.1007/s11423-019-09645-8.

Recognition of Afghanistan's General University Entrance Exam ID from the Exam Sheet

Abstract

The use and promotion of artificial intelligence (AI) in advancing industries is one of humanity's remarkable achievements in recent decades, applied in various fields. Today, governmental and non-governmental organizations face an urgent need for administrative automation and problem-solving through AI. Accordingly, this research aims to address the problem of identifying candidates' IDs for university entrance exams using AI.

In this study, the identification and verification of candidates' IDs from their exam sheets are carried out in two stages. In the first stage, the ID is recognized based on the digits handwritten by the candidates in the designated ID section of the exam sheet. In the second stage, the boldened keys marked in the ID section are utilized. Both stages are performed after scanning the exam sheets, employing image processing techniques and cropping operations to isolate the required areas.

For recognizing handwritten digits, the MNIST dataset and the TensorFlow and Keras libraries were used to develop a high-accuracy model. In the second stage, the key section is divided into nine columns, each containing a single key, and then rows are processed with precision. The boldened keys are extracted using a specific threshold value, and ultimately, the candidate's ID number is identified through both methods to ensure higher reliability.

Keywords:

handwritten digit recognition, image processing, Kankor exam, deep learning

Flexibility Approach to Incremental Housing Design

Sayed Anwarshah Mousawi[a]
[a] *Assistant professor, Khatam Al-Nabieen University*[1]

Abstract
The significant issue concerning human rights is the housing problem faced by low-income groups in society. One recurring mistake in most housing supply plans is the reliance on complete architectural designs for houses. However, a group of experts, with a better understanding of the needs of low-income residential groups, believes that instead of designing and presenting fully finished housing prototypes, authorities should focus on providing examples of incremental completion for residential units. It appears that this type of housing does not require a high budget and can be constructed quickly, making its development crucial for ensuring social and economic stability in fostering national development. Therefore, this paper aims to identify the features, dimensions, and indicators of incremental housing and propose a model designed to address the housing problems of low-income groups in developing countries. In the design and construction process of incremental housing, houses will offer various types of views and suitable layouts according to the occupants' demands. Low cost, flexibility, and the right to make informed choices about future development and the use of available building materials are among the main characteristics that will satisfy residents in this approach.
Keywords: incremental housing, low-income groups, flexibility

[1] Corresponding Author: Email: sa.mousawi@knu.edu.af

Applications of Artificial Neural Network and Deep Learning in Damage Detection and Structural Health Monitoring - a review study

Abstract

In recent years, Structural Health Monitoring (SHM) systems have been widely installed in various civil infrastructures to monitor the health status of structures and detect structural damage or abnormalities through long-term monitoring of structures. Conventional data analysis methods face challenges such as environmental noise, large amount of measurement data, computational complexity, etc., which severely limit the widespread application of SHM technology. In recent years, with the rapid development of computing hardware and image acquisition equipment, deep learning-based data processing methods provide a new channel for mining the huge data of a SHM system, towards autonomous, accurate and robust processing of monitoring information. This paper provides an overview of deep learning-based SHM, including a brief summary of the history of deep learning development, applications of deep learning-based data processing methods in SHM.

Keywords: Artificial neural network, deep learning, damage detection, structural health monitoring.

Simulating secure model of ATM machines using DEVS

Abstract

In this article, the DEVS formula of discrete event systems, which was introduced by Dr. Bernard P. Ziegler in 1976 in a book called Theory of Modeling and Simulation, the concept of DEVS is used and discussed for modeling and simulating an ATM machine. Today, ATM machine have undoubtedly become a necessity in banks, people no longer have to visit the banks to do their transactions, instead use these types of machines, however, the people who are using an ATM machine are not clear and aware of that how the secure processes work inside of the ATMs. With an ATM card and a valid PIN, the customers can use the ATM where the valid user is accepted based on a PIN (personal identification number). Authentication and authorization ensures that maximum security is maintained through ATMs. In fact, banks provide flexible functionalities for the users such as balance checking, cash deposit and money transactions, but the biggest challenge is the security of the transactions, which especially the mobility ATM machines suffer. According to the DEVS formula, in this article, an attempt has been made to provide an application model for secure data transmission using the above formula.

Key terms:

ATM machine, authentication, DEVS formula, modeling and simulation.

An Overview of Artificial Intelligence Ethics

Changwu Huang, Zeqi Zhang, Bifei Mao, and Xin Yao

Abstract

Artificial intelligence (AI) has profoundly changed and will continue to change our lives. AI is being applied in more and more fields and scenarios such as autonomous driving, medical care, media, finance, industrial robots, and internet services. The widespread application of AI and its deep integration with the economy and society have improved efficiency and produced benefits. At the same time, it will inevitably impact the existing social order and raise ethical concerns. Ethical issues, such as privacy leakage, discrimination, unemployment, and security risks, brought about by AI systems have caused great trouble to people. Therefore, AI ethics, which is a field related to the study of ethical issues in AI, has become not only an important research topic in academia, but also an important topic of common concern for individuals, organizations, countries, and society. This article will give a comprehensive overview of this field by summarizing and analyzing the ethical risks and issues raised by AI, ethical guidelines and principles issued by different organizations, approaches for addressing ethical issues in AI, and methods for evaluating the ethics of AI. Additionally, challenges in implementing ethics in AI and some future perspectives are pointed out. We hope our work will provide a systematic and comprehensive overview of AI ethics for researchers and practitioners in this field, especially the beginners of this research discipline.

Keywords: Artificial intelligence (AI), AI ethics, ethical issue, ethical theory, ethical principle.

The position (role) of teachers' professional ethics in virtual school education

Abstract

Introduction: Increasing the quality of virtual education is one of the main concerns of the education system in most countries of the world. In educational centers, virtual education has a special place, and to improve it, the professional ethics of teachers should be taken into consideration. Considering this importance, the aim of the present study was to investigate the position of professional ethics of teachers in virtual education of schools.

Materials and Methods: The research was a review method, in order to achieve the goal of the research, in addition to electronic education books and virtual education in this field, articles related to the research keywords from 2004 to 2022 from the databases of Civilica, Magiran, Sid, Researchgate, Science direct, was reviewed.

Conclusion: The results showed that the improvement of virtual education in schools, in addition to the need for information and communication technology infrastructure, the ability of teachers and students to use new technologies and the interaction between teachers and students, requires the professional ethics of teachers and the observance of ethics, is teaching therefore, one of the factors that can affect the virtual education of schools is the professional ethics of teachers. Based on this, the ethical topics of teaching and its concepts should be included in the educational courses of students of cultural universities and in – service training for teachers. At the time of application, new technologies in education should be given serious attention and emphasis, and efforts should be made to institutionalize teaching ethics among teachers.

Keywords: Professional teaching ethics, Virtual education Teachers.