

هوش مصنوعی و سیستم های خبره در کارشناسی رسمی دادگستری (۲) (جرائم هوش مصنوعی و هوش مصنوعی مولد)



رای عای
ان رسمی دادگستری

مهرداد تاج بخش (کارشناس رسمی دادگستری،

رشته رایانه و فناوری اطلاعات)

شهریورماه ۱۴۰۴

۲	چکیده
۲	مقدمه‌ای در آشنایی با هوش مصنوعی
۴	هوش مصنوعی به عنوان ابزاری در امور مرتبط با کارشناسی رسمی دادگستری
۵	هوش مصنوعی به عنوان موضوع بررسی‌های کارشناسی
۶	مقدمه‌ای بر فرآیند تحقیق و تفحص ادله الکترونیکی
۷	طبقه‌بندی انواع جرائم هوش مصنوعی
۱۰	قوانین، ملاحظات و اخلاق در هوش مصنوعی
۱۵	شفافیت و توضیح‌پذیر بودن سیستم‌های هوش مصنوعی
۱۹	پیوست یک - فرآیند تحقیق و تفحص ادله الکترونیکی هوش مصنوعی
۳۲	پیوست دو - هوش مصنوعی مولد

شورای عالی
کارشناسان رسمی دادگستری



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

چکیده

امروزه پرداختن به موضوع هوش مصنوعی^۱، چالش‌ها و ملاحظات مطرح در آن به عنوان شاخه‌ای از فناوری که به واسطه‌ی قابلیت‌های هوشمندسازی و هوش مصنوعی مولد^۲، نقش بسیار مهمی را در جنبه‌های مختلف زندگی بشری پیدا نموده، در حوزی کارشناسی رسمی دادگستری و بررسی‌های قانونی و حقوقی از یک سو و به عنوان ابزاری برای کمک در فرآیند تحلیل و بررسی کارشناسی و ارائه‌ی نظریه کارشناسی و همچنین ابزارهای مرتبط با ارجاعات کارشناسی اهمیت ویژه‌ای پیدا نموده و مورد توجه قرار گرفته است.

تا کنون بیشتر به مقوله هوش مصنوعی و کاربردهای آن به عنوان موضوع بررسی‌های کارشناسی و تحقیق و تفحص ادله‌ی الکترونیکی پرداخته شده^۳؛ این در حالی است که با ظهور فناوری‌های پیشرفته دیگر هوش مصنوعی در حوزه‌های شبکه‌های عصبی و یادگیری عمیق و در ادامه هوش مصنوعی مولد، به نظر می‌رسد که توجه و استفاده از این فناوری به عنوان ابزاری در دسترس سازمان‌های قانونی و قانون‌گذار برای انتخاب رشته و صلاحیت‌های مناسب به منظور ارجاعات کارشناسی و همچنین کمک به کارشناسان رسمی دادگستری در خصوص تحلیل، بررسی، تهیه متن و نگهداری داده‌های حائز اهمیت در گزارش‌های کارشناسی، رشد و شتاب بیشتری بگیرد.

صرف نظر از کاربرد سامانه‌های مبتنی بر هوش مصنوعی به عنوان موضوع تحقیق و تفحص و بررسی‌های کارشناسی و یا به عنوان ابزاری برای کمک به امور کارشناسی، آشنایی با این حوزه از دانش و فناوری و جنبه‌ها و چالش‌های فنی، قانونی و حقوقی، اخلاقی و ملاحظات مطرح در آن، ضروری می‌باشد. فرآیند تحقیق و تفحص ادله‌ی الکترونیکی هوش مصنوعی در پیوست یک شرح داده شده است.

با توجه به نوآوری‌های مطرح شده در هوش مصنوعی مولد و تحولات چشمگیر آن در این حوزه و ضرورت آشنایی با آن برای همکاران محترم کارشناس، در پیوست دو به هوش مولد پرداخته شده است.

مقدمه‌ای در آشنایی با هوش مصنوعی

تعاریف مختلفی از هوش مصنوعی ارائه شده است، که می‌توان به تعاریف زیر اشاره نمود:

هوش مصنوعی به سیستم‌های رایانه‌ای با قابلیت اجرای عملیات پیچیده نظیر استدلال، استنتاج، تصمیم‌گیری و حل مسائل که پیشتر تنها توسط انسان قابل انجام بود، اشاره می‌کند. امروزه واژه‌ی هوش مصنوعی به طیف گسترده‌ای از فناوری‌هایی که با استفاده از این فناوری ضمن اجرای فعالیت‌های مختلف، خدمات متنوعی را ارائه می‌نمایند، اشاره می‌کند^۴. سازمان دفاع ملی آمریکا (۲۰۱۹) تعاریف زیر را ارائه نموده است،

^۱ Artificial Intelligence

^۲ Generative AI

هوش مصنوعی مولد، قابلیت هوش مصنوعی در ایجاد متن، تصاویر، و یا محتواهای دیگر با استفاده از مدل‌های مولد می‌باشد. این مدل‌های هوش مصنوعی ضمن یادگیری الگوها و ساختار اطلاعات ورودی طبقه‌بندی شده، اقدام به تولید اطلاعات جدید با مشخصات مشابه می‌نمایند.

^۳ همان گونه که در ادامه بیشتر شرح داده خواهد شد، در این گونه موارد یا سامانه‌ی مبتنی بر هوش مصنوعی خود عامل وقوع جرم یا رخداد بوده و یا این که به نحوی در وقوع رخداد یا ارتکاب جرم تاثیر داشته و یا دارند.

^۴ Coursera website (Online education)



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

- هر سیستم مصنوعی که در شرایط گوناگون و غیرقابل پیش‌بینی و بدون نظارت و دخالت قابل توجه عامل انسانی، اقدام به اجرای کارها نموده و یا اقدام به یادگیری از طریق مجموعه‌ی داده‌های دریافت شده و ارتقای کیفیت انجام کارها می‌نماید.
 - هر سیستم مصنوعی توسعه داده توسط نرم‌افزارهای رایانه‌ای، سخت‌افزارها و یا سایر تجهیزات که به‌نوعی امکان اجرای کارها و فعالیت‌هایی را می‌دهند که به‌نوعی نیاز به درک انسان‌گونه، شناخت و تشخیص، طراحی، یادگیری، ارتباط و کارهای فیزیکی دارند.
 - یک سیستم مصنوعی که با هدف فکر و عمل کردن مشابه (ونه دقیقاً) انسان؛ از طریق الگوریتم‌ها و ساختارهای آگاهانه و شبکه‌های عصبی، طراحی شده است.
 - مجموعه‌ای از تکنیک‌ها، شامل یادگیری ماشین که برای انجام کارهایی که نیاز به شناخت و آگاهی مشابه انسان دارند، طراحی و استفاده می‌شوند.
 - سیستم مصنوعی که برای انجام کارها بصورت ادراکی طراحی شده و مورد استفاده قرار می‌گیرد. مانند عامل نرم‌افزاری هوشمند یا یک ربات توکار که برای رسیدن اهداف از طریق ادراک، استدلال و یادگیری، ارتباط و تصمیم‌گیری و اقدام، استفاده می‌کند.
- و یا تعریف ارائه شده در سند ملی هوش مصنوعی ایران^۵،
- به توانایی ماشین برای انجام عملکردهای خودکار و نظام‌مند از جمله یادگیری، درک، استنتاج، حل مسئله، پیش‌بینی، تصمیم‌گیری و اقدام از طریق به‌کارگیری دانش، اطلاعات و پردازش داده‌ها گفته می‌شود که منشاء اثرگذاری‌های گسترده برانسان و روابط انسانی در محیط فیزیکی یا مجازی و همچنین بازتاب‌های زیست‌محیطی است. هوش مصنوعی ماهیتی داده‌ای، شبکه‌ای، الگوریتمی، خوشه‌ای، لایه‌ای و یکپارچه مبتنی بر منطق‌های کلاسیک و سایر منطق‌های نوین دارد.
- با توجه به تعاریف ارائه شده، هوش مصنوعی سامانه‌ی مصنوعی مبتنی بر نرم‌افزارهای رایانه‌ای بوده که با استفاده از شرایط محیط و پارامترهای ورودی، الگوریتم‌های استنتاج، و داده‌های یادگرفته شده اقدام به حل مسئله، ارائه‌ی تصمیم و یا انجام عملی می‌کند.
- بنابراین آشنایی با ابزارها، زیرساخت‌های نرم‌افزارهای رایانه‌ای مورد استفاده در پیاده‌سازی و توسعه سامانه‌های نرم‌افزاری سیستم‌های هوش مصنوعی و داده‌های ذخیره‌شده، انتقال داده‌ها، شبکه و اینترنت و همچنین الگوریتم‌های یادگیری، جستجو و استنتاج از اهمیت ویژه‌ای برخوردار می‌باشد. بنابراین می‌توان این‌گونه مطرح نمود؛ صرف‌نظر از مورد کاربرد سیستم هوش مصنوعی، استفاده از متخصصین حوزه‌ی نرم‌افزارها و سخت‌افزارهای رایانه‌ای و همچنین شبکه‌های رایانه‌ای و سیستم‌های مبتنی بر اینترنت امری ضروری و اجتناب‌ناپذیر می‌باشد^۶.

هوش مصنوعی به‌عنوان ابزاری در امور مرتبط با کارشناسی رسمی دادگستری

^۵ سند ملی هوش مصنوعی، شورای عالی انقلاب فرهنگی - ۱۴۰۳

^۶ در حوزه‌ی کارشناسی رسمی دادگستری؛ این موضوع در صلاحیت کارشناسان رشته رایانه، فناوری اطلاعات و فضای مجازی می‌باشد.

هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

از کاربردهای هوش مصنوعی می‌توان به انتخاب رشته(ها) و صلاحیت(ها) مناسب و متناسب به منظور اجرای قرارهای کارشناسی اشاره نمود. می‌توان به موارد زیر به عنوان برخی از دلایل این کاربرد اشاره نمود:

- تنوع رشته و صلاحیت‌های کارشناسی
 - هم‌پوشانی در رشته و صلاحیت‌های کارشناسی و دشواری در تشخیص مناسب‌ترین رشته و صلاحیت
 - پیچیدگی تشخیص موضوع بررسی و انتخاب رشته و صلاحیت مناسب
 - عدم دسترسی به افراد خبره در بررسی موضوعات کارشناسی (قرار) و انتخاب رشته و صلاحیت مناسب
 - پیشنهاد برای استفاده از بیش از یک رشته و صلاحیت برای اجرای قرار کارشناسی
 - سایر موارد از این دست
- هوش مصنوعی به دلیل ماهیت هوشمندسازی و استفاده از داده‌های یادگرفته شده و همچنین امکان استفاده از سوابق پردازش داده‌ها و اضافه نمودن آن‌ها به مجموعه‌ی داده‌های یادگرفته شده^۷ می‌تواند به عنوان ابزاری قوی در انتخاب رشته و صلاحیت‌های کارشناسی مورد استفاده قرار گیرد.
- در این نوع کاربرد، سامانه‌ی هوش مصنوعی با تفکیک متن قرار کارشناسی و تبدیل آن به کلیدواژه‌ها، با استفاده از الگوریتم‌های مناسب جستجو در بانک اطلاعاتی کلیدواژه‌ها سعی در رسیدن به رشته و صلاحیت مناسب، دارد. همان‌گونه که مشاهده می‌گردد، هراندازه کلیدواژه‌های انتخاب و یادگرفته شده، از نظر معنا و ترکیب یک، دو یا سه کلمه؛ مناسب‌تر باشند، می‌توان انتظار رسیدن به نتیجه مناسب‌تری را داشت. درعین حال الگوریتم‌های تصمیم‌گیری و سیستم خبره در سامانه‌ی هوش مصنوعی نیز از اهمیت زیادی برخوردار می‌باشد.
- هوش مصنوعی می‌تواند به عنوان ابزاری برای کمک در بررسی‌های کارشناسی و انشاء نظریه‌های کارشناسی و مواردی از این دست نیز مورد استفاده قرار گیرد. از آن جمله می‌توان به موضوعات زیر اشاره نمود:
- استفاده در تحلیل موضوع کارشناسی و استفاده از نتایج کارشناسی‌های گذشته (به عنوان مثال در کاشناسی ساختمان، اطلاعات مبلغ ارزیابی، جانمایی موضوع ارزیابی و شرایط ارزیابی) با هدف ارائه‌ی نظریه‌های کارشناسی دقیق‌تر و پرهیز از اختلاف فاحش در برآوردهای کارشناسی
 - استفاده از هوش مصنوعی در انشاء نظریه‌های کارشناسی (هوش مصنوعی مولد^۸) با هدف رسیدن به قالب‌های هماهنگ در نظریه‌های کارشناسی
 - استفاده از ابزارهای هوش مصنوعی در تحلیل و بررسی مستندات، فایل‌ها و داده‌ها و استخراج ادله‌ی الکترونیکی
 - و سایر موضوعات

هوش مصنوعی به عنوان موضوع بررسی‌های کارشناسی

^۷ Reinforcement Learning

^۸ به این شاخه از هوش مصنوعی مولد (LLM (Large Language Model و GPT (Generative Pre-trained Transformer) اطلاق می‌شود.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

هوش مصنوعی به دلیل ویژگی‌های متمایز و منحصر بفردی که ارائه می‌نماید، به‌عنوان یک فناوری می‌تواند در تولید و توسعه سامانه‌ها و تجهیزات هوشمند بکارگرفته شود. از آن جمله می‌توان به کاربردهایی نظیر؛ هوش-مصنوعی به‌عنوان ابزاری برای خودکارسازی و استفاده در اتومبیل‌های خودران، هوش مصنوعی به‌عنوان ابزاری برای شبیه‌سازی حرکات، صورت و صدای انسان، هوش مصنوعی به‌عنوان ابزاری برای طراحی و ایجاد سامانه‌های هوشمند و تصمیم‌یار در حوزه‌ی مختلف مانند ساختمان، کشاورزی، پزشکی و دارویی، فناوری اطلاعات و موارد دیگر اشاره نمود.

از دیدگاه کارشناسی رسمی دادگستری، بررسی سیستم‌های هوش مصنوعی و موضوعات مرتبط با آن، صرفنظر از این‌که سامانه‌ی هوش مصنوعی خود موضوع بررسی بوده و یا به‌عنوان بخشی از رخداد (موثر در وقوع رخداد) بررسی می‌گردد، از جنبه‌های مختلف قابل طرح می‌باشد:

- ۱) حسگرها و دریافت‌کننده‌های اطلاعات و شرایط محیط در سیستم هوش مصنوعی
- ۲) برنامه‌ی نرم‌افزاری موتور جستجو، استنتاج و تصمیم‌گیری و استفاده از الگوریتم‌های مرتبط با آن در سیستم هوش مصنوعی
- ۳) برنامه‌ی نرم‌افزاری یادگیری ماشین و استفاده از الگوریتم‌های مرتبط با آن در سیستم هوش مصنوعی
- ۴) داده‌های یادگرفته شده
- ۵) سامانه‌های سخت‌افزاری بکاررفته در سیستم هوش مصنوعی (با توجه به عملکرد و وظیفه‌ی سیستم هوش مصنوعی)

یک کارشناس می‌بایست بتواند با درنظر گرفتن شرایط و نوع جرم و یا رخداد موضوع بررسی، نسبت به تعیین سیستم هوش مصنوعی به‌عنوان عامل وقوع جرم و یا به‌عنوان بخشی از جرم رخ داده شده، اظهارنظر نموده و سپس نسبت به تعیین صحنه جرم^۹ اقدام نماید. صحنه‌ی جرم در سیستم‌های هوش مصنوعی به تعیین یک یا بیش از یک عامل از عوامل پنج‌گانه موثر در بررسی سیستم‌های هوش مصنوعی اقدام نماید.

به این مجموعه اقدامات، فرآیند تحقیق و تفحص ادله‌ی الکترونیکی^{۱۰} اطلاق می‌شود و بصورت معمول از چهار اقدام، تعیین محل وقوع جرم^{۱۱}، بررسی و جمع‌آوری ادله‌ی الکترونیکی^{۱۲}، تحلیل و بررسی ادله‌ی الکترونیکی^{۱۳} و مستندسازی و ارائه‌ی ادله‌ی الکترونیکی^{۱۴} تشکیل شده است.

از آنجایی‌که صحنه‌ی جرم موضوع بررسی جرم یا رخداد مرتبط با سیستم‌های هوش مصنوعی سامانه‌های نرم-افزاری، سخت‌افزاری رایانه‌ای و شبکه‌ی اینترنت بوده، لذا ادله‌ی جمع‌آوری و بررسی شده، از نوع ادله‌ی الکترونیکی می‌باشد.

^۹ Crime scene

^{۱۰} Digital Forensic Investigation

^{۱۱} Recognizing and Preparation Crime Scene

^{۱۲} Investigation and Collection Digital Forensics

^{۱۳} Analysis and evaluation Digital Forensics

^{۱۴} Reporting and Presentation Digital Forensics



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

مقدمه‌ای بر فرآیند تحقیق و تفحص ادله الکترونیکی

فرآیند تحقیق و تفحص ادله الکترونیکی شامل چهار فعالیت اصلی، بررسی صحنه جرم^{۱۵}، جمع‌آوری داده‌ها و اطلاعات، تحلیل و مستندسازی داده‌ها و اطلاعات، ارائه مستندات در غالب ادله‌های الکترونیکی به مراجع قانونی و قانون‌گذاری می‌باشد. در این فرایند صحنه جرم موضوع بررسی و جمع‌آوری ادله الکترونیکی اغلب زیرساخت-های رایانه و سرور و همچنین فضای مجازی و اینترنت می‌باشد. فرآیند تحقیق و تفحص ادله الکترونیکی برای کاربردهایی نظیر تحقیق و تفحص جرائم فضای مجازی (مانند نقش حریم خصوصی، سرقت اطلاعات شخصی، هک کردن و آلوده کردن سامانه‌های نرم‌افزاری و نظایر آن‌ها)، تشخیص و آشکار نمودن فعالیت‌های مخرب و مجرمانه در فضای مجازی (مانند شنود غیر مجاز اطلاعات، دست‌کاری در اطلاعات و نظایر آن‌ها) و فراهم نمودن ادله استنادپذیر برای ارائه به مراجع قانونی، بکار گرفته می‌شود.

ادله الکترونیک استنادپذیر به ادله‌ای اطلاق می‌شود که ضمن مشخص بودن نحوه جمع‌آوری، بررسی و نگهداری آن ادله، شرایط حفظ جامعیت^{۱۶}، محرمانگی^{۱۷} و اصالت^{۱۸} در آن رعایت شده باشد. در آیین‌نامه استنادپذیری ادله الکترونیکی^{۱۹} مصوب مجلس شورای اسلامی، فرآیند تحقیق و تفحص ادله الکترونیکی. همچنین استنادپذیری و ملاحظات مطرح در آن بشرح زیر آورده شده است:

- جمع‌آوری ادله الکترونیکی فرآیندی است که طی آن ادله الکترونیکی به تنهایی یا به همراه سامانه‌های رایانه‌ای یا مخابراتی یا حامل‌های داده، نگهداری، حفظ فوری، تفتیش و توقیف و شنود می‌شوند.
- مجموعه اقداماتی است که ضابط دادگستری و سایر اشخاص ذیصلاح به منظور حفظ صحت، تمامیت، اعتبار و انکارناپذیری ادله الکترونیکی با بکارگیری ابزارها و روش‌های استاندارد در مراحل شناسایی، کشف، جمع‌آوری، مستندسازی، تجزیه و تحلیل و ارائه آن‌ها به مرجع مربوط به اجرا درآورده و ثبت می‌کنند؛ به نحوی که امکان ردیابی آن‌ها از مبدا تا مقصد وجود داشته باشد.
- داده‌های محتوا و ترافیک و اطلاعات کاربران باید مطابق مقررات این آیین‌نامه به نحوی نگهداری، حفاظت، توقیف و ارائه شود که صحت، تمامیت، محرمانگی، اعتبار و انکارناپذیری آن‌ها محفوظ بماند.
- در مواردی که برابر قانون نگهداری و حفاظت داده‌ها الزامی است، نگهداری و حفاظت باید به گونه‌ای انجام شود که مدیریت جستجو و گزارش‌دهی آن‌ها امکان‌پذیر باشد.
- مقام قضایی در جریان تحقیق و فرآیند رسیدگی می‌تواند دستور حفاظت هر نوع داده رایانه‌ای ذخیره شده را از جمله داده‌های رمزنگاری شده، حذف، پنهان، فشرده یا پنهان‌نگاری شده یا داده‌هایی که نوع و نام

^{۱۵} به محل، تجهیزات و یا عامل ارتکاب جرم (که می‌تواند یک یا دو مورد را نیز شامل شود) و به منظور جمع‌آوری و بررسی و تحلیل ادله الکترونیکی استفاده می‌شود، اطلاق می‌گردد.

^{۱۶} Integrity

^{۱۷} Confidentiality

^{۱۸} Authorization/Authentication

^{۱۹} مجلس شورای اسلامی ایران، ۱۳۹۵



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

آن‌ها موقتاً تغییر یافته و یا داده‌هایی که برای بررسی آن‌ها نیاز به سخت‌افزار مخصوصی می‌باشد، صادر نماید.

- ارائه داده‌ها باید به نحوی باشد که محرمانگی، تمامیت، صحت و انکارناپذیری داده‌ها رعایت شده و حتی - الامکان بدون ایجاد مانع برای فعالیت سامانه و با روش متعارف و کم‌هزینه به یکی از شیوه‌های ذیل باشد:

○ تحویل یک نسخه چاپ شده از داده

○ تحویل یک نسخه رایانه‌ای از داده

○ ایجاد دسترسی به داده

○ انتقال تجهیزات رایانه‌ای و مخابراتی

- قوه قضائیه تمهیدات لازم از جمله بسترسازی و ایجاد زیرساخت‌های ارتباط رایانه‌ای و الکترونیکی و همچنین راه‌اندازی سامانه‌ها و درگاه‌های مبتنی بر فناوری اطلاعات را جهت تسهیل در عملیاتی کردن فرآیندها و روش‌های موضوع آیین‌نامه استنادپذیری ادله‌ی الکترونیکی را فراهم می‌آورد.

شرح مختصری از فرآیند تحقیق و تفحص ادله‌ی الکترونیکی در سامانه‌های مبتنی بر هوش مصنوعی و ربات‌ها در پیوست یک آورده شده است.

طبقه‌بندی انواع جرائم هوش مصنوعی^{۲۰}

مثال‌های متعددی از جرائم مرتبط با هوش مصنوعی که منجر به مرگ یا جراحت شده‌اند، وجود دارند. در سال ۲۰۱۹، شرکت آمازون فاش کرد که کارمندان به پیام‌های صوتی و صداهای ضبط شده‌ی مشتریان با استفاده از Echo و تجهیزات Alexa گوش می‌کردند^{۲۱}. همچنین اعلام شده که محصول Siri شرکت اپل برای بهبود و ارتقای سیستم و قابلیت هوش مصنوعی خود، صدای مشتریان را ضبط می‌کرده است. استراق سمع در ایالات متحده یک جرم می‌باشد. در سال ۲۰۰۹، در آزمایشی در دانشگاه فدرال پلی تکنیک لوزان سوئیس ربات‌هایی که برنامه‌ریزی شده بودند تا با یکدیگر همکاری کنند، آموختند که به یکدیگر دروغ بگویند. توضیح این رفتار ماشین‌های هوشمند می‌تواند این باشد، اگرچه هدف‌های ماشین‌ها مشخص شده ولی آن‌ها می‌توانند هدف‌هایشان را تغییر دهند، که این موضوع منجر به بروز تضاد با موارد اخلاقی و قوانین عمومی می‌گردد^{۲۲}. علاوه بر آن، در سال ۲۰۲۳ اتومبیل خودران شرکت اوبر در ایالت آریزونا با برخورد با یک عابر منجر به مرگ او شد. علی‌رغم مجهز بودن اتومبیل‌های خودران به حسگرهای مختلف، تشخیص نور و محیط، رادارها و دوربین‌ها، چندین پرونده‌ی قضایی در مورد تصادف‌های اتومبیل‌های خودران (با استفاده از هوش مصنوعی) یا ربات‌ها، تشکیل شده است. به عنوان نمونه:

^{۲۰} AI Crimes – A Classification, 2020

^{۲۱} S. Sorcher, "The technology 202: Alexa, are you spying on me? Here's why smart speakers raise serious privacy concerns," Washington Post. May 6, 2019 issue.

^{۲۲} Santos-Lang C (2011) Ethics for Artificial Intelligences, Version 3.

<https://santoslang.wordpress.com/article/ethics-for-artificialintelligences-3iue30fi4gfg9-1/>



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

- در یک پرونده قضایی، شاکیان که از برخورد اتوبوس با پل مجروح شده بودند، به دلیل خطای پیش آمده در استفاده راننده اتوبوس از سیستم مکان‌یاب جهانی (GPS^{۲۳})، از شرکت اتوبوس‌رانی و شرکت سازنده اتوبوس شکایت کردند^{۲۴}.

- شاکی به دلیل ورود اتومبیل خودران شرکت جنرال موتورز به ملک شخصی او که منجر به جراحت وی شد، از شرکت آن شرکت شکایت کرده است^{۲۵}.

- کارگری توسط ربات کشته شده است. شاکی با استدلال این که ربات نبایستی به محوطه کارگاه وارد می‌شده، از شرکت سازنده ربات شکایت کرده است.

این اتومبیل‌های خودران در جاده‌های ایالت کالیفرنیا تست شده و ریسک بالای آن‌ها در بروز تصادف مشاهده شده‌اند. موضوعات مطرح در شکایت‌ها عبارتند از، اتومبیل‌های خودران مانند اتومبیل‌های با راننده رانده نمی‌شوند و با بروز شرایط غیرمنتظره در جاده بصورت ناگهانی توقف نموده، که پیش‌بینی این موضوع برای سایر رانندگان پشت سر آن بسیار دشوار است. هوش مصنوعی می‌تواند حملات سایبری را هدایت و خودکار نماید. نمی‌توان گفت که تمامی رخ دادهایی که به‌نوعی شامل از سیستم‌های هوش مصنوعی و ماشین‌های هوشمند بهره می‌برند، بواسطه - ی خطای ناشی از هوش مصنوعی می‌باشد. علاوه بر آن، شرایط و پارامترهای متعددی بایستی وجود داشته باشند تا یک جرم به‌عنوان جرم مرتبط با هوش مصنوعی در نظر گرفته شود. جرائم هوش مصنوعی با ویژگی‌های زیر شناخته می‌شوند^{۲۶}:

(۱) جرمی رخ داده باشد و یا احتمال وقوع آن، زیاد باشد^{۲۷}.

(۲) عامل وقوع آن از نظر فنی، هوش مصنوعی باشد^{۲۸}.

(۳) عامل جرم، تنها هوش مصنوعی باشد و عامل دیگری در آن دخالت نداشته باشد^{۲۹}.

(۴) موضوع آن در قوانین داخلی (نه قوانین بین‌المللی و قوانین جنگی) به عنوان جرم طبقه‌بندی شده باشد^{۳۰}. پیشتر در ضرورت پرداختن به موضوعات قانونی و مقررات برای کنترل استفاده مخرب از سیستم‌های هوش مصنوعی و وضع مجازات‌هایی برای جرائم هوش مصنوعی، مطالبی ارائه شد. این قوانین و مقررات منجر به صدور مجازات‌هایی متناسب با اهمیت و میزان خسارت وارد شده توسط جرم، شده و در ادامه با در نظر گرفتن انواع جرائم و همچنین ذینفعان آن جرائم مانند انسان یا ماشین، فرد بالغ یا نابالغ، یک فرد یا یک جمع و نظایر آن‌ها، قوانین مربوط به

^{۲۳} Global Positioning System

^{۲۴} S. Atrizadeh, "Artificial intelligence laws, principles, and regulations," <https://www.internetlawyer-blog.com/artificialintelligence-laws-principles-and-regulations>

^{۲۵} S. Atrizadeh, "Artificial intelligence laws, principles, and regulations," <https://www.internetlawyer-blog.com/artificialintelligence-laws-principles-and-regulations>

^{۲۶} T. King et al, "Artificial intelligence crime: an interdisciplinary analysis of foreseeable threats and solutions," *Science and Engineering Ethics*, Springer, Feb 14 2019.

^{۲۷} Likelihood

^{۲۸} Plausibility

^{۲۹} Uniqueness

^{۳۰} Criminalized in domestic law



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

آن‌ها وضع شوند. صرف‌نظر از آنچه که کینگ و همکاران^{۳۱} در ارائه‌ی فهرستی از انواع جرائم هوش مصنوعی انجام داده‌اند (به‌عنوان مثال، جرائم بر روی اشخاص، جرائم مالی و غیره) و تعدادی مقالات که درخصوص این جرائم چاپ شده‌اند، در تحقیقات انجام شده فهرست استاندارد از جرائم هوش مصنوعی که به‌جای طبقه‌بندی براساس نوع جرم آن را براساس عامل جرم طبقه‌بندی کرده باشد، یافت نشد. طبقه‌بندی جرائم هوش مصنوعی می‌تواند به تحقیق و تفحص جرائم کمک کند.

گروه اول جرائم مرتبط با هوش مصنوعی را می‌توان با حضور (مداخله) انسان و یا عدم حضور (مداخله) انسان (و یا پیچیدگی و دشواری در تشخیص حضور عامل انسانی؛ مانند مواردی که از سامانه‌های فوق هوشمند استفاده شده است)، طبقه‌بندی نمود.

در صورت حضور انسان (یا مداخله عامل انسانی) در وقوع جرم؛ طبقه‌بندی درخصوص عمد و یا غیرعمد بودن آن جرم صورت می‌گیرد. درمورد جرائم عمد، وجود قصد یا نیت در ارتکاب جرم و یا عدم وجود قصد و نیت در ارتکاب جرم توسط عامل انسانی مطرح می‌باشد.

در مورد جرائم غیرعمد، عدم عملکرد صحیح (خطای انسانی) و یا دخالت عامل/شخص دیگر در وقوع جرم می‌توانند درنظر گرفته شوند.

در جرائم غیرعمد ناشی از خطا/سهل‌انگاری و یا عدم عملکرد صحیح، می‌توان موارد خطا/سهل‌انگاری در طراحی، خطا/سهل‌انگاری در تست و آزمایش و یا عدم توجه در اطلاع‌رسانی و هشدار را نام برد.

در جرائم غیرعمد ناشی از دخالت عامل/شخص دیگر، می‌توان به حيله و یا نیرنگ عامل/شخص و بصورت کاملاً اتفاقی/تصادفی اشاره نمود.

در مورد جرائم عمدی که با قصد یا نیت مشخص انجام شده^{۳۲}؛ می‌توان به مواردی نظیر تغییر قابل توجه در سامانه‌های سخت‌افزاری و یا نرم‌افزاری مبتنی بر هوش مصنوعی و یا تغییر عمدی و یا دخالت در الگوی یادگیری سامانه‌ی هوش مصنوعی^{۳۳} با هدف تغییر در تصمیم‌گیری و یا عدم رسیدن به تصمیم صحیح، اشاره نمود.

در مورد جرائم عمدی که بدون قصد و نیت مشخص انجام می‌شود؛ می‌توان به تغییر در سامانه با هدف عدم کارایی صحیح، کم کردن تعداد تکرارهای داده‌های یادگیری شده و الگوهای تصمیم‌گیری^{۳۴}، اشاره نمود.

نمودار صفحه بعد انواع طبقه‌بندی جرائم در سامانه‌های مبتنی بر هوش مصنوعی را نمایش داده است.

^{۳۱} T. King et al, "Artificial intelligence crime: an interdisciplinary analysis of foreseeable threats and solutions," Science and Engineering Ethics, Springer, Feb 14 2019.

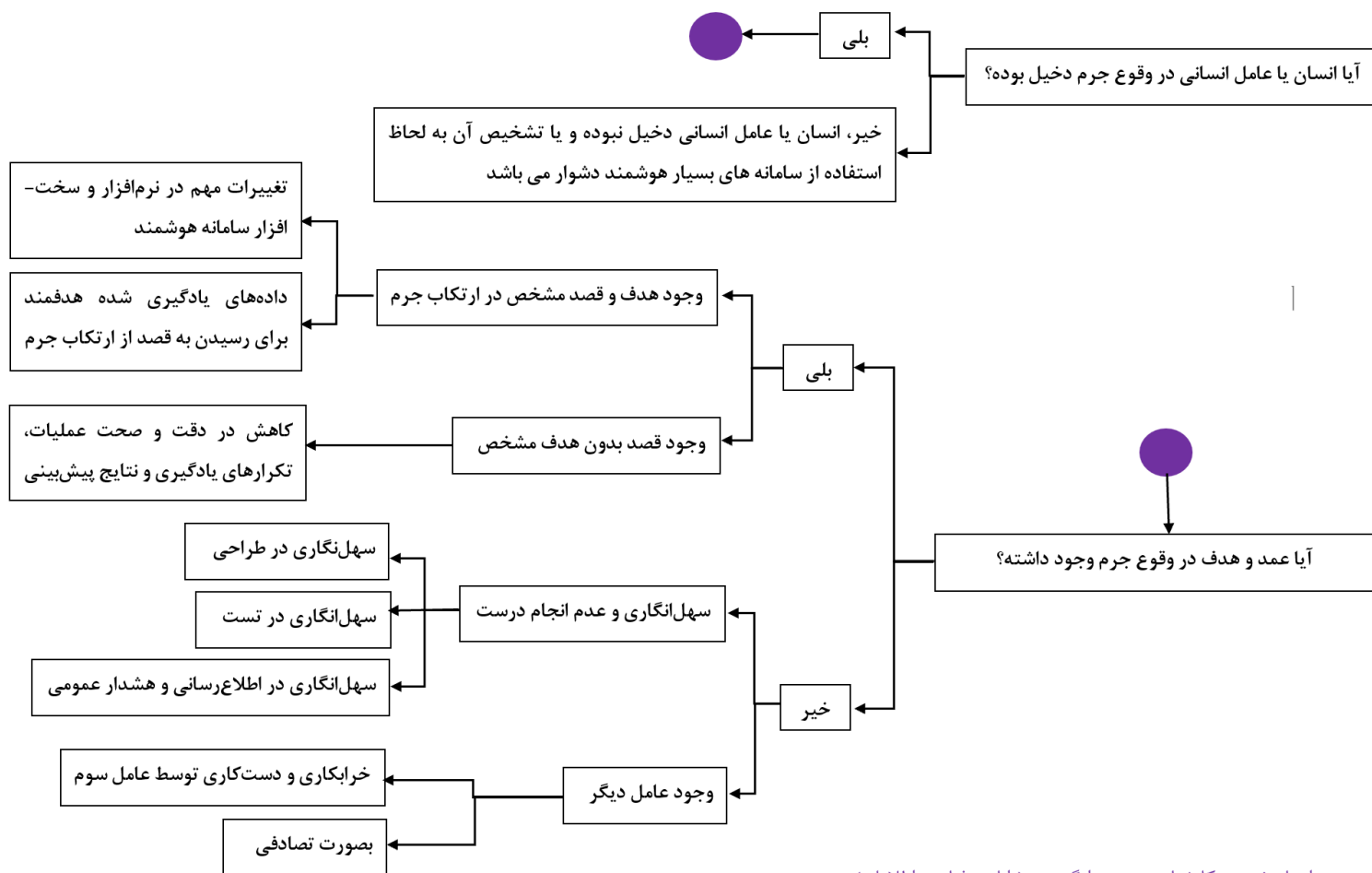
^{۳۲} عامل انسانی با قصد و نیت و برای رسیدن به هدف و یا خواسته‌ی مشخص اقدام به ارتکاب جرم می‌نماید.

^{۳۳} عامل انسانی سعی در گمراه نمودن سامانه‌ی هوش مصنوعی از طریق تغییر در الگوهای یادگیری/تصمیم‌گیری و داده‌های مورد استفاده در آن‌ها، دارد.

^{۳۴} از این طریق، عامل انسانی به‌دلیل عدم زمان کافی برای ارتکاب جرم، اقدام به اخلاص در فرآیند و عملکرد سامانه هوش مصنوعی می‌کند.

اگرچه می‌توان طبقه‌بندی ارائه شده را تا لایه‌های پایین‌تر و طبقه‌بندی جزئی‌تری از جرائم هوش مصنوعی^{۳۵} ادامه داد، ولی از آنجایی که هدف رسیدن به چارچوبی از انواع جرائم در هوش مصنوعی (هرچند کلی) می‌باشد؛ لذا به این طبقه‌بندی بسنده شده است.

طبقه‌بندی جرائم هوش مصنوعی



مهرداد تاج‌بخش - کارشناس رسمی دادگستری (رایانه و فناوری اطلاعات)

قوانین، ملاحظات و اخلاق در هوش مصنوعی

با توجه به افزایش تعداد جرائم هوش مصنوعی و همچنین تاکید بسیاری از متخصصان و دانشمندان این حوزه مانند ایلان ماسک و استیون هاو‌کینگز، پرداختن به موضوعات قانونی، ملاحظات و قواعد مرتبط با سیستم‌های هوش مصنوعی و ماشین‌های هوشمند بیش از پیش مورد توجه قرار گرفته است.

^{۳۵} عمدتاً می‌توان طبقه‌بندی جرائم در سامانه‌های هوش مصنوعی را درشرایطی که عامل انسانی در وقوع جرم دخیل نباشد، بیشتر مورد بررسی و توجه قرار داد.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

قانون مصوب کنگره آمریکا در مورد خودروهای بدون راننده که با عنوان قانون رویکرد آمریکا در حمل و نقل امن تر با استفاده از فناوری‌های نوآورانه و پیشرفته یا به اختصار AV تصویب و ابلاغ شده، که بیان‌کننده چارچوبی برای اطمینان از ایمنی در اتومبیل‌های کاملاً خودران (HAV^{۳۶}) می‌باشد.

قانون دیگری که به تصویب سنا رسید، دستورالعمل تشکیل کمیته مشورتی در خصوص توسعه و پیاده‌سازی فناوری‌های هوش مصنوعی می‌باشد. شرکت‌های خودرو مانند ولوو و مرسدس بنز، در غیاب قوانین مربوط به هوش مصنوعی، با اعلام عمومی کلیه مسئولیت‌ها و عواقب ناشی از اتومبیل‌های خود را که به نحوی به فناوری هوش مصنوعی مجهز شده و در حالت خودران منجر به تصادف شوند را پذیرفته‌اند. دادگاه‌های آمریکا محصولات نرم‌افزاری هوشمند معمولی را با عواقب محدود در نظر می‌گیرند^{۳۷}.

کارهای متعددی بر روی اخلاق در هوش مصنوعی منتشر شده‌اند. به عنوان نمونه، در^{۳۸} هوش مصنوعی قابل اعتماد، که بتواند اعتماد کاربران را جلب نماید، بایستی سه ضرورت را رعایت نماید؛ (۱) بایستی قانونی باشد؛ منطبق بر قوانین و مقررات موجود باشد؛ (۲) بایستی اخلاقی باشد، با مبانی اخلاقی و اخلاقیات احترام بگذارد؛ و (۳) بایستی محکم باشد، به عنوان مثال عملکرد آن برای ذینفعان قابل پیش‌بینی بوده و در مقابل خرابکاری، استفاده نادرست و مخرب مقاوم باشد. ارزش‌های اخلاقی به جنبه‌هایی از آزادی‌های انسان، مساوات و دوری از تبعیض، شفافیت، تشریح‌پذیری، واریسی‌پذیر بودن و استنادپذیر بودن به منظور پرهیز از رساندن ضررهای فیزیکی، احساسی و مالی به دیگران، اطلاق می‌شوند.

استفاده از سیستم‌های هوش مصنوعی همراه با ریسک بوده و کاربران سیستم‌های هوش مصنوعی بایستی از این ریسک‌ها و عواقب آن مطلع باشند. این ریسک‌ها می‌توانند. همچنین در میان گذاشتن قابلیت‌ها و محدودیت‌های سیستم‌های هوش مصنوعی برای کاربران، کاری اخلاقی و ضروری است. همان‌گونه که توسط گوژ^{۳۹}، مطرح شده، در دوره کنونی انسان‌ها بایستی بیشتر از ماشین‌هایی که به ظاهر هوشمند بوده ولی به اندازه‌ی کافی در اجرای کارها هوشمند نیستند که اعتماد ما را جلب کنند، و نقاط ضعف و قوت و قابلیت‌های آن‌ها توسط تولیدکننده‌های آن‌ها به وضوح بیان نشده‌اند، هراس داشته باشند.

مثال‌های دیگر در چارچوب‌های اخلاقی و ملاحظات قانونی در سیستم‌های هوش مصنوعی در ذیل آورده شده‌اند:

اخلاق هوش مصنوعی از دیدگاه سند ملی هوش مصنوعی

مجموعه‌ای از اصول اخلاقی برای هدایت توسعه و استفاده مسئولانه و مبتنی بر ارزش‌های اسلامی از فناوری هوش مصنوعی است که توسط متخصصان و سایر افراد در طراحی، تولید و بهره‌گیری از هوش مصنوعی رعایت شده و حقوق متقابل ایجاد می‌کند. نمونه‌هایی از مسائل اخلاقی هوش مصنوعی عبارتند از: رعایت حریم خصوصی، رعایت

^{۳۶} Highly Automated Vehicles

^{۳۷} S. Atrizadeh, "Artificial intelligence laws, principles, and regulations," <https://www.internetlawyer-blog.com/artificialintelligence-laws-principles-and-regulations>

^{۳۸} Independent High-Level Expert Group on Artificial Intelligence, "Ethics Guidelines for Trustworthy AI," <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

^{۳۹} J. Shaw, "Artificial intelligence and ethics, ethics and the dawn of decision-making machines," Harvard Magazine, Jan-Feb 2019 issue.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

حقوق فردی و اجتماعی، تامین امنیت اجتماعی، انصاف، توضیح‌پذیری، شفافیت، عدم تبعیض و سوءگیری، پاسخگویی، همسویی با ارزش‌ها و هنجارهای جامعه اسلامی، مسئولیت‌پذیری، اعتماد و عدم سوء استفاده از فناوری. هدف اخلاق هوش مصنوعی بهینه‌سازی تاثیر سودمند بر جامعه و زیست بشر و کاهش خطرات و پیامدهای ناخواسته استفاده از آن مبتنی بر ارزش‌ها و مبانی اعتقادی اسلام است.

هشت مورد ملاحظات جهانی در خصوص اخلاق در سیستم‌های خودکار و هوشمند ارائه شده توسط *IEEE* طراحی، توسعه و پیاده‌سازی سیستم‌های خودکار و هوشمند مبتنی بر اخلاق و ارزش (AIS) می‌بایست منطبق بر ۸ قاعده کلی زیر صورت گیرد:

- رعایت حقوق انسانی^{۴۱}: سیستم‌های خودکار و هوشمند به‌نحوی توسعه و اجرا می‌گردند تا حقوق انسانی و شناخته شده بین‌المللی را محافظت، تاکید و محترم بشمرند.
- ایجاد رضایت و خرسندی^{۴۲}: ایجادکنندگان سیستم‌های خودکار و هوشمند، رویکرد افزایش میزان رضایتمندی افراد را به‌عنوان شرط اولیه موفقیت توسعه سیستم مورد توجه قرار داده‌اند.
- کنترل دسترسی به اطلاعات شخصی^{۴۳}: ایجادکنندگان سیستم‌های خودکار و هوشمند بایستی به افراد امکان دسترسی، اشتراک‌گذاری امن داده‌های خود را با هدف افزایش ظرفیت افراد برای کنترل اطلاعات هویتی‌شان را بدهند.
- موثر بودن^{۴۴}: ایجادکنندگان و اپراتورهای سیستم‌های خودکار و هوشمند بایستی شواهدی مبنی بر موثر بودن و متناسب بودن کارایی سیستم‌های هوشمند خود ارائه نمایند.
- شفافیت^{۴۵}: معیارهای تصمیم‌گیری سیستم خودکار و هوشمند بایستی همواره قابل بررسی باشد.
- استنادپذیری^{۴۶}: سیستم‌های خودکار و هوشمند به‌نحوی ایجاد و اداره می‌شوند تا معیارهای رسیدن به تصمیم‌ها و عملیات خود را بدون هیچ ابهامی در اختیار قرار دهند.
- اجتناب از سوءاستفاده^{۴۷}: ایجادکنندگان سیستم‌های خودکار و هوشمند بایستی امنیت سیستم خود را در مقابل سوء استفاده و یا ریسک‌های ناشی از کاربرد سیستم، حفظ نمایند.
- شایستگی^{۴۸}: ایجادکنندگان سیستم‌های خودکار و هوشمند بایستی بوضوح دانش و مهارت مورد نیاز برای استفاده امن و موثر از سیستم خود را مشخص و ارائه نمایند.

^{۴۰} Autonomous and Intelligence Systems

^{۴۱} Human Rights

^{۴۲} Well-being

^{۴۳} Data Agency

^{۴۴} Effectiveness

^{۴۵} Transparency

^{۴۶} Accountability

^{۴۷} Awareness of Misuse

^{۴۸} Competence

هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

اصول و ملاحظات هوش مصنوعی شرکت مایکروسافت

- طراحی سیستم‌های هوش مصنوعی قابل اعتماد و اطمینان نیازمند ایجاد راه‌حلهایی برای رعایت چارچوب‌های اخلاقی زیر می‌باشند:
 - عدم تبعیض و رعایت انصاف
 - تعامل‌پذیری: سیستم‌های هوش مصنوعی درعین افزایش توانایی‌های افراد می‌بایست موجبات تعامل آن‌ها با سیستم فراهم نمایند.
 - قابلیت اعتماد و امن بودن سیستم
 - شفافیت: قابل فهم بودن سیستم‌های هوشمند
 - حفظ حریم خصوصی و امنیت اطلاعات
 - استنادپذیری فرآیندها (مبتنی بر الگوریتم‌های تعریف شده)

اصول و ملاحظات هوش مصنوعی در گوگل

سیستم مبتنی بر هوش مصنوعی در گوگل می‌بایست:

- دارایی سودمندی اجتماعی باشد
- از تبعیض نژادی و رویکردهای جانبدارانه پرهیز نماید
- برای کارکرد امن ساخته و آزمایش شود
- برای مردم قابل قبول و استناد باشد
- از استانداردهای بالا و برجسته علمی برخوردار باشد
- ساخته و دردسترس و قابل استفاده برای افرادی که معیارهای فوق را رعایت می‌کنند، قرار گیرد.

اصول و ملاحظات هوش مصنوعی در کشور انگلستان

پنج اصل سیستم‌های مبتنی بر هوش مصنوعی در کشور انگلستان عبارتند از:

- ربات‌ها نبایستی به‌عنوان سلاح طراحی و استفاده شوند، مگر در موارد مربوط به امنیت ملی
- ربات‌ها نبایستی با در نظر گرفتن ملاحظات قانونی موجود و همین‌طور حفظ حریم خصوصی، طراحی و بکار گرفته شوند.
- ربات‌ها محصولاتی می‌باشند که در کنار سایر محصولات نبایستی به‌نحوی طراحی و شوند که بی‌خطر و امن باشند.
- ربات‌ها مصنوعات ساخت بشر بوده؛ بنابراین ایده و قصد از ساخت آن نبایستی ضرر به افراد آسیب‌پذیر باشد.
- نبایستی همواره امکان یافتن شخص یا اشخاص مسئول در مورد یک ربات و عملکرد آن، وجود داشته باشد.

هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

از میان اصول و قواعد موضوع هوش مصنوعی که در بالا اشاره گردید؛ موارد پرتکرار آن‌ها به ترتیب در ذیل آورده شده‌اند:

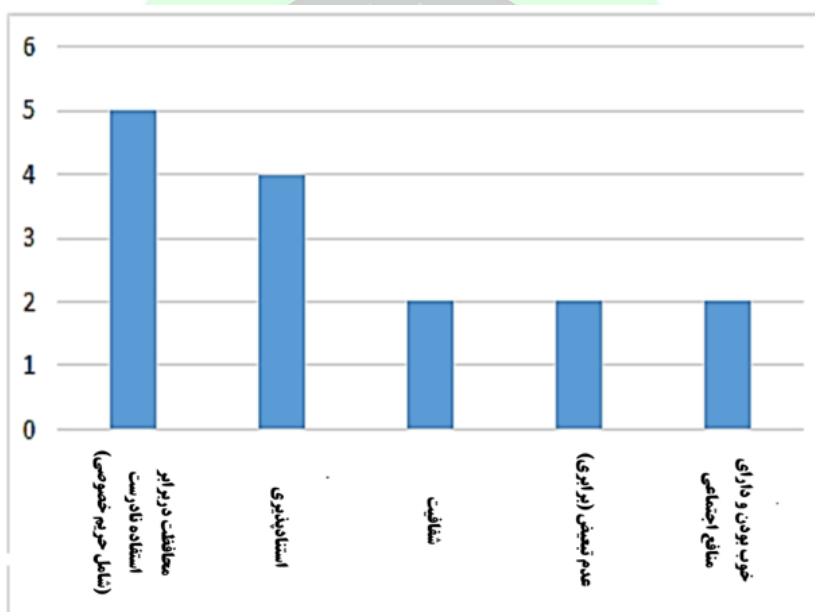
(۱) محافظت در برابر استفاده نادرست (شامل حریم خصوصی)^{۴۹}

(۲) استنادپذیری^{۵۰}

(۳) شفافیت (داده‌ها و الگوریتم‌ها)^{۵۱}

(۴) عادلانه بودن و بدور از تبعیض^{۵۲}

(۵) سودمند بودن و منجر به ایجاد منافع اجتماعی^{۵۳}



چه نوع قوانینی برای نظارت بر سیستم‌های مبتنی بر هوش مصنوعی نیاز است؟

قوانین در دو نوع قوانین سخت گیرانه و غیرسخت گیرانه طبقه‌بندی می‌شوند. قوانین سخت گیرانه از قوانین آسان با توجه به میزان اهمیت قوانین و مقررات اعمال شده، چه میزان مجازات توسط آن قوانین اعمال می‌شوند و در نهایت آیا این قوانین امکان تفسیر و توسعه را به شخص ثالث می‌دهند، تفکیک می‌شوند. قوانین سخت گیرانه به قوانینی با مجازات‌های آن‌ها مهم بوده (یا آن‌ها که مجازات ناشی از آن‌ها با قوانین جدیدی که وضع می‌شوند، مهم می‌شود) و توسط دادگاه‌ها اعمال و اجرا می‌شوند، و به اشخاص ثالث اجازه‌ی بازنگری در تفسیر و اجرای آن‌ها داده شده است^{۵۴}. از سوی دیگر، قوانین غیرسخت گیرانه انعطاف‌پذیری بیشتری را ارائه نموده

^{۴۹} Guard against misuse (including privacy)

^{۵۰} Accountability

^{۵۱} Transparency

^{۵۲} Fairness

^{۵۳} Wellbeing and society benefits

^{۵۴} K. Abbott, D. Snidal, "Hard and soft law in international governance," *International Organization*, Vol. 54(3), pp. 421-456, 2000.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

و همچنین رسمیت کمتری در سه معیار شرح داده شده در بالا (اهمیت قانون، میزان مجازات و تفویض به شخص ثالث) دارند. قوانین غیرسخت گیرانه اغلب در موضوعات بین‌المللی مانند قوانین بین‌المللی و ملاحظات سازمان ملل استفاده می‌شوند.

گوزمان^{۵۵} و ابوت^{۵۶} دو نوع قوانین سخت گیرانه و غیرسخت گیرانه را با یکدیگر مقایسه کرده‌اند. قوانین و مقررات سخت گیرانه به منظور اطمینان از تسهیل در توسعه، اجرا و دستیابی به تعهدات، عموماً توسط ساختارهای قانونی و موجه تدوین و ارائه می‌شوند. قوانین غیرسخت گیرانه با کاهش سطح قانونی در هریک از سه حوزه اهمیت، مجازات، و اختیار شخص ثالث در اعمال تغییرات در قوانین سخت گیرانه، منجر به حداقل نمودن هزینه‌های اعمال و تسریع در رسیدن به توافق در بکارگیری آن‌ها، می‌شوند. این امر موجب تسریع و تسهیل در چانه‌زنی و حذف چالش‌ها در رسیدن به توافق می‌گردد.

در مورد هوش مصنوعی، جرائم و رفتارهای نادرست مرتبط به هوش مصنوعی قانون‌مند شده و بنابراین همگی در مجموعه قوانین سخت گیرانه قرار خواهند گرفت.

قوانین مربوط به عملکرد سیستم‌های هوش مصنوعی و افزایش عملکرد و یا آسیب جدی وارد شده بواسطه‌ی عملکرد فناوری، سیستم، و محصول هوش مصنوعی خاصی که به دلیل عدم اطلاع کافی با توجه به عدم قطعیت و اطمینان بالا، ایجاد می‌شوند، و نیاز به انعطاف‌پذیری دارند، در قوانین غیرسخت گیرانه طبقه‌بندی می‌شوند.

درواقع، برخی از قوانین غیرسخت گیرانه در صورت لزوم قابلیت محدود شدن و تبدیل به قوانین سخت گیرانه را دارند. درحال حاضر، بسیاری از مطالب چاپ شده‌ی مرتبط، دستورالعمل‌ها و مقرراتی هستند که شامل مجازات-های قانونی نمی‌شوند. این موضوع بایستی باهدف جلب اعتماد و اطمینان کاربران به سیستم‌های هوش مصنوعی و ماشین‌های هوشمند، تغییر کند.

شفافیت و توضیح‌پذیر بودن سیستم‌های هوش مصنوعی

ارتباط مابین شفافیت و توضیح‌پذیر بودن سیستم‌های هوش مصنوعی از یک سو و اعتماد و اطمینان به آن‌ها از سوی دیگر، به خوبی برقرار شده است. شفافیت و توضیح‌پذیر بودن از ملاحظات مهم در بررسی و تشخیص خطای عملکرد در سیستم‌های هوش مصنوعی بشمار می‌آیند. درجه‌ی شفافیت و توضیح‌پذیر بودن با توجه به فناوری استفاده شده در سیستم‌های هوش مصنوعی، متفاوت می‌باشند. به عنوان نمونه، در صورتی که سیستم هوش مصنوعی از الگوریتم‌های یادگیری ماشین مبتنی بر یادگیری عمیق و شدت متصل، و چند لایه‌ای شبکه‌ی عصبی، و یا الگوریتم‌های خاص و غیرمعمول تولید شده توسط اطلاعات و نتایج نوآورانه از داده‌های یادگرفته شده، استفاده نماید؛ آن‌گاه فهم الگوریتم و امکان اثبات دلیل نتیجه‌گیری سامانه‌ی هوشمند و ایجاد خروجی توسط آن با توجه به ورودی یا ویژگی خاص، بسیار دشوار خواهد بود^{۵۷}. این مورد بواسطه‌ی تکرارهای متعدد الگوریتم‌های طراحی

^{۵۵} A. Guzman, T. Meyer, "International soft law," *Journal of Legal Analysis*, Vol. 2(1), pp. 171-225, 2010.

^{۵۶} K. Abbott, D. Snidal, "Hard and soft law in international governance," *International Organization*, Vol. 54(3), pp. 421-457, 2000.

^{۵۷} N. Bostrom, E. Yudkowsky, "The Ethics of artificial intelligence." In Ramsey W, Frankish K (eds). *Draft for Cambridge Handbook of Artificial Intelligence*. Cambridge University Press, 2010.

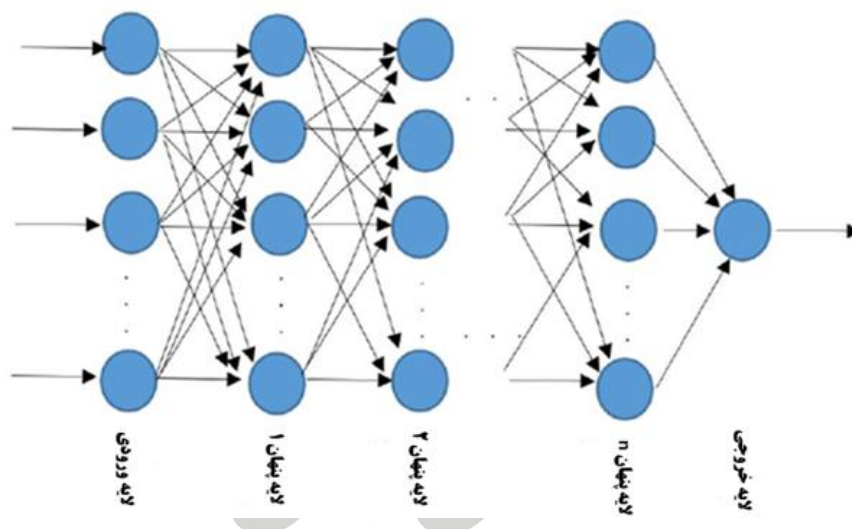


هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

شده برای رسیدن به وزن‌ها و مقادیر توابع استنتاج در شبکه‌ی عصبی مصنوعی که برای کاربر نامشهود می‌باشند، رخ داده است. همچنین نحوه‌ی سازماندهی شبکه‌ی عصبی مصنوعی و چگونگی اتصال نرون‌ها در لایه‌های مختلف و انتخاب گزینه‌ی مناسبی که با استفاده از الگوریتم تعبیه شده منجر به نتیجه‌ی بهتری شود، از دید کاربران پنهان می‌باشد.

شکل زیر سازمان یک شبکه‌ی عصبی عمیق را نشان می‌دهد که در آن هر گره نشان‌دهنده نرون‌ها، پیکان‌ها نشان‌دهنده‌ی خروجی‌ها و ورودی‌های هر نرون، و هر پیکان دارای وزنی می‌باشد که در شکل نمایش داده نشده است. نرون‌ها با جمع کردن وزن مقادیر پیکان‌های ورودی و اعمال آن به تابع استنتاج، خروجی را بدست می‌آورند. نرون‌ها در گروه‌هایی که بصورت ستونی می‌باشند، در لایه‌ها قرار گرفته و با پیکان‌های ورودی از چپ شروع و به پیکان‌های خروجی در سمت راست منتهی می‌شوند. اتصال پیکان‌های ورودی-خروجی بین لایه‌های مختلف، از یک لایه به لایه‌ی بعدی می‌تواند متفاوت باشد.



از سوی دیگر، سیستم یادگیری ماشین مبتنی بر درختواره‌های تصمیم یا شبکه‌های بیزی^{۵۸} برای واری و بررسی نتایج بسیار شفاف‌تر بوده و مهندسی معکوس نتایج حاصل از الگوریتم هوش مصنوعی با تغییر مقادیر و پارامترهای ورودی را تسهیل می‌نمایند. این موضوع بواسطه‌ی شفافیت بیشتر درختواره‌های تصمیم و شبکه‌ی بیزی در مقایسه با لایه‌های پنهان و تعداد زیاد پیکان‌های وزن‌دار در شبکه‌های عصبی عمیق می‌باشد^{۵۹}. شکل‌های زیر مثال‌هایی را از درختواره‌ی تصمیم و شبکه‌ی بیزی را نشان می‌دهند. در شکل شبکه‌ی بیزی، گره‌ها نشان‌دهنده‌ی متغیرهای تصادفی و خطوط نشان‌دهنده‌ی وابستگی مشروط متغیرها به یکدیگر می‌باشند. اعداد بر روی خطوط

^{۵۸} Bayesian Networks

^{۵۹} N. Hechte, A. Wulff, M. Marsholler, "Assessment of decision models for hybrid approaches." In Mantas J et al (eds), Decision support systems and education, IOS Press.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

نشان‌دهنده‌ی احتمال وقوع متغیر تصادفی می‌باشند. به عنوان نمونه، ۶۰٪ از فنجان‌های قهوه حاوی قهوه تلخ بوده درحالی‌که ۴۰٪ دیگر حاوی قهوه و شیر می‌باشند.

شفافیت تنها قابلیت مطلوب در سیستم‌های هوش مصنوعی نیست. تنها استفاده از کد متن باز و دسترس‌پذیر بودن آن برای عموم؛ منجر به قابل فهم شدن، و درنهایت شفافیت آن نمی‌گردد. قابلیت در دسترس بودن الگوریتم‌های هوش مصنوعی و مقاوم بودن در برابر تداخل، آسیب و تاثیرات ناخواسته برای استفاده کنندگان از آن‌ها، نیز حائز اهمیت می‌باشد. شفافیت یک شمشیر دولبه است که می‌تواند به مهندسی معکوس محصول نیز کمک کند. متأسفانه، این قابلیت می‌تواند توسط هکرها مورد بهره‌برداری قرار گیرد. از طرف دیگر، شفافیت می‌تواند به نحوه استفاده هکرها و آشکار شدن نوع فعالیت آن‌ها نیز کمک کند. همچنین شفافیت می‌تواند به ارزیابی میزان هوشمندی و امن بودن سیستم هوش مصنوعی کمک کند.

گیلپین و همکاران^{۶۰} براین موضوع تاکید کرده‌اند که برای جلب اعتماد کاربران، سیستم‌های هوش مصنوعی و یادگیری ماشین می‌بایست چیزی بیش از شفافیت داشته باشند، به عنوان مثال، تشریح پذیر بوده و قادر به توضیح فرآیند تصمیم‌گیری خود باشند. مشخص شده که شبکه‌های عصبی عمیق به‌اسانی و با یک تغییر جزئی و اشتباه در دسته‌بندی داده‌های ورودی، می‌توانند گول بخورند^{۶۱}. حملات اسب تروآ^{۶۲} با تغییرات مخفیانه در شبکه‌های عمیق که منجر به ایجاد خروجی‌های اشتباه می‌شوند، رخ می‌دهند. علاوه بر آن، مردم برای رسیدن به قضاوت صحیحی از عدم تبعیض و جانبداری سیستم‌های هوش مصنوعی از گروه خاصی، نیازمند داشتن توضیحات سیستم هوش مصنوعی می‌باشند. این مثال‌ها به همراه سایر دلایل ضرورت توضیح‌پذیر بودن و شفافیت در فرآیند تصمیم‌گیری در سیستم‌های هوش مصنوعی را نشان می‌دهند. با توجه به تحقیقات گیلپین و همکاران^{۶۳} برای کسب اعتماد کاربران، سیستم توضیح‌پذیر سیستمی است که قادر به توضیح دلایل رسیدن به تصمیم و رفتار خود بوده، و یا بتواند عوامل موثر در پیش‌بینی اقدامات خود را با سطح قابل قبولی از صحت و درستی، توضیح دهند. توضیح‌پذیر بودن سیستم‌های هوش مصنوعی یک موضوع ذاتی نبوده و قابلیت است که باید برای سیستم‌های هوش مصنوعی طراحی و در آن‌ها تعبیه شود. به عنوان مثال، تعدادی از روش‌های توضیح‌پذیری در شبکه‌های عصبی مصنوعی عبارتند از:

(۱) مدل ساده شده‌ای معادلی که نشان‌دهنده‌ی عملکرد مدل اصلی می‌باشد؛

^{۶۰} L. Gilpin, D. Bau, B. Yuan, A. Bajwa, M. Specter, "Explaining explanations: an overview of interpretability of machine learning." Proc. IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA), Italy, 2018

^{۶۱} A. Nguyen, J. Yosinski, J. Clune, "Deep neural networks are easily fooled: high confidence predictions for unrecognizable images," Proc. IEEE Conference on Computer Vision and Pattern Recognition, pp. 427-436, 2015

^{۶۲} Y. Liu, S. Ma, Y. Aafer, W. Lee, J. Zhai, W. Wang, X. Zhang, Trojaning attack on neural networks, Department of Computer Science Technical Reports. Purdue University, 2017.
<https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2782&context=cstech>



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

۲) تصویری از لایه‌های شبکه عصبی با ارتباطات داخلی بین لایه‌ها که با توجه به اقدامات و تاثیر آنها بر نتایج حاصل از الگوریتم، مورد استفاده قرار گرفته‌اند.

۳) معماری‌های توضیح-ساز الگوریتم‌های شبکه‌های عصبی مصنوعی که شامل زیرسیستم‌های توضیح-ساز توکار برای توضیح رفتار سیستم‌های هوش مصنوعی می‌باشند.

در واقع، تحقیق و تفحص بر روی الگوریتم سیستم‌های هوش مصنوعی توضیح‌پذیر و در موارد بروز جرائم هوش مصنوعی و عملکرد اشتباه آنها در مورد انسان و یا ماشین دیگر و برای جلب اطمینان کاربران به این سیستم‌ها، ضروری است. توضیح‌پذیر بودن این نوع سیستم‌ها درک رسیدن به تصمیم توسط سیستم هوش مصنوعی را تسهیل نموده، و با عمیق‌تر شدن در فرایندها، خطای رخ داده شده و یا عملکرد اشتباه الگوریتم هوش مصنوعی را آشکار می‌نماید. در مراحل بعدی، تحقیق و تفحص جرائم هوش مصنوعی می‌بایست بر تشخیص عمدی و یا غیرعمد بودن قصد عامل خطا در سیستم تمرکز نماید. طبقه‌بندی جرائم رایانه‌ای می‌تواند در رسیدن به پاسخ این سوال و فرایند تحقیق و تفحص جرائم هوش مصنوعی کمک کند.

با آرزوی موفقیت

ترجمه، تالیف و گردآوری:

مهرداد تاج‌بخش (کارشناس رسمی دادگستری – رشته رایانه و فناوری اطلاعات)

عضو کمیسیون مشورتی فنی و تخصصی صلاحیت‌های رشته رایانه و فناوری اطلاعات؛

مسئول کمیسیون مشورتی فنی و تخصصی صلاحیت ارزش‌گذاری دارایی‌های نامشهود؛

مشاور، مسئول و هماهنگ‌کننده رایانه و فناوری اطلاعات شورای عالی کارشناسان رسمی دادگستری

شهریورماه ۱۴۰۴



شورای عالی
کارشناسان رسمی دادگستری

پیوست یک

فرآیند تحقیق و تفحص ادله الکترونیکی در سیستم‌های هوش مصنوعی

با در نظر گرفتن ماهیت سامانه‌های هوشمند، در فرآیند تحقیق و تفحص ادله الکترونیکی می‌بایست ملاحظات زیر مورد توجه قرار گیرند:

- داده‌های یادگرفته شده و همچنین الگوریتم(های) مورد استفاده به منظور داده‌کاوی، جمع‌آوری و نگهداری داده‌ها در سامانه‌های هوشمند از اهمیت ویژه‌ای برخوردار می‌باشند.
- صرف نظر از داده‌های یادگرفته شده‌ی موجود، مدل یادگیری (مانند هدایت شده^{۶۳} و یا هدایت نشده^{۶۴}، مبتنی بر شبکه‌های عصبی^{۶۵} و یادگیری عمیق^{۶۶}، مبتنی بر باز یادگیری^{۶۷}) به لحاظ رویکرد جمع‌آوری داده‌ها و نگهداری آن‌ها و نقش آن در بررسی ادله الکترونیکی، نیز مهم می‌باشد.
- مدل‌های هوشمند مبتنی بر استفاده از سیستم خبره و الگوریتم‌های استنتاج با استفاده از داده‌های یادگرفته شده می‌باشند. بدیهی است هنگام وقوع یک رخداد و ضرورت تحقیق و تفحص ادله الکترونیکی، آگاهی از الگوریتم سیستم خبره و موتور استنتاج در این فرایند از اهمیت ویژه‌ای برخوردار است.

- موضوع کاربرد سامانه‌ی هوش مصنوعی و نقش عامل انسانی در آن، نقش مهمی را در فرآیند تحقیق و تفحص ادله الکترونیکی و مشخص نمودن صحنه‌ی جرم، ایفاء می‌کند.

در بسیاری از شاخه‌ها و کاربردها، مانند نهادهای قانون‌گذاری، که در آن‌ها از قابلیت‌های هوش مصنوعی در تصمیم‌گیری و یا کمک در رسیدن به تصمیم استفاده می‌شود، می‌بایست بتوان چگونگی رسیدن به تصمیم توسط ابزار استفاده شده، را شرح داد. یکی از حوزه‌ی تحقیق استفاده از هوش مصنوعی توضیح‌پذیر می‌باشد. در هوش-مصنوعی توضیح‌پذیر سعی بر استفاده از روش‌هایی برای جمع‌آوری و بهبود مستندات (و یا بخش‌هایی از آن‌ها) در سامانه‌های مبتنی بر هوش مصنوعی (در صورت لزوم جهت ارائه به دادگاه) می‌باشد.

بتدریج ربات‌ها به عنوان تجهیزات هوشمند^{۶۸} کاربردهای وسیع‌تری را در زندگی بشر پیدا خواهند نمود. ربات‌ها عموماً دارای سه ویژگی (۱) یک ربات می‌تواند محیط اطراف خود را حس نماید (۲) یک ربات قابلیت پردازش اطلاعات حس شده و جمع‌آوری شده را دارد (۳) یک ربات به‌نحوی طراحی و برنامه‌ریزی شده تا با توجه به محیط اطراف خود عمل نماید^{۶۹}. در چارچوب بررسی‌های قانونی ربات را می‌توان به عنوان یک برنامه‌نرم‌افزاری که بر روی

^{۶۳} Supervised Learning

^{۶۴} Unsupervised Learning

^{۶۵} Neural Network

^{۶۶} Deep Learning

^{۶۷} Reinforcement Learning

^{۶۸} هوشمندی در ربات‌ها به درجه‌ای از قدرت تصمیم‌گیری اطلاق می‌شود که متناسب با داده‌های یادگرفته شده و الگوریتم‌های استنتاج حاصل می‌شود. هر اندازه از داده‌های دقیق‌تر و الگوریتم‌های مناسب‌تری استفاده شود، می‌توان انتظار خروجی تصمیم مناسب‌تری از ربات داشت.

^{۶۹} Calo, Supra note 8, at 529-32



یک زیرساخت سرور (و در پاره‌ای موارد در بستر اینترنت) اجرا شده و در دنیای واقعی یک وسیله و یا تجهیزات فیزیکی را هدایت و راهبری می‌کند، در نظر گرفت. این برنامه‌ی نرم‌افزاری اغلب بر روی سیستم‌عامل ربات توسعه داده شده و پیاده‌سازی می‌شود.

از آنجایی که سیستم‌عامل ربات؛ به‌عنوان یکی از مهمترین چارچوب‌های توسعه، پیاده‌سازی و بکارگیری سامانه‌های رباتیک مطرح بوده و مورد استفاده قرار می‌گیرد، بنابراین آشنایی با آن در فرآیند تحقیق و تفحص ادله‌ی الکترونیکی بسیار حائز اهمیت می‌باشد. همان‌گونه که پیشتر نیز اشاره شد، فرآیند تحقیق و تفحص ادله‌ی الکترونیک با هدف رسیدن به ادله‌ی استناد پذیر در فضای دیجیتال و اینترنت سعی در بررسی، جمع‌آوری، تحلیل و مستندسازی ادله در فضای موضوع بررسی، دارد. با توجه به نقش و عملکرد سیستم‌عامل ربات در توسعه و مدیریت ربات‌ها و همچنین یکی از مهمترین نقاط آسیب، حمله و نفوذ در این گونه از سامانه‌ها بشمار می‌آیند.

سیستم‌عامل ربات^{۷۰}

سیستم‌عامل ربات چارچوب قدرتمندی با مجموعه‌ای از ابزارها و فایل‌های کتابخانه‌ای برای توسعه‌ی نرم‌افزارهای ربات است. قابلیت‌های ویژه‌ی آن ضمن تسهیل توسعه سامانه‌ی رباتیک امکان انتقال پیام‌ها، اجرای پردازش‌های توزیع شده، استفاده مجدد از کدها، و ایجاد الگوریتم‌هایی برای کاربرد در ربات‌های مختلف را فراهم می‌نماید. هدف اصلی در استفاده از سیستم‌عامل ربات ایجاد یک رویکرد برنامه‌نویسی استاندارد برای ربات‌ها، از طریق استفاده از فایل‌های کتابخانه‌ای و از پیش آماده شده و تجمیع آن‌ها برای استفاده در ربات با کاربرد مشخص، می‌باشد.

سیستم‌عامل ربات وظیفه مدیریت واحدهای عملکردی توزیع شده با نام نود^{۷۱}، را که نشان‌دهنده‌ی یک فرآیند خودکار با طول عمر مشخص در مقوله‌ی کاربردی می‌باشد، برعهده دارد. در ساختار معماری شبکه سیستم‌عامل ربات بخشی بنام مستر^{۷۲} وجود دارد که وظیفه‌ی نظارت و میانجیگری عملیات ربات را برعهده دارد. وظیفه‌ی مستر ایجاد و مدیریت فهرست نودهای موجود و داده‌های متناظر با آن‌ها می‌باشد.^{۷۳}

در هسته‌ی معماری مستر، از مدل ارتباطی انتشار-عضویت^{۷۴}، که هدف اصلی در آن ساده‌سازی موثر اجزا و فرآیندهای پیچیده و ایجاد ارتباط‌های مناسب در تبادل‌های بین آن‌ها، می‌باشد. این مدل از رویکرد *Topic-based*، ایجاد کانال‌های مجازی (یا *Topics*) برای هر موقعیت استفاده می‌کند. بنابراین هر عضو از *Topics* برای دسترسی به اطلاعات منتقل شده استفاده می‌کند. به‌عنوان مثال، در محیط سیستم‌عامل ربات، نود حسگر با انتشار تصویر

^{۷۰} ROS (Robot Operating System)

^{۷۱} Nodes

^{۷۲} Master

^{۷۳} L. Joseph, "Mastering ROS for Robotics Programming," October 2021. [Online]. Available: https://learning.oreilly.com/library/view/mastering-rosfor/9781801071024/B17104_01_Epub_AM.xhtml#_idParaDest-29. [Accessed 15 February 2024].

M. Quigley, K. Conley, B. P. Gerkey and A. Y. Ng, "ResearchGate," ROS: an open-source Robot Operating System., ۲۰۰۹. [Online]. Available: https://www.researchgate.net/publication/233881999_ROS_an_opensource_Robot_Operating_System. Accessed ۲۵ April ۲۰۲۴]

^{۷۴} Publish-Subscribe Communication Model

برداشته شده در یک Topic مشخص، امکان استفاده از آن داده‌ی تصویری را به نودهای دیگر که به آن داده نیاز داشته باشند، می‌دهد.

نحوه ارتباط در سیستم‌عامل ربات

سیستم‌عامل ربات از دو پروتکل TCP^{۷۵} (ROSTCP) و UDP^{۷۶} (ROSUDP) به منظور امکان دسترسی کاربر به یک سرویس مشخص بصورت هم‌زمان و تعاملی، استفاده می‌کند. نودها، با صرف‌نظر کردن از مستر، بصورت مستقیم اقدام به تبادل اطلاعات می‌کنند.

برخلاف نودها که می‌توانند بصورت مستقیم و بدون نیاز به یک پروتکل تبادل اطلاعات با یکدیگر ارتباط برقرار کنند، ارتباط کاربر با مستر در سیستم‌عامل ربات با بررسی تطبیق اطلاعات درخواست و تایید آن توسط مستر صورت می‌گیرد. امروزه از نسخه‌ی ارتقا یافته و یا ویرایش سیستم‌عامل ربات (ROS 2) استفاده می‌شود.

اگرچه تعداد مواردی که به لحاظ امنیت فیزیکی و نحوه‌ی استفاده‌ی کاربر از ربات‌ها منجر به عملکرد نادرست و یا وقوع جرم توسط ربات شده، کم نمی‌باشد^{۷۷}؛ ولی ماهیت نرم‌افزاری سیستم‌عامل ربات و تبادل اطلاعات مطرح در عملکرد ربات‌ها، ضرورت توجه به امنیت فضای مجازی، اینترنت و شبکه را بیش از پیش مورد تاکید قرار می‌دهند. ساختار سیستم‌عامل ربات در ارتباط با تعداد زیادی از ربات‌ها، شرایط مناسبی را برای آسیب‌پذیری و آلوده نمودن سیستم‌عامل و عملکرد نادرست ربات‌ها فراهم می‌نماید. براساس آمار بیش از ۹۰ درصد آسیب‌ها و حملات رخ داده شده بر روی ربات‌ها، از طریق عوامل نرم‌افزاری و حوزه‌های مرتبط با نرم‌افزارهای ربات‌ها صورت می‌گیرند. از جمله آن حملات رخ داده شده می‌توان به حمله MITM^{۷۸}، حمله‌ی غیرفعال نمودن نودی در شبکه‌ی ربات‌ها از طریق ارسال دستور اشتباه و یا دست‌کاری شده‌ی Shutdown^{۷۹}، عملکرد اشتباه ربات با ارسال دستورات آلوده، تداخل در اطلاعات ارسال و یا دریافت شده از طریق درگاه‌های غیرامن سیستم‌عامل ربات^{۸۰} و سایر موارد دیگر، اشاره نمود. در یک تحقیق آسیب‌پذیری ربات‌ها را در بخش‌های زیر مطرح کرده است^{۸۱}:

- اشکال در استفاده از رابط‌های نرم‌افزاری و برنامه‌نویسی در ربات (API misuse)
- داده‌سازی و یا استراق‌سمع داده‌ها در ربات (Data manipulation/observation)

^{۷۵} ROS Transmission Control Protocol

^{۷۶} ROS User Datagram Protocol

^{۷۷} A. Botta, S. Rotbei, S. Zinnó and G. Ventre, "Cyber security of robots: A comprehensive survey.," *Intelligent Systems With Applications*, no.18, 2023. <https://doi.org/10.1016/j.iswa.2023.200237>

^{۷۸} Man In the Middle – در این حمله عامل مخرب با قرار گرفتن در جریان عملیات سامانه، در عملکرد صحیح آن اختلال ایجاد می‌کند

^{۷۹} K. Cottrell, D. B. Bose, H. Shahriar and A. Rahman, "An Empirical Study of Vulnerabilities in Robotics.," in *IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC)*, 2021. <https://doi.org/10.1109/compsac51774.2021.00105>

^{۸۰} N. Goerke, D. Timmermann and I. Baumgart, "Who Controls Your Robot? An Evaluation of ROS Security Mechanisms," in *7th International Conference on Automation, Robotics and Applications (ICARA)*, 2021. <https://doi.org/10.1109/icara51699.2021.9376468>

^{۸۱} B. Dieber, B. Breiling, S. Taurer, S. Kacianka, S. Rass and P. Schartner, "Security for the Robot Operating System," *Journal of Robotics and Autonomous Systems*. vol. 98, 192-203, 2017.

B. Dieber, S. Kacianka, S. Rass and P. Schartner, "Application-level Security for ROS-based Applications," *IEEE International Conference on Intelligent Robots and Systems (IROS)*, pp. 4477-482, 2016.

- کسب و یا تحصیل داده‌های ترافیکی در ربات (*Data acquisition*)
 - موارد مرتبط با گردش و جابجایی داده‌ها در ربات (*Flow control*)
 - شرایط عدم عملکرد صحیح یا ارائه خدمت توسط ربات (*Race condition*)
 - موارد امنیتی در خصوص احراز هویت، احراز اصالت، کنترل دسترسی و رمزنگاری (*Security issues*)
 - عدم اعتبارسنجی داده ورودی در خصوص پارامترها و الگوهای ورودی ربات (*Unvalidated input*)
- به لحاظ اهمیت امنیت تبادل اطلاعات و جلوگیری از دسترسی غیرمجاز، هک شدن و آلوده شدن سامانه و عملکرد نادرست آن، استفاده از پروتکل‌های تبادل اطلاعات امن و بادر نظر گرفتن ملاحظات امنیتی در سیستم عامل ربات، ضروری می‌باشد.
- فرآیند تحقیق و تفحص ادله‌ی الکترونیکی در سیستم عامل ربات، بررسی ردپاهای دیجیتال برای رسیدن به دلیل آلودگی‌های امنیتی رخ داده شده در ربات می‌باشد. این بررسی‌ها با هدف‌های جرم در فضای مجازی، تشخیص خرابکاری، و تهیه ادله‌ی الکترونیکی استنادپذیر برای کاربردهای قانونی، صورت می‌گیرند.
- اگرچه امروزه به دلیل وقوع جرائم، خرابکاری‌ها و نفوذهای، ضرورت تحقیق و تفحص ادله‌ی الکترونیکی در مورد ربات‌ها و اینترنت اشیا بیش از پیش مطرح شده، ولی به لحاظ ضعیف بودن سطح آشنایی با ساختار فایل‌های سیستم عامل ربات که منجر به شناخت ناکافی و غیرموثر عوامل خرابکاری و آسیب‌ها و در نتیجه عدم حصول نتیجه‌ی مناسب در فرآیند تحقیق و تفحص شده است.
- در یک روش فرآیند تحقیق و تفحص ادله‌ی الکترونیکی سیستم عامل ربات، اقدام به دریافت داده‌های برخط از حافظه‌ی درحال اجرای ربات و سپس مقایسه‌ی داده‌ها با حرکت ربات می‌گردد. این روش کمک می‌کند که بتوان به تحلیل نسبتاً درستی از نحوه‌ی اجرای فرامین و عملکرد ربات رسید.
- از نمونه ابزارهایی که در بررسی حافظه‌ی و با هدف جمع‌آوری ادله در ربات‌ها استفاده می‌شود می‌توان به دستور *LiME*، *Volatility Memory* اشاره نمود. با بررسی محتوای حافظه در زمان واقعی و استفاده از الگوریتم *SLAM* (*Simultaneous Localization and Mapping*) امکان مکان‌یابی دقیق ربات و دستور اجرا شده در حافظه و انطباق آن با حرکت ربات، وجود خواهد داشت. این بررسی به درک بهتر نحوه تعامل ربات با محیط اطرافش کمک می‌کند. همان‌گونه که پیشتر نیز اشاره شد، هدف رسیدن به ادله‌ی استنادپذیر برای تشخیص آسیب یا حمله‌ی وارد شده و یا عملکرد نادرست ربات، و در صورت امکان رسیدن به عامل حمله یا اختلال در عملکرد ربات است.
- توجه به شرایط داده‌های جمع‌آوری شده از حافظه‌ی ربات و صحت و دقت آن‌ها از یک سو و تطابق آن‌ها با زمان‌بندی حرکت‌های ربات و رسیدن به یک ادله‌ی قابل استفاده در فرآیند تحقیق و تفحص؛ از اهمیت ویژه‌ای برخوردار است. استفاده از داده‌های رمزنگاری شده در انتقال داده‌ها و رعایت ملاحظات امنیتی در دسترسی به آن داده‌ها، می‌تواند در جمع‌آوری و بررسی ادله‌ی الکترونیک بسیار موثر باشد.

پیوست دو

مقدمه‌ای بر هوش مصنوعی مولد

هوش مصنوعی توجه‌های بسیاری را به خود در حوزه‌ی مختلف و صنایع گوناگون جلب نموده است (Hyder et al., 2019). هوش مصنوعی به عنوان "توانایی سیستم برای تحلیل و دریافت صحیح داده‌ها از خارج از سیستم، یادگیری الگوی آن‌ها، و استفاده از داده‌های یادگرفته شده برای تحقق هدف و کاربردهای متنوع مورد نظر" تعریف شده است (Kaplan & Haenli, 2019, p. 15). هوش مصنوعی از ابتدای پیدایش در دهه‌ی ۱۹۵۰ و ۱۹۶۰ مراحل مختلفی را پشت سر گذاشته و پیشرفت‌های قابل توجهی نموده است.

در سال‌های اخیر، هوش مصنوعی با یک تغییر رویکرد با معرفی AlphaGo در سال ۲۰۱۵ و سپس ChatGPT در سال ۲۰۲۲ همراه بوده است. در ادامه هوش مصنوعی با عرضه‌ی نسخه‌های هابی با کاربردهای با نام "Chat Generative Pre-trained Transformer" یا ChatGPT عرضه شده توسط شرکت OpenAI در اواخر سال ۲۰۲۲، توجهات زیادی را در سطح جهان به خود جلب نمود. بطوری که در ۳۰ نوامبر ۲۰۲۲ و با عرضه‌ی محصول ChatGPT که به گروهی از مدل‌های زبانی به‌عنوان مدل‌های زبانی بزرگ^{۸۲} که از فناوری‌های یادگیری عمیق برای بهره‌گیری از حجم بسیار زیاد داده‌ها، یافتن و یادگیری الگوی موجود در آن‌ها، رکورد بیشترین جستجوها را توسط کاربران در گوگل به خود اختصاص داد (Casella et al., 2023).

محصول ChatGPT برای کاربردهای محاوره‌ای، استخراج و تولید پاسخ‌های مشابه انسان با استفاده از دانش و اطلاعات ارزشمند یادگرفته شده توسط آن، طراحی و ارتقا داده شده است. قابلیت‌های ChatGPT با استفاده از الگوریتم‌های موسوم به هوش مصنوعی مولد و با استخراج داده‌ها از منابع مختلف و تحلیل آن‌ها، ایجاد شده، و این الگوریتم‌ها امکان تولید متن‌هایی شبیه به انسان و تولید انواع محتواها (مانند تصاویر و موسیقی) را می‌دهند (Dasborough, 2023).

شاید تا کمتر از پنج سال گذشته، کاربرد هوش مصنوعی به‌عنوان هوش مصنوعی مولد (با ویژگی تولید محتواهای مختلف و براساس دریافت دستورات و درخواست‌های کاربر در قالب زبان محاوره‌ای) امری بعید و یا در مواردی غیرمحتمل بشمار می‌آمد.

اهمان گونه که در ادامه شرح داده شده است، فناوری هوش مصنوعی مولد از سال ۲۰۱۵ و در آزمایشگاه OpenAI آغاز گردید، ولی عرضه‌ی محصول ChatGPT در نوامبر سال ۲۰۲۲ توجه بسیاری از کاربران و شرکت‌های فعال در حوزه‌ی فناوری اطلاعات و سامانه‌ها و سکوها (پلتفرم‌های) نرم‌افزاری را به خود جلب نمود. تا جایی که بالاترین رشد سهام در بورس و درحوزه‌ی کسب و کارهای فعال در هوش مصنوعی و فناوری اطلاعات را از آن خود کرد.

^{۸۲} Large Language Models (LLMs)



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

هوش مصنوعی مولد چیست؟

هوش مصنوعی مولد (^{۸۳}GenAI)، قابلیت هوش مصنوعی در ایجاد متن، تصاویر، و یا محتوای دیگر با استفاده از مدل‌های مولد می‌باشد. این مدل‌های هوش مصنوعی ضمن یادگیری الگوها و ساختار اطلاعات ورودی طبقه‌بندی^{۸۴} شده، اقدام به تولید اطلاعات جدید با مشخصات مشابه می‌نمایند.

در اوایل دهه ۲۰۲۰، پیشرفت‌های صورت گرفته در مبدل‌های مبتنی بر شبکه‌های عصبی عمیق^{۸۵} منجر به توانا نمودن سیستم‌های هوش مصنوعی مولد، به منظور دریافت دستورات ورودی از طریق زبان محاوره‌ای شدند. این توانایی‌ها شامل مدل‌های بزرگ شبیه‌ساز زبان محاوره‌ای مانند ^{۸۶}ChatGPT، ^{۸۷}Bing Chat، ^{۸۸}Bard و ^{۸۹}LLaMA و همچنین سیستم‌های هوش مصنوعی هنری برای تبدیل متن به تصویر مانند ^{۹۰}Stable Diffusion و ^{۹۱}DALL-E می‌باشند.

هوش مصنوعی مولد کاربردهای متنوعی در صنایع و مشاغل مختلف مرتبط با هنر، نوشتن، توضیح‌نویسی، توسعه نرم‌افزار، طراحی محصول، سلامت، مالی، بازی‌های رایانه‌ای، بازاریابی، و مد دارد. در اوایل دهه ۲۰۲۰ با توسعه مدل‌های مولد هوش مصنوعی توسط شرکت‌های بزرگی مانند مایکروسافت، گوگل و بایدهو و همچنین بسیاری از شرکت‌های کوچک، سرمایه‌گذاری در این حوزه افزایش یافت. اما، نگرانی‌های جدی بسیاری در استفاده نادرست از مدل‌های مولد هوش مصنوعی، مانند جرم‌های اینترنتی و فضای مجازی^{۹۲} یا ایجاد محتوای خبری جعلی^{۹۳} یا دست‌کاری تصویر اشخاص و محتواها^{۹۴} که می‌توانند برای تحت تاثیر قراردادن و کنترل رفتارها و افکار بشر استفاده شوند، بوجود آمد.

مدل‌های عملکردی هوش مصنوعی مولد

سیستم هوش مصنوعی مولد با اعمال الگوریتم‌های یادگیری ماشین هدایت‌نشده یا خودهدایت‌شونده بر روی مجموعه داده‌های طبقه‌بندی شده^{۹۵}، ساخته شده‌است. قابلیت‌های سیستم هوش مصنوعی مولد بستگی به مدل‌ها و یا نوع داده‌های طبقه‌بندی شده موضوع آن سیستم دارد.

^{۸۳} Generative Artificial Intelligence

^{۸۴} Trained Data

^{۸۵} Deep Neural Networks - به سیستم‌های هوش مصنوعی با استفاده از الگوریتم‌های یادگیری عمیق مبتنی بر شبکه‌های عصبی اطلاق می‌شود.

^{۸۶} مدل زبان مبتنی بر هوش مصنوعی توسعه داده شده توسط OpenAI با قابلیت ایجاد متن مشابه زبان محاوره‌ای با توجه به محتوای داده شده و اطلاعات یادگرفته شده پیشین مدل، می‌باشد

^{۸۷} موتور جستجوگر مایکروسافت مجهز به هوش مصنوعی و استفاده از قابلیت‌های GPT-۴ می‌باشد.

^{۸۸} ابزار مجاوره‌ای مبتنی بر هوش مصنوعی شرکت گوگل با قابلیت تحلیل و بررسی ایده‌ها و نظرها، جرقه‌های نوآوری و شتاب‌دهی بهره‌وری می‌باشد.

^{۸۹} ابزار و رابط کاربری ChatGPT پلتفرم (سکو) فیس‌بوک

^{۹۰} مدلی برای تبدیل متن ورودی به تصاویر واقعی با استفاده از قابلیت‌های هوش مصنوعی و داده‌های یادگرفته شده می‌باشد.

^{۹۱} سیستم مبتنی بر هوش مصنوعی که امکان ایجاد تصویر را با استفاده از توضیحات داده شده به زبان طبیعی دارد.

^{۹۲} Cyber Crime

^{۹۳} Fake News

^{۹۴} Deep Fake

^{۹۵} Dataset



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

هوش مصنوعی مولد می‌تواند یک مد و یا چندمد عمل نماید؛ سیستم‌های تک‌مد تنها یک نوع داده‌ی ورودی را دریافت می‌کنند، درحالی‌که سیستم‌های چندمد می‌توانند چند داده‌ی ورودی را دریافت می‌کنند. برای مثال، نسخه‌ای از GPT-4 مبتنی بر OpenAI ورودی‌های از نوع متن و تصویر را می‌پذیرند. سیستم‌های هوش مصنوعی مولد طبقه‌بندی شده منطبق بر کلمات یا عبارات‌های شناخته شده، موضوع ابزارهای GPT-3، GPT-4، BLOOM، LLaMA، LaMDA و دیگر زبان‌ها می‌باشند. آن‌ها قابلیت پردازش زبان طبیعی، عملیات ماشین، و مولد زبان طبیعی هستند و می‌توانند به‌عنوان مدل‌های پایه‌ی عملیات دیگر استفاده شوند. داده‌های طبقه‌بندی شده در این سیستم‌ها: BookCorpus، Wikipedia و دیگر مجموعه داده‌ها می‌باشند.

کدبرنامه

مدل‌هایی از زبان‌های گسترده می‌توانند علاوه بر متن حاوی زبان طبیعی برای تبدیل متن حاوی زبان برنامه‌نویسی به کد برنامه‌ی زبان‌های برنامه‌نویسی جدید، آماده شوند. به‌عنوان مثال می‌توان به ابزار متن باز مبتنی بر هوش مصنوعی ^{۹۶} Codex اشاره نمود.

تصاویر

تولید تصاویر و یا آثار هنری با کیفیت بالا، یکی از کاربردهای مهم و شناخته شده هوش مصنوعی مولد می‌باشد. تعدادی از همین آثار هنری ایجاد شده توسط این ابزارها، ضمن دریافت لوح‌های تقدیر عمومی، شناخته شده‌اند. سیستم‌های مولد هوش مصنوعی از طریق روش‌های یادگیری مجموعه‌ای از تصاویر با اسامی و عناوین آن‌ها (زیرنویس حاوی معرفی تصاویر ^{۹۷}) عمل می‌کنند. از آن جمله می‌توان به ابزار و سیستم‌های DALL-E، Imagen، Midjourney، Adobe Firefly، Stable Diffusion و سایر ابزارها اشاره نمود ^{۹۸}. عموماً این ابزارها برای تولید تصویر از متن ورودی و همچنین ایجاد محتوای تصویر براساس محتوای تصویر دیگر و براساس الگوی مشخص شده، بکار می‌روند. مجموعه داده‌های مورد استفاده در این سیستم‌ها، شامل Lion-B و سایر مجموعه داده‌ها می‌باشند ^{۹۹}.

موزیک

سیستم‌های مولد هوش مصنوعی موزیک مانند MusicLM، MusicGen می‌توانند براساس نت‌های، موسیقی‌های ضبط شده و هم‌زمان استفاده از متن توضیح موسیقی برای ایجاد نمونه‌ی جدیدی از موسیقی براساس متن توصیفی ارائه شده به آن‌ها، مانند ایجاد ملودی و بولون (موزیک) که با قطعه‌ای از موسیقی نواخته شده توسط گیتار، معرفی شده است.

ویدئو

^{۹۶} OpenAI Codex

^{۹۷} Text Caption

^{۹۸} برای اطلاعات بیشتر در این حوزه می‌توانید موضوعات Artificial Intelligence Art, Generative Art, Synthetic Media را مطالعه نمایید.

^{۹۹} برای اطلاعات بیشتر می‌توانید به Datasets in Computer Vision (مجموعه داده‌ها تصاویر کامپیوتری) مراجعه نمایید.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

سیستم هوش مصنوعی مولد فیلم که براساس داده‌های طبقه‌بندی شده از زیرنویس فیلم‌ها امکان ایجاد قطعه‌ای از فیلم با متوسط میزان مشابهت بانمونه‌ی اصلی آن را دارد. مثال‌هایی از این سیستم‌ها، Gen1 و Gen2 ایجاد شده توسط RunwayM و Make-A-Video ایجاد شده توسط Meta Platform می‌باشند.

بیولوژی و علوم زیستی

سیستم‌های هوش مصنوعی مولد می‌توانند به‌نحوی استفاده شوند که براساس ترتیب آمینه اسیدهای یا ساختار مولکولی (مانند ابزار SMILE که برای ارائه‌ی نمایش DNA مولکولی و یا پروتئین‌ها استفاده می‌شود). این سیستم‌ها، ازجمله AlphaFold، برای پیش‌بینی ساختار پروتئین مولوکی و کشف داروها بکار برده می‌شوند. مجموعه داده‌های آن‌ها شامل مجموعه داده‌های گوناگون بیولوژی می‌باشند.

ربات‌ها

سیستم‌های هوش مصنوعی مولد همچنین می‌توانند به‌نحوی بکارگرفته شوند که با استفاده از حرکت ربات‌ها، الگوریتم‌های بهینه‌سازی شده^{۱۰۰} را برای ایجاد الگوهای جدید طراحی حرکت و مسیریابی ربات‌ها، ایجاد نمایند. به‌عنوان مثال، UniPi محصول تحقیقات شرکت گوگل از فرمان‌های مانند "کاسه آبی را بردار" یا "ظرف شو با ابر زرد رنگ"، برای کنترل حرکت‌های بازوی ربات استفاده می‌کند. مدل‌ها چند بعدی^{۱۰۱} "دید-زبان-حرکت"^{۱۰۲} مانند ابزار RT-2 شرکت گوگل، که با استفاده از یک استدلال ابتدایی در پاسخ به دستورهای کاربر و ورودی‌های تصویری، با دستور "یک اسباب‌بازی دایناسور بردار"، در شرایطی که میز مملو از اسباب‌بازی‌هایی از حیوان‌ها و سایر اشیاء می‌باشد، اسباب‌بازی مورد نظر را برمی‌دارد.

برنامه‌ریزی

واژه‌ی برنامه‌ریزی توسط هوش مصنوعی مولد^{۱۰۳} یا برنامه‌ریزی مولد^{۱۰۴} طی سال‌های ۱۹۸۰ و ۱۹۹۰ و برای سیستم‌های برنامه‌ریزی توسط هوش مصنوعی^{۱۰۵}، بویژه فرآیند برنامه‌ریزی با کمک رایانه^{۱۰۶}، به‌منظور ارائه‌ی ترتیبی از عملیات و اقدامات برای رسیدن به هدف مشخصی، بکار برده می‌شد. سیستم‌های برنامه‌ریزی هوش مصنوعی مولد از روش‌های هوش مصنوعی مبتنی بر نماد^{۱۰۷} مانند جستجو فضای حالت^{۱۰۸} و ارضای محدودیت‌ها^{۱۰۹} و فناوری نسبتاً بالغ شده در اوایل دهه ۱۹۹۰، استفاده می‌کردند. از آن‌ها برای ایجاد برنامه‌های عملیات زمان بحران در مأموریت‌های نظامی، برنامه‌ریزی فرآیندها در تولید و برنامه‌ریزی مبتنی بر تصمیم، که در نمونه‌های ماشین‌های فضایی خودکار کاربرد دارند، استفاده می‌شد.

^{۱۰۰} Trajectory Motion Planning and Navigation

^{۱۰۱} Multimodal

^{۱۰۲} Vision-Language-Action

^{۱۰۳} Generative Ai Planning

^{۱۰۴} Generative Planning

^{۱۰۵} AI Planning

^{۱۰۶} Computer-Aided Process Planning

^{۱۰۷} Symbolic AI

^{۱۰۸} State Space Search

^{۱۰۹} Constraint Satisfaction



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

مدل‌های زبانی با میلیاردها پارامتر، مانند *GPT-4* یا *PaLM*، معمولاً بر روی مراکز داده و رایانه‌هایی که با مجموعه‌هایی از پردازنده‌ها تقویت شده (مانند *Nvidia H100*) (پردازشگر گرافیکی) که مجهز به مدارهای الکترونیکی خاص برای افزایش سرعت پردازش گرافیکی رایانه‌ها و پردازش‌های تصویری می‌باشد (می‌توانند بر روی کارت گرافیک مجزا و یا بر روی مادربرد، گوشی‌های همراه، رایانه‌های شخصی و کنسول‌های بازی نیز استفاده شوند)، اجرا می‌شوند. پس از طراحی اولیه‌ی این‌نوع از پردازنده‌های تصویر و با توجه به قابلیت آن‌ها در پردازش‌های موازی، استفاده از آن‌ها برای حل محاسبات غیر گرافیکی شامل مسائلی که با توجه به ساختار و ماهیت موازی آن‌ها، بصورت موازی قابل پردازش و حل می‌باشند، مورد توجه قرار گرفت. از کاربردهای غیرگرافیکی توسط این نوع از پردازنده‌ها، تحلیل و یادگیری شبکه‌های عصبی و محاسبات استخراج رمازرها توسط ماینرها، یا استفاده از آن‌ها در پردازنده‌های هوش مصنوعی مولد (مانند *Google's TPU*^{۱۱۰}) را می‌توان نام برد. *TPU* مدار مجتمع هوش مصنوعی مولد با کاربرد خاص (*ASIC*^{۱۱۱}) توسعه داده شده توسط شرکت گوگل می‌باشد که با استفاده از نرم‌افزار آن شرکت به نام *TensorFlow* برای یادگیری ماشین هوش مصنوعی^{۱۱۲} مولد بکاربرده می‌شود. شرکت گوگل استفاده داخلی از آن را از سال ۲۰۱۵ آغاز نمود و در سال ۲۰۱۸ نسخه‌ی از آن را برای کاربردهای تجاری و عرضه به مصرف‌کنندگان قابل استفاده در ساختار ابری و همچنین پردازنده‌های ابعاد کوچک برای فروش، توسعه داد. نسل پنجم پردازنده‌های *TPU* شرکت گوگل در سال ۲۰۲۱ عرضه شد و شرکت گوگل اعلام نمود که ساختار طراحی فیزیکی این نسل از پردازنده‌ها، براساس کاربرد نوآورانه‌ی بازیادگیری عمیق^{۱۱۳} صورت گرفته است. گوگل پردازنده‌ها نسل پنجم *TPU* را مشابه نسل *H100* دیده و عرضه آن را در ماه ژوئن سال ۲۰۲۳ و تست اجرایی آن را در طی سال ۲۰۲۳ انجام خواهد داد. بطور معمول، این دسته از پردازنده‌های پرقدرت در سرویس‌های ابری و از طریق اینترنت قابل دسترسی و استفاده خواهند بود.

هوش مصنوعی مولد و کاربردهای آن به‌عنوان یکی از دستاوردهای نوآورانه هوش مصنوعی در کمتر از یک دهه‌ی اخیر توانسته ضمن توسعه و گسترش کاربردی، سرمایه‌گذاری‌های زیادی را به‌خود جذب نماید. در این میان نقش آزمایشگاه تحقیقاتی هوش مصنوعی *OpenAI* بسیار حائز اهمیت می‌باشد. از این‌رو سعی شده تا در ادامه‌ی مطالب این آزمایشگاه و تعدادی از محصولات و فناوری‌های توسعه داده شده و عرضه شده توسط آن لابراتور، شرح داده شود.

OpenAI

OpenAI آزمایشگاه تحقیقاتی هوش مصنوعی آمریکا و شامل شرکت غیرانتفاعی *OpenAI* می‌باشد. شرکت زیر مجموعه آن که انتفاعی بوده و با نام *Open AI, L.P.* در دلاور آمریکا ثبت شده است. تحقیقات هوش مصنوعی *OpenAI* با هدف توسعه‌ی "امن و سودمند" هوش عمومی مصنوعی، و برای رسیدن به سیستم‌های خودکار فوق پیشرفته که از بعد اقتصادی کارهای بشری را بسیار باصرفه‌تر انجام می‌دهند، انجام می‌شود.

^{۱۱۰} Tensor Processing Unit

^{۱۱۱} AI Accelerator application-specific Integrated Circuit

^{۱۱۲} Neural Network Machine Learning

^{۱۱۳} Deep Reinforcement Learning



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

OpenAI در سال ۲۰۱۵ و بوسیله لیا ساتسکور^{۱۱۴}، گرگ بروکمن^{۱۱۵}، ترور بلکول^{۱۱۶}، ویکی چئونگ^{۱۱۷}، آندری کارپاتی^{۱۱۸}، درک کینگما^{۱۱۹}، جسیکا لیوینگستن^{۱۲۰}، جان شولمان^{۱۲۱}، پاملا وگاتا^{۱۲۲} و ویسئیچ^{۱۲۳} زرمبا، با سم آلتمن^{۱۲۴} و ایلان ماسک^{۱۲۵} به عنوان اولین اعضای هیئت مدیره آن، تشکیل گردید. شرکت مایکروسافت در سال ۲۰۱۹ مبلغ ۱ میلیارد دلار و در سال ۲۰۲۳ مبلغ ۱۰ میلیارد دلار بر روی OpenAI سرمایه گذاری نموده است.

سال ۲۰۱۹ – تبدیل لایبراتور غیرانتفاعی OpenAI به یک شرکت انتفاعی (تجاری)

در سال ۲۰۱۹، OpenAI از موسسه غیرانتفاعی به شرکت انتفاعی و با بازگشت سرمایه گذاری ۱۰۰٪ تبدیل گردید. براساس مدل بازگشت سرمایه OpenAI، آن شرکت بطور قانونی امکان جذب سرمایه از میان شرکت‌های سرمایه‌گذاری و سرمایه‌گذاران را پیدا نمود، علاوه براین، امکان جذب کارکنان در شرکت به عنوان سهامداران و با هدف اشتغال آن‌ها در شرکت و استفاده از منافع شرکت در دراز مدت برای کارکنان نیز میسر گردید. به بسیاری از محققان در شرکت‌های Google Brain، DeepMind و Facebook پیشنهادهای دریافت سهام ارائه شد که پیشتر امکان ارائه‌ی این پیشنهادها توسط یک شرکت غیرانتفاعی وجود نداشت. پیش از تبدیل آن به یک شرکت انتفاعی، از نظر قانونی اعلام عمومی منافع حاصل از شرکت برای کارکنان اصلی OpenAI ضروری بود. پس از آن شرکت ارزش خود را میان کارکنان خود تقسیم نمود و با اعلام سرمایه‌گذاری به مبلغ ۱ میلیارد دلار توسط شرکت مایکروسافت، با مایکروسافت وارد مشارکت گردید. از آن زمان، سیستم‌های OpenAI بر روی زیرساخت‌های سوپر کامپیوترهای Azeru شرکت مایکروسافت اجرا می‌شوند. در ادامه OpenAI اعلام نمود که قصد دارد تا فناوری‌های خود را ثبت تجاری نماید. OpenAI قصد دارد تا ۱ میلیارد دلار ظرف ۵ سال در شرکت سرمایه‌گذاری کند. آلتمن اعلام کرد احتمالاً ۱ میلیارد دلار کافی نخواهد بود و این آزمایشگاه در نهایت برای رسیدن به فناوری هوش عمومی مصنوعی، بیشتر از هر سازمان غیرانتفاعی دیگری به سرمایه‌گذاری نیاز خواهد داشت.

افرادی که به ماهیت شرکت‌ها و فعالیت آن‌ها به عنوان یک حقیقت در موفقیت آن‌ها باور ندارند، در خصوص فعالیت OpenAI این گونه اظهار نظر کرده‌اند، اگرچه براین موضوع که جذب محققین در یک سازمان غیرانتفاعی دشوار بوده؛ اما مشخصاً با اشاره به OpenAI و پروژه‌های کم هزینه‌ی آن و دیگر سازمان‌های مشابه، این گونه

^{۱۱۴} Ilya Sutskever

^{۱۱۵} Greg Brockman

^{۱۱۶} Trevor Blackwell

^{۱۱۷} Vicki Cheung

^{۱۱۸} Andrej Karpathy

^{۱۱۹} Durk Kingma

^{۱۲۰} Jessica Livingston

^{۱۲۱} John Schulman

^{۱۲۲} Pamela Vagata

^{۱۲۳} Wojciech Zaremba

^{۱۲۴} Sam Altman

^{۱۲۵} Elon Musk



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

اظهارنظر می‌کنند "با این نظر که یک سازمان غیرانتفاعی قادر به رقابت نمی‌باشد، موافق نیستند". "اگر قبول کنیم که سازمان‌های بزرگتر و قوی‌تر همواره بهترین هستند، پس IBM می‌بایست همچنان بهترین شرکت می‌بود" شرکت غیرانتفاعی OpenAI، سهامدار با حق کنترل در شرکت سهامی خاص OpenAI (OpenAI L.P.) می‌باشد. شرکت سهامی خاص OpenAI، صرفنظر از این که شرکت انتفاعی بوده، همواره به مسئولیت‌های حقیقی الگوی غیرانتفاعی شرکت OpenAI بطور رسمی پایبند است. بخش عمده‌ای از هیئت مدیره شرکت OpenAI از داشتن سهام در شرکت سهامی خاص OpenAI منع شده‌اند. علاوه بر آن، بخش عمده‌ای از سهامداران شرکت سهامی خاص OpenAI به دلیل تضاد منافع از داشتن حق رای‌های خاصی در شرکت منع شده‌اند. تعدادی از محققین بر این نکته که تبدیل شدن آزمایشگاه تحقیقاتی OpenAI به یک شرکت انتفاعی می‌تواند در تداوم وضعیت آن به عنوان یک سازمان برای دموکراسی هوش مصنوعی، خدشه وارد کند، تاکید دارند.

از سال ۲۰۲۰ تا کنون: ارائه‌ی محصولات ChatGPT، DALL-E و همکاری با شرکت مایکروسافت

در سال ۲۰۲۰، OpenAI محصول GPT-3، که مدل زبان محاوره‌ای آموزش داده شده بر روی مجموعه داده‌های اینترنتی بود را اعلام نمود. GPT-3 برای کمک به پاسخ سوالات مطرح شده به زبان طبیعی بوده، اما همین‌طور امکان ترجمه زبان‌های محاوره‌ای و ایجاد متن مربوط به آن‌ها را دارد. همچنین آن شرکت عرضه اولین محصول تجاری خود را که یک API کمکی بود به نام "The API" اعلام نمود.

در سال ۲۰۱۲، OpenAI محصول DALL-E را که یک مدل یادگیری عمیق با امکان ایجاد تصاویر با استفاده از توضیحات ارائه شده توسط زبان طبیعی بود، اعلام نمود.

در دسامبر ۲۰۲۲، OpenAI توانست پوشش گسترده‌ای رسانه‌ای گسترده‌ای را پس از ارائه‌ی نسخه رایگان ChatGPT بدست آورد. این محصول نمونه‌ی جدید مبتنی بر هوش مصنوعی chatbot بر روی محصول GPT-3.5 بود. براساس اعلام OpenAI، در پنج روز ابتدایی اعلام این نسخه بیش از یک میلیون تقاضای دسترسی برای آن ثبت شده است. براساس اعلام خبرگزاری رویترز از خبر سایر منابع، در دسامبر ۲۰۲۲، OpenAI درآمدی بالغ بر ۲۰۰ میلیون دلار در سال ۲۰۲۳ کسب خواهد نمود و پیش‌بینی می‌شود که این رقم در سال ۲۰۲۴ به رقمی بالغ بر ۱ میلیارد دلار برسد.

در ژانویه ۲۰۲۳، OpenAI جذب سرمایه‌گذاری بالغ بر ۲۹ میلیارد دلار را مطرح نمود، رقم بالغ بر دو برابر ارزش شرکت در سال ۲۰۲۱. در ۲۳ ژانویه ۲۰۲۳، شرکت مایکروسافت اعلام کرد سرمایه‌گذاری چندساله جدیدی بالغ بر ۱۰ میلیارد دلار را در شرکت OpenAI انجام می‌دهد. اخبار اولیه منتشر شده (تایید نشده) حاکی از آن است که مایکروسافت ۷۵٪ از سود شرکت OpenAI را تا زمان رسیدن به مرحله امن بازگشت سرمایه‌گذاری خود دریافت نموده و در نهایت ۴۹٪ از سهام آن شرکت را خواهد داشت.

این سرمایه‌گذاری بخشی از تلاش‌های شرکت مایکروسافت برای تجمیع محصول ChatGPT شرکت OpenAI در موتور جستجوگر bing آن شرکت می‌باشد. شرکت گوگل محصول مبتنی بر هوش مصنوعی مشابهی به نام Brad،



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

را پس از عرضه‌ی ChatGPT و با توجه به نگرانی از تهدید جایگاه گوگل به عنوان منبع اطلاعات مورد نیاز توسط ChatGPT، اعلام نموده است.

در ۷ فوریه ۲۰۲۳، شرکت مایکروسافت اعلام کرد فناوری مبتنی بر هوش مصنوعی برپایه‌ی ChatGPT را برای نصب بر روی Microsoft Bing، Edge و Microsoft 365 و دیگر محصولات خود، می‌سازد. در ۳ مارس ۲۰۲۳ رید هافمن برای پرهیز از تضاد منافع در هیئت مدیره OpenAI و سرمایه‌گذاری‌های او در شرکت‌هایی فناوری هوش مصنوعی که توسط او و از طریق Greylock Partners صورت گرفته بود و همچنین سمت خود به عنوان یکی از پایه‌گذاران استارت‌آپ Infection AI، از هیئت مدیره گری لاک پارتنرز استعفاء داد. هافمن همچنان در سمت خود در هیئت مدیره شرکت مایکروسافت، که یکی از سرمایه‌گذاران عمده OpenAI می‌باشد، باقی ماند.

در ۱۴ مارس ۲۰۲۳، OpenAI نسخه‌ی GPT-4 را که هم به صورت API هم بصورت نسخه‌ی جدید ChatGPT استفاده می‌شود را عرضه نمود.

در ۲۲ می ۲۰۲۳، سام آلمن، گرگ بروکمن و لیا ساتسکور توصیه‌ها و مطالبی در خصوص حکمرانی ابر هوشمندی^{۱۲۶} پست کردند. آن‌ها این گونه مطرح کردند که ابر هوشمندی می‌تواند طی ۱۰ سال آینده تحقق یافته، و "آینده‌ای سرشار از موفقیت را رقم خواهد زد" که با خود "ریسک‌های بالقوه‌ی را به همراه خواهد آورد و ما نمی‌توانیم تنها پاسخ‌گوی اتفاقات باشیم". هدف آن‌ها پیشنهاد ایجاد سازمان نظارت جهانی مشابه سازمان بین‌المللی انرژی اتمی به منظور کنترل و بررسی سیستم‌های هوش مصنوعی که از یک حدی فراتر رفته، و از طرف دیگر به تناسب آن عدم برخورد سخت‌گیرانه با سیستم‌های هوش مصنوعی ضعیف، بود. آن‌ها همچنین خواستار تحقیقات امن‌تر فنی در حوزه‌ی فناوری ابری هوشمند و هماهنگی بیشتر، به عنوان مثال، از طریق ایجاد پروژه‌ای مشترک توسط دولت‌ها که "فعالیت‌های فعلی بخشی از آن پروژه باشند"، بودند.

در آگوست ۲۰۲۳، اعلام شد که OpenAI شرکت استارت‌آپی (نوپا) واقع در نیویورک به نام Global Illumination، که شرکتی برای توسعه زیرساخت‌های دیجیتال و بکارگیری آن‌ها و ایجاد ابزارهای خلاقانه در این حوزه می‌باشد، را خریداری نموده است.

محصولات و فناوری‌های توسعه داده‌شده توسط OpenAI

از سال ۲۰۲۱، تحقیقات OpenAI بر روی بازیادگیری^{۱۲۷} متمرکز شده است. هم اکنون OpenAI یک رقیب مهم برای DeepMind^{۱۲۸} بشمار می‌رود.

Gym

^{۱۲۶} Superintelligence

^{۱۲۷} Reinforcement Learning (RL)

^{۱۲۸} شرکت انگلیسی می‌باشد که در سال ۲۰۱۰ ایجاد شد و مرکز آن در لندن می‌باشد و در سال ۲۰۱۴ توسط شرکت گوگل خریداری شد و به عنوان یکی از شرکت‌های زیرمجموعه گوگل در زمینه شبکه‌های عصبی و ماشین‌های شبکه‌ای عصبی و سیستم‌های بازیادگیرنده فعالیت می‌کند.



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

در سال ۲۰۱۶ اعلام شد، *Gym* بستری برای امکان ایجاد آسان‌تر کاربردهای عمومی هوش مصنوعی در محیط‌های مختلف، فراهم نموده است. مشابه *ImageNet Large Scale Recognition Challenge* که در تحقیقات یادگیری هدایت‌شده استفاده می‌شود، ولی بصورت گسترده‌تر از آن. این پروژه امیدوار است تا روشی استاندارد در محیطی تعریف شده برای تحقیقات چاپ شده در حوزه‌ی هوش مصنوعی ایجاد نماید، تا نتایج تحقیقات چاپ شده با سهولت بیشتری قابل تولید و عرضه باشند. این پروژه به دنبال ایجاد یک رابط کاربری ساده برای استفاده‌کنندگان خود می‌باشد. از ژوئن ۲۰۱۷، *Gym* تنها برای پایتون قابل استفاده بود. از سپتامبر ۲۰۱۷، تارنمای مستندسازی *Gym* توسعه داده نشده و کارهای جدید بیشتر بر روی صفحه *GitHub* آن متمرکز شده‌اند.

RoboSumo

در سال ۲۰۱۷ عرضه شده است، *RoboSumo* دنیای مجازی است که در آن عامل‌های رباتیک انسان‌نما با قابلیت یادگیری ماشینی^{۱۲۹} قرار دارند که حتی دانش راه‌رفتن ابتدایی را نداشته، ولی برای آن‌ها هدف‌های یادگیری چگونه حرکت کردن و خارج نمودن عامل رباتیک دیگر از رینگ، آموزش داده شده است. از طریق این فرآیند متنوع یادگیری، عامل‌های رباتیک نحوه‌ی انطباق با شرایط را یاد می‌گیرند؛ و هنگامی که عامل رباتیک از محیط مجازی فعلی به محیط مجازی جدید با ورزش باد تند منتقل شود، با توجه به این که پیشتر و بصورت کلی نحوه‌ی برقراری تعادل را آموخته است، در محیط جدید نیز می‌تواند تعادل خود را حفظ نماید. ایگور مرداچ از شرکت *OpenAI* این گونه استدلال می‌کند که رقابت میان عامل‌های رباتیک می‌تواند منجر به ایجاد هوش مبتنی بر "جنگ قدرت"^{۱۳۰} شده که این خود عامل افزایش توانایی عامل رباتیک برای حرکت، حتی در محیطی خارج از محیط رقابت است.

عامل‌های هوشمند در بازی‌ها و محیط‌های کاری هوشمند

OpenAI Five نام تیمی از پنج عامل هوشمند مبتنی بر *OpenAI* بوده که در رقابتی بصورت پنج با پنج در بازی تصویری به نام *Dota 2* استفاده شده، که یادگرفته‌اند چگونه مقابل انسان و با مهارت بالا و با استفاده از الگوریتم‌های تماماً مبتنی بر سعی و خطا^{۱۳۱}، بازی کنند. قبل از ورود تیم شامل پنج عامل، اولین ارائه‌ی بین‌المللی آن در تورنمنت قهرمانی بازی‌های رایانه‌ای آماتور در سال ۲۰۱۷ صورت گرفت، که در آن بازی دندی (یک بازیکن حرفه‌ای اهل اوکراین) در یک بازی نفر به نفر، نتیجه را به یک عامل هوشمند باخت. بعد از آن بازی، مدیر فنی شرکت گرگ برکمن^{۱۳۲} این گونه توضیح داد که عامل هوشمند برای مدت دوهفته و در زمان واقعی با بازی مقابل خودش، روش بازی را یادگرفته بود، که آن نرم‌افزار یادگیری، قدمی در جهت ساخت نرم‌افزارهایی برای کار در شرایط دشوار و پیچیده، مانند عمل جراحی بود. آن سیستم از شکلی از بازیادگیری، همان گونه که عامل‌های هوشمند در

^{۱۲۹} *Meta-Learning*، به الگوریتم‌های یادگیری ماشینی اطلاق می‌شود که در آن دانش و اطلاعات یادگرفته شده از خارج از الگوریتم‌های یادگیری کسب

می‌شوند.

^{۱۳۰} *Arms Race* – به تلاش برای برتری جویی با تولید و انبار کردن سلاح اطلاق می‌شود.

^{۱۳۱} *Trial-and-Error Algorithms*

^{۱۳۲} *Greg Berckman*



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

طول زمان و با بازی مقابل خودشان برای ماه‌ها و روزی چندصد دفعه برای انجام عملیاتی مانند کشتن دشمن و کشیدن نقشه‌ی اجسام یادگرفته‌اند، استفاده می‌کند.

تا ژوئن ۲۰۱۸، توانایی عامل‌های هوشمند برای بازی مقابل یکدیگر و در غالب یک تیم کامل پنج نفره گسترش یافت، که قادر بودند تا تیم‌های آماتور و نیمه حرفه‌ای را شکست دهند. در مسابقات بین‌المللی ۲۰۱۸، تیم چنج نفره *OpenAI* در دو مسابقه‌ی نمایشی مقابل بازیکنان حرفه‌ای بازی کرد، که البته در نهایت هردو مسابقه را باخت. در آوریل ۲۰۱۹، نیم پنج نفره *OpenAI* تیم *OG*، قهرمان مسابقات جهانی بازی‌های رایانه‌ای را در یک مسابقه نمایشی زنده در شهر سانفرانسیسکو با نتیجه ۲:۰ مغلوب کرد. آخرین ظهور عمومی عامل‌های هوشمند در روزهای پایانی همان ماه بود، جایی که در مجموع ۴۲.۷۲۹ بازی برخط را در چهار روز رقابت آزاد انجام دادند و موفق شدند ۹۹.۴٪ از بازی‌ها را پیروز شوند.

مکانیزم‌های بکاررفته در عامل‌های هوشمند *Dota 2* چالش‌های سیستم‌های هوش مصنوعی با چندبازیکن و در مسابقات برخط بازی‌های (*MOBA*) را نمایش داد و این‌که چگونه عامل‌های تیم پنج نفره *OpenAI* با استفاده از بازیادگیری عمیق موفق به دستیابی و ارائه‌ی ابر هوشمندی در مسابقات *Dota 2* شدند.

در سال ۲۰۱۸ عرضه شد، *Gym Retro* سکویی (پلتفرم) برای تحقیق بازیادگیری در بازی‌های ویدیویی می‌باشد. تمرکز تحقیق‌های پیشین در بازیادگیری بر روی عملکرد بهینه عامل‌های هوشمند در انجام یک کار بوده‌است. سکوی *Gym Retro* امکان بررسی بازی‌های با موضوع مشابه ولی ظاهر متفاوت با یکدیگر را می‌دهد.

در سال ۲۰۱۸، *OpenAI* بازی *Debate Game* را عرضه نمود، این بازی به ماشین نحوه‌ی بحث و دفاع در مقابل قاضی واقعی را از طریق ابزارهای تحلیل کیفیت الگوریتم‌های هوش مصنوعی^{۱۳۳} می‌دهد. هدف از آن تحقیق در مورد امکان استفاده از این ابزار به منظور ارزیابی الگوریتم تصمیم‌مبته بر هوش مصنوعی و توسعه‌ی ابزارها و سکوهایی تشریح مدل‌های یادگیری و درک روش‌های پیش‌بینی در الگوریتم هوش مصنوعی^{۱۳۴}، می‌باشد.

Dactyl

در سال ۲۰۱۸ توسعه داده شده است، *Dactyl* از یادگیری ماشین برای آموزش دست رباتیک شبیه انسان^{۱۳۵}، برای جابجایی اشیاء فیزیکی استفاده می‌کند. این ابزار بطور کامل از الگوریتم‌ها و برنامه‌ی رایانه‌ای آموزش بازگیری مشابه *OpenAI Five*^{۱۳۶}، استفاده می‌نماید. در این ابزار *OpenAI* چالش جابجایی اجسام را با استفاده از انتخاب تصادفی محیط‌های حرکتی^{۱۳۷}، رویکرد شبیه‌سازی که به جای سعی در ایجاد محیط واقعی سعی در انتقال تجربیات متنوعی از محیط‌های مختلف به یادگیرنده دارد، انجام می‌دهد. ساختار و قالب *Dactyl* علاوه بر استفاده از دوربین‌های رصد حرکت، همچنین شامل دوربین‌های فیلم‌برداری تصویر بصورت رنگی بوده که به ربات امکان

^{۱۳۳} به استفاده از ابزارهایی برای آزمایش و ارائه‌ی متدلوژی مورد نظر در هوش مصنوعی *Toy Problems* اطلاق می‌شود. محققان از این *Toy Problems* برای مقایسه‌ی کیفیت الگوریتم‌های مختلف هوش مصنوعی استفاده می‌کنند.

^{۱۳۴} به این مجموعه از ابزارها و سکوها در هوش مصنوعی، *Explainable AI* اطلاق می‌شود.

^{۱۳۵} *Shadow Hand*

^{۱۳۶} برنامه‌ی رایانه‌ای مبتنی بر الگوریتم‌های هوش مصنوعی مورد استفاده در بازی تصویری *Dota 2* شامل پنج عامل هوش مصنوعی می‌باشد.

^{۱۳۷} *Domain Randomization*

جایجایی اجسام دلخواه را توسط مشاهده و تشخیص آن اجسام، می‌دهد. در سال ۲۰۱۸، OpenAI نشان داد که این سیستم امکان جایجایی یک مکعب و یک حجم هشت وجهی^{۱۳۸} را دارد.

در سال ۲۰۱۹، OpenAI نشان داد که چگونه Dactyl توانایی حل مکعب روبیک را دارد. این ربات امکان حل پازل مکعب کیوبیک را در ۶۰٪ زمان داشت. اشیایی مانند مکعب کیوبیک جنبه‌های پیچیده‌تری از فیزیک را برای مدل کردن توسط ماشین مطرح می‌کند. OpenAI این مسئله را از طریق ارتقای کارایی و کیفیت Dactyl با استفاده از تکنیکی به نام ADR (Automatic Domain Randomization)، انتخاب تصادفی محیط‌ها، رویکردی که بطور مستمر و نامحدود محیط‌هایی با درجه‌ی پیچیدگی بیشتر را ایجاد می‌نماید، مرتفع نموده است. تفاوت ADR با روش‌های دستی ایجاد محیط‌های پیچیده، در عدم نیاز به انسان به‌منظور تعیین محدوده‌ی محیط‌های تصادفی است.

API

در ژوئن ۲۰۲۰، OpenAI یک رابط کاربری نرم‌افزاری با هدف دسترسی به مدل‌های هوش مصنوعی جدید توسعه داده شده توسط OpenAI را که امکان استفاده از آن‌ها را برای توسعه‌دهندگان نرم‌افزار به‌زبان انگلیسی فراهم می‌نمود، عرضه کرد.

مدل‌های مولد

اولین مقاله مدل پیش‌آموزش مولد زبان خودیادگیرنده^{۱۳۹} (مدلی که در آن با پردازش زبان طبیعی عملیات موضوع عملکرد سیستم اجرا و به سیستم آموزش داده می‌شود)، توسط الک رادفورد و همکارانش در ژوئن سال ۲۰۱۸ در تارنمای OpenAI منتشر گردید. این مقاله نشان داد که چگونه یک مدل مولد زبانی، قادر است توسط پیش‌آموزش متون مختلف و از طریق بررسی ارتباط متن‌ها، در دراز مدت دانش و ارتباط‌های فرآیندهای موجود در جهان را فراگیرد.

GPT-2

GPT-2 یک مدل زبانی خودیادگیرنده هدایت‌نشده^{۱۴۰} است که نسخه‌ی توسعه داده شده از نسخه‌ی اولیه‌ی مدل GPT (GPT-1) می‌باشد. GPT-2 برای اولین مرتبه و با عرضه‌ی عمومی نسخه‌های محدودی از آن در فوریه ۲۰۱۹ ارائه شد. نسخه‌ی کامل GPT-2 به‌دلیل وجود حساسیت‌هایی بالقوه در استفاده‌ی نادرست از آن در نوشتن اخبار جعلی، با تاخیر عرضه شد. تعدادی از متخصصان نگرانی‌هایی را در خصوص تهدیدهای جدی ناشی از عرضه GPT-2 مطرح کرده بودند.

موسسه هوش مصنوعی آلن^{۱۴۱} ابزاری را به‌منظور تشخیص اخبار جعلی ایجاد شده توسط GPT-2 عرضه نمود. محققان دیگری، مانند جرمی هووارد^{۱۴۲}، در مورد "قابلیت آن در ایجاد نوشتار و گفتار زبان عادی و فراگیر شدن

^{۱۳۸} Octagonal Prism

^{۱۳۹} Generative Pre-trained Transformer Model – GPT Model

^{۱۴۰} Unsupervised Self-Learning Language Model

^{۱۴۱} Allen Artificial Intelligence Institute

^{۱۴۲} Jeremy Howard



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

این فناوری در توییتر، رایانامه، و تارنما که می‌تواند جایگزین مطالب دیگر شده و فیلتر نمودن آن غیرممکن می‌باشد^{۱۴۳}، هشدار داده بودند. در نوامبر ۲۰۱۹، OpenAI نسخه کامل مدل زبان GPT-2 را عرضه کرد. در ماه می سال ۲۰۲۰، OpenAI زبانی خودیادگیرنده هدایت‌نشده GPT-3 خود را که بالغ بر ۱۷۵ میلیارد پارامتر را در پردازش خود بکار می‌گیرد، عرضه نمود. پردازش این تعداد از پارامترها، چالش جدید مدل‌های زبانی خودیادگیرنده به زیرساخت‌های قوی پردازنده‌ها را بیش از پیش آشکار نمود، به‌همین دلیل عرضه عمومی این محصول با تاخیر صورت گرفت. در ۲۲ سپتامبر ۲۰۲۰، لایسنس اختصاصی این محصول به شرکت مایکروسافت واگذار گردید.

در نوامبر سال ۲۰۲۲، شرکت OpenAI محصول مطرح و شناخته شده‌ی خود را به‌نام GPT-4 ارائه نمود که قابلیت دریافت درخواست کاربر از طریق زبان محاوره‌ای را داشت. شرکت OpenAI با اضافه نمودن قابلیت محاوره‌ی دوسوی با کاربر به زبان طبیعی، محصول بانام تجاری ChatGPT را عرضه عمومی نمود.

چالش‌ها و حساسیت‌ها در هوش مصنوعی مولد

توسعه‌ی هوش مصنوعی مولد منجر به ایجاد دغدغه‌هایی برای دولت‌ها، کسب و کارها، و اشخاص و در نتیجه راه‌پیمایی و اعتراض‌هایی برای توقف استفاده از این فناوری و توسعه‌ی آن و همچنین برخوردهایی از جانب دولت‌ها شده است. جولای ۲۰۲۳ آنتونیو گوترش (دبیر کل سازمان ملل متحد) در سخنانی کوتاهی در سازمان ملل این‌گونه مطرح نموده که "هوش مصنوعی مولد دارای قابلیت‌های خوبی است ولی کاربردهای گسترده از آن می‌تواند بسیار وحشتناک باشد"، از آنجایی که هوش مصنوعی می‌تواند به عنوان "شتاب‌دهنده توسعه جهانی"^{۱۴۴} و تا سال ۲۰۳۰ نقشی بالغ بر ۱۰ تا ۱۵ تریلیون دلار در اقتصاد جهانی ایفا نموده، اما استفاده مخرب از آن می‌تواند "منجر به سطح وحشتناکی از کشتار، خرابی، عوارض ناخوشایند گسترده و آسیب عمیق روانی در ابعادی غیرقابل تصور، گردد".

از دست رفتن مشاغل

از روزهای ابتدایی توسعه هوش مصنوعی مخالفت‌هایی با جوزف ویزنباوم^{۱۴۴} سازنده الیزا^{۱۴۵} و دیگران در خصوص این‌که آیا کارهای انجام شده توسط رایانه‌ها اساساً می‌بایست توسط آن‌ها انجام شوند یا نه، طرح اختلاف‌های فیمابین رایانه و انسان، و محاسبات مقداری و کیفی، و همچنین قضاوت مبتنی بر ارزش، طرح شده‌اند. در آوریل ۲۰۲۳ طی گزارشی اعلام شد که هوش مصنوعی مولد تصاویر منجر به از دست رفتن حدود ۷۰٪ از شغل‌های مربوط به سازندگان بازی‌های رایانه‌ای در چین شده است. در جولای ۲۰۲۳، توسعه‌های صورت‌گرفته در هوش مصنوعی مولد در اختلاف‌های کارگری هالیود موثر بوده‌اند. فرن درشر^{۱۴۶}، رئیس Screen Actor Guild، در ۲۰۲۳

^{۱۴۳} Turbocharge Global Development

^{۱۴۴} Joseph Weizenbaum

^{۱۴۵} ELIZA

^{۱۴۶} Fran Drescher

^{۱۴۷} SAG-AFTRA strike نشان داد که چگونه "هوش مصنوعی می‌تواند از یک تهدید بالقوه به یک عامل مخرب تبدیل شود".

تصاویر و اشخاص جعلی^{۱۴۸}

دیپ‌فیک‌ها (واژه‌ی ترکیبی از یادگیری عمیق (دیپ) و جعلی (فیک)) هوش مصنوعی مولد رسانه می‌باشند که با استفاده از شبکه‌های عصبی مصنوعی، تصویر شخص دیگری را بروی تصویر و یا فیلم شخص حاضر، نمایش می‌دهند. دیپ‌فیک‌ها به دلیل کاربردهای جنسی جعلی هنرمندان و هنرپیشه‌ها، تصاویر جنسی شخصی افراد، اخبار جعلی، شوخی‌ها و جوک‌های زننده در خصوص باورهای افراد^{۱۴۹}، و کلاهبرداری‌های مالی، توجه‌ها و حساسیت‌های زیادی را به خود جلب کرده‌اند.

جرائم فضای مجازی

توانایی هوش مصنوعی مولد در خلق محتوای جعلی در بسیاری از انواع جرائم رایانه‌ای، شامل پیام‌های جعلی^{۱۵۰}، ویدئوها و صداها، جعلی که برای ایجاد اخبار نادرست و کلاهبرداری استفاده می‌شوند، مشاهده شده است. سزار پیشین کلاهبرداری گوگل شومان قوسمایومدر^{۱۵۱}، پیش‌بینی کرده هنگامی که ویدئوهای جعلی در حال شکل‌گیری و ایجاد در رسانه‌ها می‌باشند، خیلی زود فراگیر و در نتیجه بسیار خطرناک خواهند شد. جرائم فضای مجازی مدل‌های بزرگ زبانی که مختص فعالیت‌های کلاهبرداری می‌باشند مانند WormGPT و FraudGPT را ایجاد نموده‌اند.

استفاده نادرست در روزنامه‌نگاری

در ژانویه ۲۰۲۳، تارنمای Futurism.com مطلبی را برای اولین مرتبه عنوان کرده که تارنمای CNET از یک ابزار مخفی داخلی هوش مصنوعی برای نوشتن حداقل ۷۷ عدد از داستان‌هایش استفاده نموده است، پس از آن که اخبار منتشر گردید، CNET با ارسال پستی تعداد صحیح داستان‌ها را ۴۴ عدد اعلام کرد. در آوریل ۲۰۲۳، نشریه آلمانی Die Aktuelle مصاحبه‌ی جعلی با استفاده از هوش مصنوعی مولد را با راننده مسابقات مایکل شوماخر منتشر کرد، این درحالی بود که از سال ۲۰۱۳ تاکنون و پس از جراحت مغزی ناشی از حادثه‌ای در هنگام اسکی، در دید عموم ظاهر نشده بود. داستان حاوی دو جنبه افشاگری بود. مطلب روی جلد حاوی سطری با این عنوان "به نظر واقعی"، و انتهای مصاحبه نیز حاوی مطلبی که نشان‌دهنده‌ی ایجاد شدن آن توسط هوش مصنوعی مولد بود. سردبیر مجله مدتی کوتاه پس از آن که درگیر اختلاف و شکایت گردید، اخراج شد.

قانون و مقررات در هوش مصنوعی مولد

^{۱۴۷} Screen Actors Guild – American Federation of Television and Radio Artists

^{۱۴۸} Deep fakes

^{۱۴۹} Hoaxes

^{۱۵۰} Phishing Scam - نوعی از فیشینگ برخط می‌باشد که در شبکه اینترنت و با ارسال پیام‌های جعلی برای مصرف‌کنندگان اقدام به سرقت اطلاعات

آن‌ها، می‌نمایند

^{۱۵۱} Shuman Ghosemajumder, Former Google Fraud Czar



هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

اتحادیه اروپا، دستورالعمل هوش مصنوعی شامل کلیه موارد مورد نیاز برای افشای حق نشر مطالب در استفاده برای آموزش سیستم‌های هوش مصنوعی، همچنین برچسب‌گذاری هر نوع خروجی سیستم‌های هوش مصنوعی مولد، را تدوین و ابلاغ نموده است.

در ایالات متحده، مجموعه‌ای از شرکت‌ها شامل *Meta* و *Alphabet*، *OpenAI* در جولای ۲۰۲۳ توافقنامه‌ای با کاخ سفید برای مشخص نمودن محتواهای تولید شده توسط هوش مصنوعی مولد، امضا کردند. در چین، دستورالعمل موقتی برای مدیران در استفاده از خدمات هوش مصنوعی مولد توسط سازمان نظارتی فضای مجازی چین برای کلیه موارد هوش مصنوعی مولد که در دسترس عموم قراردارند، ارائه شده است. این دستورالعمل شامل کلیه ملزومات برای مشخص نمودن تصاویر و ویدئوهای تولید شده، مقررات مرتبط با کیفیت آموزش داده‌ها و طبقه‌بندی آن‌ها، محدودیت‌های جمع‌آوری اطلاعات شخصی، و راهنمای آن‌که هوش مصنوعی می‌بایست با هدف تحقق ارزش‌های بنیادی جامعه بکار گرفته شود ^{۱۵۲} "adhere to socialist core values"

شورای عالی

کارشناسان رسمی دادگستری

^{۱۵۲} ارزشی که با ۲۴ حرف چینی نوشته شده و شامل ارزش‌های ملی "موفقیت"، "دموکراسی"، "احترام اجتماعی" و "همه‌نگی" و ارزش‌های اجتماعی "آزادی"، "برابری"، "عدالت" و "قانون" و ارزش‌های شخصی "اتحاد"، "تمرکز"، "یکپارچگی" و "دوستی" می‌باشند.

۱. Newsom, Gavin; Weber, Shirley N. (September 6, 2023). "[Executive Order N-12-23](#)" (PDF). Executive Department, State of California. Retrieved September 7, 2023.
۲. [^] [Jump up to: a b](#) Griffith, Erin; Metz, Cade (January 27, 2023). "[Anthropic Said to Be Closing In on \\$300 Million in New A.I. Funding](#)". *The New York Times*. Retrieved March 14, 2023.
۳. [^] Lanxon, Nate; Bass, Dina; Davalos, Jackie (March 10, 2023). "[A Cheat Sheet to AI Buzzwords and Their Meanings](#)". Bloomberg News. Retrieved March 14, 2023.
۴. [^] Pinaya, Walter H. L.; Graham, Mark S.; Kerfoot, Eric; Tudosiu, Petru-Daniel; Dafflon, Jessica; Fernandez, Virginia; Sanchez, Pedro; Wolleb, Julia; da Costa, Pedro F.; Patel, Ashay (2023). "Generative AI for Medical Imaging: extending the MONAI Framework". *arXiv:2307.15208 [eess.IV]*.
۵. [^] Pasick, Adam (March 27, 2023). "[Artificial Intelligence Glossary: Neural Networks and Other Terms Explained](#)". The New York Times. *ISSN 0362-4331*. Retrieved April 22, 2023.
۶. [^] Karpathy, Andrej; Abbeel, Pieter; Brockman, Greg; Chen, Peter; Cheung, Vicki; Duan, Yan; Goodfellow, Ian; Kingma, Durk; Ho, Jonathan; Rein Houthooft; Tim Salimans; John Schulman; Ilya Sutskever; Wojciech Zaremba (June 16, 2016). "[Generative models](#)". OpenAI.
۷. [^] Metz, Cade (March 14, 2023). "[OpenAI Plans to Up the Ante in Tech's A.I. Race](#)". The New York Times. *ISSN 0362-4331*. Retrieved March 31, 2023.
۸. [^] Thoppilan, Romal; De Freitas, Daniel; Hall, Jamie; Shazeer, Noam; Kulshreshtha, Apoorv (January 20, 2022). "LaMDA: Language Models for Dialog Applications". *arXiv:2201.08239 [cs.CL]*.
۹. [^] Roose, Kevin (October 21, 2022). "[A Coming-Out Party for Generative A.I., Silicon Valley's New Craze](#)". *The New York Times*. Retrieved March 14, 2023.
۱۰. [^] ["Don't fear an AI-induced jobs apocalypse just yet"](#). The Economist. March 6, 2023. Retrieved March 14, 2023.
۱۱. [^] Harreis, H.; Koullias, T.; Roberts, Roger. "[Generative AI: Unlocking the future of fashion](#)".
۱۲. [^] ["How Generative AI Can Augment Human Creativity"](#). Harvard Business Review. June 16, 2023. *ISSN 0017-8012*. Retrieved June 20, 2023.
۱۳. [^] ["The race of the AI labs heats up"](#). The Economist. January 30, 2023. Retrieved March 14, 2023.
۱۴. [^] Yang, June; Gokturk, Burak (March 14, 2023). "[Google Cloud brings generative AI to developers, businesses, and governments](#)".
۱۵. [^] Hendrix, Justin (May 16, 2023). "[Transcript: Senate Judiciary Subcommittee Hearing on Oversight of AI](#)". techpolicy.press. Retrieved May 19, 2023.
۱۶. [^] Simon, Felix M.; Altay, Sacha; Mercier, Hugo (October 18, 2023). "[Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown](#)". Harvard Kennedy School Misinformation Review. *doi:10.37016/mr-2020-127*. *S2CID 264113883*.
۱۷. [^] Crevier, Daniel (1993). *AI: The Tumultuous Search for Artificial Intelligence*. New York, New York: BasicBooks. p. 109. *ISBN 0-465-02997-3*.
۱۸. [^] Newquist, H. P. (1994). *The Brain Makers: Genius, Ego, And Greed In The Quest For Machines That Think*. New York: Macmillan/SAMS. pp. 45–53. *ISBN 978-0-672-30412-5*.
۱۹. [^] Sharkey, Noel (July 4, 2007), [A programmable robot from 60 AD](#), vol. 2611, New Scientist, [archived](#) from the original on January 13, 2018, retrieved October 22, 2019
۲۰. [^] Brett, Gerard (July 1954), "The Automata in the Byzantine 'Throne of Solomon'", *Speculum*, **29** (3): 477–487, *doi:10.2307/2846790*, *ISSN 0038-7134*, *JSTOR 2846790*, *S2CID 163031682*.
۲۱. [^] kelinich (March 8, 2014). "[Maillart's Automaton](#)". The Franklin Institute. Retrieved August 24, 2023.
۲۲. [^] SITNFlash (August 28, 2017). "[The History of Artificial Intelligence](#)". Science in the News. Retrieved December 8, 2023.
۲۳. [^] Bergen, Nathan; Huang, Angela (2023). "[A Brief History of Generative AI](#)" (PDF). Dichotomies: Generative AI: Navigating Towards a Better Future (2): 4.
۲۴. [^] Gagniu, Paul A. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. USA, New Jersey: John Wiley & Sons. pp. 2–8. *ISBN 978-1-119-38755-8*.
۲۵. [^] Grinstead, Charles Miller; Snell, James Laurie (1997). [Introduction to Probability](#). American Mathematical Society. pp. 464–466. *ISBN 978-0-8218-0749-1*.

۲۶. ^۱ Bremaud, Pierre (March 9, 2013). [Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues](#). Springer Science & Business Media. p. ix. ISBN 978-1-4757-3124-8. Archived from the original on March 23, 2017.
۲۷. ^۱ Hayes, Brian (2013). "First Links in the Markov Chain". American Scientist. **101** (2): ۹۲. doi:10.1511/2013.101.92. ISSN 0033-9966.
۲۸. ^۱ Fine, Shai; Singer, Yoram; Tishby, Naftali (July 1, 1998). "The Hierarchical Hidden Markov Model: Analysis and Applications". Machine Learning. **32** (1): 41–62. doi:10.1023/A:1007469218079. ISSN 1573-0565. S2CID 3465810.
۲۹. ^۱ Jebara, Tony (2012). Machine learning: discriminative and generative. Vol. 755. Springer Science & Business Media.
۳۰. ^۱ Cao, Yihan; Li, Siyu; Liu, Yixin; Yan, Zhiling; Dai, Yutong; Yu, Philip S.; Sun, Lichao (March 7, 2023). "A Comprehensive Survey of AI-Generated Content (AIGC): A History of Generative AI from GAN to ChatGPT". arXiv:2303.04226 [cs.AI].
۳۱. ^۱ "finetune-transformer-lm". GitHub. Retrieved May 19, 2023.
۳۲. ^۱ Radford, Alec; Wu, Jeffrey; Child, Rewon; Luan, David; Amodei, Dario; Sutskever, Ilya; others (2019). "Language models are unsupervised multitask learners". OpenAI Blog. **1** (8): 9.
۳۳. ^۱ Bubeck, Sébastien; Chandrasekaran, Varun; Eldan, Ronen; Gehrke, Johannes; Horvitz, Eric; Kamar, Ece; Lee, Peter; Lee, Yin Tat; Li, Yuanzhi; Lundberg, Scott; Nori, Harsha; Palangi, Hamid; Ribeiro, Marco Tulio; Zhang, Yi (March 22, 2023). "Sparks of Artificial General Intelligence: Early experiments with GPT-4". arXiv:2303.12712 [cs.CL].
۳۴. ^۱ Schlagwein, Daniel; Willcocks, Leslie (September 13, 2023). "ChatGPT et al: The Ethics of Using (Generative) Artificial Intelligence in Research and Science". Journal of Information Technology. **38** (2): 232–238. doi:10.1177/02683962231200411. S2CID 261753752.
۳۵. ^۱ "A History of Generative AI: From GAN to GPT-4". March 21, 2023.
۳۶. ^۱ "Explainer: What is Generative AI, the technology behind OpenAI's ChatGPT?". Reuters. March 17, 2023. Retrieved March 17, 2023.
۳۷. ^۱ Bommasani, R.; Hudson, D. A.; Adeli, E.; Altman, R.; Arora, S.; von Arx, S.; Bernstein, M. S.; Bohg, J.; Bosselut, A.; Brunskill, E.; Brynjolfsson, E. (August 16, 2021). "On the opportunities and risks of foundation models". arXiv:2108.07258 [cs.LG].
۳۸. ^۱ Chen, Ming; Tworek, Jakub; Jun, Hongyu; Yuan, Qinyuan; Pinto, Hanyu Philippe De Oliveira; Kaplan, Jerry; Edwards, Haley; Burda, Yannick; Joseph, Nicholas; Brockman, Greg; Ray, Alvin (July 6, 2021). "Evaluating Large Language Models Trained on Code". arXiv:2107.03374 [cs.LG].
۳۹. ^۱ Epstein, Ziv; Hertzmann, Aaron; Akten, Memo; Farid, Hany; Fjeld, Jessica; Frank, Morgan R.; Groh, Matthew; Herman, Laura; Leach, Neil; Mahari, Robert; Pentland, Alex "Sandy"; Russakovsky, Olga; Schroeder, Hope; Smith, Amy (2023). "Art and the science of generative AI". Science. **380** (6650): 1110–1111. arXiv:2306.04141. Bibcode:2023Sci...380.1110E. doi:10.1126/science.adh4451. PMID 37319197. S2CID ۲۵۹۰۹۵۲۷.
۴۰. ^۱ Ramesh, Aditya; Pavlov, Mikhail; Goh, Gabriel; Gray, Scott; Voss, Chelsea; Radford, Alec; Chen, Mark; Sutskever, Ilya (2021). "Zero-shot text-to-image generation". International Conference on Machine Learning. PMLR. pp. 8821–8831.
۴۱. ^۱ Desai, Saahil (July 17, 2023). "A Voicebot Just Left Me Speechless". The Atlantic. Retrieved November 28, 2023.
۴۲. ^۱ Agostinelli, Andrea; Denk, Timo I.; Borsos, Zalán; Engel, Jesse; Verzetti, Mauro; Caillon, Antoine; Huang, Qingqing; Jansen, Aren; Roberts, Adam; Tagliasacchi, Marco; Sharifi, Matt; Zeghidour, Neil; Frank, Christian (January 26, 2023). "MusicLM: Generating Music From Text". arXiv:2301.11325 [cs.SD].
۴۳. ^۱ Dalugdug, Mandy (August 3, 2023). "Meta in June said that it used 20,000 hours of licensed music to train MusicGen, which included 10,000 "high-quality" licensed music tracks. At the time, Meta's researchers outlined in a paper the ethical challenges that they encountered around the development of generative AI models like MusicGen".
۴۴. ^۱ Metz, Cade (April 4, 2023). "Instant Videos Could Represent the Next Leap in A.I. Technology". The New York Times.
۴۵. ^۱ Wong, Queenie (September 29, 2022). "Facebook Parent Meta's AI Tool Can Create Artsy Videos From Text". cnet.com. Retrieved April 4, 2023.

۴۶. ^۱ [Heaven, Will Douglas \(February 15, 2023\). "AI is dreaming up drugs that no one has ever seen. Now we've got to see if they work". MIT Technology Review. Massachusetts Institute of Technology. Retrieved March 15, 2023.](#)
۴۷. ^۱ [Yang, Sherry; Du, Yilun \(April 12, 2023\). "UniPi: Learning universal policies via text-guided video generation". Google Research, Brain Team. Google AI Blog.](#)
۴۸. ^۱ [Brohan, Anthony \(2023\). "RT-2: Vision-Language-Action Models Transfer Web Knowledge to Robotic Control". *arXiv:2307.15818 \[cs.RO\]*.](#)
۴۹. ^۱ [Jump up to: a b Altling, Leo; Zhang, Hongchao \(1989\). "Computer aided process planning: the state-of-the-art survey". The International Journal of Production Research. 27 \(4\): 553–585. doi:10.1080/00207179.1989.11988679.](#)
۵۰. ^۱ [Chien, Steve \(1998\). "Automated planning and scheduling for goal-based autonomous spacecraft". IEEE Intelligent Systems and Their Applications. 13 \(5\): 50–55. doi:10.1109/5254.722362.](#)
۵۱. ^۱ [Burstein, Mark H., ed. \(1994\). ARPA/Rome Laboratory Knowledge-based Planning and Scheduling Initiative Workshop Proceedings. The Advanced Research Projects Agency, Department of Defense, and Rome Laboratory, US Air Force, Griffiss AFB. p. 219. ISBN 155860345X.](#)
۵۲. ^۱ [Pell, Barney; Bernard, Douglas E.; Chien, Steve A.; Gat, Erann; Muscettola, Nicola; Nayak, P. Pandurang; Wagner, Michael D.; Williams, Brian C. \(1998\). Bekey, George A. \(ed.\). An Autonomous Spacecraft Agent Prototype. Autonomous Robots Volume 5, No. 1. pp. 29–45. Our deliberator is a traditional generative AI planner based on the HSTS planning framework \(Muscettola, 1994\), and our control component is a traditional *spacecraft attitude control* system \(Hackney et al. 1993\). We also add an architectural component explicitly dedicated to world modeling \(the mode identifier\), and distinguish between control and monitoring.](#)
۵۳. ^۱ [Rumiantsau, Michael \(November 16, 2023\). "Generative BI: Setting a New Standard for Business Intelligence".](#)
۵۴. ^۱ [Sabin, Sam \(June 30, 2023\). "GitHub has a vision to make code more secure by design". Axios Codebook. Retrieved August 15, 2023.](#)
۵۵. ^۱ [Vincent, James \(March 20, 2023\). "Text-to-video AI inches closer as startup Runway announces new model". The Verge. Retrieved August 15, 2023. Text-to-video is the next frontier for generative AI, though current output is rudimentary. Runway says it'll be making its new generative video model, Gen-2, available to users in 'the coming weeks.'](#)
۵۶. ^۱ [Vanian, Jonathan \(March 16, 2023\). "Microsoft adds OpenAI technology to Word and Excel". CNBC. Retrieved August 15, 2023. Microsoft is bringing generative artificial intelligence technologies such as the popular ChatGPT chatting app to its Microsoft 365 suite of business software....the new A.I. features, dubbed Copilot, will be available in some of the company's most popular business apps, including Word, PowerPoint and Excel.](#)
۵۷. ^۱ [Wilson, Mark \(August 15, 2023\). "The app's Memories feature just got a big upgrade". TechRadar. The Google Photos app is getting a redesigned, AI-powered Memories feature...you'll be able to use generative AI to come up with some suggested names like "a desert adventure".](#)
۵۸. ^۱ [Sullivan, Laurie \(May 23, 2023\). "Adobe Adds Generative AI To Photoshop". MediaPost. Retrieved August 15, 2023. Generative artificial intelligence \(AI\) will become one of the most important features for creative designers and marketers. Adobe on Tuesday unveiled a Generative Fill feature in Photoshop to bring Firefly's AI capabilities into design.](#)
۵۹. ^۱ [Michael Nuñez \(July 19, 2023\). "LLaMA 2: How to access and use Meta's versatile open-source chatbot right now". VentureBeat. Retrieved August 15, 2023. If you want to run LLaMA 2 on your own machine or modify the code, you can download it directly from Hugging Face, a leading platform for sharing AI models.](#)
۶۰. ^۱ [Pounder, Les \(March 25, 2023\). "How To Create Your Own AI Chatbot Server With Raspberry Pi 4". Retrieved August 15, 2023. Using a Pi 4 with 8GB of RAM, you can create a ChatGPT-like server based on LLaMA.](#)
۶۱. ^۱ [Kemper, Jonathan \(November 10, 2022\). "'Draw Things' App brings Stable Diffusion to the iPhone". The Decoder. Retrieved August 15, 2023. Draw Things is an app that brings Stable Diffusion to the iPhone. The AI images are generated locally, so you don't need an Internet connection.](#)

۶۲. ^۱ Witt, Allan (July 7, 2023). ["Best Computer to Run LLaMA AI Model at Home \(GPU, CPU, RAM, SSD\)"](#). To run LLaMA model at home, you will need a computer build with a powerful GPU that can handle the large amount of data and computation required for inferencing.
۶۳. ^۱ Nellis, Stephen; Lee, Jane (September 1, 2022). ["U.S. officials order Nvidia to halt sales of top AI chips to China"](#). Reuters. Retrieved August 15, 2023.
۶۴. ^۱ Shilov, Anton (May 7, 2023). ["Nvidia's Chinese A800 GPU's Performance Revealed"](#). Tom's Hardware. Retrieved August 15, 2023. the A800 operates at 70% of the speed of A100 GPUs while complying with strict U.S. export standards that limit how much processing power Nvidia can sell.
۶۵. ^۱ Patel, Dylan (October 24, 2022). ["How China's Biren Is Attempting To Evade US Sanctions"](#). Retrieved August 15, 2023.
۶۶. ^۱ ["5 free software to recognise fake AI-generated images"](#) (in Italian). October 28, 2023.
۶۷. ^۱ ["Secretary-General's remarks to the Security Council on Artificial Intelligence"](#). un.org. July 18, 2023. Retrieved July 27, 2023.
۶۸. ^۱ ["The Writers Strike Is Taking a Stand on AI"](#). Time. May 4, 2023. Retrieved June 11, 2023.
۶۹. ^۱ Tarnoff, Ben (August 4, 2023). "Lessons from Eliza". [The Guardian Weekly](#). pp. 34–39.
۷۰. ^۱ Zhou, Viola (April 11, 2023). ["AI is already taking video game illustrators' jobs in China"](#). Rest of World. Retrieved August 17, 2023.
۷۱. ^۱ Carter, Justin (April 11, 2023). ["China's game art industry reportedly decimated by growing AI use"](#). Game Developer. Retrieved August 17, 2023.
۷۲. ^۱ Collier, Kevin (July 14, 2023). ["Actors vs. AI: Strike brings focus to emerging use of advanced tech"](#). NBC News. SAG-AFTRA has joined the Writer's [sic] Guild of America in demanding a contract that explicitly demands AI regulations to protect writers and the works they create. ... The future of generative artificial intelligence in Hollywood—and how it can be used to replace labor—has become a crucial sticking point for actors going on strike. In a news conference Thursday, Fran Drescher, president of the Screen Actors Guild-American Federation of Television and Radio Artists (more commonly known as SAG-AFTRA), declared that 'artificial intelligence poses an existential threat to creative professions, and all actors and performers deserve contract language that protects them from having their identity and talent exploited without consent and pay.'
۷۳. ^۱ Wiggers, Kyle (August 22, 2023). ["ElevenLabs' voice-generating tools launch out of beta"](#). TechCrunch. Retrieved September 25, 2023.
۷۴. ^۱ Shrivastava, Rashi. ["Keep Your Paws Off My Voice: Voice Actors Worry Generative AI Will Steal Their Livelihoods"](#). Forbes. Retrieved November 28, 2023.
۷۵. ^۱ Gupta, Shalene (October 31, 2023). ["Underrepresented groups in countries around the world are worried about AI being a threat to jobs"](#). Fast Company. Retrieved December 8, 2023.
۷۶. ^۱ ["Dangers of unregulated artificial intelligence"](#). CEPR. November 23, 2021. Retrieved December 8, 2023.
۷۷. ^۱ Frackiewicz, Marcin (August 29, 2023). ["Artificial Intelligence and the Study of Social Identity"](#). TS2 SPACE. Retrieved December 8, 2023.
۷۸. ^۱ Brandon, John (February 16, 2018). ["Terrifying high-tech porn: Creepy 'deepfake' videos are on the rise"](#). Fox News. [Archived](#) from the original on June 15, 2018. Retrieved February 20, 2018.
۷۹. ^۱ Cole, Samantha (January 24, 2018). ["We Are Truly Fucked: Everyone Is Making AI-Generated Fake Porn Now"](#). Vice. [Archived](#) from the original on September 7, 2019. Retrieved May 4, 2019.
۸۰. ^۱ ["What Are Deepfakes & Why the Future of Porn is Terrifying"](#). Highsnobiety. February 20, 2018. [Archived](#) from the original on July 14, 2021. Retrieved February 20, 2018.
۸۱. ^۱ ["Experts fear face swapping tech could start an international showdown"](#). The Outline. [Archived](#) from the original on January 16, 2020. Retrieved February 28, 2018.
۸۲. ^۱ Roose, Kevin (March 4, 2018). ["Here Come the Fake Videos, Too"](#). The New York Times. [ISSN 0362-4331](#). [Archived](#) from the original on June 18, 2019. Retrieved March 24, 2018.
۸۳. ^۱ Schreyer, Marco; Sattarov, Timur; Reimer, Bernd; Borth, Damian (2019). "Adversarial Learning of Deepfakes in Accounting". [arXiv:1910.03810 \[cs.LG\]](#).
۸۴. ^۱ ["Join the Deepfake Detection Challenge \(DFDC\)"](#). deepfakedetectionchallenge.ai. [Archived](#) from the original on January 12, 2020. Retrieved November 8, 2019.
۸۵. ^۱ Clarke, Yvette D. (June 28, 2019). ["H.R.3230 – 116th Congress \(2019-2020\): Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of](#)

هوش مصنوعی و سیستم‌های خبره در کارشناسی رسمی دادگستری (۲)

(جرائم هوش مصنوعی و هوش مصنوعی مولد)

- ۲۰۱۹". www.congress.gov. Archived from the original on December ۱۷, ۲۰۱۹. Retrieved October 16, 2019.
۸۶. ["People Are Still Terrible: AI Voice-Cloning Tool Misused for Deepfake Celeb Clips"](#). PCMag Middle East. January 31, 2023. Retrieved July 25, 2023.
 ۸۷. ["The generative A.I. software race has begun"](#). Fortune. Retrieved February 3, 2023.
 ۸۸. [Milmo, Dan; Hern, Alex \(May 20, 2023\). "Elections in UK and US at risk from AI-driven disinformation, say experts". The Guardian. ISSN 0261-3077. Retrieved July 25, 2023.](#)
 ۸۹. ["Seeing is believing? Global scramble to tackle deepfakes". news.yahoo.com. Retrieved February 3, 2023.](#)
 ۹۰. [Vincent, James \(January 31, 2023\). "4chan users embrace AI voice clone tool to generate celebrity hatespeech". The Verge. Retrieved February 3, 2023.](#)
 ۹۱. [Thompson, Stuart A. \(March 12, 2023\). "Making Deepfakes Gets Cheaper and Easier Thanks to A.I." The New York Times. ISSN 0362-4331. Retrieved July 25, 2023.](#)
 ۹۲. ["A new AI voice tool is already being abused to make deepfake celebrity audio clips". Engadget. January 31, 2023. Retrieved February 3, 2023.](#)
 ۹۳. [Gee, Andre \(April 20, 2023\). "Just Because AI-Generated Rap Songs Go Viral Doesn't Mean They're Good". Rolling Stone. Retrieved December 6, 2023.](#)
 ۹۴. [Coscarelli, Joe \(April 19, 2023\). "An A.I. Hit of Fake 'Drake' and 'The Weeknd' Rattles the Music World". The New York Times. Retrieved December 5, 2023.](#)
 ۹۵. [Lippiello, Emily; Smith, Nathan; Pereira, Ivan \(November 3, 2023\). "AI songs that mimic popular artists raising alarms in the music industry". ABC News. Retrieved December 6, 2023.](#)
 ۹۶. [Skelton, Eric. "Fans Are Using Artificial Intelligence to Turn Rap Snippets Into Full Songs". Complex. Retrieved December 6, 2023.](#)
 ۹۷. [Marr, Bernard. "Virtual Influencer Noonouri Lands Record Deal: Is She The Future Of Music?". Forbes. Retrieved December 6, 2023.](#)
 ۹۸. [Thaler, Shannon \(September 8, 2023\). "Warner Music signs first-ever record deal with AI pop star". New York Post. Retrieved December 6, 2023.](#)
 ۹۹. [Sjouwerman, Stu \(December 26, 2022\). "Deepfakes: Get ready for phishing 2.0". Fast Company. Retrieved July 31, 2023.](#)
 ۱۰۰. [Sonnemaker, Tyler. "As social media platforms brace for the incoming wave of deepfakes, Google's former 'fraud czar' predicts the biggest danger is that deepfakes will eventually become boring". Business Insider. Retrieved July 31, 2023.](#)
 ۱۰۱. [Collinson, Patrick \(July 15, 2023\). "Fake reviews: can we trust what we read online as use of AI explodes?". The Guardian. ISSN 0261-3077. Retrieved December 6, 2023.](#)
 ۱۰۲. ["After WormGPT, FraudGPT Emerges to Help Scammers Steal Your Data". PCMag. Retrieved July 31, 2023.](#)
 ۱۰۳. [Gupta, Maanak; Akiri, Charankumar; Aryal, Kshitiz; Parker, Eli; Praharaj, Lopamudra \(2023\). "From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy | IEEE Journals & Magazine | IEEE Xplore". IEEE Access. 11: 80218–۸۰۲۴۵. doi:10.1109/ACCESS.2023.3300381. S2CID ۲۵۹۳۱۶۱۲۲.](#)
 ۱۰۴. [Roth, Emma \(January 25, 2023\). "CNET found errors in more than half of its AI-written stories". The Verge. Retrieved June 17, 2023.](#)
 ۱۰۵. ["A magazine touted Michael Schumacher's first interview in years. It was actually AI". NPR. April 28, 2023. Retrieved June 17, 2023.](#)
 ۱۰۶. ["ChatGPT and the EU AI Act". mhc.ie. Mason Hayes & Curran. Retrieved November 29, 2023.](#)
 ۱۰۷. [Chee, Foo Yun; Mukherjee, Supantha \(June 14, 2023\). "EU lawmakers vote for tougher AI rules as draft moves to final stage". Reuters. Retrieved July 26, 2023.](#)
 ۱۰۸. [Bartz, Diane; Hu, Krystal \(July 21, 2023\). "OpenAI, Google, others pledge to watermark AI content for safety, White House says". Reuters.](#)
 ۱۰۹. [Ye, Josh \(July 13, 2023\). "China says generative AI rules to apply only to products for the public". Reuters. Retrieved July 13, 2023.](#)