فلسفهٔ علم، پژوهشگاه علوم انسانی و مطالعات فرهنگی
دوفصل‌نامهٔ علمی (مقالهٔ علمی ـ پژوهشی)، سال دهم، شمارهٔ دوم، پاییز و زمستان ۱۳۹۹، ۲۱۷-۲۳۳

یادگیری عمیق ماشینی؛ معضلات فلسفی و ره‌یافت‌ها

رضا نیرومند*

حمید فدیشه‌ای**، **الهام محمدزاده*****

چکیده

پیشرفت شگرف بشر در تولید و ذخیره‌سازی حجم انبوه داده‌ها و نیز استفاده از این مجموعهٔ عظیم داده در ساختن ماشین استنتاج‌گر امروزه در قالب فناوری پیشرفته‌ای به‌نام «یادگیری عمیق ماشینی» ظهور یافته است. این فناوری با الهام از اتصالات موجود در ساختار مغز جان‌داران طراحی شده است و از شبکه‌های عصبی مصنوعی عمیق قدرت می‌گیرد. توانایی آن در استنتاج خبره‌گونه در زمینه‌های مختلف و تولید آثار مشابه افراد زبردست، باوجود مزایای بسیار، انسان امروز را با معضلات مختلفی روبه‌رو می‌کند.

این نوشتار تلاش می‌کند به معرفی و تحلیل معضلات فلسفی و معرفتی یادگیری عمیق ماشینی و ارائهٔ ره‌یافت مناسب درقبال هریک از این چالش‌ها بپردازد. اگرچه یادگیری عمیق ماشینی مشکلات قابل‌توجهی بر سر راه انسان قرار داده است، با آگاهی و چاره‌اندیشی دربارهٔ آن‌ها می‌توان درعین بهره‌بردن از مزایای چشم‌گیر این فناوری از ارزش‌های انسانی نیز آگاهانه محافظت کرد.

کلیدواژه‌ها: یادگیری ماشین، یادگیری عمیق، شبکه‌های عصبی مصنوعی، فلسفهٔ فناوری، معضلات فلسفی.

---

* استادیار گروه معارف اسلامی، دانشکدهٔ علوم انسانی دانشگاه بجنورد، r.niroomand@ub.ac.ir

** استادیار گروه مهندسی کامپیوتر، دانشکدهٔ مهندسی دانشگاه بجنورد، fadishei@ub.ac.ir

*** استادیار گروه معارف اسلامی، دانشکدهٔ علوم انسانی، دانشگاه بجنورد (نویسندهٔ مسئول)، e.mohammadzadeh@ub.ac.ir

تاریخ دریافت: ۱۳۹۹/۰۳/۱۳، تاریخ پذیرش: ۱۳۹۹/۰۶/۱۲

Copyright © 2018, IHCS (Institute for Humanities and Cultural Studies). This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International, which permits others to download this work, share it with others and Adapt the material for any purpose

۱. مقدمه

در سالیان اخیر انسان پیشرفت‌های شگرفی در حوزهٔ فناوری کرده است که شاید چند سال قبل دور از انتظار می‌نمود. امروزه می‌توان یک دستیار انحصاری ماشینی در گوشی همراه نصب کرد و با او به‌زبان محاوره‌ای صحبت کرد و به او فرمان داد (Sarikaya 2017). در حین خرید از فروشگاهی اینترنتی تعدادی کالای مشابه که احتمالاً به آن‌ها نیاز است به‌شکل اختصاصی به خریدار پیش‌نهاد می‌شود (Jannach and Malte 2017) که اتفاقاً پیش‌نهادهایی کاملاً مفیدند. همین امر در هنگام گوش‌کردن موسیقی از یکی از تارنماهای چندرسانه‌ای هم ممکن است به‌وقوع پیوندد (Van den Oordet al. 2013). در جایی از دنیا پزشکی با ارسال تصاویر ماموگرافی به ابر محاسباتی (computing cloud) از مشاورهٔ ماشین در تشخیص احتمال ابتلای فردی به سرطان سینه بهره می‌گیرد (Sun 2016) و در جایی دیگر احتمال ابتلای مردم به بیماری پارکینسون در آینده تنها از الگوی رفتار تایپ‌کردن کلمات در حین استفاده از گوشی همراه پیش‌بینی می‌شود (Adams 2017)! ماشینی در گوشه‌ای دیگر در چند ثانیه تابلوی نقاشی‌ای خلق می‌کند که تحسین هر بیننده‌ای را برمی‌انگیزد (Huang and Serge 2017) و ماشین دیگری درحال ساخت قطعهٔ موسیقی است (Agarwala et al. 2017)! بدیهی است که ظهور هر فناوری نوینی موج جدیدی از معضل‌ها و فرصت‌ها به‌همراه می‌آورد. جاذبه و هیجان نهفته در این فرصت‌ها معمولاً مانع تعمق و چاره‌اندیشی، و درنتیجه غلبهٔ به‌موقع بر معضل‌ها می‌شود. به همین دلیل، لبهٔ ظهور هر فناوری جدید موقعیت مهمی به‌لحاظ تصمیم‌گیری در جهت‌دهی آن برای بهره‌برداری حداکثری از فرصت‌های جدید درعین به‌حداقل‌رسانی آفت‌های ناشی از مشکلات آن است.

مثال‌هایی که پیش‌تر ذکر شد همگی مثال‌هایی از کاربردهای فناوری یادگیری عمیق ماشینی یا به‌طور خلاصه «یادگیری عمیق» (deep learning) هستند. این فناوری یکی از جدیدترین دستاوردهای شگرف بشر است که نقش مهمی در نزدیک‌کردن توان‌مندی هوش مصنوعی (artificial intelligence) به هوش انسانی ایفا کرده است. هم‌اکنون، محققان بسیاری در سراسر دنیا در دو حوزهٔ مرتبط با این فناوری درحال تحقیق و مطالعه‌اند: یکی در حوزهٔ پیشرفت دانش یادگیری عمیق و دیگری در پهنهٔ به‌کارگیری آن در کاربردهای مختلف.

مسئلهٔ این نوشتار برخی معضلات فلسفی درمورد توان‌مندی‌های وسیع این فناوری است، آن هم با تمرکز بیش‌تر بر ساحت معرفتی انسان و احتمال خروجی‌های نادرست و به‌اشتباه‌انداختن آدمی از سوی آن. به‌خطرانداختن اختیار آدمی و عدم توجه به انتخاب‌گری

او در اولویت‌دهی موضوعات ماشین و عدم تسلط و کنترل عقلانی او در یادگیری ماشینی انسان را از موجودی متعقل و تسخیرکنندهٔ طبیعت به مقلدی محض و مسخّر ماشین تبدیل خواهد کرد.

نویسندگان این نوشتار باتوجه‌به نبود پژوهشی درخور برای نخستین‌بار گام‌های دقیق و بدیع پژوهش‌گرانه‌ای برای شناساندن دقیق معضلات فلسفی و انسانی پیش‌روی این فناوری برداشته‌اند و با اتخاذ رویکرد فلسفی ره‌یافت لازم برای هر مشکلی را پیش‌نهاد داده‌اند. ساختار اصلی مقاله از سه بخش تشکیل شده است: در نخستین بخش مفهوم یادگیری ماشین و یادگیری عمیق ارائه خواهد شد؛ در بخش دوم معضلات فلسفی پیش‌روی فناوری یادگیری عمیق ماشینی به بحث گذاشته شده است؛ و سومین بخش به ارائهٔ راه‌حل و ره‌یافت‌هایی برای غلبه بر این معضلات اختصاص دارد؛ درنهایت نیز به جمع‌بندی مباحث پرداخت خواهد شد.

۲. یادگیری ماشین

یادگیری ماشین (machine learning) به‌طور کلی توانایی ماشین در بهبود در انجام کارها با تکیه بر تجربه است؛ یعنی برخلاف برنامه‌نویسی ماشین که تمام توانایی ماشین براساس آن‌چه برنامه‌نویس تعیین کرده است رقم می‌خورد، در یادگیری ماشین انجام کار محوله به آن براساس تجربیات پیوستهٔ دریافتی از محیط و بدون دخالت برنامه‌نویس صورت می‌گیرد. یادگیری عمیق نوع خاصی از یادگیری ماشین است که ایدهٔ کلی آن از سال‌ها قبل در ذهن دانشمندان وجود داشته، ولی به‌دلیل ضعف فنی لازم برای اجرای آن در عمل در سالیان اخیر متحقق شده است (Goodfellow et al. 2016). در این نوع خاص از یادگیری باوجود منابع عظیم داده که امروزه با نام کلان‌داده‌ها (big data) شناخته می‌شوند، و با تکیه بر ماشین‌هایی با توان پردازش قدرت‌مند، و با استفاده از نمونهٔ تجربه‌های وافر امکان یادگیری بسیار وسیع وجود دارد. حاصل آن ماشین‌هایی است با توانایی استنتاج (reasoning) بسیار پیچیده‌ای نزدیک به انسان خبره. تعریف رسمی یادگیری ماشین بدین قرار است که وقتی کار T باتوجه‌به معیار کیفیت P باوجود مجموعهٔ تجربیات E بتواند به‌صورت بهتری از حالت نبود تجربیات E اجرا شود، یادگیری اتفاق افتاده است (Mitchell 1997). در بسیاری موارد ماهیت کار T در حوزهٔ یادگیری تشخیص است. برای مثال، می‌توان به تشخیص لبخند انسان توسط دوربین عکاسی یا تشخیص خرابی موتور از

صدای آن یا پیش‌بینی سرطان از تصاویر ماموگرافی اشاره کرد، گرچه درعمل T محدود به این نوع خاص نیست و می‌تواند هر نوع مسئله‌ای نظیر تولید اثر هنری گرافیکی یا برقراری گفت‌وگویی شبه‌انسانی بین دستگاه و کاربر آن، با هدفی مشخص باشد.

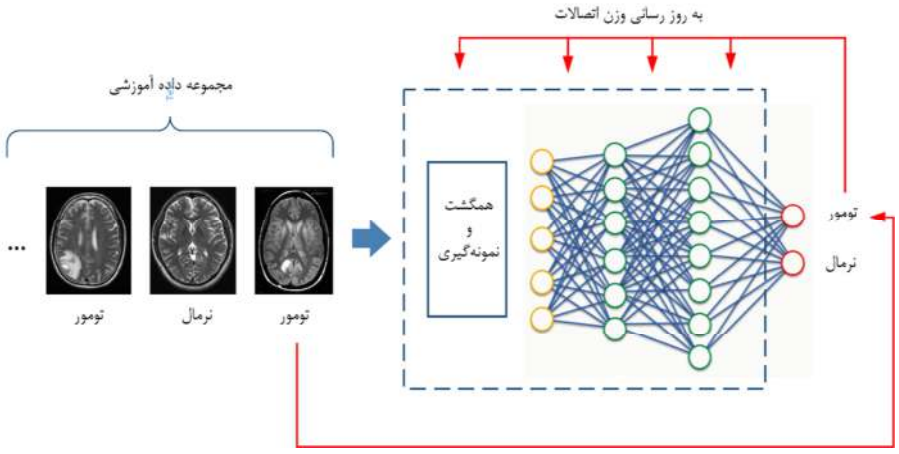
رایانه‌ها از ابتدا به‌منزلهٔ ماشین‌های محاسبه‌گر برای حل مسائل مختلف زندگی انسان (اقسام بی‌شمار عمل T) طراحی شده‌اند. آن‌چه سبب می‌شود یادگیری ماشین جهشی در رویکرد حل مسائل قلمداد شود تفاوت اساسی آن با روش دیگر حل مسئله توسط ماشین یعنی روش برنامه‌ریزی است. در حل مسائل به‌روش برنامه‌ریزی، شخص خبره در حوزهٔ مسئلهٔ T، با همکاری متخصص برنامه‌نویسی رایانه، برنامه‌ای برای حل آن طراحی می‌کند که با پیش‌بینی تمام ورودی‌ها و شرایط (تاحد ممکن) راه‌حل و گام‌های لازم برای رسیدن به پاسخ را طی می‌کند. چنین برنامه‌ای درحقیقت پیروی محض از دستورات نگاشته‌شدهٔ این دو خبره از سوی رایانه است و رایانه صرفاً مجری تک‌تک دستورات واردشده است. پس، بنابر چنین مفهومی یادگیری ماشین در بین نیست، و بهبود در انجام کار T فقط زمانی رخ می‌دهد که تیم برنامه‌نویسی با مشاهدهٔ تجربیات برنامه را به‌لحاظ کارکردهای تعریف‌شده ارتقا دهد.

یادگیری ماشین درمقابل این رویکرد مطرح می‌شود. در این پارادایم، برنامه‌نویس فقط بستری برای انجام کار و ارزیابی کیفیت انجام آن را (که به الگوریتم یادگیری شناخته می‌شود) بر روی ماشین پیاده می‌کند. سپس این ماشین است که با دریافت تجربه‌های جدید کیفیت انجام وظیفه‌اش را بدون نیاز به برنامه‌نویسی مجدد بهبود می‌بخشد. گرچه هوش مصنوعی در هردو رویکرد برنامه‌ریزی و یادگیری ماشین در کار است، پیشرفت‌های دهه‌های اخیر این حوزه یقیناً به‌میزان زیادی مدیون یادگیری ماشین است.

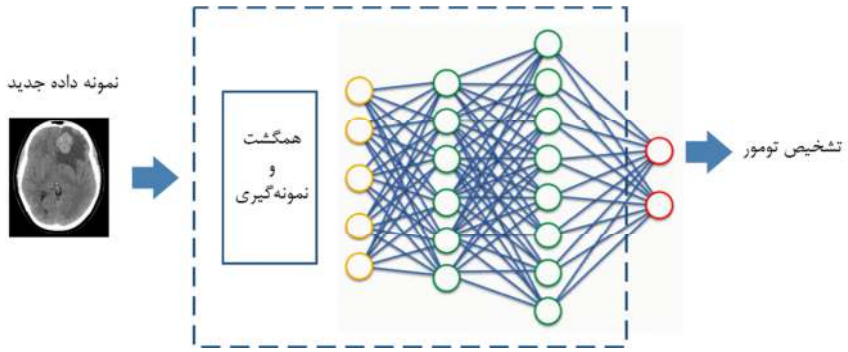
۳. یادگیری عمیق

یادگیری عمیق (Goodfellow et al. 2016) نوع خاصی از یادگیری ماشین است؛ بنابراین مطالبی که درمورد یادگیری ماشین گفته شد درباب آن نیز صادق است. وجه‌تمایز این نوع خاص یادگیری استفاده از تعداد زیادی نمونهٔ آموزشی با تکیه بر فناوری مرتبط دیگری به‌نام کلان‌داده است. بستر مورداستفاده در یادگیری عمیق شبکه های عصبی مصنوعی عمیق یا به‌طور خلاصه «شبکه‌های عصبی عمیق» (deep neural networks) است(.ibid) که با الهام از ساختار اتصالات موجود در مغز جان‌داران طراحی شده‌اند. کارکرد و ساختار داخلی

اتصالات بین اجزای این شبکه‌ها (نورون‌های مصنوعی) با استفاده از نمونه‌های آموزشی فراوان موجود و طی فرایند یادگیری استخراج می‌شود و پس‌ازآن این شبکه‌ها قادر به انجام استنتاج برای نمونه‌های جدید داده‌ها خواهند بود. شکل ۱ مثالی است از ساختار کلی و کاربرد شبکهٔ عصبی عمیق در حل مسئلهٔ پیش‌بینی تومور مغزی از روی تصاویر ام‌آرآی. فرایند یادگیری ماشین شامل دو مرحله است: در مرحلهٔ اول (شکل الف) با استفاده از تعدادی نمونهٔ آموزشی که برچسب تشخیص وجود یا عدم وجود تومور قبلاً توسط متخصص به آن‌ها زده شده است به شبکه به‌اصطلاح آموزش داده می‌شود؛ در مرحلهٔ استقرار (شکل ب) شبکه آماده می‌شود برای هر تصویر جدید تشخیص خود را ارائه دهد.



الف) آموزش شبکه با استفاده از مجموعهٔ تصاویر برچسب‌خورده



ب) استقرار شبکه برای انجام تشخیص بر روی تصاویر جدید

شکل ۱. نمونه‌ای از کاربرد شبکه‌های عصبی عمیق در تشخیص تومور مغزی

۴. معضلات یادگیری عمیق

در این قسمت به بررسی معضلات فلسفی پیش‌روی فناوری یادگیری عمیق می‌پردازیم و در بخش بعد کوششی عقلانی می‌کنیم برای دست‌یابی به راه‌حل و ره‌یافت‌های هریک از این مشکلات.

۵. احتمال استنتاج نادرست

یادگیری ماشین به‌طور عام و یادگیری عمیق به‌طور خاص متکی بر داده‌ها و اعداد و ارقام است. برای توسعهٔ سامانهٔ استنتاج مبتنی‌بر یادگیری عمیق، لازم است انبوهی از داده‌های ازقبل آماده تهیه شود. رویکرد کلی الگوریتم‌های یادگیری عمیق بر کشف هم‌بستگی‌های (correlation) نهفته در انبوه داده‌ها میان متغیرهای ورودی و متغیر هدف پیش‌بینی استوار است. این درحالی است که هم‌بستگی و هم‌راهی میان دو چیز لزوماً به‌معنای وجود رابطهٔ علیت (causation) میان آن دو نیست (Mitchell 1997). از لحاظ فلسفی، ضرورت علی رابطه ای قطعی میان علت و معلول است که باوجود علت انقطاعی در ظهور معلول قابل‌تصور نیست، درحالی‌که در هم‌بستگی این رابطه کاملاً متغیر است. برای مثال، ممکن است الگوریتم یادگیری ماشین با تحلیل داده‌های یک جامعهٔ آماری هم‌بستگی قابل‌توجهی بین میزان فقر و میزان وقوع جرم در جامعه گزارش کند، درحالی‌که این هم‌بستگی لزوماً به‌معنای این نیست که فقر عامل اصلی ارتکاب جرم است. ممکن است متغیر پنهان دیگری عامل هردو (فقر و جرم) باشد که از دید این استنتاج‌گر ماشینی پنهان مانده است، مانند بی‌کاری یا بی‌سوادی. درنتیجه، باوجود سادگی کشف هم‌بستگی و هم‌راهی توسط ماشین، استنتاج روابط علی و استدلال عقلانی از توان یادگیری عمیق ماشینی خارج است.

اتکا به دقت تحلیل‌های داده‌ای ماشینی و اعتقاد به این‌که این نوع استنتاج می‌تواند به‌کلی جای‌گزین انسان خبره شود، باوجود نادرست‌بودن، حامیان خودش را دارد. این اعتقاد اصطلاحاً با نام دانش عاری از نظریه (theory-free Science) و بینش شناخته می‌شود. جملهٔ معروف «اعداد خود به‌تنهایی سخن می‌گویند» از زبان اندرسون (Wright 1921) در مقاله‌ای در حمایت از این دیدگاه در زمان خود منتقدان زیادی را برآشفت. یکی از مثال‌های معروف شکست این دیدگاه در داستان واقعی «گوگل و دانش» اتفاق افتاد. زمانی مهندسان گوگل با تحلیل الگوی جست‌وجوهای اینترنتی مراجعه‌کنندگان به وب‌گاه این موتور جست‌وجو، بدون داشتن دانش پزشکی، توانستند الگوی گسترش آنفلونزا در شهرهای مختلف ایالات متحد را پیش‌بینی و در این زمینه به مردم اطلاع‌رسانی کنند. مهم‌تر این‌که

فرایند انجام این پیش‌بینی خیلی سریع‌تر از روش سنتی به‌نتیجه رسید که توسط تیمی از متخصصان خبره با بررسی‌های میدانی انجام می‌شد. این‌جا بود که این جمله بر سر زبان‌ها افتاد که «دانش از گوگل چه می تواند بیاموزد؟»؛ اما جالب این جاست که همین روش سریع مبتنی‌بر تحلیل صرف داده‌ها بعد از چند بار پیش‌بینی به نتایج غلط و شکست انجامید، درحالی‌که روش سنتی کماکان توسط متخصصان به نتایج صحیح می‌رسد.

۶. تفسیرناپذیری خروجی

همان‌گونه‌که گذشت، الگوهای یادگیری عمیق که متکی بر شبکه‌های عصبی عمیق‌اند براساس نمونه‌های یادگیری (تجارب ازپیش مشاهده‌شده) شکل می‌گیرد. بر این اساس، مدل با ورود اطلاعات نمونه‌های جدید به پیش‌بینی لازم می‌پردازد و نتیجه را اعلام می‌کند. مسئله این است که نتیجهٔ اعلام‌شده معمولاً قابل‌تحلیل و تفسیر (interpretability) به‌معنای رایج انسانی آن نیست (Anderson 2008) و این امر کاربرد مدل را با مشکل مواجه می‌سازد. برای روشن‌شدن موضوع بهتر است به مثال ذیل توجه شود. مدیران یک بانک قصد دارند برای بهبود نظام تسهیلات دست‌به‌دامان فناوری یادگیری عمیق شوند تا موفقیت یا عدم موفقیت وام‌گیرنده را در بازپرداخت تسهیلات اعطایی پیش‌بینی، و در فرایند تصمیم‌گیری دخیل کنند. فرایند مدل‌سازی یادگیری عمیق بدین نحو عمل می‌کند که سوابق نمونه‌های زیادی از وام‌گیرندگان قبلی اعم از خوش‌حسابان و بدحسابان بانک را که شامل اطلاعات مختلفی از این افراد نظیر شغل، جنسیت، درآمد، وضعیت تأهل، تعداد فرزندان، و غیره است به‌عنوان تجربه (یا نمونه‌های یادگیری) استفاده می‌کند. براساس این نمونه‌ها، شبکهٔ عصبی عمیقی مختص این مسئله شکل می گیرد که می‌تواند موفقیت یا عدم موفقیت متقاضیان جدید در بازپرداخت را پیش‌بینی کند. معضل مذکور در این مرحله نمایان می‌شود: این‌که این مدل درمقام توضیح راجع به علت عدم موافقت با درخواست متقاضی خاصی عاجز از تفسیر و پاسخ انسانی است؛ این درحالی است که اگر همین پیش‌بینی به‌روش سنتی و توسط تیم خبرهٔ انسانی انجام می‌شد، معمولاً امکان ارائهٔ توضیح وجود داشت.

۷. اولویت‌دهی نامناسب و نادرست در به‌کارگیری

فناوری یادگیری عمیق در حوزه های مختلفی مانند سلامت، امنیت، فرهنگ و هنر، اقتصاد، و سایر عرصه‌های مرتبط با زندگی انسان کاربرد دارد. برخی از این حوزه‌های کاربردی از

قبیل تشخیص پیش از موعد بیماری‌ها اهمیت فراوان دارند. ازسوی‌دیگر، برخی کاربردهای کم‌فایده نیز برای این دانش مدرن متصور است؛ ازجمله می‌توان به کاربردی تحت عنوان «جعل عمیق» (deep fake) اشاره کرد که در آن چهرهٔ بازیگر یک فیلم ویدئویی با استفاده از یادگیری عمیق جای‌گزین چهرهٔ شخص دیگری می‌شود (Zhang and Chun Zhu 2018). حال مسئله این است که باتوجه‌به وجود درجات مختلف از میزان فایده مندی دانش یادگیری عمیق در حوزه‌های مختلف، اولویت‌دهی سرمایه‌گذاری فکری و مالی در این حوزه‌ها باید به چه نحو انجام پذیرد. به‌نظر می‌رسد آن‌چه امروز در جهت‌دهی این اولویت اثرگذار است میزان سودآوری آن برای افراد یا شرکت‌های خاص در کسب‌وکارهای نوین بشری است؛ درحالی‌که چنان‌که در ره‌یافت‌ها بدان اشاره خواهد شد، براساس فلسفهٔ فناوری، به‌طور کلی، و یادگیری عمیق، به‌طورخاص، محور این اولویت‌گذاری را باید کرامت انسانی (human dignity) دانست.

۸. مداخله در ارادهٔ آزاد انسان (اختیار)

بسیاری از فیلسوفان ارادهٔ آزاد (free will) آدمی را وجه‌تمایز او از حیوان دانسته‌اند، تاآن‌جاکه انسان را حیوان مختار تعریف کرده‌اند. ازاین‌رو، اختیار به‌عنوان یکی از ذاتیات انسانی در اخلاق، حقوق، و سایر شئونات زندگی او اهمیت فوق‌العاده‌ای دارد. هرچیزی که با این اصل انسانی در تعارض قرار گیرد می‌تواند تهدیدی برای انسانیت انسان به‌حساب آید و ازاین‌رو، مردودبودن آن از نگاه جامعهٔ انسانی قطعی است.

فناوری یادگیری عمیق این توانایی را به برخی افراد و گروه‌ها می‌دهد تا مخاطب و جامعه را به سمت‌وسوی خاصی سوق دهند. برای نمونه تبلیغات نمایش‌یافته در حاشیهٔ تارنمایی خبری می‌تواند به‌گونه‌ای آرایش یابد که مخاطب را به رأی‌دادن به نامزد انتخاباتی خاصی سوق دهد؛ یا دنبالهٔ ویدئوهای پیش‌نهادی توسط ماشین در تارنمایی چندرسانه‌ای می‌تواند به‌گونه‌ای طرح‌ریزی شود تا به‌تدریج فرد را به گرایش جنسی خاصی متمایل کند. البته این مداخله در انتخاب‌گری ممکن است لزوماً در موارد منفی و غیراخلاقی رخ ندهد؛ مثلاً ممکن است هدف از دنبالهٔ ویدئوها در مثال قبل علاقه‌مندسازی شخص به محیط‌زیست باشد. اما به‌هرصورت مداخله در اختیار انسانی در این حالت نیز به‌وقوع پیوسته است. این امر از گذشته با عنوان «گرایاندن نامحسوس» (nudge) یا «پدرخواندگی نرم» (soft paternalism) شناخته شده است و از

مصادیق آن می‌توان به گذاشتن سالاد در کنار غذا برای ارتقای سلامتی اشاره کرد. البته در دوران یادگیری عمیق با نوع جدید و بسیار قدرت‌مندتری از این امر مواجهیم که برایش واژهٔ «ابَرگرایاندن» (hypernudge) پیش‌نهاد شده است (Harris 2018). این قدرت آن‌چنان عظیم است که اکنون این سؤال برای بشر متفکر شکل گرفته است که آیا دموکراسی از مهلکهٔ هوش مصنوعی و کلان‌داده‌ها جان سالم به‌در خواهد برد یا نه (Yeung 2017).

ازدیگرسو، آن‌چه اختیار آدمی را تحت‌الشعاع قرار می‌دهد فراگیری دستیارهای دیجیتال است. این دستیارها برنامه‌های کاربردی قابل‌اجرا بر روی گوشی هم‌راه افرادند که قادرند در امور روزمرهٔ زندگی انتخاب‌ها و پیش‌نهادهای بهینه‌ای را در تصمیم‌گیری‌هایی نظیر رژیم غذایی، مسیر حرکت، برنامهٔ تفریح، و نظایر آن به شخص ارائه دهند. غلبهٔ این دستیارها ازسویی و پیشرفت تصمیم‌های آن‌ها در سایهٔ یادگیری ماشین ازسوی‌دیگر عملاً افراد را به عادت به حضور و اتکا به وجود آن در زندگی سوق می دهد و منجر به کاهش تجربهٔ انتخاب در افراد می‌شود (Helbing 2019).

۹. اثرگذاری نامطلوب در توسعه و رشد فکری و اجتماعی

ورود ماشین‌های هوش‌مند در ابعاد مختلف زندگی متضمن نوعی کنترل روزافزون بر انسان است. این روند با بلوغ فناوری به‌سمتی پیش می‌رود که از ارتکاب حتی کوچک‌ترین اشتباهات نیز جلوگیری شود. برای نمونه، امروزه ممکن است راننده‌ای در کشوری پیشرفته به‌دلیل سرعتی معادل یک کیلومتر بر ساعت بیش‌تر از حد مجاز جریمه شود، یا شخصی توسط دستیار دیجیتال خود از خوردن وعدهٔ غذایی جدید در موقعیتی خاص به‌دلیل عدم تطابق با رژیم مناسبش منع شود. این تصمیمات پشتوانه کاملاً علمی و فنی دارند، نتیجهٔ علم بشری‌اند، و برای رسیدن به نتایج ایدئال طراحی شده‌اند. درعین‌حال، این میزان دقت علمی و فنی وقتی‌که با نوعی تلاش برای به‌صفررساندن خطاهای بشر و تصمیمات نادرست هم‌راه است عملاً باعث حرکت به‌سمت جامعهٔ هم‌گون و یک‌سان‌سازی جامعه و افراد خواهد شد. این نوع هم‌گونی، علاوه‌براین‌که مانعی بر سر راه توسعهٔ فکری افراد و خلاقیت آن‌هاست، عاملی برای ازبین‌رفتن خرد جمعی و توسعهٔ اجتماعی بشر خواهد بود، چراکه خرد جمعی حاصل دگراندیشی و برایند تفاوت در دیدگاه‌ها و افکار جامعه است (Helbing 2019).

۱۰. نقض حریم خصوصی

همان‌گونه‌که گفته شد، یادگیری معادل بهبود از طریق تجربه تعریف می‌شود. در یادگیری ماشین می‌توان تجربه را معادل داده و در یادگیری عمیق معادل کلان‌داده دانست؛ مثلاً برای پیش‌بینی وقوع جرم در محدوده‌ای جغرافیایی با استفاده از این فناوری نیاز به داشتن مجموعه‌ای از داده‌های مرتبط با تاریخچهٔ جرم و جنایت در آن منطقه است. چنین پایگاهی می‌تواند حاوی اطلاعات خصوصی شهروندان باشد و دست‌رسی همگانی به آن نقض حریم خصوصی (privacy) افراد را در پی داشته باشد و زمینهٔ سوءاستفاده را فراهم سازد.

آن‌چه این موضوع را پررنگ‌تر می‌سازد این حقیقت است که الگوریتم‌های یادگیری عمیق عمدتاً پیچیدگی محاسباتی فراوانی دارند و ازآن‌جاکه رایانه‌های معمولی قادر به پردازش آن‌ها نیستند اجرایشان به سرورهای قدرت‌مندی محول می‌شود که در فضای ابر محاسباتی توزیع شده‌اند. این امر سبب پراکندگی و نقل و انتقالات متعدد داده‌ها از مرحلهٔ جمع‌آوری تا تحلیل در فضای اینترنت خواهد شد که احتمال سوءاستفاده و دست‌یابی نفوذگران را به داده‌های مذکور بیش‌تر می‌سازد.

توجه به این نکته لازم است که مسئلهٔ حریم خصوصی معمولاً نه به‌صورت صفر و صد (مخفی‌کاری کامل اطلاعات یا آشکارسازی کامل اطلاعات)، بلکه چیزی بین این دو حالت تعریف می‌شود. تن‌دادن به نگاه صفر و صدی سبب می‌شود تا مجبور به پذیرش این موضوع شویم که باید از بین مزایای فناوری یادگیری عمیق و حفظ حریم خصوصی انسان‌ها یکی را انتخاب کنیم. حریم خصوصی درحقیقت وجود سطوح تعریف‌شده در دست‌رسی به جریان اطلاعات است (فدیشه‌ای و دیگران ۱۳۹۵).

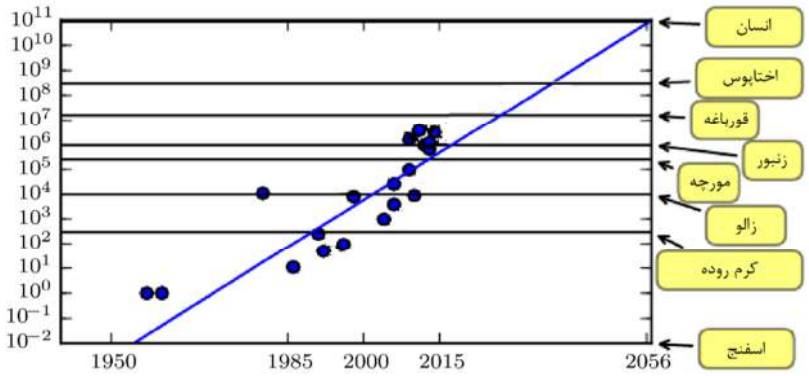
۱۱. ره‌یافت‌های پیش‌رو

اگرچه این نوشتار درشمار نخستین گام‌های جدی جامعهٔ علمی ما در مواجهه با معضلات فناوری پیش‌روندهٔ یادگیری عمیق ماشینی است، به‌نظر می‌رسد باید به‌منظور حل آن‌ها تلاش عقلی، فلسفی، و علمی صورت گیرد. ازاین‌رو، نویسندگان بر آن شدند تاحد امکان برای حل آن ره‌یافت‌های زیر را پیش‌نهاد کنند. باید توجه کرد که گرچه ممکن است نتوان ره‌یافتی اساسی و راه‌حلی نهایی برای این معضلات یافت، اما گام‌برداشتن در این مسیر نیز بی‌فایده نخواهد بود.

۱۲. احتمال استنتاج نادرست

همان‌گونه‌که گذشت، معضل موجود در یادگیری ماشینی عمیق در عدم توجه به تفاوت ماهوی هم‌بستگی و علیت است که نتیجهٔ آن احتمال خروجی نادرست را رقم می‌زند. درمقابل معضل مذکور ممکن است این پرسش به‌ذهن خطور کند که آیا باتوجه‌به روند پیشرفت فناوری یادگیری ماشین دست‌یابی به توانایی استنتاج علی توسط ماشین، آن‌چنان‌که از یک خبرهٔ انسانی توقع می‌رود، در آیندهٔ نزدیک دسترس‌پذیر نیست؟ به‌لحاظ فنی، رسیدن یادگیری ماشین به توان استنتاج علی نیازمند دو عامل است: وجود ماشین یادگیر توان‌مند در پردازش داده‌های فراوان؛ و وجود داده‌های دقیق، مرتبط، و کافی برای استفادهٔ ماشین.

تاریخ پیشرفت فناوری نشان‌دهندهٔ پیشرفت ماشین‌های یادگیر در طول زمان است. شکل ۲ این پیشرفت را به‌صورت روند افزایش پیچیدگی شبکه‌های عصبی مصنوعی ساخته‌شده توسط انسان (که زیرساخت اصلی در یادگیری ماشینی عمیق اند) در طول زمان نشان می‌دهد (Goodfellow 2016). درضمن، این تصویر مقایسه‌ای از این تعداد بین شبکه‌های عصبی مصنوعی و مغز برخی جانوران مختلف را به‌تصویر کشیده است. تصویر نشان می‌دهد که درصورت ادامهٔ روند فعلی، تخمین زمان دست‌یابی به شبکهٔ عصبی مصنوعی با تعداد نورون برابر با مغز انسان سال ۲۰۵۶ خواهد بود. البته این بدان معنا نیست که قدرت استنتاج چنین شبکه‌ای لزوماً برابر با مغز انسان خواهد بود، اما به‌هرحال در پاسخ به پرسش مذکور عامل پیشرفت ماشین قابل‌دست‌یابی است و دور از انتظار نیست؛ اما عامل دوم، یعنی وجود دادهٔ دقیق و کافی که مرتبط با مسئله باشد، نیازمند حضور انسان خبره در فرایند یادگیری و نظارت بر آن است. این حضور هم‌اکنون درقالب پژوهش‌هایی در زمینهٔ یادگیری ماشینی و در حوزه‌هایی نظیر تفسیرپذیرسازی مدل‌های یادگیری و ایجاد مدل‌های دوگانه با امکان داخل‌کردن صریح تجربه‌های خبره در مدل توسط محققان علوم و مهندسی رایانه در جریان است. بنابراین، بعید نیست که باتوجه‌به فراز و فرودی که در پیشرفت فناوری به‌دست بشر شاهد بوده‌ایم، روزی فرارسد که تمام آن‌چه نیازمند انتقال قدرت استنتاج خبرهٔ انسانی به ماشین است فراهم شود. درعین‌حال، اگرچه این انتقال قدرت استنباط به ماشین از لحاظ فلسفی ممکن است، اما نباید این نکتهٔ اساسی را از نظر دور داشت که این جریان انتقال امری پیوسته است؛ بدین معنا که همیشه درکنار پیشرفت ماشین پیشرفت و پرورش خبرهٔ انسانی ضروری است.



شکل ۲. پیشرفت شبکه‌های عصبی مصنوعی ساختهٔ بشر در طول زمان به‌لحاظ تعداد نورون تعبیه‌شده در آن‌ها. هر نقطهٔ آبی‌رنگ مربوط به یکی از شبکه‌های عصبی مصنوعی ساختهٔ محققان است (Goodfellow 2016).

۱۳. تفسیرناپذیری خروجی

اگرچه اغلب مدل‌های یادگیری عمیق درحال‌حاضر فاقد توان ارائهٔ تفسیر بر استنتاج‌های خویش‌اند، نیاز به رفع این نقص نمایان شده است و محققانی درحال تحقیق با هدف تفسیرپذیرکردن مدل‌های یادگیری عمیق‌اند. برای نمونه، در مدلی با هدف تشخیص بیماری سرطان از تصاویر رادیوگرافی ممکن است مدل یادگیری عمیق به‌گونه‌ای طراحی شود که قادر به تجزیهٔ عوامل اثرگذار تصویر (مثلاً محل توده‌های مشکوک به سرطان) و متمایزکردن آن‌ها و اثرهای آن‌ها بر استنتاج باشد.

به‌علاوه، نکتهٔ حائز اهمیت این است که تلازم پررنگی میان تفسیرپذیربودن خروجی ماشین و قدرت استنتاج و درنتیجه حضور خبره در فرایند یادگیری و استنتاج وجود دارد. تفسیرپذیربودن به‌معنای انسانی مدنظر ما زمانی محقق می شود که مفسر از قدرت استنتاج برخوردار باشد و ازآن‌جاکه در بحث قبلی قدرت استنتاج با اشکال مواجه شده و نیاز به حضور خبره نمایان شده بود، بنابراین با ره‌یافت ذکرشده در مسئلهٔ قدرت استنتاج، می‌توان مشکل تفسیرناپذیری خروجی را نیز تاحدی حل کرد.

۱۴. اولویت‌دهی نامناسب و نادرست در به‌کارگیری

همان‌گونه‌که پیش‌تر ذکر شد، یادگیری عمیق در زمینه‌های مختلفی کاربرد دارد که این امر معضل اولویت‌بندی در هریک از زمینه‌های ممکن را رقم می‌زند. به‌طور کلی، اولویت‌بندی

مؤثر باید در سطح سیاست‌گذاری کلان جوامع انجام شود. اما این پرسش مطرح می‌شود که معیار مناسب این اولویت‌بندی چیست؟

به‌باور نگارندگان، کرامت انسانی می‌تواند محور اصلی این اولویت‌بندی قرار گیرد. به‌عبارت‌دیگر، در توسعهٔ زمینه‌های مختلف این فناوری، هر مورد که هم‌سویی بیش‌تری با کرامت انسانی دارد باید در اولویت بالاتری از سایر زمینه‌ها برای تحقیق و رشد قرار گیرد. بر این اساس، در توسعهٔ فناوری یادگیری عمیق، به‌صورت موردی، امکان تدوین سنجه‌های ثانوی از این محور اصلی وجود خواهد داشت. مثلاً، در بودجه‌ریزی برای حمایت از تحقیقات دانشگاهی می‌توان ابتدا به‌دنبال این مسئله رفت که دستاورد تحقیق مدنظر در نجات جان چند انسان یا رفع فقر چند نفر اثرگذار خواهد بود؛ یا در انتخاب بین دو تحقیق که یکی به تشخیص زودهنگام سرطان و دیگری به جعل عمیق چهرهٔ افراد در ویدئوها می‌پردازد مورد اول در اولویت اقدام قرار گیرد.

۱۵. مداخله در ارادهٔ آزاد انسان (اختیار)

همان‌گونه‌که گذشت، اختیار و ارادهٔ آزاد انسانی وجه‌ممیز انسان است که در ذات او به‌ودیعه گذاشته شده است. فرایند عملی اختیار انسان چنان نیست که به‌راحتی دست‌خوش نابودی شود یا حتی موردخدشه و مداخله قرار گیرد. تأثیرگذاری و سمت‌وسودادن عادت و سلیقه و حتی اندیشهٔ افراد در تصمیم‌گیری و اختیار آدمی طبیعی است و طبیعتاً انسان باتوجه‌به عوامل متعدد و مختلف تصمیم‌گیری می‌کند. باید توجه داشت که خارج از فناوری یادگیری عمیق، انسان متأثر از ژنتیک، نیازهای فیزیولوژیک، عقاید بنیادین و دینی، جامعه و محیط تربیتی، و بسیاری عوامل دیگر دست به انتخاب می‌زند. از میان عوامل بی‌شماری که در اختیار آدمی اثرگذارند آخرین این عوامل خود فرد است.

برای مواجهه با این معضل دو راه‌کار اساسی پیش‌نهاد می‌شود: راه‌کار نخست تغییر رویه از «گرایاندن نامحسوس» به «گرایاندن محسوس» است (Helbing et al. 2015). نمود این پیش‌نهاد در مثالی که قبلاً ذکر شد این است که مثلاً ذیل یک تصویر ویدئویی ذکر شود که آگاه باشید دیدن این ویدئو ممکن است شما را به فردی علاقه‌مند به محیط‌زیست تبدیل کند؛ راه‌کار دوم توسل به مواجه‌کردن کاربران با تضادها و استفاده از اثر شگرف این امر در رشد انسان است. اگر ماشین‌ها جهات مختلفی به‌خصوص دو سوی مقابل را پیش‌روی آدمی بگذارند و پیش‌نهاد کنند، درواقع این خود مصداقی برای ظهور کامل اختیار آدمی در

این فضاست. برای مثال، زمانی‌که شخصی در سامانهٔ محتوای ویدئویی با دنباله‌ای از ویدئوها با مضمون‌های محرک جنسی مواجه می‌شود در این میان به‌ناگهان ویدئویی دربارهٔ اهمیت کنترل این نیرو و جهت‌دهی آن در مسیر صحیح گنجاند شود. این تضادها به‌مثابهٔ خروجی‌های تعبیه‌شده در بزرگ‌راه عمل خواهند کرد تا افراد از جبر راه‌سپردن در مسیری خاص رهایی یابند.

۱۶. اثرگذاری نامطلوب در توسعه و رشد فکری و اجتماعی

پیش‌فرض این معضل مبتنی‌بر وجود جبر آدمی در دنیای فناوری یادگیری عمیق است. این پیش‌فرض براساس ره‌یافتی که درباب ارادهٔ آزاد آدمی داده شد مبنای صحیح و دقیق ندارد و بیش از آن‌که علمی و منطقی باشد براساس نوعی بدبینی به فناوری‌های جدید طرح شده است. درعین‌حال، همان‌گونه‌که گذشت، جامعه‌ای که در چشم‌اندازش ابَرهوش مصنوعی و یادگیری عمیق را قرار داده است جامعه‌ای است که افراد در آن (اغلب به‌شیوه‌های نامحسوس) تحت نظارت‌اند و از ارتکاب کوچک‌ترین خطاها نیز جلوگیری می‌شود. در چنین جامعهٔ یک‌دستی، که هم‌اکنون نیز علائمی از آن درحال ظهور و مشاهده است، همهٔ شهروندان کمابیش شبیه یک‌دیگرند و خرد جمعی که نتیجهٔ تفاوت آراست رنگ می‌بازد و توسعهٔ فکری و اجتماعی بشر کند می‌شود.

برای مقابله با این مشکل، اولین قدم به‌رسمیت‌شناختن و تذکردادن وجود ارادهٔ آزاد به شهروندان جامعه است. کاستن از میزان اتکای افراد به ماشین و استنتاج‌های آن نیز چارهٔ مهم دیگری در این مسیر است. همان‌طورکه برای حفاظت از طبیعت روز بدون خودرو پیش‌نهاد می‌شود، می‌توان در دنیای مدرن هوش مصنوعی روزی یا ساعتی بدون هوش ماشینی حداقل برای جزئی از ابعاد نفوذ این فناوری در زندگی افراد پیش‌نهاد کرد تا انسان بتواند لحظاتی خاص و در مواردی خاص به خود و قدرت تصمیم خود بازگردد. نباید فراموش کرد که در روزهای نه‌چندان گذشته شهروندان شمارهٔ تماس نزدیکان خود را در ذهن داشتند، مسیریابی حرکت خود در خیابان‌های شهر را در ذهن خود رقم می‌زدند، و در انتخاب قطعهٔ موسیقی موردعلاقهٔ خود متکی به نظر ماشین نبودند. پس این امر ناممکن نیست. البته ممکن است این امر سبب ارتکاب اشتباهات یا تجربیات ناخوشایندی شود، اما باید توجه داشت که بسیاری از دستاوردهای مهم در تاریخ پیشرفت بشر از ابتدا حاصل یک اشتباه بوده است.

۱۷. نقض حریم خصوصی

چنان‌که در مقدمهٔ این نوشتار ذکر شد، فرایند یادگیری عمیق نیاز به نمونه‌های زیادی از داده‌ها برای یادگیری دارد. به همین دلیل، این فناوری ارتباط تنگاتنگی با فناوری کلان‌داده دارد. گرچه جمع‌آوری داده از فعالیت‌های روزمرهٔ انسان‌ها از حضور آن‌ها در شبکه‌های اجتماعی گرفته تا رفت‌وآمد و سوابق پزشکی آن‌ها می‌تواند منجر به دستاوردهای مفیدی با کمک یادگیری عمیق شود، مخاطرهٔ نقض حریم خصوصی آن ها را نیز به‌دنبال دارد. فهرستی مختصر از رهیافت‌هایی در مواجهه با نقض حریم خصوصی به‌شرح زیر پیش‌نهاد می‌شود (فدیشه‌ای و دیگران ۱۳۹۵):

- شفافیت: آگاهی از جزئیات اطلاعات ذخیره‌شدهٔ کاربران و اختیار آن‌چه ذخیره می‌شود و آن‌چه مجاز به ذخیره‌شدن نیست باید جزء حقوق کاربران تلقی شود؛
- سطوح دست‌رسی: دست‌رسی کاربران به اطلاعات باید در سطوح مختلفی باشد، به‌نحوی‌که در هر سطح به‌میزان کافی و نه بیش‌تر از آن دست‌رسی وجود داشته باشد؛
- توسعهٔ قوانین: باید قوانین پویا و کارآمدی برای حمایت از حریم خصوصی افراد تدوین شود؛
- توسعهٔ فنی: محققان باید تلاش کنند محیط‌های کاربری امنی مهیا کنند. باید برخی روش‌های پردازش داده توسعه یابد که با داده‌های بی‌نام‌شده، مبهم‌شده، تکه‌تکه، و توزیع‌شده به‌جای داده‌های اصلی کار می‌کنند.

۱۸. نتیجه‌گیری

همان‌طورکه به‌تفصیل گذشت، فناوری یادگیری عمیق ماشینی مانند هر پدیدهٔ نوظهوری که محصول پیشرفت علم و فناوری است به‌طبع با مسائل و معضلات فلسفی، اخلاقی، و اجتماعی روبه‌رو است. مشکلات مذکور که عمدتاً رنگ‌وبوی فلسفی، منطقی، و معرفتی دارند ساحت شناختی و درنتیجهٔ آن ساحت رفتاری انسان را از جهت درست و نادرست هدف قرار داده‌اند. علم و معرفت و به‌طریق اولی فناوری از نگاه انسان‌شناختی زمانی ارزش‌مند محسوب می‌شود که آدمی را از جهل به علم و از بیهودگی به فایده‌مندی برساند. ازاین‌رو، پژوهش حاضر تلاش کرد به دو مسئلهٔ اساسی پاسخ دهد:

معضلات فلسفی این فناوری پیشرفته چیست؟ چه راه‌کارهایی در رویارویی با این معضلات وجود دارد؟ در این مقاله مشکلات موجود از دیدگاهی فلسفی تبیین شدند و درادامه راه‌کارهایی برای آماده‌شدن در مقابله با آن‌ها ارائه شد. به‌عنوان نتیجهٔ حاصل از این پژوهش باید گفت باتوجه‌به راه‌حل‌هایی که گذشت می‌توان معضلات نهفته در این فناوری را همواره با کوشش عقلانی و علمی حل کرد و درنتیجه باعث شد از سوق‌دادن آدمی به خطا جلوگیری شود تا با نتایج و پی‌آمدهای مثبت آن باعث بهبود زندگی بشری شود.

باتوجه‌به این‌که مسیر مشکلات و ره‌یافت‌ها مستمر و تمام‌نشدنی است، مقابله با این معضلات نیازمند کوشش هر دو طیف پژوهش‌گران در حوزهٔ نظریه‌پردازی فلسفی و علوم فنی و کاربردی رایانه برای تبیین این مشکلات و کوشش برای یافتن راه‌حل‌های نظری و عملی و ایجاد بستر مناسب برای مقابله با آن‌هاست. بنابراین، به‌جای دو راه‌برد آسان و غیرمعقول بدبینی، محدودسازی، و تعطیلی فناوری ازسویی و انکار مشکلات انسانی و معرفتی فناوری‌ها ازدیگرسو، باید مدام به تعقل، تفلسف، و کوشش علمی و فنی به‌منظور شناسایی و حل مسائل و مشکلات مذکور پرداخت.

کتاب‌نامه

فدیشه‌ای، حمید، رضا نیرومند، و الهام محمدزاده (۱۳۹۵)، «چالش‌های اخلاقی بزرگ‌داده‌ها»، *اخلاق در علوم و فناوری*، س ۱۱، ش ۴، زمستان.

Adams, Warwick R. (2017), "High-Accuracy Detection of Early Parkinson's Disease Using Multiple Characteristics Of Finger Movement While Typing", *Plos One*, vol. 12, no. 11.

Agarwala, Nipun, Yuki Inoue, and Axel Sly (2017), "Music Composition Using Recurrent Neural Networks",
<https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1174/reports.html >

Anderson C. (2008), "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete", *Wired* Magazine; vol. 16, no. 7.

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016), *Deep learning*, Verlag: MIT Press.

Harris, Douglas (2018), "Deepfakes: False Pornography Is Here and the Law Cannot Protect You", *Duke L. & Tech*, vol. 17, no. 1.

Helbing, D. et al (2019), "Will Democracy Survive Big Data and Artificial Intelligence?", in: *Towards Digital Enlightenment*, Helbing (ed.), Zürich: Springer.

Helbing, D (2019), "Societal, Economic, Ethical and Legal Challenges of the Digital Revolution: From Big Data to Deep Learning, Artificial Intelligence, and Manipulative Technologies", *Towards Digital Enlightenment*, Helbing (ed.), Zürich: Springer.

Huang, Xun, and Serge Belongie (2017), "Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization", Proceedings of the *IEEE* International Conference on Computer Vision, <https://ieeexplore.ieee.org/document/8237429>.

Jannach, Dietmar and Malte Ludewig (2017), "When Recurrent Neural Networks Meet the Neighborhood for Session-Based Recommendation", Proceedings of the Eleventh *ACM* Conference on Recommender Systems.

Mitchell, Thomas M. (1997), *Machine learning* (McGraw Hill Series in Computer Science), New York: McGraw-Hill Education Ltd.

Sarikaya, Ruhi (2017), "The Technology behind Personal Digital Assistants: An Overview of the System Architecture and Key Components", *IEEE Signal Processing*, vol. 34, no.1.

Sun, Wenqing (2016) "A Preliminary Study On Breast Cancer Risk Analysis Using Deep Neural Network", International Workshop on Breast Imaging, Springer, Cham.

Van den Oord, Aaron, Sander Dieleman, and Benjamin Schrauwen (2013), "Deep Content-Based Music Recommendation", in: *Advances in Neural Information Processing Systems*, Burges and Bottou and Welling and Ghahramani and Weinberger (eds.), <https://papers.nips.cc/paper/2013 https://papers.nips.cc/paper/2013>.

Wright, Sewall (1921, "Correlation and Causation", Journal of *Agricultural Research*, vol. 20, no.7.

Yeung, K (2017), "Hypernudge: Big Data as a Mode of Regulation by Design", *Information, Communication, and Society*, vol. 20, no. 1.

Zhang, Quan-shi and Song-Chun Zhu (2018), "Visual Interpretability for Deep Learning: A Survey", *Frontiers of Information Technology and Electronic Engineering*, vol. 19, no. 1.