کلان‌داده

چگونه انقلاب اطلاعاتی زندگی ما را دگرگون می‌کند

برایِن کِلِگ

ترجمهٔ
پری آزرموند (مختاری)

فرهنگ‌نشرنو
با همکاری نشر آسیم
تهران- ۱۳۹۹

دربارهٔ نویسنده

تازه‌ترین کتاب‌های براین کِلِگ عبارت‌اند از قالب واقعیت (*The Reality Frame*) (آیکان، ۲۰۱۷)، خورشید چه رنگ است (*What Colour is the Sun*) (آیکان، ۲۰۱۶)، و ده میلیارد فردا (*Ten Billion Tomorrows*) (سنت مارتینز پرس، ۲۰۱۶). دو کتاب او، دنیای مهره (*Dice World*) و تاریخچهٔ بیکرانگی (*A Brief History of Infinity*) برای «جایزهٔ انجمن سلطنتی برای کتاب‌های علمی» نامزد دریافت جایزه بودند. کِلِگ برای نشریات بی‌شماری از جمله وال استریت جورنال، نیچر (*Nature*)، بی‌بی‌سی فوکاس (*BBC Focus*)، دنیای فیزیک، تایمز، آبزرور، *Good Housekeeping* و پلی‌بوی مطلب نوشته است. کِلِگ ویراستار popularscience.co.uk و بلاگ‌های brianclegg.blogspot.com و www.brianclegg.net است.

فهرست مطالب

۱

ما می‌دانیم شما چه می‌اندیشید

کار کارستانِ کلان‌داده

مشکل بتوان از «کلان‌داده» چشم پوشید. واژه‌ها تمام مدت از گزارش‌های خبری و فیلم‌های مستند بر سر و روی ما می‌بارند. اگر چه دهه‌هاست که ما در عصر اطلاعات زندگی می‌کنیم. پس چه چیزی تغییر کرده است؟

به یک نمونهٔ موفق عصر کلان‌داده نگاهی بیفکنید: نتفلیکس[1]. شرکتی که زمانی کارش کرایه دادن دی‌وی‌دی بود از برکت کلان‌داده تغییر ماهیت داده است ــ و این تغییر خیلی بیشتر از صرف انتقال از دی‌وی‌دی به اینترنت است. ارائهٔ خدمات ویدیوی درخواستی ناگزیر در گرو به کار گرفتن مقادیر زیادی داده است. هر چند کرایه دادن دی‌وی‌دی هم به همین منوال بود. همهٔ کاری که دی‌وی‌دی می‌کند ذخیرهٔ چندین گیگابایت داده روی یک لوح فشردهٔ بصری است. در هر دو مورد با داده‌پردازی در یک مقیاس بزرگ سر و کار داریم. اما کلان‌داده معنایی به‌مراتب بیشتر از این دارد. کلان‌داده مربوط می‌شود به کاربرد تمامی طیف داده‌هایی که برای دگرگون کردن یک سرویس یا سازمان در دسترس است.

1. Netflix

نتفلیکس نشان می‌دهد که چگونه شرکتی که کارش ویدیو هنگام درخواست[۱] است می‌تواند کلان‌داده را محور کار خود قرار دهد. خدماتی مانند خدمات نتفلیکس بیشتر از یک ایستگاه سخن پراکنی متعارف متضمن ارتباط دوسویه‌ای است. چنین شرکتی می‌داند چه کسی چه چیزی را در چه زمانی و در کجا تماشا می‌کند. سیستم‌های آن می‌توانند فهرست راهنمایی از معیارهای علائق بیننده، همراه با بازخورد آنان، تهیه کنند. ما در مقام بیننده پیامد این تحلیل را در توصیه‌های نتفلیکس می‌بینیم، و این توصیه‌ها گاه عجیب به نظر می‌رسند زیرا سیستم سعی دارد پسند و ناپسند هر آدمی را پیش‌بینی کند. اما از دیدگاه نتفلیکس جفت‌وجور کردن سلیقه‌های شمار بسیاری از جمعیت نفع خیلی زیادتر و کاراتری دارد: این کار ممکن است روند سفارش سریال‌های جدید را دگرگون کند.

مثلاً، نخستین اقدام نتفلیکس را که موفقیت سریال عظیم خانهٔ پوشالی[۲] باشد در نظر بگیرید. اگر این پروژه به شبکهٔ تلویزیونی متعارفی واگذار شده بود، نخست یک برنامهٔ آزمایشی تهیه و آن را روی مخاطبان مختلف آزمایش می‌کرد، شاید خطر می‌کرد و بودجهٔ یک فصل کوتاه را هم تأمین می‌کرد (که ممکن بود در اواسط کار لغو شود) و آن‌گاه فقط پس از همهٔ این کارها تهیهٔ سریال را تمام‌وکمال تعهد می‌کرد. نتفلیکس به برکت کلان‌داده این روند را دور زد.

تهیه‌کنندگان این سریال، مردخای ویژیک و آصف سچو، در سال ۲۰۱۱ با شبکه‌های امریکایی تماس گرفتند و کوشیدند بودجهٔ تهیهٔ یک برنامهٔ آزمایشی را از آنها بگیرند. چون از زمان اتمام سریال بال غربی[۳]،

۱. on-demand video؛ سیستم‌هایی که به کاربران این امکان را می‌دهند که محتوای صوتی یا تصویری را هر زمان که تمایل دارند گوش یا تماشا کنند.
(همهٔ پانوشت‌ها از ویراستار است، مگر آنهایی که با «ــ م.» تصریح شده است).

۲. *House of Cards*؛ سریالی داستانی با محتوای سیاسی، محصول سال ۲۰۱۳، در ۷۳ قسمت، به کارگردانی دیوید فینچر، برگرفته از سریالی بریتانیایی به همین نام، ساختهٔ بی‌بی‌سی و نیز رمانی به همین نام نوشتهٔ مایکل دابز.

۳. *The West Wing*؛ مجموعهٔ تلویزیونی امریکایی، ساختهٔ آرون سورکین، در ژانر سیاسی، که از سپتامبر ۱۹۹۹ پخش شد. پخش‌کنندهٔ اصلی آن شبکهٔ ان‌بی‌سی بود و در مجموع چند جایزهٔ گولدن گلوب و امی را ربود.

در سال ۲۰۰۶، هیچ سریال سیاسی موفقی تهیه نشده بود و کسانی که اختیار پول را در دست داشتند احساس می‌کردند ساختن خانهٔ پوشالی متضمن خطر خیلی بالایی است. ولی نتفلیکس از طریق داده‌های انبوه مربوط به مشتریان خود می‌دانست که چنین سریالی با اقبال مشتریان فراوانی روبه‌رو خواهد شد؛ مشتریانی که شوخ‌طبعی و ایهام و سیاهی سریال اصلی بی‌بی‌سی را که سریال تازه بر اساس آن قرار داشت و در کتابخانهٔ نتفلیکس هم موجود بود می‌فهمند و ارج می‌نهند. در ضمن نتفلیکس مشتریان زیادی داشت که کار کارگردان، دیوید فینچر، و بازیگر، کوین اسپیسی، را که در ساخت این سریال نقشی محوری بازی کرد، دوست داشتند.

نتفلیکس، با در دست داشتن شواهد قوی دال بر این‌که مخاطبان مشتاقی در انتظارند، به جای تولید یک برنامهٔ آزمایشی پیشاپیش ۱۰۰ میلیون دلار برای تهیهٔ دو مجموعهٔ اولیه ــ که جمعاً ۲۶ قسمت می‌شد ــ اختصاص داد. این بدان معنا بود که سازندگان خانهٔ پوشالی می‌توانستند با اطمینان خاطر کار خود را گسترش دهند و ژرفای خیلی بیشتری به سریال ببخشند که در غیر این صورت امکان‌پذیر نبود. و حاصل کار بسیار موفقیت‌آمیز از کار درآمد. البته هر نمایش نتفلیکس به اندازهٔ خانهٔ پوشالی موفق نیست. ولی خیلی از آنها پرصرفه بوده‌اند، و حتی وقتی که استقبال عامه کندتر شد، مانند نمایش ۲۰۱۶ تاج[1] نتفلیکس، با توجه به دو فصل اولیه پرهزینهٔ مشابه، در مقایسه با وقتی که نمایش به صورت متعارف پخش می‌شود مدت خیلی بیشتری طول می‌کشد تا نمایش‌ها موفق از آب درآیند. این الگو تاکنون منتج به چندین موفقیت بزرگ شده است چون تصمیم‌ها بر پایهٔ کلان‌داده‌ها بوده است نه غریزهٔ مدیران اجرایی صنعت که این بدنامی را با خود یدک می‌کشند که خیلی بیش از آن‌که درست عمل کرده باشند غلط عمل کرده‌اند.

۱. *The Crown*؛ مجموعهٔ تلویزیونی، ساخته و نوشتهٔ پیتر مورگان. این مجموعه را دو شرکت لفت بانک پیکچرز و سونی پیکچرز تلویژن برای نتفلیکس ساختند و اول بار در نوامبر ۲۰۱۶ منتشر شد.

قابلیت درک مخاطبان بالقوهٔ یک سریال تازه تنها راهی نبود که کلان‌داده کمک کرد تا خانهٔ پوشالی به موفقیت برسد. بهره‌گیری هوشمندانه از داده به این معنی است که مثلاً برنامه‌های آیندهٔ سریال‌های گوناگون بتوانند در دسترس مخاطبان مختلف نتفلیکس قرار گیرند. و نکتهٔ تعیین‌کننده این‌که نتفلیکس بر خلاف شبکه‌های متعارف که هر سریالی را قسمت به قسمت، آن هم هفتگی پخش می‌کنند، عمل می‌کند و تمام قسمت‌های یک فصل را یک‌جا در دسترس مخاطبان قرار می‌داد. به بیان دیگر، نتفلیکس بی‌آن‌که برای جلب بیننده تبلیغات کند، توانست اختیار دیدن سریال را به بینندگان واگذارد. این روش اکنون متداول‌ترین راهکار پخش سریال‌های سیال شده است ــ الگویی که فقط با به کارگیری کلان‌داده امکان‌پذیر است.

ولی البته کلان‌داده فقط تماماً در خدمت کسب‌وکار نیست. از جمله توانایی و امکان آن را دارد که با پیش‌بینی نقاط احتمالی وقوع جرم حفظ نظم را دگرگون کند؛ قابلیتش را دارد که عکس ساکن را به حرکت درآورد؛ نخستین ابزار را برای مردم‌سالاری راستین ارائه دهد؛ کتاب پرفروش بعدی نیویورک تایمز را پیش‌بینی کند؛ به ما درکی از ساختار بنیادین طبیعت بدهد؛ و در پزشکی تحولات اساسی ایجاد کند.

اما آنچه جذابیت کمتری دارد این است که به شرکت‌ها و دولت‌ها این امکان بالقوه را می‌هد که چیزهای خیلی بیشتری دربارهٔ شما بدانند اعم از این‌که بخواهند چیزی به شما بفروشند یا در صدد نظارتتان برآیند. تردید نکنید ــ کلان‌داده ماندگار است، پس لازم است که هم منافع و هم خطرهای آن را بدانیم و درک کنیم.

کلید

درست همان‌گونه که در مورد تحلیل نتفلیکس از مخاطبان بالقوهٔ خانهٔ پوشالی دیدیم قدرت کلان‌داده ناشی از این است که مقادیر زیادی اطلاعات را گردآوری و به شیوه‌هایی تحلیل می‌کند که انسان هرگز

نمی‌تواند بدون کامپیوتر در انجام این کار که ظاهراً غیرممکن است توفیق یابد.

داده مدت درازی است که با ماست. به ۶۰۰۰ سال پیش یعنی مراحل آغازین جوامع کشاورزی باز خواهیم گشت تا شروع مفهوم داده را ببینیم. داده در درازای زمان، از طریق حساب کردن و واژهٔ مکتوب، ستون فقرات تمدن شد. خواهیم دید که چگونه داده در سده‌های هفدهم و هجدهم تحول یافت تا ابزاری برای تلاش در جهت گشودن روزنه‌ای به سوی آینده شود. اما این تلاش همیشه به دلیل گسترهٔ تنگ داده‌های موجود و قابلیت‌های ضعیف ما برای تحلیل آنها محدود بود. اکنون، برای نخستین بار، کلان‌داده دنیای تازه‌ای را به روی ما می‌گشاید. گاه به گونه‌ای خیره‌کننده با کامپیوترهایی مانند آمازون اکو[1] که فقط با به کار گرفتن گفتار تعامل برقرار می‌کنیم. گاهی هم در ظاهر دیده نمی‌شود چنان‌که در مورد کارت‌های وفاداری سوپرمارکت‌ها انجام می‌شود. قدر مسلم این‌که کاربردهای کلان‌داده به‌سرعت افزایش می‌یابند و امکان بالقوه عظیمی دارند که خواه‌ناخواه بر ما اثر بگذارند.

چگونه این همه قدرت نهفته می‌تواند در چیزی مقدماتی چون «داده» باشد؟ برای پاسخ به این پرسش باید بهتر درک کنیم که کلان‌داده به‌راستی چیست و چطور می‌توان از آن استفاده کرد. بگذارید بحث را با واژه‌ای ادامه بدهیم که با حرف «د» آغاز می‌شود.

۱. Amazon Echo؛ بلندگویی هوشمند که سایت آمازون طراحی کرده و توسعه داده است. این دستگاه از یک بلندگوی ۲۳/۵ متری با هفت قطعه میکروفن تشکیل شده است و قابلیت‌های تعامل صوتی، پخش موسیقی، ساخت فهرست کار، تنظیم آلارم، پخش پادکست، پخش کتاب‌های صوتی و اعلام وضعیت آب‌وهوا در آن تعبیه شده است. امکان کنترل دستگاه‌های هوشمند دیگر را نیز دارد و می‌شود از آن به عنوان یک هاب اتوماسیون خانگی بهره برد. ــ و.

۲

اندازه مهم است

داده یعنی...

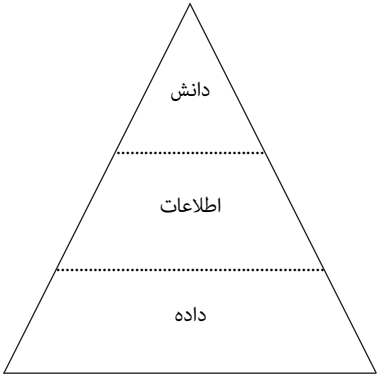
واژهٔ «داده‌ها» (data) در لغت‌نامه‌های انگلیسی از واژهٔ جمع لاتینی datum مشتق می‌شود، به معنای «چیزی که داده می‌شود». بیشتر دانشمندان چنین وانمود می‌کنند که گویی ما به زبان لاتینی حرف می‌زنیم، و می‌گویند «داده» (data) باید جمع باشد و باید بگوییم «داده‌ها قانع‌کننده هستند» نه «داده قانع‌کننده است». ولی فرهنگ انگلیسی آکسفورد که معمولاً محافظه‌کار است می‌پذیرد که امروزه کاربرد داده (data) به عنوان اسم جمع مفرد ــ که به یک مجموعه اشاره دارد ــ «به طور کلی استاندارد تلقی می‌شود».

«چیزی که داده می‌شود» خود قدری حالت رمزی دارد و معمولاً بر اعداد و اندازه‌ها دلالت می‌کند، هرچند که ممکن است هر چیزی باشد که بشود آن را ضبط و ثبت کرد و بعداً مورد استفاده قرار داد. مثلاً واژه‌های همین کتاب داده هستند.

از داده اطلاعات می‌سازیم. یعنی مجموعه‌هایی از داده‌های مرتبط را کنار هم می‌گذاریم تا چیزی معنی‌دار دربارهٔ جهان به ما بگوید. اگر واژه‌های این کتاب داده هستند، طریقی که من واژه‌ها را در جمله‌ها، بندها و فصل‌ها تنظیم کرده‌ام آنها را تبدیل به اطلاعات می‌کند. از اطلاعات هم دانش

می‌سازیم. دانش ما تعبیر و تفسیری است از اطلاعات برای به کار گرفتن آن ــ با خواندن کتاب و پردازش اطلاعات به منظور شکل‌دادن به اندیشه‌ها، نظرات و اقدامات آتی دانش حاصل می‌شود.

در هرم زیر می‌توان داده را به صورت قاعدهٔ هرم در نظر گرفت:



در مثالی دیگر، داده ممکن است مجموعه‌ای از اعداد باشد. تنظیم این اعداد در یک جدول برای نشان دادنِ مثلاً تعداد ماهی در ناحیهٔ مشخصی از دریا، ساعت به ساعت، به ما اطلاعات می‌دهد. و کسی که از این اطلاعات استفاده می‌کند تا تصمیم بگیرد بهترین زمان ماهیگیری چه موقع است دارای دانش است.

بالا رفتن از هرم

از آغاز تمدن بشر، ما دانش فنی خود را برای به کار گرفتن داده و بالا رفتن از این هرم افزایش داده‌ایم. این کار دست‌کم ۴۰۰۰ سال پیش با الواح گلین شروع شد که در بین‌النهرین از آنها استفاده می‌شد. به جای این‌که داده در ذهن نگه داشته یا روی دیوار غار کنده شود، الواح این امکان را به وجود می‌آوردند که داده به صورتی عملی و قابل استفاده حفظ شود.

تقریباً در همان زمان‌ها بود که نخستین داده‌پرداز به صورت چرتکه‌ای ساده، ولی به گونه‌ای شگفت‌انگیز پرقدرت، به‌تدریج ساخته شد. این ابزار ــ که ابتدا با استفاده از نشانه یا سنگ به صورت ستون، و بعدها مهره‌هایی بود که مفتولی از داخل‌شان رد می‌شد ــ به کارگیری داده‌های عددی ساده را امکان‌پذیر کرد. ولی با این‌که توانایی انسان برای استفادهٔ درست و خوب از داده در طی سده‌ها افزایش یافته بود، مفاهیم کلان‌داده تازه در پایان سدهٔ نوزدهم و به سبب مشکلی که در سرشماری پیش آمد، ظهور کردند.

از همان اوایلی که در امریکا سرشماری باب شد، احتمال می‌رفت منابع برقراری ارتباط با داده از عهدهٔ ذخیره و پردازش مقدار فزایندهٔ داده‌ها برنیایند. به نظر می‌رسید که کل جریان محکوم به شکست است. فاصلهٔ سرشماری‌ها بازه‌ای ده ساله بود ــ ولی با رشد جمعیت و پیچیدگی داده‌ها، مدت‌زمانی که صرف جدول‌بندی داده‌های سرشماری می‌شد مرتباً طولانی‌تر می‌شد. چیزی نمانده بود که تحلیل کامل یک سرشماری حتی پیش از رسیدن زمان سرشماری بعدی هم به سرانجام نرسد. این مشکل با مکانیزه کردن حل شد. ابزارهای الکترومکانیک امکان استفاده از کارت‌های منگنه‌شده را فراهم کردند که هر کدام نشانگر پاره‌ای از داده‌ها بود که می‌شد به طور خودکار خیلی سریع‌تر از هر آدمی از آن به‌درستی استفاده کرد.

تا اواخر دههٔ ۱۹۴۰، با پیدایش کامپیوترهای الکترونیکی، تجهیزات به مرحلهٔ دوم هرم رسیدند. داده‌پردازی جای خود را به فناوری اطلاعاتی داد. انبار اطلاعاتی از زمان اختراع خط وجود داشت. کتاب یک انبار اطلاعاتی است که فضا و زمان را شامل می‌شود. اما فناوری جدید این امکان را به وجود آورد که از این اطلاعات طوری استفاده شود که پیش‌تر هرگز به کار گرفته نشده بود. کامپیوترهای غیرانسانی (واژهٔ کامپیوتر در اصل اشاره به ریاضی‌دانانی بود که کار محاسبات روی کاغذ را به‌عهده داشتند) نه تنها می‌توانستند داده را به کار گیرند بلکه می‌توانستند آن را تبدیل به اطلاعات کنند.

مدت‌ها این‌طور به نظر می‌رسید که انگار مرحلهٔ نهایی خودکار کردن هرم ــ تبدیل اطلاعات به دانش ارزنده ــ نیاز به «سیستم‌های دانش‌بنیان» دارد. این برنامه‌های کامپیوتری در صدد برآمدند قواعدی را که انسان‌ها برای به کار گرفتن دانش و تفسیر داده‌ها استفاده می‌کردند جذب و ذخیره کنند. ولی سیستم‌های خوب دانش‌بنیان به سه دلیل دست‌نیافتنی از آب درآمدند. اول این‌که، متخصصان انسانی چندان رغبتی نداشتند که خود را از کار کنار بکشند و به‌ندرت همکاری می‌کردند. دوم این‌که، متخصصان انسانی اغلب خود نمی‌دانستند چگونه اطلاعات را تبدیل به دانش می‌کنند و حتی اگر می‌خواستند هم نمی‌توانستند این قواعد را برای متخصصان فناوری اطلاعات بازگو کنند. دست آخر این‌که، آن جنبه‌هایی از واقعیت که به این شیوه ساخته و پرداخته می‌شدند خیلی پیچیده‌تر از آن بودند که نتیجهٔ سودمندی به دست دهند.

دنیای واقعی در مفهوم ریاضی غالباً درهم‌برهم است. منظور این نیست که آنچه اتفاق می‌افتد تصادفی و درهم است ــ درست برعکس. منظور این است که تعداد تعامل‌های میان اجزایی از جهان که مورد مطالعه قرار می‌گیرند آن‌قدر زیاد است که یک تغییر بسیار کوچک در وضعیت فعلی می‌تواند تغییری عظیم در پیامد آتی ایجاد کند. پیش‌بینی آینده در حدی که به کار آید، عملاً غیرممکن می‌شود.

اما حالا که ما از طریق دسترسی به اینترنت و کامپیوتر همراه انقلاب کامپیوتری دیگری را از سر می‌گذرانیم، کلان‌داده روش جایگزین عملگرایانه‌تری برای رسیدن به سطح بالای هرم داده ــ اطلاعات ــ دانش به دست می‌دهد. سیستم کلان‌داده مقادیر زیادی داده را می‌گیرد ــ داده‌ای که معمولاً بی‌نظم و به‌سرعت در جریان است ــ و با استفاده از آخرین فناوری‌های اطلاعاتی این داده را به کار می‌گیرد و تحلیل می‌کند، آن هم به شیوه‌ای که انعطاف‌ناپذیری کمتر و تأثیرپذیری و جوابگویی بیشتری دارد. تا همین اواخر این کار غیرممکن بود. به‌کارگیری داده در این مقیاس عملی