



جمهوری اسلامی ایران  
شورای عالی فضای مجازی  
مرکز ملی فضای مجازی

فضای  
مجازی



# کلان داده و آینده علوم انسانی-اجتماعی

۱۴



## کلان داده و آینده علوم انسانی-اجتماعی

گزارش شماره ۴

خرداد ماه ۱۳۹۸

---

تهیه شده در: پژوهشگاه مرکز ملی فضای مجازی-گروه مطالعات فرهنگی و اجتماعی  
تهیه کننده: یحیی شعبانی (دانشجوی دکترای پژوهشگاه علوم انسانی و مطالعات فرهنگی)  
ناظر علمی: امیررضا باقرپورشیرازی

---

نشانی: تهران، میدان آرژانتین، خیابان بیهقی، نش خیابان ۱۶ غربی، پلاک ۲۰، کدپستی ۱۵۱۵۶۷۴۳۱۱

شماره تماس: ۸۶۱۲۱۰۶۱ <http://www.majazi.ir>

حقوق مادی و معنوی این اثر متعلق به مرکز ملی فضای مجازی است و استفاده از مطالب آن صرفاً با ذکر مأخذ  
بلامانع است.

محتوای انتشاریافته در این گزارش الزاماً بیانگر دیدگاه مرکز ملی فضای مجازی نیست

## سخن نخت

فضای مجازی با شتاب شگرف و روبه‌ترایدی که در حال بط و گسترش است تمام ساحات اجتماعی، اقتصادی، سیاسی و فرهنگی زندگی بشر را در نور دیده و هر روز بخش بزرگی از زندگی واقعی را در خود فرو برده و حیات متفاوت و جدیدی به آن می‌دهد. لذا به نظرمی رسد دو نگاه کلان به فضای مجازی وجود دارد: نگاه اول که بالانص در ابتدای رشد و تکوین فضای مجازی مسلط شده بود، آن را همچون ابزارهای کنارسای ابزارهای بشری تصویر می‌کرد که تنها طریقت داشت. اما نگاه دوم، در نتیجه رشد تحولات خیره‌کننده فضای مجازی و سایه‌گستری آن در حوزه‌ها و شئون بشر در یک دهه اخیر آن را چون سکویی می‌داند که بسیار فراتر از شان ابزارهای حیات انسان‌ها را سامان جدیدی داده و ادعای تمدن‌نویسی را دارد. رویکردی که از قضا از چشمان بصیر رهبران انقلاب نیز دور نمانده و انتظاری تمدنی از فضای مجازی در ایران را مطالبه داشته‌اند.

در همین راستا گزارش‌های عصر فضای مجازی تلاش می‌کند تا فهم سازمان‌ها و دستگاه‌های مرتبط با حوزه‌ی فضای مجازی را ارتقاء بخشد و آن‌ها را برای مواجهه فعال و خردمندانه با تحولات این عرصه همیار سازد.

سید ابوالحسن فیروزآبادی  
 دبیر شورای عالی و رئیس مرکز ملی فضای مجازی

## چکیده

داده‌ها همه جا حضور دارند. ظهور ابزارهای دیجیتالی و استفاده از آن‌ها در فعالیت‌های یومیه موجب شده است تا اغلب این فعالیت‌ها ردپاهای دیجیتالی از خود بر جا بگذارند. در دنیای امروز بصورت روزانه حجم بسیار بزرگی از داده‌ها (کلان‌داده‌ها) تولید می‌شود که علاوه بر نیاز به ذخیره‌سازی، نیازمند پردازش و مدیریت نیز هستند. کلان‌داده‌ها امروزه به عنصری جدایی‌ناپذیر در حوزه فناوری اطلاعات و ارتباطات تبدیل شده‌اند علی‌الخصوص با توسعه روزافزون شبکه‌های اجتماعی. قابلیت‌ها و توانایی‌های ما برای اندازه‌گرفتن پدیده‌های طبیعی و اجتماعی از طریق کلان‌داده‌ها در مقیاسی بی‌سابقه به سرعت در حال افزایش است. در علم، قابلیت برای گردآوری و تحلیل حجم عظیمی از داده‌ها و ردپای دیجیتالی آن‌ها موجب پیشرفت‌های فراوانی شده و در بسیاری از حوزه‌های پژوهشی تغییرات صریحی ایجاد نموده است. جهان جدید در حال تجربه‌ی بحران‌های فراگیر و گسترده‌ای است و با چالش‌های جدیدی مواجه است؛ به عنوان مثال انفجار جمعیتی مسائل جدیدی را برابر ما قرار می‌دهد. برای حل این چالش‌ها نیازمند فهم دینامیک اجتماعی و نیز ساختار سازمان‌های اجتماعی هستیم. سیل داده‌ها امکان رویکرد کاملاً جدیدی برای تحلیل اجتماعی فراهم می‌آورد و علوم اجتماعی با کمک گرفتن از همین رویکرد جدید می‌توانند با چالش‌های فوق‌الذکر مواجه شوند. کلان‌داده‌ها فرصت‌های جدیدی برای پژوهشگران علوم اجتماعی ایجاد کرده‌اند تا عمیقاً نحوه‌ی مطالعه‌ی پدیده‌های اجتماعی را دگرگون کنند. به نظر می‌رسد ظهور کلان‌داده و مناقشات پیرامون آن به تغییراتی رادیکال در علوم خواهد انجامید. علوم، به‌ویژه علوم اجتماعی و انسانی، در مواجهه با کلان‌داده‌ها به معرفت‌شناسی، هستی‌شناسی و روش‌شناسی جدیدی نیاز دارند. تولید گسترده کلان‌داده‌ها همراه با توسعه‌ی معرفت‌شناسی‌های جدید بسیاری را متقاعد کرده که یک انقلاب داده‌ای در راه است و این انقلاب پیامدهای فراگیری دارد از جمله اینکه نحوه تولید معرفت دگرگون خواهد شد، نحوه انجام کسب و کار تغییرات شدیدی را از سر خواهد گذراند و نحوه حکمرانی نیز اساساً دچار دگردیسی خواهد شد. در تحلیل نهایی، کلان‌داده امکان یک پارادایم جدید برای پژوهش را فراهم می‌آورد. برخی متفکران معتقدند کلان‌داده‌ها و ابزارهای تحلیلی ملازم با آن حوزه‌ی جدیدی از تولید معرفت را اعلام می‌کنند که مشخصه‌ی آن پایان نظریه است. به عبارتی، پدیده‌ها برای سخن گفتن از خود دیگر نیازی به نظریه نخواهند داشت.

**واژگان کلیدی:** کلان‌داده، ردپای دیجیتالی، انقلاب داده‌ای، علوم اجتماعی محاسباتی، انقلاب علمی، پایان نظریه.

## فهرست مطالب

- ۱- مقدمه: سیل داده‌ها ..... ۴
- ۲- تغییر پارادایم به سوی علوم اجتماعی محاسباتی ..... ۸
- ۲-۱ حیات/داده ..... ۱۱
- ۲-۲ ذهن/ماشین ..... ۱۲
- ۲-۳ استقراء/استنتاج ..... ۱۳
- ۳- تغییر روش‌های آماری با ظهور کلان داده‌ها: نمونه‌برداری ایستا و دینامیک ..... ۱۴
- ۴- انقلاب داده‌ای و پیامدهای آن: تغییر پارادایم در علم ..... ۱۷
- ۴-۱ پایان نظریه: باززایی تجربه‌گرایی ..... ۱۸
- ۵- علم داده-رهنمون: از علوم معرفت-محور به علوم داده-محور ..... ۲۰
- ۶- کلان داده‌ها و رویکرد پوزیتیویستی در علوم انسانی ..... ۲۱
- ۷- علوم انسانی دیجیتال و از دست رفتن لحن و رتوریک انسانی ..... ۲۲
- ۸- آرشیو و روایت: کلان داده و علوم اجتماعی تاریخی ..... ۲۴
- ۹- نتیجه‌گیری: فرآیند به جای ساختار ..... ۲۸
- ۱۰- منابع ..... ۲۹

## ۱- مقدمه: سیل داده‌ها

داده‌ها همه جا هستند، آن‌ها به درون جهان ما نفوذ کرده‌اند. فعالیت‌های روزمره‌ی ما سرشار از داده‌ها است، فعالیت‌هایی مانند فرستادن ای-میل، فراخواندن دوستان، استفاده از کارت‌های اعتباری، آزمایش‌هایی که با ابزارهای اندازه‌گیری اختریفی یکی انجام می‌دهیم، بررسی اخبار ترافیک، خرید روزانه و... امروزه بخاطر ظهور ابزارهای دیجیتال و سرریز کردن آن‌ها در فعالیت‌های یومیه، اغلب این فعالیت‌ها ردپاهای دیجیتال از خود بر جا می‌گذارند. مادامیکه قابلیت‌ها و توانایی‌های ما برای اندازه‌گرفتن پدیده‌های طبیعی و اجتماعی در مقیاسی بی‌سابقه به سرعت در حال افزایش است، شاهد رشدی نمایی در همه‌ی این ردپاهای دیجیتال هستیم. رشد اطلاعات دیجیتال همان چیزی است که آن را کلان‌داده (Big Data) می‌خوانیم. داده‌هایی که بسیار سریع‌تر از میزان پردازش، تحلیل و استخراج آنها را تولید و کسب می‌کنیم.

به عنوان مثال در دهه گذشته ما انسان‌ها قریب به پنج اگزابایت (۱۰<sup>۱۸</sup>) داده تولید کرده‌ایم، اما امروزه این مقدار داده در کمتر از دو روز تولید می‌شود و هر دو سال نیز حجم تولید آن دوبرابر می‌شود. طبیعی است که این میزان رشد در تولید داده‌ها چالشی برای صنایع، علوم، سیاست، اقتصاد و دیگر حوزه‌های انسانی به حساب می‌آید هم از حیث محاسبه‌پذیری و هم از حیث ذخیره‌سازی آن‌ها. به عنوان مثال، کلان‌داده‌ها شیوه‌های بی‌سابقه‌ای برای کسب و کار ایجاد می‌کنند اگرچه در تحلیل نهایی تأثیر داده‌ها بسیار فراتر از حوزه کسب و کار می‌رود.<sup>۱</sup>

کلان‌داده‌ها امروزه به عنصری جدا ناپذیر در حوزه فناوری اطلاعات و ارتباطات تبدیل شده است علی‌الخصوص با توسعه روزافزون شبکه‌های اجتماعی. در دنیای امروز بصورت روزانه حجم بسیار بزرگی از داده‌ها تولید می‌شود که علاوه بر نیاز به ذخیره‌سازی، نیازمند پردازش و مدیریت نیز هستند. فناوری‌های کلاسیک معمولاً در مواجهه با حجم بسیار بزرگ داده‌ها یا بسیار کند عمل می‌کنند و یا اینکه توانایی مدیریت حجم وسیعی از داده‌ها با انواع گوناگون (ساخت‌یافته، نیمه ساخت‌یافته، ساخت نیافته)<sup>۲</sup> را نخواهند داشت. عبارت Big Data مدت‌ها است که برای اشاره به حجم‌های عظیمی از داده‌ها که توسط سازمان‌های بزرگی مانند گوگل یا ناسا ذخیره و

<sup>1</sup> Seref Sagiroglu and Duygu Sinanc, "Big data: A review", in: *Collaboration Technologies and Systems (CTS), 2013 International Conference on, IEEE, 2013, pp. 42-47.*

<sup>2</sup> Rob Kitchin and Gavin McArdle, What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets, p.1.

تحلیل می‌شوند مورد استفاده قرار می‌گیرد. اما به تازگی، این عبارت بیشتر برای اشاره به مجموعه‌های داده‌های بزرگی استفاده می‌شود که به قدری بزرگ و حجیم هستند که با ابزارهای مدیریتی و پایگاه‌های داده سنتی و معمولی قابل مدیریت نیستند. مشکلات اصلی در کار با این نوع داده‌ها مربوط به برداشت و جمع‌آوری، ذخیره‌سازی، جستجو، اشتراک‌گذاری، تحلیل و نمایش آن‌ها است. پردازش کلان‌داده‌ها عموماً توسط دانشمندان در علوم مانند هواشناسی، ژنتیک، شبیه‌سازی‌های پیچیده فیزیک، تحقیقات زیست‌شناسی و محیطی، جستجوی اینترنتی، تحلیل‌های اقتصادی و مالی و تجاری مورد به کار گرفته می‌شود. حجم داده‌های ذخیره‌شده در مجموعه‌های کلان‌داده‌ها، عموماً به خاطر تولید و جمع‌آوری داده‌ها از مجموعه بزرگی از تجهیزات و ابزارها مانند گوشی‌های موبایل، حسگرهای محیطی، لاگ نرم‌افزارهای مختلف، دوربین‌ها، میکروفون‌ها، دستگاه‌های تشخیص یا سامانه بازشناسی با امواج رادیویی (RFID)، شبکه‌های حسگر بیسیم و غیره با سرعت خیره‌کننده‌ای در حال افزایش است.

در علم، قابلیت برای گردآوری و تحلیل حجم عظیمی از داده‌ها و ردپای دیجیتال آن‌ها موجب پیشرفت‌های فراوانی شده است و در کنار این پیشرفت‌ها بی‌تردید در بسیاری از حوزه‌های پژوهشی تغییرات صریحی ایجاد نموده است. دامنه این تغییرات از فیزیک و زیست‌شناسی<sup>۱</sup> تا علوم کامپیوتر و اقتصاد<sup>۲</sup> را در برمی‌گیرد اما این پیشرفت‌ها در هیچ جا به اندازه‌ی مطالعه‌ی نظام‌های اجتماعی اهمیت ندارد. جهان جدید در حال تجربه‌ی بحران‌های فراگیر و گسترده‌ای است و به طور کلی با چالش‌های جدیدی مواجه می‌شود؛ به عنوان مثال انفجار جمعیتی و حتی اشکال جدید منازعه در بافت‌های اجتماعی مسائل جدیدی را برابر ما قرار داده است. برای حل این چالش‌ها نیازمند فهم دینامیک اجتماعی و نیز ساختار سازمان‌های اجتماعی هستیم. برخی از این چالش‌ها عبارتند از: مهاجرت جهانی، ناپایداری اقتصادی، انتشار بیماری‌های اپیدمیک، جرایم سازمان‌یافته و... سیل داده‌ها امکان رویکرد کاملاً جدیدی برای تحلیل اجتماعی فراهم می‌آورد و علوم اجتماعی با کمک گرفتن از همین رویکرد جدید می‌توانند با چالش‌های فوق‌الذکر

<sup>1</sup> Elena Aronova, Karen S Baker, and Naomi Oreskes, "Big science and big data in biology: from the international geophysical year through the international biological program to the long term ecological research (LTER) network, 1957–present", in: (2010).

<sup>2</sup> Liran Einav and Jonathan D Levin, *The data revolution and economic analysis*, tech. rep., National Bureau of Economic Research, 2013.

مواجه شوند.<sup>۱</sup> علاوه بر این، توسعه‌ی قابلیت‌های محاسباتی و ذخیره‌سازی داده‌ها این امکان را در اختیار دانشمندان قرار می‌دهد تا کلان‌داده‌ها را مدیریت و کنترل کنند و از طریق آن رفتار انسان در جامعه را بفهمند، مدل‌سازی و پیش‌بینی کنند.<sup>۲</sup> دانشمند علوم اجتماعی در مواجهه با کلان‌داده‌ها بصیرت‌های ارزشمندی درباره‌ی دینامیک اجتماعی به دست می‌آورد. به عنوان مثال، داده‌هایی که از طریق شبکه‌ی موبایل به دست می‌آیند منبعی غنی و ارزشمند برای فهم و رصد فعالیت‌های انسان در ابعاد بسیار بزرگ است. می‌توان از این داده‌ها اطلاعات جغرافیایی استخراج کرد و از آن برای تخمین فعالیت‌های جمعیتی استفاده نمود و حتی می‌توان حرکات کاربران را بازسازی کرد. این ردپاهای جغرافیایی به پژوهش بر روی نقشه جمعیتی و نیز تحرک انسانی کمک شایانی می‌کنند و نتیجه‌ی این پژوهش‌ها می‌تواند در برنامه‌ریزی شهری،<sup>۳</sup> پیش‌بینی ترافیک<sup>۴</sup> و حتی جلوگیری از انتشار بیماری‌های اپیدمیک<sup>۵</sup> کمک کند. داده‌های استخراج شده از موبایل می‌تواند منبع مناسبی برای بررسی تعاملات اجتماعی نیز باشند. دانشمندان علوم اجتماعی و رفتاری با مطالعه‌ی همین داده‌های موبایل می‌توانند الگوهای ارتباطی و نیز آرایش گروه‌های

---

<sup>1</sup> David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al., "Life in the network: the coming age of computational social science", in: *Science (New York, NY)* 323.5915 (2009), p. 721.

<sup>2</sup> Claudio Cioffi-Revilla, "Computational social science", in: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.3 (2010), pp. 259–271.

<sup>3</sup> Carlo Ratti, S Williams, D Frenchman, and RM Pulselli, "Mobile landscapes: using location data from cell phones for urban analysis", in: *Environment and Planning B Planning and Design* 33.5 (2006), p. 727.

<sup>4</sup> Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava, "Using mobile phones to determine transportation modes", in: *ACM Transactions on Sensor Networks (TOSN)* 6.2 (2010), p. 13.

<sup>5</sup> Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza, "On the use of human mobility proxies for modeling epidemics", in: *PLoS computational biology* 10.7 (2014), e1003716.



اجتماعی را در مقیاس‌های بسیار کلان به دست آورند.<sup>۱</sup> شناخت ما از جمعیت انسانی و توزیع آن در بسیاری از مناطق دچار فقیر اساسی است هر چند این شناخت برای سیاست‌گذاری و تصمیمات مربوط به حکمرانی اهمیت فراوانی دارد. اگرچه سرشماری یک راه حل برای این معضله به شمار می‌رود اما سرشماری منبعی کمیاب و بسیار گران برای داده‌های جمعیتی است. در بسیاری از سرشماری‌ها نیز تخمین دقت لازم را ندارد و اصطلاحاً رزلوشن این سرشماری‌ها بنا به دلایل بسیار زیاد کم و نادقیق است. سرشماری منبعی گسسته در مطالعات جمعیت‌شناختی نیز است زیرا تداومی در زمان ندارد و ما سرشماری پیوسته از جمعیت نداریم، اصطلاحاً سرشماری فقط عکسی لحظه‌ای از یک رودخانه‌ی جاری می‌گیرد. برخی پژوهش‌ها نشان می‌دهند که داده‌های موبایلی تا حد زیادی می‌توانند این شکاف را پر کنند و تخمینی قابل اعتماد از جمعیت و توزیع آن به دست دهند. ترسیم نقشه جمعیتی از طریق داده‌های موبایلی بسیار ارزانتر از روش‌های قدیمی و سرشماری است و پیوستگی و سیلان بیشتری دارد. علاوه بر تسهیلات روشی و هزینه‌ی کمتر این داده‌ها، مزایای دیگری نیز افزون بر روش‌های قدیمی در استفاده از آن‌ها به چشم می‌خورد که عموماً برای اندازه‌گیری دینامیک جمعیت به کار می‌روند.<sup>۲</sup> آنچه از حیث روش‌شناختی در این پژوهش‌ها اهمیت دارد طراحی الگوریتم‌های مناسب برای تحلیل کلان داده است. طراحی و معرفی الگوریتم مناسب می‌تواند پژوهشگر را بسیار به مقصود مورد نظرش نزدیک کند.<sup>۳</sup> به نظر می‌رسد همین الگوریتم‌ها قرار است جای نظریه‌های کلاسیک را بگیرند. پژوهشگر به جای سر و کله زدن با نظریه‌های مختلف در حوزه‌ی کارش باید الگوریتم و ابزاری برای دسته بندی و تحلیل بیشمار داده بیابد تا این داده‌ها خودشان سخن بگویند. نکته‌ی مهم در اینجا آن است که دانشمند با مسائل جدیدی مواجه می‌شود. مجموعه داده‌ها عموماً حجم و اندازه‌ی عظیمی دارند و به همین دلیل، هم به تکنیک‌ها و سخت‌افزارهای جدیدی برای ذخیره‌ی آن‌ها

<sup>1</sup> Markus Schläpfer, Luís MA Bettencourt, Sébastien Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B West, and Carlo Ratti, "The scaling of human interactions with city size", in: *Journal of The Royal Society Interface* 11.98 (2014), p. 20130789.

<sup>2</sup> DL Balk, U Deichmann, G Yetman, F Pozzi, SI Hay, and A Nelson, "Determining global population distribution: methods, applications and data", in: *Advances in parasitology* 62 (2006), pp. 119–156.

<sup>3</sup> Budhendra Bhaduri, Edward Bright, Phillip Coleman, and Marie L Urban, "LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics", in: *GeoJournal* 69.1-2 (2007), pp. 103–117.

نیاز داریم و هم به ابزارها و قابلیت‌های جدیدی برای دستیابی، محاسبه و تحلیل داده‌ها. بخش مهمی از کار علمی برای دانشمندی که با کلان‌داده‌ها سر و کار دارند اندیشیدن به همین موارد است. واضح است که دانشمندان در عصر کلان‌داده‌ها وظایفی بر عهده دارند که به هیچ وجه نشانی از آنها در علوم کلاسیک نیست. در یک کلام دانشمند در جهان جدید باید فهم مکفی از «محاسبه» داشته باشد.

## ۲- تغییر پارادایم به سوی علوم اجتماعی محاسباتی

کلان‌داده‌ها فرصت‌های جدیدی برای پژوهشگران علوم اجتماعی ایجاد کرده‌اند تا عمیقاً نحوه‌ی مطالعه‌ی پدیده‌های اجتماعی را دگرگون کنند. با ظهور تکنولوژی‌های گردآوری داده‌ها و شیوه‌ها و ابزار تحلیل داده‌ها به نظر می‌رسد تغییرات عمیقی در پرسش‌هایی که می‌توان در پژوهش‌های اجتماعی پرسید و روش‌هایی که می‌توان در آنها بکار برد ایجاد خواهد شد. کلان‌داده‌ها فقط در علم اثرگذار نخواهند بود بلکه به تحلیل‌گران حوزه کسب و کار نیز اجازه خواهد داد تا مشاهداتی مکرر، کنترل‌شده و معنادار درباره‌ی پدیده‌های اقتصادی در جهان واقعی کسب کنند.<sup>۱</sup> فرصت‌هایی که کلان‌داده‌ها برای پژوهش فراهم آورده موجهی از هیجان ایجاد کرده است. با این حال، پرسش‌ها و ابهاماتی نیز در این فرآیند وجود دارد. به عنوان مثال می‌توان از نقش عامل انسانی در این رویکرد جدید به علم پرسید. آیا انسان‌ها هنوز عاملان اصلی علم خواهند بود یا جای خود را به ماشین‌ها خواهند داد؟<sup>۲</sup> طبیعتاً هنوز پاسخ محکم و متقنی به این پرسش‌ها وجود ندارد. با وجود این، برخی گمان می‌کنند که انسان‌ها هنوز هم فرآیندها را طراحی خواهند کرد و با ترکیب تحلیل ماشین‌محور داده‌ها و شهود انسانی بصیرت‌ها به بهترین شکل به دست خواهند آمد. پیشرفت در محاسبه این امکان را فراهم می‌آورد که داده‌ها را در مقیاس جمعیت جهان تحلیل کنیم. اما مسئله بر سر این است که چگونه می‌توان عمق پژوهش در علوم اجتماعی را با مقیاس و بزرگی آمار و علوم کامپیوتر ترکیب کرد. نقش یادگیری ماشین در علوم اجتماعی از چه قرار است؟ پیشرفت در محاسبات تحلیل داده‌ها را در مقیاسی ممکن ساخته است که در دسترس پژوهشگران قبلی قرار نداشت. وب و شبکه به بزرگترین رصدخانه‌ی علوم اجتماعی

<sup>1</sup> R.M. Chang, et al., Understanding the paradigmshift to computational social science in the presence of big data, Decision Support Systems (2013), p.1. <http://dx.doi.org/10.1016/j.dss.2013.08.008>

<sup>2</sup> Prabhakar Raghavan, It's time to scale the science in the social sciences, pp. 1-4.

تبدیل شده است. اما همزمان نظریه‌های رفتار انسان که ای بسا قرن‌ها سابقه دارند پشت صحنه‌ای برای این ظرفیت‌های اخیر تشکیل می‌دهند. به طور سنتی این نظریه‌ها متکی بر برآوردها و آزمایش‌های مفصلی هستند که در مقیاس متوسط انجام شده‌اند و عمق این نظریات و مطالعات در تقابل با نتایج نیرومند اما ساده‌ی تحلیل داده‌های بزرگ قرار دارد. یک پیشنهاد ممکن برای پژوهش این است که درباره‌ی پرسش‌ها و روش‌شناسی‌های علمی در پرتو روش‌های کسب داده و داده‌هایی که اکنون در دسترس ما قرار دارند بازاندیشی کنیم. موشکافی دیگری نیز در اینجا لازم است. ما زبانی برای به رمز درآوردن رفتار انسان در اختیار نداریم. چه می‌شد اگر دانشمندان می‌توانستند رفتار انسان را به رمز درآورند و عناصر رفتاری او را تشخیص دهند، به نحوی که بتوان قوانین مقدماتی فیزیک را ترکیب کرد تا مأموریت‌های پیچیده‌ی انسان را تولید کند؟<sup>۱</sup>

دو پرسش برای شاخه‌ی در حال ظهور علوم اجتماعی محاسباتی محوریت دارند. (۱) چگونه می‌توان عمق پژوهش در علوم اجتماعی را با مقیاس و قدرتمندی آمار و علوم کامپیوتر ترکیب کرد؟ با مشاهده‌ی میلیون‌ها کاربر می‌توان دقیقاً پیش‌بینی کرد که چه کسری از کاربران بر روی یک لینک کلیک خواهند کرد اما تقریباً فهمی از این نکته نداریم که چرا آن‌ها کلیک می‌کنند. آیا می‌توان به تلفیقی از عمق مطالعات میدانی و دقت در تحلیل داده‌ها با مقیاس بزرگ به پرسش‌های علوم اجتماعی پاسخ داد؟ (۲) آیا می‌توان پرسش‌های پیچیده در علوم اجتماعی را به فرضیه‌های ساده‌تری تجزیه کرد که آزمودن آن‌ها نیرومندتر باشد؟ این کار بدان معناست که پژوهشگران می‌توانند بر روی سازه‌ای از فرضیات نیرومند کار کنند که نظریه‌های اجتماعی پیچیده را می‌توان با آن‌ها مونتاژ کرد. همه‌ی رفتارهای انسان به آن نحوی که نظریه بازی‌ها مدل‌بندی می‌کند عقلانی نیست. پیشرفت در یادگیری ماشین که رشته‌ای در پیوند آمار و علوم کامپیوتر است، این امر را ممکن ساخته است که بسیاری از مسائل مربوط به پیش‌بینی را صرفاً با تحلیل تعداد عظیمی از نمونه‌های تاریخی با برنامه‌ی یادگیری-ماشین حل کنیم.<sup>۲</sup> نکته‌ی مهم در این زمینه نه الگوریتم‌های جدید یا نظریه‌های اجتماعی نوین بلکه روش‌شناسی‌های جدید برای تجزیه‌ی یک پرسش دشوار و سخت در علوم اجتماعی به سلسله‌ای از تحلیل‌های نیرومند است که تکرارپذیر و قابل ساختن هستند. اگر این کار را انجام دهیم نه تنها فهم خود از علوم اجتماعی کلاسیک را یکپارچه کرده‌ایم بلکه در تولید علوم اجتماعی جدید نیز موفق شده‌ایم،

<sup>۱</sup> Ibid.

<sup>۲</sup> R.M. Chang, et al., Understanding the paradigmshift to computational social science in the presence of big data, Decision Support Systems (2013), p.4. <http://dx.doi.org/10.1016/j.dss.2013.08.008>

یعنی علمی که به تجربه‌های ترکیبی مربوطند و این تجربه‌ها در جوامع انسانی بیش از پیش رشد خواهند نمود. البته پاسخ دادن به پرسش‌های فوق به بررسی‌های موشکافانه‌ای نیازمند است. باید مفصلاً به پرسش‌ها، موضوعات، چالش‌ها، مسائل و نویدهایی که کلان‌داده‌ها در علوم اجتماعی به ارمغان می‌آورند پرداخت. برخی از این پرسش‌ها از این قرارند: جایگاه و سرشت حیات انسان در روش‌های تحلیل کلان‌داده‌ها چیست؟ آیا سرشت تفسیر پدیده‌های اجتماعی و انسانی در تحلیل کلان‌داده‌ها دچار تغییر می‌شود؟ آیا شاهد دگرگونی در برساخت‌های مقولی هویت‌ها و عوامل فردی خواهیم بود؟ آیا سرشت و ارتباط بافت‌ها و زمانمندی پدیده‌های اجتماعی و انسانی دچار تحولات معناداری خواهد شد؟ آیا اساساً با مفهومی جدید از علیت مواجه خواهیم شد؟ اندیشمندانی که بر روی کلان‌داده‌ها و تأثیر آن‌ها بر علوم اجتماعی و انسانی فکر می‌کنند هنوز پاسخ‌های متقن و صریحی به این پرسش‌ها نداده‌اند، اما طرح این پرسش‌ها به نوبه خود خبر از تحولاتی مهم در این علوم می‌دهد.<sup>۱</sup>

فرضیاتی که در کنش‌های روش‌شناختی علوم محاسباتی وجود دارند نیازمند بررسی بنیادی تأثیر کلان‌داده‌ها بر روی علوم اجتماعی است. وضعیت‌های جدید روش‌شناختی و هستی‌شناختی در حال تغییر چهره‌ی قلمروهای علوم اجتماعی هستند. علاوه بر آنکه کلان‌داده‌ها فرض‌های علمی را مضمحل و دگرگون می‌کنند، بلکه فرضیات قدرتمند خود درباره‌ی ماهیت امر اجتماعی را نیز القا می‌کنند، فرضیاتی که باید آن‌ها را کاوید و ای‌سا خنثی کرد. کار با کلان‌داده‌ها فرضیات محوری موجود در علوم اجتماعی را به چالش کشیده است. این چالش معادل است با چالشی هستی‌شناختی. هستی‌شناسی در اینجا یعنی: نظامی از اعمال، علل، زمانمندی‌ها و بافت‌ها، مقولات، معانی و هویت‌هایی که عوامل و کنش‌ها در درون آن مستقر شده‌اند. بسیاری از این مؤلفه‌ها باید به نحوی در شرایط کلان‌داده‌ها مورد بازبینی قرار بگیرند.<sup>۲</sup> حتی فرضیات مربوط به ماهیت پژوهش علمی نیز دستخوش تغییر شده‌اند. از حیث روش‌شناختی می‌توان تأثیر کلان‌داده‌ها را در چارچوب دوگانه‌های زیر سنجید:<sup>۳</sup>

<sup>۱</sup> Prabhakar Raghavan, It's time to scale the science in the social sciences, pp. 1-4.

<sup>۲</sup> Robin Wagner-Pacifioci, John W Mohr and Ronald L Breiger, Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences, p.1.

<sup>۳</sup> Ibid, p.2.

## ۲-۱ حیات/داده

کلان داده‌ها چگونه می‌توانند مسئولیت خود زندگی را برعهده بگیرند و نماینده‌ی آن باشند؟ برخی معتقدند که کلان داده‌ها تصویری خنثی از واقعیت اجتماعی عرضه می‌کنند. برخی نیز معتقدند ناتوانی در پیش‌پردازش داده‌ها در تحلیل کلان داده‌ها می‌تواند پیامدهای منفی داشته باشد. قبل از آنکه کلان داده‌ها بتوانند برای خودشان حرف بزنند تحلیلگر باید با موضوعاتی مثل کیفیت داده‌ها و اینکه نتایج چگونه قرار است بافت‌مند شوند مواجه شود. به عنوان مثال در مورد ویکیپدیا برخی جامعه‌شناسان و پژوهشگران با بررسی سیاست داخلی ویکیپدیا نشان داده‌اند که سوگیری‌های متعددی در فرآیند تولید و خود پایگاه داده‌ها وجود دارد. به عنوان مثال بازنمایی ناقص و ناکافی دانشوران زن در کاتالوگ ویکیپدیا با عنوان «جامعه‌شناسان زنده آمریکایی» یکی از این سوگیری‌ها است. یا حتی در تحلیل داده‌های توییتر عنوان شده است که چند نوع انحراف ممکن است به وجود آید. به عنوان مثال نشان داده می‌شود که چگونه حضور کاربران فعال‌تر مجموعه‌ای از سوگیری‌های سیستماتیک در شیوه‌ی خلق اطلاعات توییتر و تحلیل آن‌ها به وجود می‌آورد. به همین خاطر کلان داده‌ها ممکن است به خطاهای معنی‌داری منتهی شود. تقطیع داده‌ها یکی از تکنیک‌ها در تحلیل داده‌ها است. مقصود از این تکنیک نشان دادن این امر است که چگونه زیر-جمعیت‌های ناهمگن در اینترنت می‌توانند شناسایی شوند تا بفهمیم که چگونه گروه‌های متمایز و متعدد شرکت‌کنندگان اینترنت را می‌سازند. این نکات به خوبی نشان می‌دهد که در کلان داده‌ها سوگیری و انحرافات می‌تواند وجود داشته باشد و پژوهشگر باید با طرح ابزارها و روش‌هایی آن‌ها را تشخیص دهد و خنثی نماید. وفور داده‌ها لزوماً به معنی درستی داده‌ها نیست. از این حیث بخش مهمی از روش‌شناسی کار با کلان داده‌ها طراحی و ساخت الگوریتم‌ها و ابزارهای تحلیلی برای شناخت انحرافات و اعوجاجات در کلان داده‌ها است. همین نکته‌ی ساده نشان می‌دهد که دانشمند علوم اجتماعی باید فهم خوبی از آمار و تحلیل‌های آماری داشته باشد نه اینکه مثل علوم اجتماعی کلاسیک صرفاً به برخی فرایندهای آن آگاهی داشته باشد.<sup>۱</sup>

---

<sup>۱</sup> Ibid, pp. 2-3.

## ۲-۲ ذهن/ماشین

در فرآیند قرائت داده‌ها، انسان‌ها و تکنیک‌های محاسباتی (ماشین‌ها) چگونه معنا را تفسیر می‌کنند؟ در روش مدل‌سازی موضوعی<sup>۱</sup> (Modeling Topic) از اطلاعات مربوط به وقوع کلمات در اسناد و متون استفاده می‌کنیم تا خوشه‌های مفهومی بسازیم. این خوشه‌های مفهومی قرار است به داده‌های متن و سنت ساخت بدهند و از خلال همین ساخت‌یافتن است که معنا پدیدار می‌شود. اما مدل محاسباتی کامپیوتر هیچ چیز درباره‌ی گرامر، معناشناسی یا فونولوژی نمی‌داند. به همین دلیل در اینجا شاهد نوعی ابزار هرمنوتیکی هستیم که در معنای دقیق کلمه «کودن» است و به تعبیر یکی از پژوهشگران ما را به یاد «لوح سفید» جان لاک می‌اندازد. این

<sup>۱</sup> با افزایش حجم اسناد و اطلاعات و از طرفی نمایش آن‌ها به صورت دیجیتال در قالب اخبار، وبلاگ‌ها، مقالات علمی، کتاب‌های الکترونیکی، عکس، صوت و تصویر و شبکه‌های اجتماعی پیدا کردن مطالبی که به دنبال آن هستیم مشکل می‌شود. از این رو نیاز به ابزار محاسباتی جدیدی برای سازمان‌دهی، جستجو و درک این حجم عظیم اطلاعات هستیم. تا به حال کار ما روی اسناد آنلاین به دو طریق انجام می‌گرفت: یکی جستجو و دیگری پیوندها. کلمه‌ی کلیدی را در یک موتور جستجو وارد و مجموعه‌ای از اسناد مرتبط با آن را مشاهده می‌کردیم. در این اسناد اگر پیوند مفیدی به اسناد دیگر می‌دیدیم با کلیک روی آن به مجموعه اسناد دیگر دسترسی پیدا می‌کردیم. این دو روش‌های قدرتمندی برای کار کردن با آرشیوهای الکترونیکی بودند اما مشکلاتی نیز وجود داشت. روش جستجویی که در بالا شرح داده شد مبتنی بر ظاهر یک کلمه بود و هنگام انجام عمل جستجو تمامی اسنادی که حاوی آن کلمه‌ی کلیدی هستند به عنوان نتیجه برگردانده می‌شوند. این اسناد اگرچه دربردارنده‌ی آن کلمه هستند ولی ممکن است متعلق به موضوعات مختلفی باشند در حالی که ما احتمالاً فقط به دنبال یک موضوع خاص هستیم. برای اینکه جستجوی دقیق‌تری داشته باشیم بایستی ابتدا موضوع مورد علاقه‌ی خود را در بین اسناد جستجو کرده و سپس جستجوی خود را محدود به این اسناد جدید کنیم و می‌توانیم در این اسناد جدید دوباره موضوع خود را محدودتر کنیم و همین‌طور پیش برویم تا جایی که به طور دقیق به اسناد مورد نیاز خود دسترسی پیدا کنیم. به عنوان مثال فرض کنیم می‌خواهیم در آرشیو یک روزنامه به دنبال یک موضوع باشیم. موضوعات در دسته‌های سیاسی، اقتصادی، فرهنگی، ورزشی و حوادث قرار دارند. مثلاً می‌خواهیم در موضوع سیاسی جستجو کنیم. داخل این موضوع زیرموضوع‌های سیاست داخلی و خارجی وجود دارد که یکی را انتخاب می‌کنیم و همین‌طور پیش می‌رویم تا دقیقاً به اسناد مورد نیازمان دسترسی پیدا کنیم. البته این‌گونه کار کردن با اسناد به‌سادگی امکان‌پذیر نیست؛ زیرا هر چه حجم اسناد و اطلاعات افزایش می‌یابد دسته‌بندی فوق برای انسان کار مشکل و یا غیرممکن می‌شود؛ بنابراین نیاز به تکنیک‌های یادگیری ماشین داریم تا بتوانیم از طریق کامپیوتر دسته‌بندی فوق را انجام دهیم. پژوهشگران حوزه‌ی یادگیری ماشین برای این کار مجموعه‌ای از الگوریتم‌ها تحت عنوان مدل‌سازی موضوعی آماری را توسعه داده‌اند. الگوریتم‌های مدل‌سازی موضوعی روش‌های آماری هستند که کلمات داخل یک متن را تحلیل کرده و از این طریق موضوعات داخل متن را استخراج می‌کنند. هم‌چنین ارتباط این موضوعات با یکدیگر و نیز تغییر آن‌ها در طول زمان را مشخص می‌کنند. این الگوریتم‌ها نیازی به هیچ فرض اولیه‌ای در مورد موضوعات متن و یا برچسب‌گذاری متن ندارند. بلکه ورودی آن‌ها متن اصلی است. الگوریتم‌های مدل‌سازی موضوعی به ما این امکان را می‌دهند تا سازمان‌دهی و خلاصه‌سازی آرشیوهای الکترونیکی مان را در ابعادی که از عهده‌ی انسان بر نمی‌آید انجام دهیم.

هرمنوتیک بهای سنگین استفاده از تکنیک‌هایی است که البته مزایای بسیار زیادی دارند. در این تکنیک‌ها به سرعت و به صورت اتوماتیک پیکره‌های عظیمی از متون برحسب موضوعات اولیه رمزگذاری و نشان‌دار می‌شوند و با شکل‌گیری خوشه‌های مفهومی نظامی از دانش ظاهر می‌شود و چنین حجم و دقتی در تاریخ بشری بی‌سابقه بوده است.<sup>۱</sup>

## ۲-۳ استقراء/استنتاج

تحلیل در علوم انسانی و اجتماعی چگونه باید انجام شود؟ آیا در اینجا توازنی بین دو روش استقرایی و استنتاجی به وجود خواهد آمد؟ علوم اجتماعی در مکاتب کلاسیک به واسطه‌ی فقر داده‌ها معمولاً استنتاجی‌تر است. در این علوم فرض بر این است که انسان‌ها عملاً رونوشت یک نوع ایده‌آل (انسان معدل) هستند و با کمک روش‌شناسی آماری می‌توان برخی خطاها را خنثی کرد. به زبان ساده فرض بر این است که انسان‌ها عموماً مثل هم فکر و رفتار می‌کنند و با شروع از همین اصل می‌توان معارفی را درباره انسان و اجتماع استنتاج و استخراج کرد. البته خلاف‌آمدهایی هم وجود خواهد داشت که می‌توان با ابزارهای روش‌شناختی آنها را خنثی نمود. اما با ورود کلان‌داده‌ها رویکردی ظاهر می‌شود که تعلق بیشتری به استقراء دارد. در واقع، غنای داده‌ای این امکان را فراهم می‌آورد که استراتژی کاملاً متفاوتی در پژوهش اتخاذ کنیم. در این رویکرد جمعیت‌ها را می‌توان بلوک‌بندی کرد و به صورتی انعطاف‌پذیر آنها را کاوید. در اینجا می‌توان به جای یک پرسش نامتغیر برحسب انسان معدل پرسش‌های متفاوتی پرسید. در این رویکرد اساساً تصور دیگری از حوزه‌های اجتماعی و امر اجتماعی وجود دارد که جامعه را سیستمی پیچیده‌تر در نظر می‌گیرد. این روش بیشتر به نقشه‌برداری شباهت دارد و پستی و بلندی‌های جامعه را به خوبی منعکس می‌کند نه اینکه از دور از آن عکس‌برداری کند. در این رویکرد تصور می‌شود که نظریه و فرضیه نوعی نزدیک‌بینی ایجاد می‌کند و این نزدیک‌بینی تخیل جامعه‌شناختی را به قهقرا می‌برد و آن را فقیر می‌سازد. برخی مفسران این نزدیک‌بینی را نزدیک‌بینی مقولی یا مفهومی نامیده‌اند. حتی برخی از رویکردی شبیه پزشکی قانونی در علوم اجتماعی سخن می‌گویند. در این رویکرد قرار است از روی ردپاهای دیجیتالِ غیرساخت‌یافته شواهد به دقت و با احتیاط گردآوری شوند و بدین وسیله حتی نظریه‌های جدیدی ایجاد شود. به عبارتی در اینجا نظریه پس از داده به وجود خواهد آمد نه قبل از آن. برخی نیز گمان می‌کنند

<sup>۱</sup> Ibid, pp 3-4.

در رویکرد برآمده از کلان‌داده‌ها، روش نه کاملاً استقرایی است نه کاملاً استنتاجی بلکه Abductive است. این روش بین نظریه و اکتشاف استقرایی عقب و جلو و نوسان می‌کند. با ظهور کلان‌داده ما دیگر نیازی نداریم به اصطلاح با ایده‌ای درباره‌ی هویت قاتل به صحنه‌ی جنایت برویم. در اینجا باید فقط به این بیندیشیم که چگونه کوهی از داده‌ها را درون مقولات شناخت سامان‌مند کنیم و به آن‌ها ساختار بدهیم. ظاهراً با فرهنگ جدیدی در مدل‌سازی آماری مواجهیم.<sup>۱</sup>

کلان‌داده‌ها می‌توانند تصورات هستی‌شناختی علوم اجتماعی را نیز دگرگون کنند. ظهور کلان‌داده‌ها پرسش‌هایی درباره‌ی پنج مقوله‌ی هستی‌شناختی ایجاد می‌کند که باید به آن‌ها پرداخت و در اینجا فقط عناوین آن‌ها ذکر می‌شود. هر یک از این پرسش‌ها می‌توانند مسائلی هستی‌شناختی برای علوم اجتماعی طرح کنند و پاسخ به آن‌ها مستقیماً روی روش‌شناسی این علوم تأثیر خواهد گذاشت:

(۱) شیء چیست؟

(۲) عامل چیست؟

(۳) زمان چیست؟

(۴) بافت چیست؟

(۵) علیت چیست؟<sup>۲</sup>

### ۳- تغییر روش‌های آماری با ظهور کلان‌داده‌ها: نمونه‌برداری ایستا و دینامیک

ظهور کلان‌داده و مناقشات پیرامون آن به تغییراتی رادیکال در علوم انجامیده است. یکی از این تغییرات عوض شدن تلقی ما از پژوهش در علوم و به‌ویژه علوم انسانی و اجتماعی است. به عنوان مثال در سطوح معرفت‌شناختی و اخلاق علم تغییراتی بنیادی حاصل شده است. تغییرات حاصل شده و دگردیدی‌هایی که احتمال آن می‌رود به پرسش‌های جدیدی درباره‌ی تقویم شناخت، فرایند پژوهش، نحوه‌ی مواجهه ما با اطلاعات و حتی سرشت و مقوله‌بندی واقعیت منتهی شده است. در این تلقی جدید از سرشت واقعیت حوزه‌های جدیدی از ابژه‌ها، روش‌های

<sup>۱</sup> Ibid, pp 4-5.

<sup>۲</sup> Ibid, pp 5-7.



فهم و تعاریف حیات اجتماعی ظاهر شده است.<sup>۱</sup> کلان داده مثل بسیاری از مفاهیم نوظهور تعاریف متعددی دارد و به شیوه‌های مختلفی عملیاتی می‌شود. به عنوان مثال ممکن است بگوییم کلان داده عبارت است از مجموعه داده‌های بسیار بزرگی که نمی‌توان آن‌ها را در صفحه‌ی اکسل نمایش داد یا در یک ماشین منفرد ذخیره کرد، یا از حیث هستی‌شناختی چنان پیچیده‌اند که ابزار سنتی برای ذخیره و تحلیل آن‌ها کافی به نظر نمی‌رسد. یا ممکن است کسی کلان داده را صرفاً برحسب اندازه‌ی آن تعریف کند و آن را شامل چندین ترابایت یا پتابایت داده در نظر بگیرد. یا صرفاً آن‌ها را پرسرعت در نظر بگیرد یا از این حیث که سامان‌مند یا غیر سامانمند هستند آن‌ها را متنوع لحاظ کند. یا ممکن است فراگیر بودن آن مؤلفه‌ای برای تعریف کلان داده باشد. با نظر به همه‌ی موارد فوق به سادگی می‌توان حکم کرد که کلان داده‌ها صرفاً به حجم داده‌ها ارجاع ندارند بلکه در شکل‌گیری آن‌ها مؤلفه‌های دیگری نیز اهمیت دارند.<sup>۲</sup>

بخش‌های صنعتی، دولتی و دانشگاهی در طی مدتی طولانی حجم عظیمی از داده‌ها را تولید کرده‌اند، مثلاً در سرشماری‌های ملی، اما این داده‌ها با استفاده از تکنیک‌های نمونه‌گیری محدود به دامنه، زمان و اندازه، به شیوه‌هایی شدیداً کنترل‌شده تولید شده‌اند.<sup>۳</sup> برای گردآوری داده‌های قابل مدیریت آن‌ها را هر ۵ یا ۱۰ سال و با پرسیدن ۳۰ تا ۴۰ پرسش جمع‌آوری می‌کنند و خروجی‌های آن‌ها معمولاً از حیث وضوح و رزولوشن ظرافت چندانی ندارد. روش‌های بکار رفته برای ایجاد این داده‌ها نیز کاملاً غیر منعطف هستند مثلاً وقتی سرشماری انجام شد دیگر نمی‌توان پرسش‌هایی به آن افزود یا آن‌ها را تعدیل کرد. سرشماری و دیگر اشکال تولید داده‌ها عموماً جامعه‌ی اصلی را خلاصه می‌کنند و جمعیت را به نمونه‌هایی کوچک تقلیل می‌دهند. برخلاف آن، کلان داده به طور پیوسته تولید می‌شوند، فراگیرند و ریزدانه‌اند (دامنه‌ی خود را به بلوک‌های کوچکتری تقسیم می‌کنند)، و منعطف و قابل تعدیل‌اند. یک نمونه از چنین داده‌هایی عبارت است از تلویزیون مداربسته‌ی دیجیتال، ثبت خریدهای جزئی، وسایل دیجیتالی که تاریخ استفاده‌ی خودشان را ثبت و اعلام می‌کنند (مثل موبایل)، ثبت و نگارش تراکنش‌ها و تعاملات

<sup>1</sup> Boyd D and Crawford K (2012) Critical questions for big data. *Information, Communication and Society* 15(5): 662–679.

<sup>2</sup> Rob Kitchin, "Big Data, new epistemologies and paradigm shifts", pp. 1-2.

<sup>3</sup> Miller HJ (2010) The data avalanche is here. Shouldn't we be digging? *Journal of Regional Science* 50(1): 181–201.

در شبکه‌های اجتماعی (ای‌میل یا بانکداری آنلاین)، داده‌های مربوط به کلیک استریم<sup>۱</sup> (جریان کلیک) که نوبری از خلال یک وبسایت یا اپ را ثبت می‌کنند، اندازه‌گیری‌های سنسورهایی که درون یک ابژه یا محیط قرار داده شده‌اند و... همه‌ی این موارد و موارد مشابه در حال تولید سیلانی عظیم و دینامیک از داده‌های متنوع، ریزداده و مرتبط با هم هستند. فقط در نظر بگیرید که مدیران فیسبوک چگونه باید ۲,۵ میلیارد محتوا، ۲,۷ میلیارد لایک و ۳۰۰ میلیون عکس را در هر روز (سال ۲۰۱۲) پردازش کنند. تولید، تجزیه و تحلیل چنین داده‌هایی بسیار متفاوت از سرشماری‌هایی است که در هر ۵ سال انجام می‌شوند. به طور سنتی تکنیک‌های تحلیل داده به نحوی طراحی شده‌اند که از مجموعه داده‌های کمیاب، ایستا، بی‌نقص و با ارتباط اندک بصیرت‌هایی را استخراج کنند. این داده‌ها باید به نحو علمی نمونه‌گیری شده باشند و به پیش‌فرض‌های اکیدی (مثلاً استقلال، ایستایی و بهنجار بودن) وفادار باشند و بر اساس یک یا چند پرسش خاص در ذهن ایجاد و تحلیل شوند.<sup>۲</sup> چالش تحلیل کلان‌داده‌ها عبارت است از فراوانی، فراگیری، تنوع، بی‌زمانی، دینامیسم، درهم‌برهمی، عدم قطعیت و ارتباط بسیار زیاد داشتن داده‌ها با هم. در کنار این چالش به این نکته نیز باید توجه نمود که کلان‌داده‌ها برخلاف داده‌های آماری معمولاً به هیچ پرسش مشخصی در ذهن ربط ندارند و محصول جانبی فعالیت‌های دیگری هستند. اگرچه این دشواری‌ها تاکنون بسیار پیچیده و چالش‌برانگیز بوده‌اند اما با ظهور محاسبات بسیار قدرتمند و تکنیک‌های جدید تجزیه و تحلیل کم‌کم وارد ادبیات علمی می‌شوند. این تکنیک‌ها ریشه در پژوهش‌های مرتبط با هوش مصنوعی دارند و بیشتر با حوزه‌ی یادگیری ماشین مرتبط هستند که می‌تواند به صورت محاسباتی و اتوماتیک الگوها را بیابد و استخراج کند و در نهایت مدل‌هایی را نیز پیش‌بینی کند و خروجی‌ها را بهینه سازد. این تکنیک‌ها می‌توانند صدها الگوریتم مختلف را بر روی یک مجموعه داده‌ها به کار بگیرند تا بهترین تبیین و مدل ممکن را بیابند. چنان که مشخص است این تکنیک‌ها اساساً متفاوت از تکنیک‌های سنتی هستند و رویکرد معرفت‌شناختی کاملاً جدیدی برای معنادادن به جهان در اختیار ما قرار می‌دهند. در این رویکرد دیگر قرار نیست یک نظریه خاص را با تحلیل داده‌های مقتضی بسنجیم بلکه این رویکرد بصیرت‌هایی را کسب می‌کند که از دل خود داده‌ها زاده می‌شوند. در مدل‌های کلاسیک برای نمونه‌برداری آماری و تحلیل آن‌ها، ایستابودن و فقر داده‌ها موجب می‌شد تا از پیش با

<sup>1</sup> Clickstream

<sup>2</sup> Jeremy Mennis, "Generating Surface Models of Population Using Dasymetric Mapping?", in: *The Professional Geographer* 55.1 (2003), pp. 31–42.

فرضیات یا نظراتی به تحلیل آن‌ها بپردازیم. اما در کلان داده‌ها دینامیک، پویایی و وفور داده‌ها اساساً نیاز به فرضیه یا نظریه را منتفی نموده است. در اینجا به وضوح با یک انقلاب داده‌ای مواجه می‌شویم.<sup>۱</sup>

#### ۴- انقلاب داده‌ای و پیامدهای آن: تغییر پارادایم در علم

تولید بسیار زیاد کلان داده‌ها همراه با توسعه‌ی معرفت‌شناسی‌های جدید بسیاری را متقاعد کرده که یک انقلاب داده‌ای در راه است و این انقلاب پیامدهای فراگیری دارد از جمله اینکه نحوه تولید معرفت دگرگون خواهد شد، نحوه انجام کسب و کار تغییرات شدیدی را از سر خواهد گذراند و نحوه حکمرانی نیز اساساً دچار دگردیسی خواهد شد. کلان داده امکان یک پارادایم جدید برای پژوهش را فراهم می‌آورد که نه به یک رشته بلکه به چندین رشته علمی تعلق دارد.<sup>۲</sup> در واقع می‌توان مدل تغییر پارادایم کوهن را با محوریت داده‌ها بازخوانی و تعدیل کرد. در مدل کوهن تغییر پارادایم به این علت اتفاق می‌افتد که پارادایم غالب نمی‌تواند برخی پدیده‌های خاص را تبیین کند یا به برخی پرسش‌های کلیدی پاسخ دهد. اما جیم گری<sup>۳</sup> الگویی برای تغییر پارادایم عرضه کرده است<sup>۴</sup> که گذار از یک پارادایم به پارادایم دیگر بیش از هر چیز مبتنی بر توسعه در فرم‌های داده و پیشرفت در روش‌های تحلیلی جدید است. وی معتقد است که علم به طور کلی در حال حاضر وارد چهارمین پارادایم خود شده است و این پارادایم بر اساس دسترسی به کلان داده‌ها و ابزار تحلیلی جدید است. گری چهار پارادایم برای علم قائل است:<sup>۵</sup>

<sup>1</sup> Rob Kitchin, "Big Data, new epistemologies and paradigm shifts", pp. 1-4.

<sup>2</sup> Kitchin R (2014) The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences. London: Sage. Pp. 224-232.

این کتاب با نظر به جنبه‌های مختلف کلان داده‌ها تلاش می‌کند با بررسی نقش کلان داده‌ها در اقتصاد، حکمرانی، علوم انسانی و اجتماعی و سیاست به فهم انقلاب داده‌ای نزدیک شود.

<sup>3</sup> Jim Gray

<sup>4</sup> Hey T, Tansley S and Tolle K (2009) Jim Grey on eScience: A transformed scientific method. In: Hey T, Tansley S and Tolle K (eds) The Fourth Paradigm: Data-Intensive Scientific Discovery. Redmond: Microsoft Research, pp. xvii-xxxii.

به نقل از منبع زیر:

<sup>5</sup> Rob Kitchin, "Big Data, new epistemologies and paradigm shifts", p. 3.

پارادایم	ماهیت	فرم	زمان
۱	علم تجربی	تجربه‌گرایی؛ توصیف پدیده‌های طبیعی	پیشا-رسانس
۲	علم نظری	مدل‌سازی و تعمیم	پیشا-کامپیوتر
۳	علم محاسباتی	شبیه‌سازی پدیده‌های پیچیده	پیشا-کلان‌داده
۴	علم اکتشافی	پر-داده؛ اکتشاف آماری و داده‌کاوی	اکنون

وی مدعی است که در پارادایم چهارم معرفت‌شناسی اساساً متفاوتی در حال خلق شدن است اگرچه درخصوص فرم این معرفت‌شناسی مناقشه وجود دارد. البته برخلاف جیم‌گری دانشمندی وجود دارند که گمان می‌کنند پارادایم چهارم اساساً حوزه‌ی جدیدی از تجربه‌گرایی است و فقط حجم داده‌ها و تکنیک‌های تحلیل آن‌ها چنان است که به داده‌ها اجازه می‌دهد رها از نظریه برای خودشان سخن بگویند. نکته‌ی جالب آن است که دیدگاه تجربه‌گرا بیرون از آکادمی در محافل کسب و کار مورد استقبال واقع شده است اما به هر حال ایده‌هایش ریشه در قلمرو جدیدی از علوم داده‌ای دارد. در مقابل، در رشته‌های سنتی آکادمیک وضعیت جدیدی از علوم داده‌محور در حال ظهور است. به طور کلی این دو رویکرد دو نیروی محرکه‌ی متفاوت دارند که ریشه‌ی آن‌ها در قلمرو کسب و کار و آکادمی قرار دارد. اولی اساساً بیشتر به بازار و فرصت‌های تجاری می‌اندیشد تا توسعه‌ی معرفت و دومی به بهترین شیوه‌ی معنادار به جهان و توضیح فرآیندها و پدیده‌های جهان می‌اندیشد.<sup>۱</sup>

#### ۴-۱ پایان نظریه: باززایی تجربه‌گرایی

به طور خلاصه، برخی متفکران معتقدند کلان‌داده‌ها و ابزارهای تحلیلی ملازم با آن حوزه‌ی جدیدی از تولید معرفت را اعلام می‌کنند که مشخصه‌ی آن پایان نظریه است و به طور کلی سیل داده‌ها روش علمی را منسوخ می‌کند. فرض مهم این نگاه آن است که الگوها و روابط مشمول در کلان‌داده‌ها ذاتاً معرفتی معنادار و بینش‌آفرین درباره‌ی پدیده‌های پیچیده تولید می‌کند. به همین

<sup>۱</sup> Ibid.

دلیل این متفکران نتیجه می‌گیرند که کلان داده‌ها امکان نوعی تولید معرفت را فراهم می‌آورند که اساساً تجربه‌گرایانه است. از چشم‌انداز علم آمار و با صرف در نظر گرفتن همبستگی آماری می‌توان داده‌ها را بدون در نظر گرفتن فرضیه‌ای تحلیل کرد. اگر اعداد و مفاهیم را در بزرگترین خوشه‌های محاسباتی قرار دهیم و به الگوریتم‌های آماری اجازه دهیم تا الگوها را بیابند الگوهایی ظاهر می‌شوند که ماهواً از دسترس علوم نظری به دورند. در اینجا همبستگی آماری جای علیت را می‌گیرد و علم می‌تواند بدون مدل‌های همگن، نظریه‌های یکپارچه و حتی تبیین مکانیستی رشد کند. بدیهی است که در این پارادایم اساساً پیش‌بینی بر تبیین غلبه می‌کند و وظیفه‌ی علم دیگر تبیین امور نیست بلکه پیش‌بینی آنهاست. نکته‌ی دیگر حذف جنبه‌ی انسانی در تحلیل کلان داده‌ها است. ماشین‌های محاسبه و الگوریتم‌های آماری دیگر اجازه‌ی دخالت سوگیری‌های انسانی در تحلیل را نمی‌دهند. به عنوان مثال در علم کلاسیک برای تحلیل داده‌ها ناچار به تدوین پرسش‌هایی بودیم که اساساً با محوریت انسان تنظیم می‌شدند اما در پارادایم جدید دیگر نیازی به این پرسش‌ها نیست و کفایت از الگوریتم‌ها برای یافتن الگوها استفاده کنیم. نکته‌ی دیگر بافت‌زدایی از علم است. در تحلیل کلان داده‌ها به الگوهایی می‌رسیم که از بافت و دامنه‌های خاص معرفتی فراتر می‌روند. طرز عمل این حالت جدید از علم ماهواً استقرایی است. البته منتقدان نیز می‌گویند که هر داده‌ای چون مبتنی بر مدل‌گزینش داده است نمی‌تواند خنثی باشد و به همین دلیل سوگیری و لحن خاصی نسبت به جهان دارد. حتی در این زمینه برخی معتقدند که کلان داده‌ها از ناکجا نمی‌آیند. یعنی از نیروی تنظیم‌گر فلسفه‌رهایی ندارند. علاوه بر این، سیستم‌ها به گونه‌ای طراحی شده‌اند تا انواع خاصی از داده‌ها را ثبت کنند و الگوریتم‌ها و ابزار تحلیلی بکار رفته در تحلیل نهایی مبتنی بر استدلال علمی است و از خلال آزمون‌های علمی اصلاح و بهسازی می‌شوند. استقرایی بودن این علوم نیز نوعی استراتژی است که در خلاء علمی رخ نمی‌دهد و یافته‌ها و نظریه‌های قبلی به آنها چارچوب می‌بخشند. برخی نیز این پرسش را طرح می‌کنند که آیا اساساً داده‌های بی‌جان می‌توانند از سوی خودشان سخن بگویند یا همواره چارچوبی مفهومی را به دوش می‌کشند. به طور کلی برخی معتقدند که معنادادن به داده‌ها همواره در چارچوبی قبلی اتفاق می‌افتد و همواره زمینه‌مند است.<sup>۱</sup>

<sup>۱</sup> Ibid. pp. 3-5.

## ۵- علم داده-رهنمون: از علوم معرفت-محور به علوم داده-محور

برخلاف فرم‌های جدید تجربه‌گرایی، علم داده-رهنمون انگاره‌های روش علمی را حفظ می‌کند اما برای فهم پدیده‌ها گشودگی بیشتری نسبت به استفاده از ترکیب دورگه‌ی رویکردهای استقرایی و استدلالی دارد. این نوع علم با رویکرد سنتی که مبتنی بر آزمایش و استدلال است فرق دارد. در علم سنتی از دل نظریه فرضیه ایجاد می‌شود اما در علم داده-رهنمون فرضیات از دل داده‌ها ایجاد می‌شوند. ایجاد فرضیه در این نوع علم اساساً قبل از آن است که رویکرد استنتاجی را اعمال کنیم. به عبارتی، در این مدل داده‌ها به هر وسیله‌ی ممکن تولید نمی‌شوند یا با استفاده از هر نوع تکنولوژی و چارچوبی برای نمونه‌برداری در دسترس قرار نمی‌گیرند، بلکه انواع و اقسام استراتژی‌ها برای تصمیم‌گیری در کارند تا انواع خاصی از داده‌ها را برای ما به ارمغان بیاورند نه هر داده‌ای را. در واقع، فرض‌های ما به ما می‌گویند که از کدام تکنیک استفاده کنیم برای تمهید داده‌های معنادار، پردازش و مدیریت و تحلیل آن‌ها. داده‌ها در معرض هر نوع چارچوب‌دادن هستی‌شناختی محتمل قرار ندارند. از هر نوع تکنیک داده‌کاوی نیز نمی‌توان استفاده کرد تا حقایق پنهان در داده‌ها آشکار شوند. نهایتاً نوعی استدلال قیاسی (Abduction) وجود دارد که بر اساس آن تصمیم گرفته می‌شود که از چه روشی برای تولید داده استفاده کنیم. به عبارتی علوم داده‌ای نسخه‌ای بازسازی‌شده از روش علمی سنتی است که شیوه جدیدی برای ساخت نظریه فراهم می‌آورد.<sup>۱</sup> در اینجا برخلاف تجربه‌گرایی و پایان نظریه ادعا بر آن است که این علوم داده‌ای به پارادایم جدیدی برای روش علمی در عصر کلان‌داده‌ها تبدیل خواهند شد و معرفت‌شناسی موجود در این پارادایم اساساً بصیرت‌هایی به بار خواهد آورد که در دسترس علوم سنتی قرار نداشت. به عبارتی در اینجا ادعا بر آن است که علوم از معرفت-محوری به داده-محوری رسیده‌اند. علوم معرفت-محور با استفاده از رویکرد استدلالی و سراسر و در شرایطی که اصولاً داده‌ها در آن کمیابند و محاسبات بسیار ضعیف و کند است اندکی در فهم و تبیین جهان به ما کمک می‌کنند و اکنون که تکنولوژی و روش‌های داده‌ای بسیار پیشرفت نموده‌اند و شاهد وفور داده‌ها و سرعت محاسبات هستیم نوع جدیدی از علم وارد عرصه می‌شود.<sup>۲</sup>

<sup>۱</sup> Andrew Iliadis and Federica Russo, *Critical data studies: An introduction*, pp. 1-7.

یادداشت فوق علاوه بر بررسی علم و تجربه‌ی داده-رهنمون تلاش می‌کند به مفهوم حکمرانی داده-رهنمون نیز نزدیک شود و آن را در چارچوب انقلاب داده‌ای طرح کند.

<sup>۲</sup> Rob Kitchin, *“Big Data, new epistemologies and paradigm shifts”*, pp. 5-7.

## ۶- کلان داده‌ها و رویکرد پوزیتیویستی در علوم انسانی

نکته‌ی مهم در ادامه‌ی مباحث فوق آن است که اگرچه معرفت‌شناسی‌ها و تجربه‌گرایی کلان داده‌ها و علوم داده‌ای ظاهراً تغییرات زیادی در علوم طبیعی به بار می‌آورند اما تأثیر آن‌ها در علوم انسانی و اجتماعی محل مناقشه است. آنچه مسلم است، برای رویکرد پوزیتیویستی در علوم اجتماعی کلان داده‌ها فرصتی فراهم می‌آورند تا مدل‌هایی برای حیات انسان به دست دهند که پیچیده‌تر، ریزبافت‌تر و فراگیرتر هستند. مطالعات اجتماعی در شکل کلاسیک آن در شرایطی شکل گرفته است که اساساً در فقر داده‌ای به سر می‌برد و اصطلاحاً داده‌ها در آن کمیاب بود، اما کلان داده‌ها این شرایط را دگرگون کرده‌اند و در شرایط کنونی ما شاهد غنای داده‌ای هستیم و این می‌تواند علوم اجتماعی را به حیطه‌هایی وارد کند که تاکنون از دسترس خارج بوده است. از تصاویر لحظه‌ای تا تصاویر دینامیک، از گردآیه‌های درشت‌بافت تا رزولوشن‌های بسیار بالا، از مدل‌های نسبتاً ساده تا شبیه‌سازی‌های پیچیده‌تر همگی با ورود کلان داده‌ها و ابزارهای تحلیلی جدید ممکن شده‌اند و امکانی برای تکوین نوعی علوم اجتماعی محاسباتی فراهم آورده‌اند که از حیث عمق و دامنه سابقه نداشت.<sup>۱</sup> از طرفی این امکانات جدید به رویکرد پوزیتیویستی اجازه می‌دهد تا بر بسیاری از نقص‌های کلاسیک غلبه کند و فهم بهتری از جهان اجتماعی به دست دهد. بدین ترتیب برای دانشمندانی که رویکرد پسا-پوزیتیویستی دارند کلان داده‌ها هم فرصت به حساب می‌آیند و هم چالش. فرصت‌ها عبارتند از: تکثیر، دیجیتالی شدن و اتصال مجموعه‌ی متنوعی از داده‌های آنالوگ و ساخت‌نیافته و نیز منابع اطلاعاتی که بسیاری از آن‌ها تا پیش از این دسترس‌پذیر نبودند. دیجیتالی شدن منابع در علوم اجتماعی به تنهایی می‌تواند موجب تحولاتی معنادار شود. همین دیجیتالی شدن منابع به نوبه خود به ایجاد ابزار جدیدی برای مدیریت و دسترسی و تحلیل آن‌ها انجامیده است که می‌توانند سیل داده‌ها را برای کاوش علمی مدیریت کنند. اکنون می‌توان تعداد زیادی از منابع را در ارتباط با هم پژوهید و این امکان تا پیش از این فراهم نبود. این فرصت‌ها در حوزه‌ی نوظهوری مثل علوم انسانی دیجیتال (Digital Humanities) ارزیابی می‌شوند و نتایج این ارزیابی روز به روز در حال ازدیاد است.<sup>۲</sup>

<sup>۱</sup> Ibid, pp. 7-9.

<sup>۲</sup> Michael Ulliot, Digital Humanities Projects, Renaissance Quarterly, Vol. 66, No. 3 (Fall 2013), pp. 937-947

## ۷- علوم انسانی دیجیتال و از دست رفتن لحن و رتوریک انسانی

علوم انسانی دیجیتال در ابتدا عبارت بود از سازماندهی و تحلیل داده‌هایی که در جهان دیجیتال به دنیا آمده بودند و دیجیتال‌سازی و آرشیوکردن پروژه‌هایی که متون آنالوگ و ابژه‌های مادی را به فرم‌های دیجیتال تبدیل می‌کردند. ویژگی این منابع آن بود که می‌شد آن‌ها را به نوعی سازماندهی کرد و تحلیل فراگیر و اتومات آن‌ها ممکن می‌شد. متعاقب همین امر بود که دو نگاه متفاوت به وجود آمد. نخستین دسته گمان می‌کنند که تکنیک‌های جدید علوم انسانی دیجیتال (شمردن، نمودارسازی، نگاشتن و دور-خوانی)<sup>۱</sup> موجب دقت روش‌شناختی و عینیتی در رشته‌های علوم انسانی شده است که تا پیش از این به صورت سیستماتیک وجود نداشت. در مقابل، دسته‌ی دوم معتقدند که تکنیک‌های جدید به جای کنار گذاشتن روش‌های سنتی یا تمهید رویکردی تجربه‌گرا یا پوزیتیویستی به علوم انسانی، روش‌های موجود در علوم انسانی را تکمیل و تقویت می‌کنند و موجب تسهیل در فرم‌های سنتی تفسیر و نظریه‌سازی می‌شوند. این امر اخیر به مطالعات اجازه می‌دهد تا دامنه‌ی وسیعتری را پوشش دهند و به پرسش‌هایی پاسخ دهند که بدون محاسبات سریع و گسترده پاسخ دادن به آنها ممکن نبود. البته چالش‌هایی نیز در مقابل این علوم وجود دارد. برخی معتقدند که علوم انسانی دیجیتال با استفاده از کامپیوترها به عنوان ماشین‌های قرائت و توسل به دور-خوانی سطح مواجهه با پدیده‌ها را عوض کرده‌اند و دیگر مواجهه نزدیک (نزدیک-خوانی) وجود ندارد. نزدیک-خوانی یعنی توجه و تمرکز بر این نکته که معنا چگونه تولید و هدایت می‌شود و برای آرشیو کردن نیازمند چه نوع استراتژی ادبی یا رتوریکی است؛<sup>۲</sup> دور-خوانی اصولاً قادر به چنین تشخیصی نیست.<sup>۳</sup> برای این دسته از منتقدان از دست رفتن فرآیند خواندن و قرائت در علوم انسانی به شخصیت‌زدایی از آن‌ها می‌انجامد و آن‌ها را به ورطه‌ی بی‌معنایی و پوچی نزدیک می‌کند، چنانکه گویی دستی نامرئی و غیرشخصی در این علوم در کار است. به تعبیر این منتقدان نمی‌توان محصولات فرهنگی را داده‌های صرف در نظر گرفت و حذف کردن لحن شخصی و انسانی از آن‌ها می‌تواند آن‌ها را نابود کند. یک فقره

<sup>1</sup> S. Jänicke , G. Franzini , M. F. Cheema and G. Scheuermann, On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges, pp. 1-15.

<sup>۲</sup> البته ناگفته نماند که بر سر این نکته مناقشات فراوانی نیز وجود دارد. به عنوان مثال در کتاب زیر می‌توان برخی از پیشنهادات برای استراتژی قرائت رتوریک در علوم انسانی دیجیتال را یافت:

J I M R I D O L F O AND W I L L I A M H A R T - D A V I D S O N ( E D I T ) , *Rhetoric and the Digital Humanities*, The University of Chicago Press, 2015.

<sup>3</sup> Rob Kitchin, "Big Data, new epistemologies and paradigm shifts", p. 8.



نوشتاری فقط نظم حروف و کلمات نیست بلکه در بافتی حاضر می‌شود و حامل معانی و کیفیاتی است که به سادگی نمی‌توان آن‌ها را تصریح کرد.<sup>۱</sup> الگوریتم‌ها نمی‌توانند به سادگی بافت و معنا را به چنگ آورند و به سادگی ممکن است لحن و جنبه‌ی انسانی علوم انسانی را زایل کنند. تحلیل الگوریتمی رمان‌ها و روزنامه‌ها ضرورتاً از نظر این منتقدان تقلیل‌گرا است و تمایزات معنادار ادبی را نادیده می‌گیرد، ذائقه ادبی و انسانی را کور می‌کند و امکان پالایش اخلاقی به واسطه این علوم را نابود می‌سازد. مضاف بر اینکه این رویکرد اساساً تاریخ اقبال به آثار علوم انسانی از سوی خوانندگان را که بخش مهمی از معنای آن است نمی‌بیند.<sup>۲</sup> در یک کلام، این منتقدان گمان می‌کنند علوم انسانی دیجیتال منجر به تحلیل‌هایی ضعیف و سطحی می‌شود که دیگر بصیرت‌های عمیق و نافذ در اختیار ما قرار نخواهند داد. به نظر می‌رسد در اینجا نیز شاهد همان نوع نقدهایی هستیم که در تاریخ علوم انسانی بر رویکرد پوبتیویستی و کمی وارد شده است.

تردیدی وجود ندارد که توسعه کلان‌داده‌ها و ابزارهای جدید تحلیلی امکانی فراهم می‌آورد برای چارچوب‌بندی مجدد معرفت‌شناسی علوم، علوم اجتماعی و علوم انسانی و چنین چارچوب‌هایی به صورت فعال محدودیت‌های رشته‌های علمی را پشت سر می‌گذارد و تحلیل‌هایی در اختیار ما قرار می‌دهد که از خلال چند رشته عبور می‌کنند و دیگر محدود به تک رشته‌ها نمی‌مانند. ظهور کلان‌داده‌ها به شکل‌گیری پایگاه داده‌های عظیم، سیال، پویا و متنوعی انجامیده است که مسائل علمی را در سطح دیگری طرح می‌کنند. در این سطح جدید نیازمند ابزارهای مدیریت داده‌ها و تکنیک‌های داده‌کاوی جدیدی هستیم که بیشتر مبتنی بر یادگیری ماشین، هوش مصنوعی و بصری‌سازی هستند. ظهور کلان‌داده‌ها هم‌زمان است با ظهور پارادایمی از علوم که ریشه در وفور و انفجار داده‌ها دارند و رویکرد استنتاجی کلاسیک علمی را پشت سر می‌گذارند. اینکه کلان‌داده‌ها پس از تخریب علوم کلاسیک چه رویکردی خواهند داشت هنوز به نوبه خود جای کاوش دارد. دو مسیر بالقوه ممکن است ظاهر شوند که به دو معرفت‌شناسی متفاوت می‌انجامند: (۱) تجربه‌گرایی که در آن داده‌ها می‌توانند رها از نظریه از سوی خود سخن بگویند، (۲) علوم داده‌رهنمون که اساساً روش‌های موجود علمی را تعدیل می‌کنند و به جای علوم کلاسیک معرفت‌محور می‌نشینند. با در نظر گرفتن چارچوب‌های متنوع فلسفی وضعیت در علوم انسانی و اجتماعی کمی پیچیده‌تر است. به علت وضعیت فلسفی این علوم، روش‌های جدید هنوز

<sup>1</sup> Ibid.

<sup>2</sup> Marche S (2012) Literature is not data: Against digital humanities. Los Angeles Review of Books, 28 October 2012. Available at <https://lareviewofbooks.org/article.php?id=1040> (accessed 4 April 2013).

نمی‌توانند به طور کامل جای روش‌های کلاسیک را بگیرند. بخشی از این وضعیت معلول تفاوت ماهوی کلان‌داده‌ها در علوم طبیعی و انسانی نیز هست. تولید کلان‌داده در علوم انسانی و اجتماعی سخت‌تر است و نیازمند مطالعات بیشتر در ابژه‌ها و قلمروهای این علوم است. با این حال، کلان‌داده‌ها فرصت‌های زیادی برای علوم انسانی و اجتماعی فراهم آورده است که فقط به وفور و انفجار داده‌های اجتماعی، فرهنگی، سیاسی و تاریخی محدود نیست. اما چالش‌هایی نیز پیش روی این علوم قرار دارند. کمبود مهارت برای تحلیل و معنادادن به داده‌های فراوان و خلق رویکردی معرفت‌شناختی که فرم‌های پسا-پوزیتیویستی در اختیار علوم اجتماعی محاسباتی قرار می‌دهند می‌تواند این علوم را دچار دردسرهایی کند. به همین دلیل نیازمند بررسی بیشتر استلزامات معرفت‌شناختی و روش‌شناختی کلان‌داده‌ها در علوم انسانی و اجتماعی هستیم.<sup>۱</sup>

#### ۸- آرشیو و روایت: کلان‌داده و علوم اجتماعی تاریخی

کلان‌داده‌ها می‌توانند موجب انقلابی در علوم اجتماعی تاریخی شود به شرط آنکه از بافت‌هایی برآمده باشند که اساساً مهم هستند و معطوف به پاسخ دادن به پرسش‌هایی باشند که اهمیت حیاتی دارند. کلان‌داده‌ها می‌توانند نقش ویژه‌ای در یافتن پاسخ به پرسش‌های قبلی ما ایفا کنند به ویژه آن پرسش‌هایی که به زمان‌بندی، توالی و کرانه‌های این رویدادها مربوط‌اند و در علوم سنتی پرسش‌هایی رام‌نشده محسوب می‌شوند. برخی گمان می‌کنند که کلان‌داده‌ها برای دانشمندان و مورخان که تفسیرشان مبتنی بر جملات روایی است تفاوت چندانی ایجاد نمی‌کند و ممکن است در برآورد تأثیر کلان‌داده‌ها اغراق شده باشد.<sup>۲</sup> با گسترش فضای مجازی مورخان و عموم مردم دسترسی بی‌سابقه‌ای به مواد و مدارک تاریخی یافته‌اند. این نکته خود را در گسترش آرشیوها و بانک‌های اطلاعاتی نشان می‌دهد. تکنیک‌های جستجو در این آرشیوها کار مورخان را بسیار سهل نموده است. آیا آرشیوهای کلان‌داده‌ها فهم ما از گذشته و بنابراین اکنون را تغییر می‌دهند؟ آیا دیجیتالیزه شدن منابع قدیمی و دسترسی آنلاین به آن‌ها در حجم‌های بسیار بزرگ می‌تواند به پرسش‌های قدیمی ما پاسخ‌های جدیدی بدهد؟ آیا این مخازن دیجیتالی ممکن است تغییری در نحوه‌ی شکل‌گیری نخبگان، ادغام گروه‌های قومی یا حتی شیوع

<sup>۱</sup> Rob Kitchin, "Big Data, new epistemologies and paradigm shifts", p. 10.

<sup>۲</sup> Peter Bearman, *Big Data and historical social science*, p.1.

بیماری و تغییر آب و هوا داشته باشند؟ ما تازه در ابتدای انقلاب آرشیوی قرار داریم. آیا همه‌ی این موارد تأثیری بر روی کار مورخ و علوم اجتماعی دارد یا نه؟ پاسخ هم بله است و هم خیر.<sup>۱</sup> معروف است که «وب همه چیز را می‌بیند و هیچ چیز را فراموش نمی‌کند.»<sup>۲</sup> این شعار برخی محققان است اما این ربط چندانی به کار مورخ ندارد چون مورخ روایت می‌کند نه اینکه فقط شرح وقایع را بنویسد. به عنوان مثال در شرح وقایع ممکن است بگوییم که نیوتون در سال ۱۶۴۲ به دنیا آمد. اما مورخ می‌گوید در سال ۱۶۴۲ «پدر فیزیک مدرن» به دنیا آمد. در روایت مورخ این نیز در نظر گرفته می‌شود که سالها پس از تولد نیوتون فیزیک مدرن توسط او ایجاد می‌شود، یعنی روایت او از این رخداد در کلیت سرگذشت نیوتون اتفاق می‌افتد. در روایت مورخ نوعی در هم آمیزی زمان و مکان دیده می‌شود حال آنکه در وقایع‌نگاری خیر.<sup>۳</sup> درست است که وب همه چیز را می‌بیند و هیچ چیز را فراموش نمی‌کند اما این به معنی در هم آمیزی مکان و زمان و به تعبیری روایت کردن نیست. وب روایت نمی‌کند. از این چشم‌انداز ساختمان داده‌های دیجیتال و جدید کوچکترین اهمیتی ندارند.<sup>۴</sup> اما برای تاریخ علم اجتماعی که نوشتن چنین جملاتی اهمیت ندارد، این ساختمان داده‌ها ممکن است تخیل جدیدی از گذشته به ارمغان آورند.<sup>۵</sup> به عنوان مثال ممکن است کسی در ارزیابی ظهور جوامع مدرن برای این پدیده‌ها نقش محوری قائل شود: توسعه دولت مرکزی، ظهور جوامع ملی، توسعه جامعه بازاری، انحلال روابط اجتماعی سنتی (مبتنی بر شأن و مقام) و ظهور ساختاری اجتماعی مبتنی بر طبقات اجتماعی، ظهور روش‌های

<sup>۱</sup> Niels Brügger, *The Archived Web Doing History in the Digital Age*, The MIT Press, 2018, pp. 11-17.

<sup>۲</sup> Peter Bearman, *Big Data and historical social science*, p.2.

همچنین نگاه کنید به یادداشت جالب زیر در مجله‌ی نیویورک تایمز:

JEFFREY ROSEN, *The Web Means the End of Forgetting*,  
<https://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html>

<sup>۳</sup> Peter Bearman, *Big Data and historical social science*, p.2.

<sup>۴</sup> در خصوص نسبت علوم انسانی دیجیتال و روایت نگاه کنید به:

David M. Berry, *Understanding the Digital Humanities* in: David M. Berry, *Understanding Digital Humanities*, PALGRAVE MACMILLAN, 2012, pp. 1-21.

<sup>۵</sup> برخی پژوهشگران علوم داده‌رهنمون را به معنای عجز و درماندگی تخیل می‌دانند، نگاه کنید به:

Petr Keil, *Data-driven science is a failure of imagination*,  
<http://www.petrkeil.com/?p=302>

انطباقی برای کنترل و قدرت. می‌توان به این لیست چیزهای دیگری نیز افزود اما آنچه مهم است یافتن نسبتی علی بین آنها است. یافتن روابط علی و تقدم و تأخر زمانی بین این پدیده‌ها می‌تواند به پژوهش اجتماعی تاریخی ما جهت و سویه‌ای روش‌شناختی بدهد. مسئله در اینجا بر سر زمان‌بندی این پدیده‌ها و رویدادها است. تمرکز بر کلان‌داده‌ها و روش‌های اخذ شده از آن می‌تواند با کاوش در آرشیوهای تاریخی به نوعی در پی زمان‌بندی بین پدیده‌های فوق بگردد و ابزاری در دست مورخان قرار دهد. روش‌های برآمده از کلان‌داده‌ها می‌توانند دینامیک علی مستتر در بنیاد خوشه‌ی پیچیده‌ای از فرایندهای در هم تنیده را برای ما آشکار کنند.<sup>۱</sup> در اینجا کلان‌داده‌ها می‌توانند ابزاری مهم در فهم دنباله‌ی رویدادهای تاریخی باشد. از این حیث به دوره‌بندی محکمی در باب تاریخ دست می‌یابیم که به خودی خود فهم ما را از تاریخ دگرگون می‌کند. دسته‌ی کاملاً جدیدی از مسائل ظاهر می‌شود که فقط دسترسی به کلان‌داده‌ها می‌تواند آن‌ها را بیان و حل کند. یک نکته‌ی مهم در بحث کلان‌داده‌ها امکان انجام پژوهش‌هایی است که تا پیش از این یک پژوهشگر نمی‌توانست به تنهایی انجام دهد، یعنی یک پژوهشگر نمی‌توانست در طی عمرش به چنین پژوهشی دست بزند، اما تحول در آرشیو و ظهور کلان‌داده‌ها این امکان را برای پژوهشگر فراهم می‌آورد.

انقلاب در علم اغلب در پی انقلاب در اندازه‌گیری رخ می‌دهد. اگر بپذیریم که ظهور کلان‌داده‌ها متر و معیارهای جدیدی در سنجش پدیده‌های علوم اجتماعی با خود به ارمغان آورده است در این صورت باید اذعان کرد که در آستانه‌ی تحولی عظیم در این علوم قرار داریم. دسترسی به کلان‌داده‌ها به همراه رویکردهای جدید در تجزیه و تحلیل داده‌ها معرفت‌شناسی‌های موجود در علوم و علوم اجتماعی را به چالش می‌کشد و موجب جابجایی در پارادایم‌هایی می‌شود. در واقع کلان‌داده‌ها فرم‌هایی جدیدی از تجربه‌گرایی را به ظهور می‌رساند که با معرفی ابزارها و الگوریتم‌های جدیدی برای اندازه‌گیری و سنجش امور «پایان نظریه» را اعلام می‌کنند و به خلق علمی می‌انجامند که داده‌محورند نه شناخت‌محور. کلان‌داده‌ها عرصه را برای انجام محاسبات در علوم اجتماعی باز می‌کنند و همین محاسبات و الگوریتم‌ها نیاز به مفاهیم و نظریات در این علوم را مرتفع می‌سازند. کلان‌داده‌ها به توسعه‌ی علوم انسانی دیجیتال و علوم اجتماعی محاسباتی دامن می‌زنند و این علوم شیوه‌هایی اساساً متفاوت برای معنادادن به فرهنگ، تاریخ، اقتصاد و جامعه وضع می‌کنند. به طور کلی کلان‌داده‌ها و شیوه‌های جدید تجزیه و تحلیل داده‌ها ابداعاتی

<sup>۱</sup> Peter Bearman, *Big Data and historical social science*, pp. 2-3.

آشوبگر نیز به شمار می‌آیند که در حال شکل دادن به شیوه‌های جدید انجام پژوهش هستند. به همین دلیل به تأملاتی انتقادی نیازمندیم تا استلزامات معرفت‌شناختی این انقلاب داده‌ای را بکاوند. در واقع به نظر می‌رسد انقلاب داده‌ای و کلان داده‌ها به صورت بالقوه امکان توسعه‌ی نوعی از معرفت‌شناسی را در خود دارد که برخلاف معرفت‌شناسی‌های کلاسیک تأملی و بازتابی است و بسته به بافت دارای زیر و بم‌هایی است. در این معرفت‌شناسی‌های جدید اساساً تلقی ایستا و منجمد از معرفت جای خود را به نگاهی پویا<sup>۱</sup> و منعطف به داده‌ها می‌دهد که دیگر اسیر مفاهیم و نظریات صلب و از پیش تعیین شده نیست.

البته چنان که دیدیم، مشارکت کلان داده‌ها در علوم اجتماعی محدود به دسترس‌پذیری به داده‌ها نیست بلکه شامل پیشنهاد رویکردهای تحلیلی نیز می‌شود که در علوم کامپیوتری و یادگیری ماشین توسعه یافته‌اند و فرهنگ جدیدی برای مدل‌سازی آماری به همراه می‌آورند. این فرهنگ پتانسیل بسیار فراوانی برای علم اجتماعی به همراه می‌آورد.<sup>۲</sup> یادآوری این نکته اهمیت دارد که روش‌های محاسباتی، به ویژه در شکل شبیه‌سازی، در تعارض با توسعه‌ی نظریه‌ها قرار ندارند بلکه مجال برای بهبود و اصلاح آن‌ها فراهم می‌کنند و آن‌ها را از شکل دلخواهی و پیشینی بودن خارج می‌کنند. اشتباه است اگر گمان کنیم کلان داده‌ها قرار است نسل نظریه را در علوم اجتماعی منقرض کنند بلکه باید تصدیق کنیم که کلان داده‌ها تصور ما از نظریه و فرضیه در علوم را اصلاح و تعدیل می‌کنند. نظریه در اینجا دیگر چشم‌بندی برای ندیدن برخی وجوه رخدادهای اجتماعی نیست بلکه اتفاقاً زوایا و جنبه‌های نهان و در سایه قرار گرفته‌ی این پدیده‌ها را جلو چشم ما قرار می‌دهد. نظریه در شکل کلاسیک آن فقط یک چشم‌انداز به پدیده‌ها است. به نظر می‌رسد اساساً ادعای کلان داده‌ها فراتر رفتن از این چشم‌انداززدگی در علوم اجتماعی است. البته باید نگاهی انتقادی به این ادعا داشته باشیم و بسنجیم که آیا در واقع کلان داده‌ها می‌توانند تصویری پانوراما از رویدادها و پدیده‌های اجتماعی در اختیار دانشمندان قرار دهند یا خیر.

<sup>1</sup> David Beer, How should we do the history of Big Data?

<sup>2</sup> Giuseppe Alessandro Veltri, Big Data is not only about data: The two cultures of modeling, pp. 1-5.

## ۹- نتیجه‌گیری: فرآیند به جای ساختار

عطش برای نوآوری روش‌شناختی در علوم اجتماعی که احتمالاً به جار و جنجال کنونی پیرامون «انقلاب داده‌ای» انجامیده است، می‌تواند معلول بحران دیرپایی باشد که مشخصه‌ی اغلب تکنیک‌های موجود و شایع در علوم اجتماعی است. روش‌شناسی پشتیبان در رشته‌های بسیار مختلف از بحرانی دیرپا رنج می‌برند، بحرانی که معلول دشواری‌های روزافزون در واکنش به سنجش‌ها، چارچوب‌های نمونه‌گیری و ظرفیت محدود در کنترل متغیرهایی است که در اطلاعات جامعه‌شناختی روزبه‌روز مهم‌تر می‌شوند، به عنوان مثال موقعیت جغرافیایی دقیق. درباره مصاحبه عمقی نیز ملاحظات مشابهی وجود دارد؛ مصاحبه عمقی ابزار مهمی در جمع‌آوری داده‌ها در علوم اجتماعی است. یک انتقاد مربوط است به ترجمه‌ی تکنیک‌های توسعه یافته پیش از ظهور رسانه‌های دیجیتال و مسائل مربوط به استلزامات مصاحبه‌های انجام شده از طریق ارتباطاتی که به واسطه کامپیوتر انجام شده‌اند. درحالی‌که اغلب بر روی سوگیری مصاحبه‌کننده تمرکز شده است، چیزی درباره افزایش ادبیات علمی درباره سوگیری انسانی در یادآوری خاطره، حساسیت به عناصر بافتی در فعال‌سازی هیوریستیک‌ها و فرآیند عقلانی‌سازی وجود ندارد. این عوامل اخیر مصاحبه‌شونده را تحت تأثیر قرار می‌دهند. برآوردها و مصاحبه‌هایی که انگیزه‌ها و رفتار انسان را اندازه می‌گیرند شدیداً ذیل موشکافی قرار دارند و با منابع داده‌ای ارگانیک‌تر مقایسه می‌شوند. این بدان معنا نیست که کلان‌داده‌ها هیچگونه نگرانی اساسی ایجاد نمی‌کنند. مناقشه‌ی کنونی بر سر نوعی علم اجتماعی انتقادی است که باید همراه هر روشی باشد که دانشمندان اجتماعی استفاده می‌کنند. شاید به دلایل تاریخی، ماهیت مصنوعی روش‌های سنتی مدتهای مدید تا همین اواخر نادیده گرفته شده باشد، اما در تحلیل نهایی ظرفیت‌شان برای تولید داده‌های کیفی به شدت پروبلماتیک شده است. اگر دو جنبه‌ی دیگر را نیز در نظر بگیریم چنین محدودیت‌هایی حتی روشن‌تر می‌شوند: اولاً بخش اعظم داده‌های علوم اجتماعی که از پیمایش‌ها و مصاحبه‌ها به دست می‌آید مقطعی هستند و بعد زمانی طولی ندارند. ثانیاً اغلب مجموعه داده‌های علوم اجتماعی به دلیل محدودیت‌های تاریخی و تکنیکی انبوهه‌های زمختی از متغیرها هستند که در بسیاری از موارد نه از رزولوشن بالا برخوردارند و نه ریزبافت هستند.<sup>۱</sup> کلان‌داده‌ها از هر دو حیث فوق به نوآوری نیرو و جهت می‌دهند: حرکت از تصاویر لحظه‌ای ایستا به سمت سیلانی دینامیک و حرکت از توده‌های زمخت به سمت وضوح بالا در داده‌ها. این نوآوری‌ها در علوم اجتماعی

<sup>1</sup> Ibid. p. 2.

محصول فرعی جالبی دارند که می‌توان آن را «امکان تمرکز بیشتر بر فرآیندها به جای ساختارها»<sup>۱</sup> نامید. برای نخستین بار می‌توانیم هنجارهایی برای مبنای طولی، واریانس و رفتار دوری به دست بیاوریم. این امر نیازمند آن است که فراسوی علیت ساده‌ی روش‌های علمی سنتی بیندیشیم و وارد مدل‌های سیستمیک همبستگی آماری و پیوند آماری شویم. تحلیل شبکه نمونه‌ی خوبی است: دسترس‌پذیری به داده‌های عقلانی طولی در نوآوری‌های روش‌شناختی و نظری درباره‌ی دینامیک شبکه‌ها غوغایی به راه انداخته است. اما تمرکز بر فرآیندها به جای ساختارها همزمان خوف و رجای کلان‌داده‌ها در علوم اجتماعی نیز می‌تواند باشد. از طرفی می‌توان برخی محدودیت‌های روش‌شناختی در علوم کلاسیک را پشت سر گذاشت و از طرفی ممکن است تعریف ما از علم به طور کلی دچار دگرگونی شود. بخش مهمی از پژوهش در تأثیر کلان‌داده‌ها بر روی علوم به طور کلی و علوم اجتماعی باید معطوف به «بازتعریف» علم در شرایط کنونی باشد.

#### ۱۰- منابع

Seref Sagiroglu and Duygu Sinanc, “**Big data: A review**”, in: *Collaboration Technologies and Systems (CTS), 2013 International Conference on*, IEEE, 2013.

Rob Kitchin and Gavin McArdle, **What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets.**

Elena Aronova, Karen S Baker, and Naomi Oreskes, “**Big science and big data in biology: from the international geophysical year through the international biological program to the long term ecological research (LTER) network, 1957–present**”, in: (2010).

Liran Einav and Jonathan D Levin, *The data revolution and economic analysis*, tech. rep., National Bureau of Economic Research, 2013.

David Lazer, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler,

<sup>۱</sup> Ibid.

Myron Gutmann, et al., “**Life in the network: the coming age of computational social science**”, in: *Science (New York, NY)* 323.5915 (2009).

Claudio Cioffi-Revilla, “**Computational social science**”, in: *Wiley Interdisciplinary Reviews: Computational Statistics* 2.3 (2010).

Carlo Ratti, S Williams, D Frenchman, and RM Pulselli, “**Mobile landscapes: using location data from cell phones for urban analysis**”, in: *Environment and Planning B Planning and Design* 33.5 (2006).

Sasank Reddy, Min Mun, Jeff Burke, Deborah Estrin, Mark Hansen, and Mani Srivastava, “**Using mobile phones to determine transportation modes**”, in: *ACM Transactions on Sensor Networks (TOSN)* 6.2 (2010).

Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza, “**On the use of human mobility proxies for modeling epidemics**”, in: *PLoS computational biology* 10.7 (2014), e1003716.

Markus Schläpfer, Luís MA Bettencourt, Sébastien Grauwin, Mathias Raschke, Rob Claxton, Zbigniew Smoreda, Geoffrey B West, and Carlo Ratti, “**The scaling of human interactions with city size**”, in: *Journal of The Royal Society Interface* 11.98 (2014), p. 20130789.

DL Balk, U Deichmann, G Yetman, F Pozzi, SI Hay, and A Nelson, “**Determining global population distribution: methods, applications and data**”, in: *Advances in parasitology* 62 (2006).

Budhendra Bhaduri, Edward Bright, Phillip Coleman, and Marie L Urban, “**LandScan USA: a high-resolution geospatial and temporal modeling approach for population distribution and dynamics**”, in: *GeoJournal* 69.1-2 (2007).

R.M. Chang, et al., **Understanding the paradigmshift to computational social science in the presence of big data**, *Decision Support Systems* (2013), p.1.  
<http://dx.doi.org/10.1016/j.dss.2013.08.008>



Prabhakar Raghavan, **It's time to scale the science in the social sciences.**

Robin Wagner-Pacifici, John W Mohr and Ronald L Breiger, **Ontologies, methodologies, and new uses of Big Data in the social and cultural sciences.**

Boyd D and Crawford K (2012) **Critical questions for big data.** *Information, Communication and Society* 15(5): 662–679.

Rob Kitchin, **“Big Data, new epistemologies and paradigm shifts”.**

Miller HJ (2010) **The data avalanche is here. Shouldn't we be digging?** *Journal of Regional Science* 50(1): 181–201.

Jeremy Mennis, **“Generating Surface Models of Population Using Dasymetric Mapping?”**, in: *The Professional Geographer* 55.1 (2003), pp. 31–42.

Kitchin R (2014) **The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences.** London: Sage. Pp. 224-232.

Hey T, Tansley S and Tolle K (2009) **Jim Grey on eScience: A transformed scientific method.** In: Hey T, Tansley S and Tolle K (eds) *The Fourth Paradigm: Data-Intensive Scientific Discovery.* Redmond: Microsoft Research, pp. xvii–xxxii.

Andrew Iliadis and Federica Russo, **Critical data studies: An introduction**

Michael Ulliyot, **Digital Humanities Projects,** *Renaissance Quarterly*, Vol. 66, No. 3 (Fall 2013), pp. 937-947

S. Jänicke , G. Franzini , M. F. Cheema and G. Scheuermann, **On Close and Distant Reading in Digital Humanities: A Survey and Future Challenges,** pp. 1-15.

J I M R I D O L F O AND W I L L I A M H A R T - D A V I D S O N (Edit), ***Rhetoric and the Digital Humanities,*** The University of Chicago Press, 2015.

Marche S (2012) **Literature is not data: Against digital humanities.** Los Angeles Review of Books, 28 October 2012. Available at <https://lareviewofbooks.org/article.php?id=1040>

Niels Brügger, *The Archived Web Doing History in the Digital Age*, The MIT Press, 2018, pp. 11-17.

Jeffrey Rosen, **The Web Means the End of Forgetting**,  
<https://www.nytimes.com/2010/07/25/magazine/25privacy-t2.html>

Peter Bearman, *Big Data and historical social science*.

David M. Berry, *Understanding the Digital Humanities in: David M. Berry, Understanding Digital Humanities*, PALGRAVE MACMILLAN, 2012, pp. 1-21.

Petr Keil, **Data-driven science is a failure of imagination**,  
<http://www.petrkeil.com/?p=302>

Giuseppe Alessandro Veltri, **Big Data is not only about data: The two cultures of modeling**.

## کلان داده و آینده علوم انسانی-اجتماعی

حوزه فضای مجازی به اندازه انقلاب اسلامی اهمیت دارد. این فضا مثل یک رودخانه پر از آب و خروشان است که می آید و دائماً هم بر آب آن افزوده و خروشان تر می شود. اگر ما بر این رودخانه تدبیر کنیم و برنامه داشته باشیم، زه کشی کنیم و هدایت کنیم، این رودخانه را تا به سد بریزد، می شود فرصت. اگر رهاپیش کنیم و برنامه ای برای آن نداشته باشیم می شود یک تهدید.

عماد  
۱۳۹۱/۷/۲

