

The Multi-Layered and Transformative Impact of Big Data, Artificial Intelligence, and the Internet of Behaviors on the Social Sciences

Bekir EMİROĞLU*

Abstract

This study examines the multi-layered and transformative impact of big data, artificial intelligence (AI), and the Internet of Behaviors (IoB) on the social sciences. It argues that these technologies represent not merely new methodological tools, but a fundamental paradigm shift that challenges the field's core epistemological assumptions, ethical norms, and societal role. The analysis progresses through four key areas. First, it addresses the epistemological rupture, focusing on the "End of Theory" debate and the tension between correlation and causality in data-driven science. Second, it evaluates the methodological transformation, contrasting new research opportunities like predictive modeling and social simulation with significant risks, including algorithmic bias and the 'black-box' problem. Third, the study explores the profound ethical and political dimensions, such as surveillance, algorithmic authority, and power asymmetries. Finally, it analyzes the transition from the Internet of Things (IoT) to IoB, highlighting the shift from passive observation to active behavioral intervention, which poses significant threats to individual autonomy. The paper concludes that social scientists must adopt a critical role, moving beyond technical application to question the underlying power structures of these technologies and guide the development of a more just and democratic digital future.

Keywords: *Algorithmic authority; artificial intelligence (AI); big data; internet of behaviors (IoB); paradigm shift; social sciences.*

1. Introduction

This study argues that big data, artificial intelligence (AI), and the Internet of Behaviors (IoB) not only present new methodological tools for the social sciences but also create a paradigm shift that fundamentally questions the field's basic epistemological assumptions, ethical norms, and societal role. The main argument is built upon the necessity of establishing a critical balance between the immense opportunities offered by this digital transformation and the serious risks it brings, and the obligation for social scientists to assume a critical role in this new ecosystem [1,2].

Big data, the cornerstone of the digital age, refers to massive datasets composed of the digital traces of billions of people; while artificial intelligence defines systems that learn from this data to perform human-like cognitive tasks [3]. Machine learning (ML), as a sub-field of artificial intelligence, enables algorithms to learn patterns from data. The integration of these technologies into the physical world is realized through the Internet of Things (IoT); continuous data is collected via sensors and smart devices [4]. The Internet of Behaviors (IoB), the next step in this process, uses this collected data to understand and even influence human behavior. This section, centered around this conceptual framework, will analyze the deep and multi-layered impact of the digital age on the social sciences by proceeding in four main parts, from epistemological impacts to methodological transformation, and from societal and political reflections to the rise of IoB [5].

1.1. Methodology

This study is designed as a conceptual and theoretical paper. The methodological framework relies on a critical literature synthesis, examining the intersection of big data, AI, and IoB through a multidisciplinary lens that encompasses sociology, philosophy of technology, and political science. Rather than relying on empirical fieldwork, the study utilizes a critical theory approach to deconstruct the epistemological shifts and power dynamics introduced by digital technologies. The analysis is conducted by synthesizing key theoretical debates—such as the "End of Theory" and "Algorithmic Authority"—to develop a holistic understanding of how these technologies are reshaping the ontology of social sciences.

*Corresponding author

Bekir EMİROĞLU; Izmir Katip Celebi University, Izmir Katip Celebi University, Faculty Of Theology, Türkiye; e-mail: bekir.emiroglu@ikc.edu.tr
 0000-0002-3395-5722

2. Redefining Knowledge: The Epistemological Impacts of Big Data

2.1. The transition from traditional epistemology to computational thinking

Epistemology, in its most basic sense, is the branch of philosophy concerned with the nature, source, and limits of knowledge. Traditional epistemology, pioneered in Western philosophy by thinkers such as René Descartes and John Locke, was largely seen as an individual endeavor. According to this approach, knowledge is a justified true belief that the individual obtains through their own reasoning and sensory experiences, abstracted from their social environment. The social sciences, too, were long shaped within this individual-centric and theory-oriented understanding of knowledge. However, the emergence of big data is challenging these deep-rooted epistemological assumptions [6].

Big data has endowed the social sciences with capabilities such as “quantitativeness” and the “ability to predict” events in advance, which are considered fundamental characteristics of the natural sciences. This situation has strengthened the status of social and humanistic research as “science,” but it has also transformed the definition of what knowledge is [7]. Knowledge is no longer just a phenomenon obtained through the understanding of causal processes, but also an outcome produced by the discovery of patterns in massive datasets. This new epistemological framework is called “data-driven science”. Data-driven science aims to generate new scientific hypotheses by uncovering correlations and patterns in large data pools, rather than testing predetermined hypotheses. This approach fundamentally changes the process of knowledge production, reframing the epistemology of the social sciences [8].

2.2. “The End of Theory” debate: correlation versus causality

The most radical expression of the epistemological rupture brought by big data was put forth in the 2008 article titled “The End of Theory,” written by Chris Anderson, the then-editor-in-chief of *Wired* magazine. Anderson claimed that in the age of petabytes of data, theories developed to explain human behavior, such as sociology, psychology, or linguistics, are no longer needed. According to him, when a sufficient amount of data is available, “the numbers speak for themselves,” and identifying statistical correlations between data is more valuable and sufficient than understanding the underlying causal mechanisms of phenomena. According to this provocative thesis, “correlation replaces causality,” and science can progress without needing consistent models or unified theories [9].

However, this view was met with serious criticism. Philosopher Byung-Chul Han argues that there is no such thing as “data-driven thought,” only “computation.” According to Han, massive piles of information devoid of theory and thought produce only “noise” instead of knowledge. Science aims to explain not only what happens, but why it happens; therefore, abandoning the search for causality means renouncing the fundamental mission of science. It is a classic statistical principle that correlation does not imply causation. For example, a strong correlation between ice cream sales and drowning incidents does not mean one causes the other; there is a third variable (hot weather) underlying both [10].

Today, this debate is evolving toward a more balanced synthesis. It is accepted that big data approaches yield the most productive results when integrated with existing theories, rather than eliminating theory entirely. Big data offers the possibility of testing existing theories on an unprecedented scale or developing new theories based on previously unnoticed patterns. Therefore, it is more accurate to speak of the beginning of a more dynamic and mutually nourishing relationship between theory and data, rather than “the end of theory” [11].

This debate is, in fact, a new manifestation in the digital age of the social sciences' long-standing “quantitative-qualitative” divide. Anderson’s position represents an extreme positivist and quantitative stance, while the counter-arguments reflect the importance that the interpretative tradition places on “understanding” and “explanation.” The proliferation of the “correlation is sufficient” paradigm carries the potential to transform the field's intellectual identity through a chain reaction. In the first step, this approach relegates the effort to understand the structural causes underlying social phenomena, such as inequality or power relations, to a secondary plane. This situation weakens the critical potential of the social sciences; because it promotes a “technocratic social science” understanding that focuses on predicting and managing symptoms rather than delving into the root of problems. Ultimately, the social sciences face the risk of transforming from an intellectual force that encourages social change into a “social engineering” tool used to manage the existing order more efficiently. This points to a profound shift in meaning regarding the field's political and societal role [12].

3. Digital Transformation Tools and Social Science Research

3.1. New research horizons: application areas of artificial intelligence, analytics, and machine learning

Big data analytics, artificial intelligence, and machine learning offer research possibilities previously unimaginable for the social sciences. These tools make it possible to extract meaningful conclusions from massive and unstructured datasets, enriching our ways of understanding the social world [13, 14].

Text and Sentiment Analysis: Billions of texts obtained from sources like social media platforms, news sites, blogs, and political discourse can be analyzed using AI-powered natural language processing (NLP) techniques. In this way, public opinion trends on specific topics, emotional tone (sentiment analysis), and dominant ideological frameworks become quantitatively measurable [15, 16].

Prediction and Classification: Machine learning models, especially supervised learning algorithms, are powerful tools for predicting and classifying social phenomena. For example, election results can be predicted using past election data and demographic information, crime rates in specific areas can be foreseen by analyzing crime data, or purchasing behaviors can be modeled with consumer data [17, 18].

Social Simulations: Large language models (LLMs) like GPT-4 are breaking new ground in social science research. These models can simulate human subjects with different demographic characteristics in survey experiments, thus giving researchers the opportunity to test their hypotheses and conduct pilot studies at a low cost. This adds an innovative stage to the research design process [19, 20].

Political Science and Public Policy: These fields are among the social science disciplines where data science is most intensively used. Applications such as targeting voters in political campaigns, modeling election results, analyzing public opinion dynamics on social media, and shaping public policy decisions based on data have become widespread [21, 22].

3.2. The researcher's dilemma: opportunities and risks

These new digital tools, alongside the opportunities they offer, also bring significant risks and methodological challenges.

Opportunities: The greatest advantage is the potential to uncover complex, multi-dimensional, and non-linear relationships within large datasets that the human mind or traditional statistical methods cannot detect. The automation of analysis processes saves researchers time and offers the opportunity to investigate hypotheses that were previously impossible to test [23].

Risk 1: Algorithmic Bias: Machine learning models unconsciously learn and reproduce existing societal prejudices (racial, gender, class-based, etc.) present in the data they are trained on. For example, a “predictive policing” algorithm trained on police records historically focused on certain neighborhoods may reinforce existing inequalities by unfairly labeling these neighborhoods as more “risky.” This situation fundamentally undermines assumptions about the objectivity of algorithms [24].

Risk 2: “Black-Box” Problem and Interpretability: Complex neural network models, especially deep learning, struggle to explain how they arrived at their conclusions, even though they can make highly accurate predictions. This situation, where the model's inner workings are not transparent, referred to as the “black-box,” makes it difficult to scientifically validate, theoretically frame, and legitimize the results [25].

Risk 3: Data Privacy and Ethics: The data used in social science research often contains personal and sensitive information about individuals. Protecting the privacy of individuals, anonymizing data, and obtaining informed consent during the processes of collecting, storing, and analyzing this data are critical ethical obligations [26].

The “Opportunities” and “risks” brought by the use of artificial intelligence and machine learning in the social sciences actually represent two different faces of the same technical and philosophical problem. A model's high “predictive power” (opportunity) is generally achieved by increasing the model's “complexity.” This complexity,

in turn, leads to a decrease in the model's "interpretability" (risk). Therefore, social scientists must constantly strike a balance between "accuracy" and "interpretability." This is not just a technical choice, but also reflects an epistemological stance on the nature of knowledge and the purpose of science.

The chain effects of this situation may have significant consequences for the future of the field. As social scientists increasingly adopt easy-to-use but "black-box" AI tools, the research process may focus on technical steps like data collection and model selection. This shift in focus can lead to a weakening of the ability to understand and interpret the mechanisms underlying the results. When the researcher has difficulty explaining "why" the results produced by the model are correct, the "scientific authority" of social science research faces the risk of erosion. If researchers cannot place their own findings within a theoretical and causal framework, the validity and reliability of the "knowledge" they produce become questionable. In this scenario, social science faces the danger of being reduced to a "data reporting" discipline that merely reports complex patterns.

The table below summarizes the fundamental differences between traditional and big data-based research paradigms, revealing the dimensions of this transformation more clearly [27].

Table 1. *Comparison of Research Paradigms in Social Sciences.*

Criterion	Traditional Research Methods	Big Data And Ai-Assisted Methods
Data Source	Surveys, Interviews, Observation, Archival Documents	Digital Traces, Sensor Data, Social Media, Administrative Records
Data Scale	Small And Medium-Scale, Sample-Based	Large-Scale, Often The Entire Population (N=All)
Methodology	Hypothesis-Testing (Deduction)	Pattern Discovery, Correlation Analysis (Induction)
Primary Goal	Causal Explanation, Understanding (Verstehen)	Prediction, Classification, Behavioral Foresight
Analysis Tool	Statistical Software (Spss, Stata)	Machine Learning Algorithms, Cloud Computing Platforms (Hadoop, Spark)
Key Challenges	Generalizability, Representation Issues, Researcher Bias	Data Quality, Algorithmic Bias, Privacy, "Black-Box" Problem

4. The Societal Dimensions of Data: Philosophical, Ethical, and Political Reflections

4.1. The ethics of the algorithm: privacy, property, and justice

The proliferation of big data and artificial intelligence technologies is creating new problem areas that profoundly shake society's fundamental ethical norms and understanding of justice [2, 28].

Violation of Privacy: Artificial intelligence applications and big data systems collect massive amounts of personal data by tracking users' every step in the digital environment. This data is often processed and used without the users' explicit and informed consent. This situation leads to a fundamental erosion of the individual's right to privacy as an autonomous sphere. The collection and analysis of highly sensitive information, such as health data, further heighten these concerns [29].

The Problem of Data Ownership: The question of who owns the data individuals produce as a result of their digital activities is one of the most complex legal and philosophical problems of the 21st century. Does the ownership of the data lie with the individual, the technology platform that collects the data, or the state that regulates this data? Although regulations like the European Union's General Data Protection Regulation (GDPR) try to strengthen data ownership in favor of the individual, the uncertainties and power struggles on this issue continue [26].

Algorithmic Discrimination and Justice: Perhaps the most serious ethical problem is the potential for algorithms to reinforce existing social inequalities. Algorithms learn the prejudices (race, gender, socio-economic status, etc.) present in the historical datasets they are trained on. Consequently, algorithms that evaluate a credit application, filter candidates in a recruitment process, or calculate a defendant's risk of re-offending can systematically discriminate against certain demographic groups. This "algorithmic bias" damages the principle of impartial justice and increases the pressure on marginalized groups [30, 31].

4.2. Data as power: surveillance, power asymmetries, and political impacts

Big data is not just a technology, but also a new form of power. Whoever controls the data also holds the power to shape society.

Surveillance Society: States and large technology companies have obtained an unprecedented capacity for surveillance over citizens and users by using big data analytics. This situation creates a “surveillance society” where individuals are constantly monitored and their behaviors recorded. This surveillance dramatically shifts the balance of power against individuals and in favor of institutions [32].

Anticipatory Governance: States are increasingly using data analytics to predict events such as crime, terrorism, or social unrest before they occur and to make preventive interventions. Practices like “predictive policing” can lead to certain neighborhoods or demographic groups being unfairly labeled as “risky” and subjected to more intense police scrutiny. This situation threatens the presumption of innocence and social justice [33].

Information Asymmetry and Power: The data economy operates on “winner-take-all” dynamics. The dominance of the data market by a few technology giants like Google, Meta, and Amazon creates a massive asymmetry in access to and processing of information. This “big data divide” gives these companies disproportionate power not only economically, but also in setting the social and political agenda [1].

The ethical and political problems in this section (privacy, discrimination, surveillance) are not independent of each other; on the contrary, they form a self-reinforcing cycle. The process begins with the violation of privacy (mass data collection). This collected data is fed into models that lay the groundwork for algorithmic discrimination (biased data processing). The results produced by these models then legitimize applications such as mass surveillance and anticipatory governance (data use). This shows that each stage of the data life cycle (collection, processing, use) harbors its own unique ethical and political dangers.

The most alarming consequence of this cycle is the emergence of a new form of power that can be called “algorithmic authority.” As states and corporations place predictive algorithms at the center of social life (justice, health, employment) in the name of efficiency and security, the assumption that these systems are “objective” and “data-driven” makes it difficult to oppose their decisions. When individuals suffer a loss of rights, such as a job application being rejected or a loan not being approved, with the justification “the algorithm decided so,” they become unable to question the logic behind the decision due to the “black-box” problem. This situation creates an authority that operates outside of democratic accountability mechanisms, is non-transparent, and is difficult to appeal. This new form of power reduces citizenship from being an active participant to a passive “data subject” and erodes the foundations of the social contract by shifting them from negotiation to computation [34].

5. The Internet of Behaviors: From the Observed Individual to the Guided Individual

5.1. The transition process from the internet of things (IoT) to the internet of behaviors (IoB)

The newest and perhaps most impactful phase of the digital transformation is represented by the transition from the Internet of Things (IoT) to the Internet of Behaviors (IoB). This transition signifies a radical shift from passive data collection to active behavioral intervention [4].

IoT as the Foundation: IoT is an infrastructure that enables billions of physical objects—such as smartwatches, smart home devices, industrial sensors, and even clothing—to connect to the internet, continuously producing and sharing data. These devices cause individuals and their environments to leave a massive data trail called “digital dust.” IoT essentially forms the data collection layer [5].

The Rise of IoB: The Internet of Behaviors (IoB) is a natural extension and a higher level of IoT. The fundamental difference separating IoB from IoT is its purpose. While IoT collects data, IoB combines this collected data with behavioral psychology, data analytics, and artificial intelligence techniques to not only understand human behaviors but also to influence, change, and guide these behaviors. This means the individual transitions from being passively observed to becoming a subject whose behaviors are actively intervened upon [35].

5.2. The central role of behavioral data analysis

At the center of IoB's operation lies behavioral data analysis. This process aims to uncover the motivations behind individuals' actions and to predict their future behaviors.

Data Collection and Analysis: IoB systems collect data from a wide variety of sources, such as social media interactions, health data from wearable devices (heart rate, sleep patterns), location data from smartphones, and online purchasing history. This data is analyzed using machine learning algorithms, producing in-depth conclusions about individuals' behavioral patterns, preferences, habits, and even psychological tendencies [36].

Application Areas: The results of these analyses are used to guide behaviors in a wide variety of fields. For instance, behavior-shaping algorithms on platforms like TikTok and YouTube do not merely recommend content but actively curate user engagement loops that reinforce specific behavioral patterns to maximize screen time. Similarly, the integration of Google Health with Fitbit data exemplifies how biometric monitoring transcends simple health tracking to become a tool for granular behavioral profiling and potential insurance risk assessment. For example, an e-commerce site provides personalized product recommendations based on the user's past browsing data. A health app sends motivational messages based on the user's activity level. A car insurance company can determine the insurance premium by monitoring the driver's driving habits (sudden braking, acceleration). As a more controversial example, China's social credit system monitors citizens' behaviors, assigns them scores, and these scores affect individuals' access to certain services.

5.3. Challenges Encountered in Behavioral Data Analysis and Future Projections

The proliferation of IoB brings significant challenges, both technical and ethical.

Technical and Methodological Challenges: Behavioral data is often scattered, unstructured, and of poor quality. Integrating this data coming from different systems (data silos) is technically complex. During the analysis process, risks such as a harmless activity being perceived as a threat (false positive) or a real threat being overlooked (false negative) can lead to serious problems, especially in areas like cybersecurity.

Ethical and Privacy Challenges: The continuous monitoring and analysis of individuals' behaviors penetrate the deepest layers of privacy. The biggest concern is the use of this collected data to manipulate individuals, weaken their autonomous decision-making abilities, and “nudge” them toward certain behaviors.

Future Projections: Projections from the technology research firm Gartner reveal the scale of IoB's future societal impact. Gartner predicts that by the end of 2025, more than half of the world's population will be subject to an IoB program (either commercial or government-supported). According to another projection, as of 2023, the behaviors of 40% of the global population will be digitally monitored through these technologies. These striking figures show that IoB is moving from being a niche technology to becoming a fundamental mechanism regulating social life.

The transition from IoT to IoB represents a fundamental transformation in the meaning of the concept of "data." In the IoT era, data was a “record” or “representation” of a state (e.g., a room's temperature being 22 degrees). In the IoB era, data becomes an “intervention tool” and a “behavioral trigger” (e.g., the smart system turning on the air conditioner because the room temperature is 22 degrees, or a company showing the user an ad for winter clothing). Data is evolving from a reactive (responding) nature to a proactive (taking action) nature [37]. The chain effects of this transformation pose serious risks to the social fabric. As IoB systems become widespread with the aim of optimizing individuals' behaviors (living healthier, consuming less energy, working more efficiently), this “optimization” logic creates constant pressure on individual choices and deviations from the norm. Behaviors labeled as “undesirable” or “inefficient” by the system are attempted to be “corrected” through subtle incentives (discounts) or penalties (higher insurance premiums, lower social credit scores). Ultimately, this process gives rise to a risk of “behavioral homogenization” at the societal level. While the space for individual autonomy, spontaneous action, and creativity narrows, individuals are constantly encouraged to conform to algorithmic norms. This situation poses a serious threat not only to individual freedoms but also to social diversity and the potential for democratic dissent [1].

6. Conclusion

The analysis conducted throughout this section has revealed the multi-layered and transformative impact of big data, artificial intelligence, and the Internet of Behaviors on the social sciences. The findings show that these technologies are not just new methodological tools, but also a paradigm shift that questions the most fundamental assumptions of the social sciences. Big data has re-ignited epistemological debates about the nature of knowledge along the “correlation-causality” axis. While artificial intelligence and machine learning offer immense opportunities for analyzing and predicting social phenomena, they have also brought serious ethical risks such as algorithmic bias, lack of transparency, and privacy. The societal and political reflections of these technologies are shaking our understanding of justice and power with new concepts like surveillance, power asymmetries, and algorithmic authority. Finally, the transition from IoT to IoB has shown that this process has moved from passively observing the individual to actively guiding their behaviors, and that this poses profound threats to individual autonomy and social diversity. In this new digital ecosystem, the task falling to social scientists is not just to be technicians who use these tools effectively. The real and more important task is to be “critical theorists” who question the epistemological assumptions underlying these systems, the power relations they internalize, and their potential societal consequences. The social sciences must expose the biases hidden behind the veil of algorithmic neutrality, lead the creation of ethical and legal frameworks regarding data ownership and use, and defend the human values being sacrificed in the name of “efficiency.” Looking to the future, the social sciences must play a leading role both in using these technologies responsibly and ethically, and in developing a critical “data literacy” awareness in society regarding these technologies. This is a vital necessity not only for the future of the social sciences themselves, but also for the goal of building a more just, democratic, and humane society in the digital age. The solution to these complex problems transcends the boundaries of a single discipline; therefore, strong interdisciplinary collaborations to be established between different fields such as law, philosophy, ethics, and computer science will be the most fertile ground for future research.

To translate this critical perspective into practice, this study proposes three concrete strategies for social scientists and policymakers: First, the establishment of independent "Algorithmic Auditing Boards" comprising social scientists to test public-sector AI systems for bias before deployment. Second, the integration of "Critical Data Studies" into standard social science curricula to ensure future researchers possess both technical competence and ethical literacy. Third, the development of policy frameworks that legally recognize and protect "cognitive liberty"—the right of individuals to be free from manipulative behavioral interference by IoB systems.

Declaration of Interest

The author declare that there is no conflict of interest.

References

- [1] Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- [2] Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.
- [3] Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- [4] Ashton, K. (2009). That ‘Internet of Things’ thing. *RFID Journal*, 22(7), 97–114.
- [5] Gartner. (2020). *Top strategic technology trends for 2021: Internet of behaviors*. Gartner Research.
- [6] Audi, R. (2010). *Epistemology: A contemporary introduction to the theory of knowledge* (3rd ed.). Routledge.
- [7] Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.
- [8] Frické, M. (2015). Big data and its epistemology. *Journal of the Association for Information Science and Technology*, 66(4), 651–661. <https://doi.org/10.1002/asi.23212>
- [9] Anderson, C. (2008, June 23). *The end of theory: The data deluge makes the scientific method obsolete*. *Wired*. <https://www.wired.com/2008/06/pb-theory/>
- [10] Han, B.-C. (2017). *Psychopolitics: Neoliberalism and new technologies of power*. Verso Books.
- [11] Leonelli, S. (2020). Philosophy of data: Why data science is not a replacement for the scientific method. *Harvard Data Science Review*, 2(1). <https://doi.org/10.1162/99608f92.74f2d3f6>
- [12] Savage, M., & Burrows, R. (2007). The coming crisis of empirical sociology. *Sociology*, 41(5), 885–899. <https://doi.org/10.1177/0038038507080443>
- [13] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabási, A.-L., Brewer, D., ... Van Alstyne, M. (2009). Computational social science. *Science*, 323(5915), 721–723. <https://doi.org/10.1126/science.1167742>
- [14] Salganik, M. J. (2019). *Bit by bit: Social research in the digital age*. Princeton University Press.
- [15] Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/1500000011>
- [16] Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welp, I. M. (2010). Predicting elections with Twitter: What 140 characters reveal about political sentiment. *ICWSM*, 10(1), 178–185.
- [17] Breiman, L. (2001). Statistical modeling: The two cultures. *Statistical Science*, 16(3), 199–231.

- <https://doi.org/10.1214/ss/1009213726>
- [18] Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, 105(5), 491–495. <https://doi.org/10.1257/aer.p20151023>
- [19] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. *Minds and Machines*, 30(4), 681–694. <https://doi.org/10.1007/s11023-020-09548-1>
- [20] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv*. <https://doi.org/10.48550/arXiv.2108.07258>
- [21] Tufekci, Z. (2014). Engineering the public: Big data, surveillance and computational politics. *First Monday*, 19(7). <https://doi.org/10.5210/fm.v19i7.4901>
- [22] Grimmer, J. (2015). We are all social scientists now: How big data, machine learning, and causal inference work together. *PS: Political Science & Politics*, 48(1), 80–83. <https://doi.org/10.1017/S1049096514001784>
- [23] Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1). <https://doi.org/10.1177/2053951714528481>
- [24] Lum, K., & Isaac, W. (2016). To predict and serve? *Significance*, 13(5), 14–19. <https://doi.org/10.1111/j.1740-9713.2016.00960.x>
- [25] Lipton, Z. C. (2018). The mythos of model interpretability. *Queue*, 16(3), 31–57. <https://doi.org/10.1145/3236386.3241340>
- [26] Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR): A practical guide*. Springer.
- [27] Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv*. <https://doi.org/10.48550/arXiv.1702.08608>
- [28] Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). <https://doi.org/10.1177/2053951716679679>
- [29] Sharon, T. (2016). The Googlization of health research: From disruptive innovation to disruptive ethics. *Personalized Medicine*, 13(6), 563–574. <https://doi.org/10.2217/pme-2016-0057>
- [30] Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671–732. <https://doi.org/10.2139/ssrn.2477899>
- [31] Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.
- [32] Andrejevic, M. (2014). Surveillance and alienation in the online economy. *Surveillance & Society*, 12(3), 381–397. <https://doi.org/10.24908/ss.v12i3.5113>
- [33] Ferguson, A. G. (2017). *The rise of big data policing: Surveillance, race, and the future of law enforcement*. NYU Press.
- [34] Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication & Society*, 20(1), 118–136. <https://doi.org/10.1080/1369118X.2016.1186713>
- [35] Sunstein, C. R. (2016). *The ethics of influence: Government in the age of behavioral science*. Cambridge University Press.
- [36] Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.
- [37] Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8), 56–59. <https://doi.org/10.1109/MC.2018.3191268>