NEXT-GENERATION DATA GOVERNANCE

KIMBERLY A. HOUSER,[†] JOHN W. BAGBY[††]

ABSTRACT

The proliferation of sensors, electronic payments, click-stream data, location-tracking, biometric feeds, and smart home devices, creates an incredibly profitable market for both personal and non-personal data. It is also leading to an amplification of harm to those from or about whom the data is collected. Because federal law provides inadequate protection for data subjects, there are growing calls for organizations to implement data governance solutions. Unfortunately, in the U.S., the concept of data governance has not progressed beyond the management and monetization of data. Many organizations operate under an outdated paradigm which fails to consider the impact of data use on data subjects due to the proliferation of third-party service providers hawking their "check-the-box" data governance systems. As a result, American companies suffer from a lack of trust and are hindered in their international operations due to the higher data protection requirements of foreign regulators. After discussing the pitfalls of the traditional view of data governance and the limitations of suggested models, we propose a set of ten principles based on the Medical Code of Ethics. This framework, first encompassed in the Hippocratic Oath, has been evolving for over one thousand years advancing to a code of conduct based on stewardship. Just as medical ethics had to evolve as society changed and technology advanced, so too must data governance. We propose that a new iteration of data governance (Next-Gen

[†] Fulbright Specialist, University of Lisbon Law School; Visiting Scholar, Ostrom Workshop on Data Management and Information Governance, Indiana University; and Advisory Board Member, Digital Democracy Lab, William & Mary Law School.

[††] Professor Emeritus, Colleges of Information Sciences & Technology and Smeal College of Business, The Pennsylvania State University. The authors wish to thank the participants at the 2022 Data Law and AI Ethics Research Colloquium co-hosted by the Institute for Computational and Data Sciences and Penn State Law at Penn State; Monash Law at Monash University (Australia); Pamplin School of Business at Virginia Tech; and the Kelley School of Business at Indiana University for their helpful comments.

Data Governance) can mitigate the harms resulting from the lack of data protection law in the U.S. and rebuild trust in American organizations.

INTRODUCTION

Every time you go online, use an app on your phone, drive through a toll booth, ask Alexa to turn on your lights, or buy something online, organizations are tracking you. While you may understand that the website or app you access collects information about you or the toll authority records your license plate number, few consider what happens to this data afterwards. Behind the scenes, our interactions with technology are stored, analyzed, shared, and sold. The use of this data can be highly intrusive and unexpected. For example, your browsing habits can be used to label you for the purpose of ad targeting as "working class," "African American," "debtor," or "seeking medical care."[1] Your location data can be rounded up with a geofence warrant resulting in law enforcement interrogation simply because your phone was tracked near the scene of a crime[2] or a Planned Parenthood office.[3] A *New York Times* investigation revealed that location data is tracked by nearly all of the apps on your smartphone "in startling detail, accurate to within a few yards and in some cases updated more than 14,000 times a day."[4] Such information provides enormous value to those who obtain it.[5]

Monetizing data is big business. Data sharing and big data

[1] *See generally* FED. TRADE COMM'N, A LOOK AT WHAT ISPS KNOW ABOUT YOU: EXAMINING THE PRIVACY PRACTICES OF SIX MAJOR INTERNET SERVICE PROVIDERS, AN FTC STAFF REPORT (2021).

[2] *See* John Holden & Kimberly A. Houser, *Taboo Transactions: Selling Athlete Biometric Data*, 49 FLA. S. U. L. REV. 103, 130 n.186 (2022) (discussing how someone using an exercise-tracking app was located and investigated by the police for a burglary committed on his usual bike route).

[3] Aziz Z. Huq & Rebecca Wexler, *Digital Privacy for Reproductive Choice in the Post-Roe Era*, 97 N.Y.U L. REV. 555, 581--584 (forthcoming 2023) (explaining how the police can surveil pregnant people and abortion providers through digital apps).

[4] Jennifer Valentino-DeVries et al., *Your Apps Know Where You Were Last Night, and They're Not Keeping It Secret*, N.Y. TIMES (Dec. 10, 2018), https://www.nytimes.com/interactive/2018/12/10/business/location-data-privacy-apps.html.

[5] It is estimated that Meta's (formerly Facebook) profit from their user's personal data was $56.5 billion of the $108.6 billion in total revenue in 2018. Robert J. Shapiro, *What Your Data Is Really Worth to Facebook*, WASH. MONTHLY (Jul 12, 2019), https://washingtonmonthly.com/2019/07/12/what-your-data-is-really-worth-to-facebook/; *see also* Kean Birch et al., *Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech*, 8 BIG DATA & SOC'Y 1 (2021).

analytics based on artificial intelligence enhance a business's competitive potential by maximizing the value of information.[6] One company, Reveal Mobile, whose location data code has been embedded in over 500 apps to enable data harvesting, justified this technology explaining that it allows app developers to provide a free service, such as news, to the smartphone user while monetizing the location data about the user.[7]

Of particular concern is individual health data. The Health Insurance Portability and Accountability Act of 1996, commonly referred to as "HIPAA" provides illusory protection to prevent the collection and sharing of this data without your consent. Consider how often you read the HIPAA notice at your doctor's office.[8] Data brokers can purchase this health data from doctors, assign it an identification number, and combine it with information from other sources.[9] Although your doctor's office strips your identity from the data, the information can be re-identified and re-connected back to you.[10] Pharmacies also make money by selling your

[6] Tianshu Sun et al., *The Value of Personal Data in Internet Commerce: A High-Stake Field Experiment on Data Regulation Policy*, NET INST. WORKING PAPER No. 21-10 (2021), https://ssrn.com/abstract=3962157.

[7] Valentino-DeVries, *supra* note 4 (providing that targeted advertising is a most common use of this data).

[8] *See generally* Nathaniel Good et al., *Stopping Spyware at The Gate: A User Study of Privacy, Notice and Spyware*, 93 PROC. OF THE 2005 SYMP. ON USABLE PRIV. AND SEC. 43 (2005), https://www.law.berkeley.edu/files/Spyware_at_the_Gate.pdf; *see also*, Marie C. Pollio, *The Inadequacy of HIPAA's Privacy Rule: The Plain Language Notice of Privacy Practices and Patient Understanding*, 60 N.Y.U ANN. SURV. AM. L. 579 (2004) (arguing the plain language regime in law as applied to HIPAA privacy notices fails to adequately assure understanding). For a discussion of harms related to health data and predictive analytics, see Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L.J. 251 (2016); Philip M. Nichols, *Bribing the Machine: Protecting the Integrity of Algorithms as the Revolution Begins*, 56 AM. BUS. L.J. 771 (2019).

[9] Adam Tanner, *How Data Brokers Make Money Off Your Medical Records*, SCI. AM., (Feb. 1, 2016), https://www.scientificamerican.com/article/how-data-brokers-make-money-off-your-medical-records/.

[10] Luc Rocher et al., *Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models*, 10 NATURE COMMC'N 3069 (2019) (creating a model that could correctly re-identify Americans from anonymized data sets 99.98% of the time); Deanonymization development relies on two pillars. First, lax security investment degrades the performance of data handlers' privacy commitments. John W. Bagby, *White Paper on Cloud Computing: Overcoming Challenges of Integrating Robust Competition with Security* (Sept.13, 2023). Available at SSRN: https://ssrn.com/abstract=4570963 or http://dx.doi.org/10.2139/ssrn.4570963 Submission to Federal Trade Commission, Rulemaking Docket - "Request for Information-Solicitation for Public Comments on the Business Practices of Cloud Computing Providers," Comment #FTC-2023-0028-

prescription data.[11] Similar to credit reporting agencies, there are three main agencies that collect and sell medical reports about you.[12] Insurance companies can then purchase these reports to determine rates for health and life insurance.[13] Some of these agencies also provide ancillary services to their clients such as supplying criminal records, traffic violations, and mortality predictions.[14] Aside from some restrictions in California and Vermont,[15] there are few legal limitations on the data brokers who sell your data as they conduct these activities outside of any federal regulatory scheme in the U.S.[16]

There are several types of data that pose differing risks to individual privacy. Data is simply a collection of observations. Raw data is regularly collected and includes observations about people, animals, things, or conditions (e.g., location, speed, temperature). However, it is not just this raw data that is of concern, but also the new data created from its

0023, Track.No. lhz-0pzs-bo1wp (May 21, 2023) (arguing the cloud is essential as feedstock to all big data analytics such as AI and ML; competitive market failures in (cloud) security investment produces sub-optimal privacy protections). Second, increasingly the relentless data search and comparisons among discrete databases eventually identifies "data fingerprints" enabling corroboration with outside/auxilliary databases. Once numerous similarities are identified between two or more partial, but unidentical profiles of one individual's data, confidence grows that the two describe the same person. If either database reveals personal identifiers, the anonymization fails. This frequently accomplished revelation encourages the "deanonymizer" to combine the two seemingly diverse records into one. Thus, as big data types and sources grow, reidentification becomes easier. *See, e.g.,* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L.Rev. 1701 (2010) https://www.uclalawreview.org/pdf/57-6-3.pdf (arguing how reidentification or deanonymization undermines basic assumptions of privacy law and practice).

[11] Kalev Leetaru, *How Data Brokers and Pharmacies Commercialize Our Medical Data*, FORBES (Apr. 2, 2018), https://www.forbes.com/sites/kalevleetaru/2018/04/02/how-data-brokers-and-pharmacies-commercialize-our-medical-data/?sh=6185dc1b11a6.

[12] CONSUMER REPORTS & Connie Thompson, *More People Have Access to Your Prescription Medicine History Than You May Realize*, KOMONEWS (May 9, 2019), https://komonews.com/news/consumer/more-people-have-access-to-your-prescription-medicine-history-than-you-may-realize (identifying the three agencies as Exam One, Millian IntelliScript, and the Medical Information Bureau).

[13] Valentino-DeVries et al., *supra* note 4.

[14] *About MIB's Ancillary Services*, MIB, https://www.mib.com/ancillary_services.html (last visited Nov. 10, 2023).

[15] *See* CAL. CIV. CODE § 1798.140 (West 2020) (passed as the California Consumer Privacy Act); 9 V.S.A. § 2430(4)(A).

[16] Louise Matsakis, *The WIRED Guide to Your Personal Data (and Who Is Using It)*, WIRED (Feb. 15, 2019), https://www.wired.com/story/wired-guide-personal-data-collection.

analysis.[17] Derived data (or imputed data) is "information that can be developed from multiple data points about an individual or from an individual's relationship to a group."[18] Because data subjects are unaware of this analysis on their data, they are oblivious as to how this new derived data is being used and shared. Derived data is used to make predictions about people, such as what movie Netflix should recommend, or more concerningly, how likely someone is to commit a crime, skip bail, or reoffend.[19] Although this data is collected and analyzed by private industry, data brokers make it available to government agencies.[20]

In a similar vein, there are also the risks associated with linked data.[21] Given the multiple sources from which information and data emerges, and the incredible value in linking the divergent data in order to create an accurate dossier of a potential consumer, despite promises of anonymity, linking data makes it even easier to deanonymize.[22] Because

[17] Eric González, *The Harms of Data Abuse*, ACLU (Jan. 29, 2021), https://www.aclu-wa.org/docs/harms-data-abuse (describing some harms from the analysis of data from apps - derived data).

[18] Kimberly A. Houser & John Bagby, The Data Trust Solution to Data Sharing Problems, 25 VAND. J. ENT. & TECH. L. 113, 124 (2023) (citations omitted).

[19] *See generally* Rainer Mühlhof, *Predictive Privacy: Towards an Applied Ethics of Data Analytics*, 23 ETHICS & INFO. TECH. 675 (2021) (explaining the harms resulting from predictive analytics using derived data); Kate Crawford & Jason Schultz, *Big Data and Due Process: Toward a Framework to Redress Predictive Privacy Harms*, 55 B.C. L. REV. 93, 117 (2014); Rashida Richardson et al., *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*, 94 N.Y.U. L. REV. ONLINE 192 (2019).

[20] Bennett Cyphers, *How the Federal Government Buys Our Cell Phone Location Data*, EFF (Jun. 13, 2022), https://www.eff.org/deeplinks/2022/06/how-federal-government-buys-our-cell-phone-location-data*; see also* Dori Rahbar, Note, *Laundering Data: How the Government's Purchase of Commercial Location Data Violates Carpenter and Evades the Fourth Amendment*, 122 COLUM. L. REV. 713 (2023) (explaining this loophole in the case law).

[21] Leora Eisenstadt, *Data Analytics and the Erosion of the Work/Nonwork Divide*, 56 AM. BUS. L.J. 445, 448 (2019) (explaining how employers can link an employee's data with their data from social media and online profiles to "screen for the most productive teams," or "track their employees' family planning thoughts and health-care concerns").

[22] Boris Lubarsky, *Re-Identification of "Anonymized" Data*, 1 GEO. L. TECH. REV. 202 (2017) (explaining that only a small amount of data is needed to re-identify an individual. This linked data can be gender, data of birth, and zip code. In fact, 63% of the population can be uniquely identified on that data alone); Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1734 (2010) ("The accretion problem is this: Once an adversary has linked two anonymized databases together, he can add the newly linked data to his collection of outside information and use it to help unlock other anonymized databases. Success breeds further success. Narayanan and

linked data is unknown to the data subject and there is no legal requirement that permits a data subject to request that the data be "unlinked," there is little control over these use of these dossiers.[23] With the emerging ubiquity of the Internet of things (IoT) and the Internet of Bodies (IoB), linkages will only increase. Your Fitbit, smart phone, and smart car in combination can provide a highly detailed picture of your life, health, movements, interests, habits, and connections.

The risks from the unregulated use of data are well-documented in legal scholarship.[24] As aptly stated by Georgetown law professor, Paul

Shmatikov explain that 'once any piece of data has been linked to a person's real identity, any association between this data and a virtual identity breaks the anonymity of the latter.'").

[23] Elizabeth R. Pike, *Defending Data: Toward Ethical Protections and Comprehensive Data Governance*, 69 EMORY L. J. 687, 703 (2020) (explaining how collecting data from multiple data points can identify information about a person that they "themselves had not known or wanted shared")

[24] *See e.g.*, Salome Viljoen, *A Relational Theory of Data Governance*, 131 YALE L.J. 573 (2021) (arguing the widespread extant theorizing about data governance is producing varying legislative "datafication" proposals, many enhancing the rights and remedies of subject individual but most appear to ignore the social informational harm that should, instead, be addressed by democratic institutions of data governance for societal betterment); Danielle Keats Citron & Daniel J. Solove, *Privacy Harms*, 102 B. U. L. REV. 793 (2022) (describing physical, economic, reputational, discrimination, relationship, psychological and autonomy harms from data privacy violations); David Nersessian & Ruben Mancha, *From Automation to Autonomy: Legal and Ethical Responsibility Gaps in Artificial Intelligence Innovation*, 27 MICH. TECH. L. REV. 55 (2020) (describing the legal and ethical concerns surrounding AI); Margaret Hu, *Algorithmic Jim Crow*, 86 FORDHAM L. REV. 663 (2017) (describing how algorithmic screening devices can result in discriminatory disparate impact); Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 CALIF. L. REV. 671, 677 (2016); Eric Siegel, PREDICTIVE ANALYTICS: THE POWER TO PREDICT WHO WILL CLICK, BUY, LIE, OR DIE (2d ed. 2016) (describing how predictive analytics are currently being used by the government and business to identify preferences and risks and noting that the use of data about groups that have been historically discriminated against can result in discriminatory outcomes); Kate Crawford, *The Hidden Biases in Big Data*, HARV. BUS. REV. (Apr. 1, 2013), https://hbr.org/2013/04/the-hidden-biases-in-big-data; Deborah Hellman, *Patterned Inequality, Compounding Injustice, And Algorithmic Prediction*, 1 AM. J.L. & EQUAL. 252 (2022) (explaining how predictive analytics can result in data harms); Niklas Eder, *Beyond Automation: Machine Learning-Based Systems and Human Behavior in the Personalization Economy*, 25 STAN. TECH. L. REV. 1 (2021) (explaining how algorithms are used to target ads and manipulate consumers); Kimberly Houser, *Artificial Intelligence and The Struggle Between Good and Evil*, 60 WASHBURN L.J. (SPECIAL ISSUE ON A.I.) 475 (2021) [hereinafter Houser, *Artificial Intelligence*] (describing potential harms relating to the collection and use of biometric data); Kimberly Houser and

Ohm, "[F]or almost every person on earth, there is at least one fact about them stored in a computer database that an adversary could use to blackmail, discriminate against, harass, or steal the identity of him or her."[25] Not only does existing federal law provide insufficient protection for subjects of data collection,[26] but there is surprisingly little legal scholarship on data governance as a means to address this shortcoming.[27] This article seeks to fill the gap by examining how data governance can evolve to protect data subjects and garner trust in organizations. The article begins by explaining how traditional data governance focuses on the management and monetization of data, not the prevention of harms. After examining the components of current data governance and its limitations, it explores various pathways to more robust governance systems, settling on the stewardship model underlying the Medical Code of Ethics.

Just as medical ethics had to grow and adjust as society changed and technology advanced, so too must data governance. The tipping point in medical ethics occurred in the late 60s and early 70s when organ harvesting became possible, and the public learned about the horrific Tuskegee Syphilis Experiment.[28] Ethics in medicine quickly evolved over the next decade from a paternal model to one of patient agency. Additionally, ethical opinions were drafted in conjunction with the revisions to the Medical Code of Ethics to guide physician decision-making, and organizations began requiring approval for conducting medical studies.[29]

The harms resulting from the unregulated use of data are not going away and the federal government is unlikely to pass any comprehensive data protection law in the near future.[30] As cybersecurity concerns hit the

Debra Sanders, *The Use of Big Data by the IRS: Efficient Solution or the End of Privacy As We Know It?*, 19 VAND. J. ENTERTAIN. TECH. L. 817 (2017) (describing harms relating to the government's use of data analytics).

[25] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1748 (2010).

[26] Kimberly A. Houser & W. Gregory Voss, *GDPR: The End of Google and Facebook or a New Paradigm in Data Privacy?*, 25 RICH. J.L. & TECH. 1 (2018); Pike, *supra* note 23, at 710–16.

[27] Melanie McCaig & Davar Rezania, *A Scoping Review on Data Governance*, 2021 PROC. INT'L CONF. ON IoT BASED CONTROL NETWORK & INTEL. SYS. 1, 2, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3882450 ("Data governance remains an under-researched and under-practiced field despite its documented high importance.").

[28] *See* discussion *infra* Part IV A.

[29] AM. MED. ASS'N, HIST. CODE, https://www.ama-assn.org/sites/ama-assn.org/files/corp/media-browser/public/ethics/ama-code-ethics-history.pdf (last visited Dec. 28, 2022).

[30] Although Congress did pass a law protecting the personal data of judges, such

boardroom[31] and various antitrust and privacy violations are targeting tech companies,[32] a new generation of data governance is needed, not only to protect data subjects, but for the benefit of organizations as well. This next iteration of data governance principles requires that those collecting and using data engage in the trustworthy and responsible stewardship of data (Next-Gen Data Governance).

The paper proceeds as follows. Part I provides a descriptive account of traditional data governance and its components, describing how hard, de jure law, soft law, and institutional policies and procedure intertwine to create a haphazard set of rules. Part II explores the limitations of the narrow view of data governance as "asset management." Part III examines different pathways to expand the notion of data governance. Part IV proposes a new model of data governance (Next-Gen Data Governance) loosely based on the Code of Medical Ethics, replacing the concept of data management with data stewardship, concluding with ten actionable principles.

I. TRADITIONAL DATA GOVERNANCE

According to *Data Governance: The Definitive Guide, "*Data governance is, first and foremost, a *data management function* to ensure the quality, integrity, security, and usability of the data collected by an organization."[33] To corporations, data governance is a set of procedures and policies designed to both manage and monetize data. [34] To

protections do not extend to the rest of the U.S. population. *Congress Passes the Daniel Anderl Judicial Security and Privacy Act*, U.S. COURTS (Dec. 16, 2022), https://www.uscourts.gov/news/2022/12/16/congress-passes-daniel-anderl-judicial-security-and-privacy-act.

[31] *See SEC Proposes Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies*, U.S. SEC. AND EXCH. COMM'N (Mar. 9, 2022), https://www.sec.gov/news/press-release/2022-39 (requiring public companies to explain how they are managing cybersecurity risks)*; see also* Virginia Harper Ho, *Nonfinancial Risk Disclosure and the Costs of Private Ordering*, 55 AM. BUS. L.J. 407 (2018) (discussing the expanding board requirements regarding cybersecurity and other risks).

[32] Joe Panettieri, *Big Tech Antitrust Investigations: Amazon, Apple, Google, Meta/Facebook and Microsoft Updates*, CHANNEL E2E (Dec. 5, 2022), https://www.channele2e.com/business/compliance/big-tech-antitrust-regulatory-breakup-updates/.

[33] Evren Eryurek et al., *Data Governance: The Definitive Guide,* O'REILLY, https://www.oreilly.com/library/view/data-governance-the/9781492063483/ch01.html (emphasis added).

[34] *See, e.g.*, *Data Monetization Capabilities, Governance, and Planning,* IDC, https://www.idc.com/itexecutive/planning-guides/data-monetization-capabilities-governance-planning (associating data governance and data monetization).

governments, data governance similarly involves "managing information as a strategic resource."[35]

Data governance, as it is currently used, serves as an umbrella term for a collection of several related data management sub-disciplines. It is a complex system[36] deserving of, first, understanding of its parts in isolation, then second, integrative analysis of data governance as an interactive system. This conceptualization requires an analysis of each major component in isolation to determine its results and sensitivities to controls. Then, second, a holistic integration is needed.[37] Only a systems approach could successfully identify interactions, reinforcements, or balances among individual elements. Otherwise, such analysis will be based mostly on speculative hypotheses or the analysis of failure. Political economists are learning to adapt network analytics and theories of evolutionary selection to build better models of large-scale socio-economic processes. A complex systems approach offers the conceptual tools to unify these efforts to understand large systems and their macroscopic properties and typical behaviors. The challenges a complex systems approach poses for standard economic analysis are explored with reinterpretations of several major institutional transitions in the development of European society.[38]

In this conception, data governance involves an organization's management of the collection, storage, and maintenance, in a usable format, of data for both internal and commercial purposes. The focus is on data quality and accessibility. Generally, IT departments are tasked with data governance. However, as organizations began amassing more data than could be stored on their own servers due to the Internet and smartphones, third party service providers appeared to provide storage and

[35] Revision of OMB Circular No. A-130, "Managing Information as a Strategic Resource," 81 Fed. Reg. 49689 (July 28, 2016).

[36] *See*, *e.g.*, Charles B. Keating & Joseph M. Bradley, *Complex System Governance Reference Model*, 6 INT'L. J. SYS. OF SYS. ENGR. 33–52 (Apr. 14, 2015); *see also* Hilton L. Root, *The Role of Complex Networks and Selection in Political Economy*, SSRN (Sept. 13, 2023). Political economists are learning to adapt network analytics and theories of evolutionary selection to build better models of large-scale socio-economic processes. A complex systems approach offers the conceptual tools to unify these efforts to understand large systems and their macroscopic properties and typical behaviors. The challenges a complex systems approach poses for standard economic analysis are explored with reinterpretations of several major institutional transitions in the development of European society. *Id.* at 1.

[37] *See generally* ROBERT F. SMALLWOOD, INFORMATION GOVERNANCE: CONCEPTS, STRATEGIES, AND BEST PRACTICES (John Wiley & Sons, 2019).

[38] *Id.*

data management services.[39]

Data governance comprises a loosely aggregated set of concerns that accrete slowly enough to defy comprehensive accumulation as a code of coherent rules, practices, or constraints. This broad palette of data governance constraints complicates the risk of compliance prediction given the risk analyst's need for expertise in such disparate authorities. Data governance concepts stem from three major components of authority: (1) hard, de jure law, (2) soft and de facto law, and (3) organizational procedures and policies.

A. Hard, De Jure Law

To many, law is the essential driver of data governance.[40] Hard, de jure law consists of laws and regulations enacted and enforced by governmental bodies.[41] The most prominent example regarding data use is the General Data Protection Regulation (GDPR). [42] While other jurisdictions (and several U.S. states)[43] have sought to replicate its purpose through their own regulations, the U.S. federal government is not among them. This leaves a hodgepodge of federal and state laws for domestic companies and conflicting foreign regulations for U.S. multinationals to sift through.

[39] The cloud has become an essential Internet component spurring the growth of data accumulations produced from Internet traffic and use of mobile personal devices. For the foreseeable future, increasing amounts of record retention, transaction processing, and big data availability (feedstock for AI), will rely on cloud computing. *See generally*, Keith D. Foote, *A Brief History of Cloud Computing*, DATAVERSITY https://www.dataversity.net/brief-history-cloud-computing/ (Dec.17, 2021); Esther Shein, *A Brief History of Cloud Computing*, TECHREPUBLIC (Oct.26, 2022) https://www.techrepublic.com/article/brief-history-cloud-computing/.

[40] *What Drives Data Governance*, PWC (Nov. 2019), https://www.pwc.in/consulting/technology/data-and-analytics/govern-your-data/insights/what-drives-data-governance.html.

[41] WBG, WORLD DEVELOPMENT REPORT 2017: GOVERNANCE AND THE LAW 83 (2017).

[42] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ 2016 L 199/1 [hereinafter GDPR].

[43] Anokhy Desai, *US State Privacy Legislation Tracker*, IAPP (Oct. 20, 2023), https://iapp.org/resources/article/us-state-privacy-legislation-tracker/ (identifying California, Utah, Colorado Virginia, and Connecticut). *See, e.g.*, Cal. Civ. Code § 1798.140 (West 2022) (passed as the California Consumer Privacy Act) [hereinafter CCPA].

Understanding data governance through law alone is difficult. The sectoral approach to data protection law in the U.S. differs significantly from the European omnibus approach. The EU's method is essentially a uniform, pan-EU approach applying data protection law uniformly to data users (with some exceptions for governmental use of data) granting strong rights for data subjects located with the EU.[44] Although this legal regime provides certain rights to those in the EU, despite the efforts (and expectations) of EU authorities,[45] these protections have not been adopted globally and remain mostly unavailable in the U.S. Exacerbating the problem is that the U.S. federal regulatory scheme is woefully outdated.[46]

An example of a law that addresses data governance is the EU Data Governance Act which will become applicable in September 2023.[47] Although it aims to increase the sharing of data and proposes the use of data intermediaries, it has been criticized for violating the World Trade Organization's prohibition on requiring data sharing services to maintain a local office and may overly restrict data flows "resulting in billions of euros in lost trade."[48] China has also acknowledged the need for data governance with respect to artificial intelligence (AI) that rely on large data sets or being used to make automated decisions which could impact an individual's rights.[49] China's May 2019 Beijing AI Principles indicate that:

[44] W. Gregory Voss & Kimberly A. Houser, *Personal Data and the GDPR: Providing a Competitive Advantage for U.S. Companies*, 56 AM. BUS. L.J. 287, 294–95 (2019).

[45] *Id.* at 8–9.

[46] *See* Houser & Sanders, *supra* note 24 (explaining how the Privacy Act of 1974 has not been updated since before the widespread use of the internet and social media).

[47] *Proposal for a Regulation of the European Parliament and of the Council on European data governance (Data Governance Act)*, COM (2020) 767 final (Nov. 25, 2020).

[48] *Data Governance Act: Eain Elements and Business Implications*, DR2 CONSULTANTS (Apr. 12, 2022), https://dr2consultants.eu/data-governance-act-main-elements-and-business-implications/.

[49] In its AIDP China set forth a goal of becoming a world leader in ethical standards for AI. In 2019, the National New Generation Artificial Intelligence Governance Expert Committee released its eight principles that included: "AI development should begin from enhancing the common well-being of humanity. Respect for human rights, privacy and fairness were also underscored within the principles. Finally, they highlighted the importance of transparency, responsibility, collaboration, and agility to deal with new and emerging risks." Huw Roberts, Josh Cowls, Jessica Morley, Mariarosaria Taddeo, Vincent Wang & Luciano Floridi, *The Chinese Approach to Artificial Intelligence: An Analysis of Policy, Ethics, and Regulation*, 36 AI & SOC 59, 68 (2021).

> The development of Artificial Intelligence (AI) concerns the future of the whole society, all mankind, and the environment. The principles below are proposed as an initiative for the research, development, use, *governance* and long-term planning of AI, calling for its healthy development to support the construction of a community of common destiny, and the realization of beneficial AI for mankind and nature.[50]

Although the U.S. 2020 AI Initiative indicates that "[t]he United States must foster public trust and confidence in AI technologies and protect civil liberties, privacy, and American values in their application in order to fully realize the potential of AI technologies for the American people," to date Congress has not taken legislative action on AI technologies.[51]

Organizations must interpret not only the regulations of multi-source authorities, but they must also determine the legitimacy or supremacy of each jurisdiction and regulator's commitment to enforcement. Because data flows cannot generally be bounded territorially, questions of jurisdiction become somewhat befuddled.[52] Additionally, an organization's compliance costs necessarily increase due to the multiple sources for authoritative rules amalgamating into disintegrated controls.[53] Such a plethora of sources undermines coherent understanding of applicable regulations to certain entities.[54] It is well documented that

[50] *Beijing Artificial Intelligence Principles*, INT'L RSCH. CTR. FOR AI ETHICS AND GOVERNANCE (2019), https://ai-ethics-and-governance.institute/beijing-artificial-intelligence-principles/ (emphasis added).

[51] Exec. Order No. 13859, 3 C.F.R. 254 (2020). See also, Executive Order (E.O.) 14110 on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[52] *See* Roxana Vatanparast, *Data Governance and the Elasticity of Sovereignty*, 46 BROOK. J. INT'L L. 1, 5 (2020), (explaining that the extraterritoriality of laws regarding data are not unique, but rather a "creative reimagining of the elasticity of sovereignty").

[53] Jennifer Huddleston, *The Price of Privacy: The Impact of Strict Data Regulations on Innovation and More*, AM. ACTION F. (June 3, 2021), https://www.americanactionforum.org/insight/the-price-of-privacy-the-impact-of-strict-data-regulations-on-innovation-and-more/ ("A 2018 EY and International Association of Privacy Professionals report found companies reported spending an average of $1.3 million per year on GDPR compliance costs. These costs are undertaken not only by European companies but also by U.S.-based companies with an EU presence.").

[54] *See generally* HOWARD BEALES, ET AL., GOVERNMENT REGULATION: THE GOOD, THE BAD, & THE UGLY (2017) (arguing poorly designed regulations may cause more harm than good; stifle innovation, growth, and job creation; waste

differing regulatory regimes create enormous headaches for both the organizations wishing to comply as well as the enforcement officers.[55]

Additionally, creating new hard, de jure law is insufficient due to the speed with which technology advances.[56] Even the most technologically adept state legislatures have difficulty in keeping up. Finally, even when laws do protect data subjects from harm, they may not be enforced. The FTC is notoriously understaffed and underbudgeted.[57] Hard law is that not only is it incomplete and outdated, but the few rights provided to data subject are being whittled away by the judiciary.[58] Because of the piecemeal uncoordinated set of hard laws throughout the world, they provide insufficient guidance for data governance.

B. Soft Law

While hard law, such as the GDPR, tends to focus on restricting activities,[59] soft law is viewed as aspirational. At the macro level, soft law is gaining favor as a governance device, particularly in international and transnational areas.[60] Information mobility, ubiquity, and ease of transfer across borders suggest a number of soft law sources of data governance.

Soft Law arguably holds more promise to harmonize data governance than potential treaty-driven, regulatory, legislative, or judicial

limited resources; undermine sustainable development; inadvertently harm the people they are supposed to protect; and erode the public's confidence in our government).

[55] Pravin Kothari, *Multinationals Face Unique Challenges for Data Privacy and Security compliance,* CPO MAG. (Sept. 19, 2018), https://www.cpomagazine.com/data-protection/multinationals-face-unique-challenges-for-data-privacy-and-security-compliance/.

[56] John W. Bagby & Nizan G. Packin, *RegTech and Predictive Lawmaking: Closing the RegLag Between Prospective Regulated Activity and Regulation*, 10 MICH. BUS. & ENTREPRENEURIAL L. REV. 127 (2021).

[57] Mary Ashley Salvino, *Analysis: How Will the FTC Get Its Privacy Mojo Back in 2022?*, BLOOMBERG L. (Nov. 1, 2021), https://news.bloomberglaw.com/bloomberg-law-analysis/analysis-how-will-the-ftc-get-its-privacy-mojo-back-in-2022.

[58] Daniel J. Solove & Danielle Keats Citron, *Standing and Privacy Harms: A Critique of TransUnion v. Ramirez*, 101 B.U. L. REV. ONLINE 62 (2021) (describing how the Supreme Court decision in TransUnion v. Ramirez further complicates the issue of standing to sue for violations by industry of privacy law, effectively nullifying the few privacy rights that do exist under federal law).

[59] Houser & Voss, *supra* note 26, at 90 (explaining how the EU's GDPR may "severely restrict the use of machine learning algorithms").

[60] Kumaravadivel Guruparan & Jennifer Zerk, *Influence of Soft Law Grows in International Governance*, CHATHAM HOUSE (Jun. 17, 2021), https://www.chathamhouse.org/2021/06/influence-soft-law-grows-international-governance.

adoption of uniform hard laws across the globe.[61] According to Duke law professor Steven L. Schwarcz,

> [S]oft law generally refers to nonstate rules that may be aspirational or reflect best practices but are not yet legally enforceable. For this reason, soft law sometimes is called non-state law. It contrasts with standard, or "hard," law, which is legally enforceable (citations omitted).[62]

Often associated with international law,[63] soft law consists of resolutions, regulatory guidance, and the nonbinding rules or instruments that interpret binding legal rules (hard law). Although soft law lacks statutory, regulatory rulemaking, and other formal procedural steps, it still exerts influence.[64] It also finds sources in western constitutional law, administrative law, and even congressional resolutions and practices.[65]

Soft law can serve to inform the public and political institutions about policy preferences, influencing decision-making and behaviors of the public, various organizations, and all levels of government.[66] Under this conception, soft law is a form of "choice architecture"[67] that presents decision-making points, often implemented under "nudge" theory.[68] Nudge theory is the incentivization of desired behavior through encouragement or suggestion.[69] This contrasts with hard law that imposes

[61] *See, e.g.*, Cary Coglianese, *Environmental Soft Law as a Governance Strategy*, 61 JURIMETRICS 19 (2020).

[62] Steven L. Schwarcz, *Soft Law as Governing Law*, 104 MINN. L. REV. 2471, 2472 (2020).

[63] *See generally* Pierre-Marie Dupuy, *Soft Law and The International Law of the Environment*, 12 MICH. J. INT'L L. 420 (1910) (arguing soft law seems paradoxical because the classical view is binary: the rule of law is considered "hard" and compulsory, otherwise there is no law at all); Andrew Guzman & Timothy L. Meyer, *Explaining Soft Law*, 2 J. LEG. ANALYSIS 171 (2010) .

[64] Tax incentives serve as a major tool of industrial policy. Financial statement disclosures, ESG with particular contemporary impacts, have long incentivized disciplines in diverse fields such as governance, sustainability, and foreign criminal activity. *See, e.g.*, Kimberly A. Houser & Kathryn Kisska-Schulze, *Disrupting Venture Capital: Carrots, Sticks & Artificial Intelligence*, 13 U.C. IRVINE L. REV. 901 (2023) (describing how tax incentives have been successful in promoting investment in certain industries and the hiring of employees from neglected groups).

[65] *See* Jacob E. Gersen & Eric A. Posner, *Soft Law: Lessons from Congressional Practice*, 61 STAN. L. REV. 573 (2008).

[66] *Id.*

[67] RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: THE FINAL EDITION (2021).

[68] *See generally* Cass R. Sunstein, *The Ethics of Nudging*, 32 YALE J. ON REG. 413 (2015).

[69] *See* RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS

mandatory sanction or penalty under laws or regulations.

The EU has been especially active in creating soft laws through guidance documents. There is the Digital Markets Act,[70] which sets up parameters around data sharing, the Platform to Business Regulation,[71] which requires platforms and search engines to provide understandable terms and conditions, the Artificial Intelligence Act,[72] which addresses the risks involved with the use of AI, and the Coordinated Plan on Artificial Intelligence,[73] whose goal is to create EU global leadership in human-centered AI. Although some of these proposals may eventually become law, currently they serve as guidelines, a key soft law mechanism.

Of particular interest with respect to governance are the 61 Guidelines, Recommendations, and Best Practices put out over the years by the European Data Protection Board.[74] Additionally, the GDPR provides 173 Recitals which serve as guidance for the specific provisions of the GDPR.[75] The main limitation of soft law is the lack of clear enforceability. It is also subject to many interpretations, further frustrating a harmonized understanding of data governance.

C. Institutional Policies and Procedures

While soft law is instructive in signaling how hard law might be enforced, institutional policies and procedures also provide a source for

ABOUT HEALTH, WEALTH, AND HAPPINESS 6 (2008).

[70] *See generally* LUÍS CABRAL ET AL., THE EU DIGITAL MARKETS ACT: A REPORT FROM A PANEL OF ECONOMIC EXPERTS (2021).

[71] *See generally Platform-to-Business Trading Practices*, EUR. COMM'N, https://digital-strategy.ec.europa.eu/en/policies/platform-business-trading-practices (last visited Feb. 24, 2022).

[72] *Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, COM (2021) 206 final (Apr. 4, 2021).

[73] *See generally Coordinated Plan on Artificial Intelligence 2021 Review*, EUR. COMM'N, https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review. With respect to EU soft law data governance, the EU created the High-Level Expert Group on Artificial Intelligence (AI HLEG) as part of their AI strategy which has created a slew of guidelines for "trustworthy AI." *High-Level Expert Group on Artificial Intelligence*, EUR. COMM'N (June 7, 2022), https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai.

[74] *See generally Guidelines, Recommendations, Best Practices*, EUR. DATA PROT. BD, https://edpb.europa.eu/our-work-tools/general-guidance/guidelines-recommendations-best-practices_en (last visited Jan. 15, 2024).

[75] GDPR, *supra* note 40, at 1–31; *see also* Tadas Klimas & Jurate Vaiciukaite, *The Law of Recitals in European Community Legislation*, 15 ILSA J. INT'L & COMP. L. 61 (2008) explaining the purpose of recitals in EU legislation).

traditional data governance. In its earliest stages, data governance was created as professional strategic standards in the records management industry.[76] DAMA International is one such organization that provides resources for information and data management, defining data governance as "the "planning, oversight, and control over management of data and the use of data and data-related sources."[77] The Data Governance Institute defines it as "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."[78]

In a report by World Economic Forum in collaboration with Deloitte, the authors point out that new technologies, such as AI, are challenged by "a lack of regulation, misuse of technology and challenges in addressing cross-border differences."[79] Without some type of data governance, there is the risk of harms from automated decision-making, biased data, the revealing of sensitive data, unfair targeting, or exclusion of certain groups.[80] In the Global Data Governance Project report created by Thomas Stuart and Susan Ariel Aaronson, they indicate that no one is even close to creating a comprehensive data governance system which they define as a "systemic and flexible approach to govern different types of data use and reuse."[81]

Due to this lack of an authoritative data governance code, several major companies have created their own rules for "Ethical AI," "Trustworthy AI," or "Responsible AI." Unfortunately, most read like a

[76] *See, e.g.*, INT'L ORG. FOR STANDARDIZATION, ISO 15,489-1:2016 INFORMATION AND DOCUMENTATION — RECORDS MANAGEMENT — PART 1: CONCEPTS AND PRINCIPLES (2021) (describing how the standard defines "the concepts and principles from which approaches to the creation, capture and management of records are developed").

[77] Thor Olavsrud, *What Is Data Governance? Best Practices for Managing Data Assets*, CIO (Mar. 18, 2021), https://www.cio.com/article/202183.

[78] *Definitions of Data Governance,* DATA GOVERNANCE INST., https://datagovernance.com/the-data-governance-basics/definitions-of-data-governance/ (last visited Nov. 10, 2023).

[79] William D. Eggers & Ruth Hickin, *Foreward* to GLOBAL TECHNOLOGY GOVERNANCE REPORT 2021: HARNESSING FOURTH INDUSTRIAL REVOLUTION TECHNOLOGIES IN A COVID-19 WORLD 3, 4 (Dec. 2020).

[80] Sylvie Delacroix & Jessica Montgomery, *From Research Data Ethics Principles to Practice: Data Trusts as a Governance Tool*, HANDBOOK OF BEHAV. DATA SCI. (forthcoming 2023).

[81] THOMAS STRUETT & SUSAN ARIEL AARONSON, THE GLOBAL DATA GOVERNANCE PROJECT 2, https://cpb-us-e1.wpmucdn.com/blogs.gwu.edu/dist/c/3127/files/2021/01/The-Global-Data-Governance-Project-Executive-Summary-Jul23.pdf (last visited Dec. 21, 2022).

marketing brochure, with no explanation of how their promises of "fairness" or "non-discrimination" will be achieved. The lack of agreement among those proposing standards creates a situation where organizations rely on data governance mechanisms offered by third-party consultants which do not generally consider the potential harms to data subjects. Additionally, like soft law, institutional policies and procedures seldom include a true enforcement mechanism. In the following section we break down the shortcomings of traditional data governance.

II. The Misadventures of Data Governance

Although traditional data governance may arise from hard, de jure law, soft law, and institutional policies and procedures, these sources are not interested in protecting data subjects, but rather making sure data is accessible and usable for the organization's purposes. Within each organization, it is the CIO or IT department that is given the responsibility to "govern" the data.[82] This ensures a limited viewpoint of data collection, sharing, and analytics, which presents several issues. First, there is a lack of understanding and coordination between technologists and lawyers within an organization. Second, there is no consistency between organizations due to the lack of universal data governance guidelines. Third, there is a lack of investment in omnibus data governance by organizations.

A. Lack of Coordination

The first issue that organizations face is that there is generally no overarching team responsible for monitoring data use by the organization. This is compounded by the fact that there is often a disconnect between the legal and IT departments. While the lawyers may focus on meeting the minimum legal requirements, technologists desire to push the envelope on technological development, many times without consideration of how such development will impact data subjects, let alone society. Additionally, the lack of coordination between the law department and technologists exposes the lack of understanding between each domain's specific expertise and scholarly perspectives. For instance, technologists have the expertise to understand how data is collected and used, but this knowledge may evade clear understanding by the attorneys assessing data use risks. Consequently, a coordinated effort often eludes effective internal data governance schemes. To balance against technological myopia,

[82] Randy Bean, *The CDO/CIO Dynamic: The Business-Of-Data Meets the Technology-Of-Data*, Forbes (Jan. 12, 2022, 8:47 AM), https://www.forbes.com/sites/randybean/2022/01/12/the-cdocio-dynamic-the-business-of-data-meets-the-technology-of-data/?sh=d409f7b43807.

interdisciplinary teams are necessary, but often lacking.[83] Even relying on a coordinated effort between law and tech is not enough. To counter ineffective data governance, robust input is needed from multiple stakeholders from throughout the organization.[84]

The outsourcing of data storage and management compounds the lack of coordination between in-house attorneys and technologists within a firm. As amassed data drastically increases within organizations, data governance is often outsourced to third parties, [85] and internal organizational policy moves to the background, [86] resulting in organizations having less control over data governance responsible data use.[87] When such efforts are outsourced, the company shifts control of data governance to third parties.

[83] *See* Tim Fountaine et al., *Building the AI-Powered Organization*, HARV. BUS. REV., Jul.–Aug. 2019, at 62 (explaining the benefit to interdisciplinary teams); *see also* Urs Gasser, *Recoding Privacy Law: Reflections on the Future Relationship Among Law, Technology, and Privacy*, 130 HARV. L. REV. F. 61, 65–66 (2016) ("A growing body of interdisciplinary research demonstrates the theoretical and practical promise of holistic analytical frameworks for a modern privacy analysis that incorporates recent research from fields such as computer science, statistics, law, and the social sciences.").

[84] Because data may be kept separately in different departments and used differently, internal input from all departments is needed to create a data governance program that can be applied throughout the organization.

[85] *45 Marketing Data Management Statistics 2020*, DATA SERVICES INC. (Aug. 12, 2020), https://www.dataservicesinc.com/newsletter/marketing-data-management-statistics/ (reporting that in in 2019, business spent $5.5 billion on in-house data management and $11.9 billion on third party data management).

[86] Kristen E. Eichensehr, *Data Extraterritoriality*, 95 TEX. L. REV. 145, 146 (2017) ("The era of cloud computing— where data crosses borders seamlessly, parts of a single file may exist in multiple jurisdictions, and data's storage location often depends on choices by private companies—raises new and difficult questions for States exercising enforcement authority.").

[87] See Christopher Tozzi, *When – and When Not – To Outsource Data Center Operations*, DATA KNOWLEDGE CENTER (Sept. 23, 2023), https://www.datacenterknowledge.com/solutions-and-suppliers/when-and-when-not-outsource-data-center-operations#close-modal (explaining how data outsourcing impacts data governance); George Lawton, *Data Governance vs. Information Governance: What's the Difference?*, TECHTARGET, (Nov. 30, 2021), https://whatis.techtarget.com/feature/Data-governance-vs-information-governance-Whats-the-difference (describing data governance as a framework for accountability for the management of data and related resources, including data ownership, quality, architecture, tooling, access, and security and information governance as a framework for accountability to ensure appropriate behavior and regulatory compliance in the creation, storage, use, sharing, protection, archiving, and deletion of information).

B. Lack of Uniform Standards

External policy influences especially lack standardization. The proliferation of unique data management practices in different departments of an organization limits interoperability and access by those within the organization. For example, records of individual human subjects such as employees, customers, or supplier representatives could be recorded or secured differently among various business units, such as by human resources (HR), sales, supply chain or other internal repositories. [88] Such a silo approach seems predictable given organizational sub-unit autonomy and performance metrics may suppress sharing. Repositories running different software or data standards fail to integrate quickly or accurately, complicating easy compliance with data governance requirements.[89]

Many scholars urge restraint in treating almost any new subject matter as too unique to borrow wholesale from stable, comprehensible, existing law when building new fields. By contrast, others urge early integration to avoid inward-focused myopia fostered by intra-disciplinary silo thinking. [90] Commentators express mounting concerns that the prospective risks in data and information operations require a more integrated approach to data governance - less sectoral and more omnibus.[91] This is typical with how emerging technologies create new disciplines. The Internet in the 1990s is a related case in point. The Internet suffered what Professor/Judge Easterbrook observed was no more than a transitory, "Law of the Horse" approach to integration.[92]

Despite optimistic statements by Ursula von der Leyen, data governance concepts are actually fragmenting, rather than coalescing. As scholars Douglas W. Arner, Giuliano G. Castellano, and Eriks K. Selga

[88] *See e.g.*, Craig Stedman & Jack Vaughan, *What is Data Governance and Why Does it Matter?*, TECHTARGET (Feb. 20, 2020), https://searchdatamanagement.techtarget.com/definition/data-governance.

[89] *See e.g.*, Uday S. Murthy et al., *The Effects of Information Systems Compatibility on Firm Performance Following Mergers and Acquisitions*, 34 J. INFO. SYS. 211 (2020) (discussing the data processing and integration issues that arise in the context of mergers and acquisitions).

[90] *See e.g.*, John W. Bagby, *Cyberlaw: A Forward*, 39 AM. BUS. L. J. 521 (2002).

[91] *See e.g.*, Nuala O'Connor, *Reforming the U.S. Approach to Data Protection and Privacy*, COUNCIL ON FOREIGN REL. (Jan. 30, 2018), https://www.cfr.org/report/reforming-us-approach-data-protection (describing the sectoral approach as narrowly designed to apply to particular industrial sectors).

[92] Frank H. Easterbrook, *Cyberspace and the Law of the Horse*, 1996 U. CHI. LEGAL F. 207, 207 (1996). *But see* Lawrence Lessig, *The Law of the Horse: What Cyberlaw Might Teach*, 113 HARV. L. REV. 501, 548 (1999) (disagreeing with Easterbrook).

explain "[data governance] fragmentation stems from the emergence of distinct data governance styles in the three largest economies, the United States, the European Union, and China."[93] There are also organizations that have created data protection standards, but no one paramount authority has emerged.[94]

C. Lack of Investment

A third issue is the failure of organizations to invest in data governance development. Although investment in AI and data analytics in general are on the uptick, [95] there is no corresponding increase in

[93] Douglas W. Arner et al., *The Transnational Data Governance Problem*, 37 BERKELEY TECH. L. J. 623, 628 (2022).

[94] *See Privacy*, OECD, https://www.oecd.org/sti/ieconomy/privacy.htm (last visited Nov. 9, 2023) (claiming the council's recommendation of privacy guidelines are recognized as "the global minimum standard for privacy and data protection"); *Data Protection and Privacy Laws*, WORLD BANK'S IDENTIFICATION FOR DEV. INITIATIVE, https://id4d.worldbank.org/guide/data-protection-and-privacy-laws, (regarding the GDPR as "setting a new threshold for international good practices") (last visited Nov. 9, 2023); G-20, *G-20 G20 Digital Economy*, at 7 (Aug. 24, 2018), https://www.g20.org/content/dam/gtwenty/about_g20/previous_summit_documents/2018/Digital_economy_ministerial_declaration.pdf (describing the G20 Digital Government Principles as principles to "facilitate an inclusive and whole-of-government approach to the use of [information and communication technology] and assist governments in reshaping their capacities and strategies, while respecting the applicable frameworks of different countries, including with regards to privacy and data protection"); Asia-Pacific Economic Cooperation [APEC], *APEC Privacy Framework*, at 4 (Dec. 2005), https://www.apec.org/docs/default-source/publications/2005/12/apec-privacy-framework/05_ecsg_privacyframewk.pdf?sfvrsn=d3de361d_1 (saying that to enable global organization in APEC economies to "develop and implement uniform approaches within their organizations for global access to and use of personal information"); CROSS BORDER PRIV. RULES SYS., http://cbprs.org/ (a certification program for organizations within the member countries which identifies best practices for data transfers) (last visited Nov. 9, 2023); *Privacy Framework*, NAT'L INST. OF STANDARDS AND TECH., https://www.nist.gov/privacy-framework/privacy-framework (a voluntary set of privacy standards created by the national Institute of Standards and Technologies released on Jan. 16, 2020) (last visited Nov. 9, 2023); ISO/IEC 27701:2019 (2019), https://www.iso.org/standard/71670.html (international standards for privacy information management systems); *ISO 27001, the International Information Security Standard*, INT'L INFO. SECURITY STANDARD, https://www.itgovernance.eu/en-ie/iso-27001-ie ("international standard that describes best practice for an ISMS (information security management system).") (last visited Nov. 9, 2023).

[95] Michael Chui et al., *The State of AI in 2021*, MCKINSEY (Dec. 8, 2021), https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/global-survey-the-state-of-ai-in-2021 (analyzing a survey of 1,843 companies

investment in data governance.[96] Traditional data governance continues to focus on data management and monetization. This short-term, managerial myopia is incentivized almost solely by two things: first, corporate ethos to maintain stock price as presumably driven by earnings per share (EPS), and second, quarterly bonus formulas largely dependent on near-term quarterly financial performance. These incentives necessarily deemphasize data governance investment because immediate benefits are not manifest.[97] The quarterly EPS focus is likely preventing data governance investment in development and deployment. Managerial myopia plagues advocates of investment in security, privacy protection, information infrastructure enhancement, and the other components of data governance.[98] Because the risks associated with lack of data governance, such as a data breach, are remote and uncertain, organizations disregard them.[99] When future benefits remain uncertain or less tangible, they make for a less compelling or sizable present investments.[100] As such, organizations do not prioritize investment in governance components, such as preventing cybersecurity incidents.[101]

Additionally, companies are often not held accountable for their data governance failures. The lack of omnibus privacy law in the U.S. and underfunding of regulatory agencies significantly skew the probability of being caught in the favor of the organization, unless the organization

revealed that "[n]early two-thirds [of those surveyed] say their companies' investments in AI will continue to increase over the next three years").

[96] Tami Frankenfield et al., *The AI Era Is Here. Is Your Data Governance Ready?*, WALL ST. J., (April 6, 2021) https://deloitte.wsj.com/cmo/2021/04/06/the-ai-era-is-here-is-your-data-governance-ready (arguing definition of data governance practice lags change in the competitive landscape; concluding interdisciplinary improvements needed given that data scientists are rarely involved in early data strategy discussions and advocating D/I/Gov team composition decisions should consider representation from C-suite stakeholders for sponsorship; subject matter experts with domain expertise who can validate data and resulting insights; data analysts to consult on data structures and data processing tools; data scientists for expertise in science, IT, math, and statistics as well as domain knowledge; IT to consult on data infrastructures and security; and legal counsel).

[97] *See generally* Mei Cheng et al., Earnings Guidance and Managerial Myopia (Nov. 2005) (unpublished), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=851545; Arthur G. Kraft et al., *Frequent Financial Reporting and Managerial Myopia*, 93 ACCT'G REV. 249 (2018).

[98] Jay P. Kesan & Carol M. Hayes, *Liability for Data Injuries*, 2019 U. ILL. L. REV. 295, 357–58 (2019).

[99] *Id.*At 316.

[100] Brigitte C. Madrian, *Applying Insights from Behavioral Economics to Policy Design*, 6 ANNU. REV. ECON. 663, 680 (2014).

[101] John W. Bagby, *Regulation and Public Policy for Accounting Professionals*, SECURITY4ACCOUNTANTS, 1, 10 (2021).

suffers a significant data breach.

Besides the lack of legal mandates for aspects of data governance, there are also few consequences for companies who violate the laws.[102] In *TransUnion v. Ramirez*, for example, the Supreme Court gutted the possibility of a private cause of action for data breaches that do not result in individual financial loss, exacerbating the pact of the much-criticized *Spokeo* decision.[103] By continuing to believe a plaintiff has not suffered harm unless their identity is stolen and they suffer financial loss, the court demonstrates its complete lack of understanding of how technology, the internet, and social media work. Apparently, being identified as a terrorist is not harmful. The nonexistence of a legal mandate or possible relief from most data breaches undermines organizations' incentives to invest in a robust data governance program as plaintiffs have little recourse against a company following a data breach. In the following section we discuss the potential pathways to advance the concept of data governance.

III. PATHWAYS TO DATA GOVERNANCE

As explained, part of the reason the United States is trailing other regions in providing appropriate protections for data subjects is the absence of omnibus federal data protection laws. Different regions, industries, and scholars diverge on what data governance should look like and what it should entail. As explained, because data governance is often left to IT or third parties, potential harms to data subjects are left out of the governance equation.[104] The following sections explore various data governance regimes, including notice and consent, AI governance models

[102] Elizabeth Earle Beske, *Charting a Course Past Spokeo and TransUnion*, 29 GEO. MASON L. REV. 729 (2022). This is because plaintiffs lack standing for violations of their privacy or data security breaches until they can show harm.

[103] TransUnion LLC v. Ramirez, 141 S.Ct. 2190, 2200 (2021) (ruling that even though an incorrect credit report identified the plaintiff as a terrorist, there was no "concrete harm." The only plaintiffs incorrectly identified as a terrorist who had standing under the Fair Credit Act were the ones whose reports were disclosed to third parties by TransUnion.*)*; *see also* Joshua Briones et al., *Supreme Court Decision May Have Significant Implications for Data Breach and Privacy Class Actions*, SECURITY (July 2, 2021) ("Beyond the facts of Ramirez, the Court's decision will impact data security and privacy class action litigation by providing defendants with a more powerful defense in cases where alleged privacy and security violations do not result in a disclosure of information resulting in any tangible harm. In the data breach context, if private information was not published, or if the data was not used in any fraudulent way, defendants may be able to argue that the class has not been harmed and, therefore, lacks standing."); Beske, *supra* note 102.

[104] *See* discussion *supra* Part II.A.

and ethics and fair information practices (FIPs).

A. Notice and Consent

Although the GDPR and CCPA are eons ahead of United States federal law, as discussed in Part II, they rely on a notice and consent model of governance. However, as the World Economic Forum's report titled *Redesigning Data Privacy* explains: "Consent has become illusory and, through its current design and deployment, does not always operate in expected, or at times even logical, ways."[105] Once a data subject grants consent, the data collector becomes the "gateway for everything that happens in the future . . . far beyond what could be envisioned."[106] While a company may indicate what information is collected and the data subject consents to the use of their data, it is doubtful that the data subject truly understands what they are agreeing to. A 2019 Pew Research survey revealed that only nine percent of people in the United States read privacy policies prior to agreeing to them.[107] One Technology columnist reported that when he looked up all of the privacy polices relating to the apps on his phone, they totaled nearly 1 million words, quipping "'War and Peace' is about half as long."[108] He also explained that laws requiring consent may have made things worse as people will click on every pop-up window to get through to the information or technology they need.[109] FTC Commissioner Rebecca Kelly Slaughter agrees that the current system of notice and consent does not provide meaningful choice as to how one's data is shared.[110]

The GDPR, which uses a notice and consent model, also provides

[105] WORLD ECON. F., REDESIGNING DATA PRIVACY: REIMAGINING NOTICE & CONSENT FOR HUMAN-TECHNOLOGY INTERACTION 1, 4 (2020) [hereinafter *Redesigning Data Privacy*].

[106] *Id.* An FTC Privacy Con presentation noted that companies like CNN, Bloomberg, and Wells Fargo collect information from your phone such as motion, orientation and light. Anupam Das, *The Web's Sixth Sense: A Study of Scripts Accessing Smartphone Sensors*, PRIVACYCON (2019), https://www.ftc.gov/system/files/documents/public_events/1415032/privacycon2019_das_study_of_scripts_accessing_smartphone_sensors.pdf.

[107] Brooke Auxier et al., *Americans' Attitudes and Experiences with Privacy Policies and Laws*, PEW RSCH. CTR. (Nov. 15, 2019), https://www.pewresearch.org/internet/2019/11/15/americans-attitudes-and-experiences-with-privacy-policies-and-laws/.

[108] Geoffrey A. Fowler, *I Tried to Read All My App Privacy Policies. It Was 1 Million Words*, WASH. POST (May 31, 2022, 7:00 A.M.), https://www.washingtonpost.com/technology/2022/05/31/abolish-privacy-policies/.

[109] *Id.*

[110] *Id.*

data subjects with a set of rights.[111] However, reliance on the notice and consent model of data protection is misplaced for a number of reasons. First, as discussed, consent essentially involves clicking on a box to quickly get to the website the data subject is seeking to view.[112] The notice and consent model places the burden on the data subject to determine what the data user is collecting and with whom they are sharing the data.[113] Second, neither the GDPR nor the CCPA seem to address the ease with which individuals can be identified from data sets, permitting an enormous loophole in the regulation for de-identified information.[114] Third, state statutes, like the CCPA, only address the sale of data, not the sharing of data.[115] The inability of regulation to provide robust protections for data subjects makes it inadequate guidance for data governance.

The World Economic Forum report also notes that the problem with notice is not just the length of the privacy policies, but their understandability, the sheer amount of privacy policies encountered by users, and their take-it-or-leave-it terms.[116]The conclusion that notice is ineffectual and consent is not meaningful is also echoed by privacy scholars Daniel Solove and Woodrow Hartzog.[117] Notice and consent also does not address issues with emerging technologies such as AI or predictive analytics.[118] In terms of data governance, it is an insufficient model, despite its widespread use. In the next section we explore AI governance models as a potential source.

[111] Scott Jordan, *Strengths and Weaknesses of Notice and Consent Requirements under the GDPR, the CCPA/CPRA, and the FCC Broadband Privacy Order*, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3894553 at 4.

[112] Pike, *supra* note 23, at 717 ("Notice and check-the-box consent are generally considered lacking as a meaningful consumer protection.").

[113] *GDPR & CCPA: Opt-Ins, Consumer Control, and the Impact on Competition and Innovation: Hearing Before the S. Comm. on the Judiciary*, 116th Cong. 13 (2019) (statement of Michelle Richardson, Director of Privacy & Data, Center for Democracy & Technology) ("Existing privacy regimes including the GDPR and CCPA rely too heavily on the concept of notice and consent, placing an untenable burden on consumers and failing to rein in harmful data practices.").

[114] *Id.* at 717, 719 (explaining that these regulations only apply to identifiable data).

[115] *Id.* at 719 ("[The CCPA] protects only those instances where a consumer's personal data are sold and not when personal data are given away for free.").

[116] WORLD ECON. F., *supra* note 108, at 8.

[117] *Id.* at 10.

[118] Lori Cameron*, Artificial Intelligence and Consent: Navigating The Ethics of Automation and Consumer Choise,* IEEE COMPUTER SOCIETY*,* https://www.computer.org/publications/tech-news/research/ai-and-the-ethics-of-automating-consent *(last visited Nov. 19, 2023).*

B. AI Governance Models

AI governance models receive a lot of attention because AI has been reported on in a way that has created anxiety and fear.[119] Model AI governance regimes, termed "Responsible AI," "Trustworthy AI," or "Ethical AI," often center around algorithms being explainable and transparent. While the EU has proposed the Artificial Intelligence Act[120] to achieve these goals, not every jurisdiction will be willing to put such stringent limitations on their burgeoning tech industries. Although the GDPR may serve as a model for data subject rights, it is doubtful that the Artificial Intelligence Act will do the same.[121]

A second issue with AI governance models is that the difference between the legal and technical understanding of what it means for AI to be explainable, transparent, and ethical presents an enormous barrier. For example, a legal definition of *explainability* may be based on concepts of obviousness, accessibility, and clarity concerning how AI actually produced particular results actually used in specific decision-making.[122] The technical definition of explainability would more likely involve describing the data accessed, steps taken, and computations made by the software code when fed particular data, thereby subjecting it to a potential

[119] Faiz Siddiqui, *Elon Musk Debuts Tesla Robot, Optimus, Calling It a 'Fundamental Transformation*,' WASH. POST (Oct. 1, 2022), https://www.washingtonpost.com/technology/2022/09/30/elon-musk-tesla-bot/ ("Musk has said he fears artificial intelligence could one day outsmart humans and endanger us, citing AI as the biggest threat to civilization.").

[120] *Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence,* COM (2021) 206 final (Apr. 21, 2021).

[121] Martin Anderson, *The Failings of the Draft EU Artificial Intelligence Act*, UNITE AI (Sept. 10, 2021), https://www.unite.ai/the-failings-of-the-draft-eu-artificial-intelligence-act/ ("A new legal critique of the European Union's draft 'AI Act' levels a wide array of criticisms at the proposed regulations released in April, concluding that much of the document is 'stitched together' from scarcely applicable 1980s consumer regulation; that it actually promotes a deregulated AI environment in Europe, rather than bringing the sector under coherent regulation; and – among a slew of other criticisms – that the proposals map out a future regulatory AI framework that has 'little sense and impact.'").

[122] *See* Ashley Deeks, *The Judicial Demand for Explainable Artificial Intelligence*, 119 COLUM. L. REV. 1829 (2019) (defining explainable artificial intelligence (xAI) as capable of revealing to judges how algorithms reach their conclusions or predictions thus adequately informing judges who are ruling on the fairness of implemented recommendations or actual decision-making based on AI systems; alternatively *explaining* this to individual subjects who are impacted by AI decision-making, thereby overcoming black box opacity; and *advocating* judges demand explainability when ruling on algorithm fairness, accuracy and reasonable bases for decision-making).

forensic audit.[123] In addition, different industries and public endeavors may need to employ different interpretations of explainability. For example, in industries where the costs of failure from AI decision-making are high, such as in medicine, the testing of nuclear weapons, or capital punishment sentencing, the explainability regime may need to be more robust and would likely be subject to early and repeated revelations.[124]

While *transparency* means understanding how and why a decision was made, a technician might design an algorithm using different taxonomies than would be used when reviewed under a legal review or regime.[125] As with almost every inter-disciplinary endeavor, each contributing field brings to the table differing perspective and unique language derived from their primary interests and activities. Generally, technical AI developers are less accustomed to accommodating social impacts of AI decision-making than they are to achieving innovative design, performance efficiencies, and advanced application of their coding expertise.[126]

AI governance models have arisen from a variety of sources. The UN, EU and a number of countries have proposed models as policy statements, guidance, or enforceable regulations.[127] Non-profits, academic centers, and standards organizations have also suggested how AI could be ethically managed.[128] Even Big Tech corporations have published their own AI guidelines.[129] One of the first AI governance frameworks came out of Singapore in 2019 and was described as "a sector-, technology- and algorithm-agnostic framework, which converts relevant ethical principles to implemental practices in an AI deployment process so that organizations

[123] *See generally* Adrien Bibal et al., *Legal Requirements on Explainability in Machine Learning*, 29 A.I. & L. 149–69 (2021).

[124] *See generally* Milda Pocevičiūtė et al., *Survey of XAI in Digital Pathology*, A.I.. AND MACH. LEARNING FOR DIGIT. PATHOLOGY 56 (2020).

[125] Łukasz Górski & Shashishekar Ramakrishna, *Explainable Artificial Intelligence, Lawyer's Perspective*, 2021 PROC. 18TH INT'L CONF. A.I. & L. 60.

[126] *See* Diane Coyle, *The Tensions Between Explainable AI and Good Public Policy*, BROOKINGS (Sept. 15, 2020), https://www.brookings.edu/techstream/the-tensions-between-explainable-ai-and-good-public-policy/ (arguing baked-in bias from non-representative data learned by algorithms assures unfair and biased outcomes, AI always represents a trade-off between performance and explainability, and the policy principles are nearly always compromises that AI algorithms do not utilize well because such dilemmas permit correlative phenomena to dominate the recommendations produced).

[127] See *supra* notes 68–73 and accompanying text.

[128] Brain John Aboze, *Demystifying AI Governance*, MLCON2.0, https://cnvrg.io/ai-governance/ (last visited Nov. 10, 2023).

[129] Sebastian Klovig Skelton, *AI Experts Question Tech Industry's Ethical Commitments*, COMPUT. WKLY. (Oct. 31, 2022), https://www.computerweekly.com/feature/AI-experts-question-tech-industrys-ethical-commitments.

can operationalize these principles."[130]

In January 2020, Singapore's Personal Data Protection Commission released its second edition of the *Model Artificial Intelligence Governance Framework* (*Model Framework*) to promote responsible AI deployment and use. The Model Framework 1st ed. was initially released a year earlier for critical exposure at the World Economic Forum in Davos. The *Model Framework* establishes principle-based standards for AI development, expressed as *aspirational* guiding principles that define terms and aspire towards universal applications with appropriate protections. The two overarching principles are that (1) Organizations using AI in decision-making should ensure that the decision-making process is explainable, transparent and fair, and (2) AI solutions should be human-centric.[131] Hard, de jure law like the EU's Artificial Intelligence Act and soft law like Singapore's Model Framework may not have the intended effect. "The results [of these initiatives] are statements of principles or values based on abstract and vague concepts, for example commitments to ensure AI is 'fair' or respects 'human dignity', which are not specific enough to be action-guiding."[132]

A number of academic institutions have also begun issuing AI governance models and reports. The AI Governance Research Group out of Oxford has put together some guidance on AI.[133] The Berkman Klein Center at Harvard Law has also issued several reports on AI governance. Of particular note is their *Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI*, which examines 36 "prominent" AI Guidelines to visualize commonalities and 47 principles supporting these themes.[134] The goal was to provide a

[130] Personal Data Protection Commission. (2020). *Model Artificial Intelligence Governance Framework – Second Edition, https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/resource-for-organisation/ai/sgmodelaigovframework2.pdf.*

[131] *Id.* at 15.

[132] Brent Mittelstadt, *Principles Alone Cannot Guarantee Ethical AI*, 1 NATURE MACH. INTEL. 501, 505 (2019); JESS WHITTLESTONE ET AL., ETHICAL AND SOCIETAL IMPLICATIONS OF ALGORITHMS, DATA, AND ARTIFICIAL INTELLIGENCE: A ROADMAP FOR RESEARCH (2019).

[133] Carina Prunkl et al., *Institutionalizing Ethics in AI Through Broad Impact Requirements*, 3 NATURE MACH INTEL. 104 (2020).

[134] JESSICA FJELD ET AL., PRINCIPLED ARTIFICIAL INTELLIGENCE: MAPPING CONSENSUS IN ETHICAL AND RIGHTS-BASED APPROACHES TO PRINCIPLES FOR AI (2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3518482. This source expends considerable detailed discussion of several principles underlying each of the eight themes that comprise the bulk of this 71 page report. For example, the first theme of privacy is evaluated by themes such as consent, control over the use, ability to restrict processing, right to rectification, right to erasure, privacy by design, and recommends data protection laws, et.al. As the authors discuss in their

collection of information to "push the fractured, global conversation on the future of AI toward consensus."[135] While these efforts should be commended, fractured adoption of AI governance practically assures inconsistent understanding and widely varying compliance.

Interestingly, the Google report on AI governance indicates that, "To date, self- and co-regulatory approaches informed by current laws and perspectives from companies, academia, and associated technical bodies have been largely *successful at curbing inopportune AI use* [emphasis added]."[136] Despite this statement, it is very clear that self-regulation is not in the best interests of data subjects. As with many of the theoretical schema in data protection, AI and other precatory frameworks, the terminology varies by time, nationality or culture producing the report and the perceived relevance of the matter addressed. This discipline is developing such that later studies refine, define and distinguish using more detail including exemplars derived from real incidents and speculative dialog. Although Google acknowledges the need for some governmental regulation, unsurprisingly, it makes no clear recommendations. These AI models make broad statements using undefined terms, leading different organizations to reach varying conclusions. They also provide lengthy lists of principles, but do not provide implementation examples.[137] They do, however, serve an important function in normalizing the function of governance as protecting data subjects, not just managing data. The following subsection explores ethics and FIPs as a potential model for data governance.

Principle and Theme Selection Methodology section at 15, a dataset of 36 documents were sampled and normed by the research team. Generalization of the relationship between themes and principles appears to be that themes are major top-level issues and illuminated by component subjects that suggest particular implementations of protective strategies. This is exemplified in the privacy theme above and its implementation themes.

[135] *Id.* There are also a number of other non-profit and academic organizations with valuable information. *See, e.g.*, *Featured Research*, A.I. CTR. FOR THE GOVERNANCE OF A.I., https://www.governance.ai/research (NFP Wales and England); *About*, INFO. SOC'Y PROJECT https://law.yale.edu/isp/about (Yale Law School); *Research*, STANFORD UNIV. HUMAN-CENTERED A.I., https://hai.stanford.edu/research (multiple departments across Stanford University).

[136] *Perspectives on Issues in AI Governance*, GOOGLE, https://ai.google/static/documents/perspectives-on-issues-in-ai-governance.pdf (last visited Nov. 9, 2023).

[137] Mittelstadt, *supra* note 137, at 1 ("AI Ethics initiatives have thus far largely produced vague, high-level principles and value statements which promise to be action-guiding, but in practice provide few specific recommendations and fail to address fundamental normative and political tensions embedded in key concepts (e.g. fairness, privacy).") (citation omitted).

C. Ethics and FIPs

As discussed, existing law is insufficient to protect data subjects (and society in general) from data use abuses.[138] In addition, the lack of standardization and understanding of the components of data governance has created a gap in protection. Ethical frameworks regarding the use of data were created, like AI governance frameworks, to address this deficiency. Most ethical data frameworks evolved from the Fair Information Practices (FIPs) developed by the Department of Health Education and Welfare in the 1970s. When government agencies began using computers, there was a concern that this would lead to privacy issues for U.S. citizens.[139] The FIPs were designed to ensure proper ethical boundaries were in place to address the collection, use and sharing of personal information about people by the government. These principles were the basis of the U.S. Privacy Act of 1974 (regarding the government's collection and storing of data) and the GDPR (which applies to both government and private industry).[140] The FIPs have also been adopted by the FTC as the five core principles of privacy protection.[141] However, the FIPs, and later regulations which were based on the notice and consent mechanism, provides insufficient protection in today's data economy.[142]

While these mechanisms stem from the Fair Information Practice Principles (FIPs) crafted in the 1970s in the U.S.,[143] the U.S. has failed to

[138] *See supra* Part 1.A.

[139] U.S. DEP'T. OF HEALTH, EDUCATION AND WELFARE, RECORD COMPUTERS AND THE RIGHT OF CITIZENS 48–50 (1973), https://epic.org/wp-content/uploads/2021/11/1973-hew-report.pdf.

[140] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 U.C.L.A. L. REV. 1701, 1733–34 (2010) ("The FIPS have been enormously influential, inspiring statutes, law review articles, and multiple refinements.").

[141] Houser & Sanders, *supra* note 24, at 834. ("The main tenets of the FIPs are that (1) there should be no secret data collection systems; (2) there should be a way for data subjects to find out what information is in their records and how it is used; (3) data collected for one purpose should not be used for another without user permission; (4) the data subject should have the ability to correct inaccuracies; and (5) the data collector should keep reliable records and protect them.").

[142] Ewa Janiszewska-Kiewra et al., *Ethical Data Usage in an Era of Digital Technology and Regulation*, MCKINSEY (Aug. 26, 2020), https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/tech-forward/ethical-data-usage-in-an-era-of-digital-technology-and-regulation ("The European Union's General Data Protection Regulation (GDPR), for instance, works well as a breach-notification system but has not been consistent in imposing penalties to deter company behavior that violates customer data privacy.").

[143] U.S. DEP'T. OF HEALTH, EDUCATION AND WELFARE, *supra* note 140.. See also, Houser & Sanders, supra note 24 at 834-835.

amend its privacy laws. This has resulted in enormous gaps and loopholes in U.S. privacy law of which both government and private industry take advantage.[144] The EU, on the other hand, has continually been involved in debating and crafting directives, guidance, and regulations to address risks from advances in technology.[145] As Professor Voss explains, these data protection principles, which have been incorporated into EU laws, include: data quality, purpose limitation, integrity and confidentiality, transparency, rights of the data subject, accountability, and lawfulness of processing.[146]

Experts in the field have offered a number of ethical models, recognizing that complying with the law and ensuring the ethical use of data are not the same thing. According to Harvard Professor Dustin Tingley, "Data ethics asks, 'Is this the right thing to do?'"[147] Rather than focusing on how to create a new technology, it asks "*should* we create this new technology?" In 2016, Professors Luciano Floridi and Mariarosaria Taddeo described data ethics as a "new branch of ethics."[148] This broader viewpoint is more helpful than the hundreds of AI governance models for several reasons. First, it acknowledges that a high-level (macro) approach is vital, while clarifying that balancing the development of data sciences with the protection of human rights is no easy task.[149] The article goes on to predict, quite accurately, that "failing to advance both the ethics and the science of data, would have regrettable consequences."[150] Research into

[144] See generally, Houser & Voss, supra note 26 (regarding how private industry's business model flourishes in the U.S. due to the lack of data use restrictions) and Houser & Sanders, *supra* note 24 (regarding how the failure to update data use law results in the government's use of citizen data inconsistent with the 1974 Privacy Act).

[145] *See* W. Gregory Voss, *Cross-Border Data Flows, the GDPR, and Data Governance*, 29 WASH. INT'L L.J. 485, 520 (2020).

[146] *See* W. Gregory Voss, *Obstacles to Transatlantic Harmonization of Data Privacy Law in Context,* 2019 U. ILL. J.L. TECH. & POL'Y 405*,* 421–22 (2019).

[147] Catherine Cote, *5 Principles of Data Ethics for Business*, HARVARD BUS. SCH. ONLINE (Mar. 16, 2021), https://online.hbs.edu/blog/post/data-ethics#:~:text=What%20Is%20Data%20Ethics%3F,and%20how%20it%20affects%20individuals.

[148] Luciano Floridi & Mariarosario Taddeo, *What Is Data Ethics?*, 374 PHIL. TRANSACTIONS OF THE ROYAL SOC'Y A: MATHEMATICAL, PHYSICAL AND ENG'G SCIS. 1, 4 (2016) ("This theme issue has the founding ambition of landscaping data ethics as a new branch of ethics that studies and evaluates moral problems related to data (including generation, recording, curation, processing, dissemination, sharing and use), algorithms (including artificial intelligence, artificial agents, machine learning and robots) and corresponding practices (including responsible innovation, programming, hacking and professional codes), in order to formulate and support morally good solutions (e.g. right conducts or right values).").

[149] *Id.*

[150] *Id.* at 2. Given what we have witnessed with election interference, Cambridge

data ethics reveals a burgeoning, yet ill-defined, multi-disciplinary field drawing some of its early roots from privacy law and professional confidentiality requirements. Most groundbreaking work, like the data ethics discussed here, must initially struggle to overcome a widespread lack of consensus, enlist serious interdisciplinary participants that do not initially share understanding many basic concepts and then persist until some success is achieved.

In 2021, Professor Dennis Hirsch, along with several of his colleagues from Ohio State University, surveyed 21 organizations regarding their conceptions of data ethics due to the recognition that current law fails to provide adequate guidance.[151] For example, the use of de-identified data by corporations arguably skirts current privacy law[152] and the complexity of AI often defies effective forensic inspection. While major tech companies may adopt ethical frameworks that transcend privacy law, it is likely that this is an effort to engender trust with those who use their services while also seeking to preempt future regulation.

The study of ethics has been around for millennia and is understood as aspirational rather than prescriptive. However, because most ethical models stem from academia, specifically from fields like theology and philosophy, they tend to be behavioral signals rather than actionable information. The lack of specificity in these modelsfail to provide appropriate guidance to a technologist charged with ethical data use. Telling a technologist, "Do No Evil," such as Google did,[153] has no true impact. The problem is not creating evil algorithms or sharing data with the intent to harm, but rather the failure to take into consideration the ***potential*** harms implicit in the technologies being created. Certainly, ethical evaluation should not be left to technologists. Instead, ethical considerations should be a component of data governance. The following section forecasts the next generation of data governance as a stewardship model.

Analytica, the unlawful detention of those based on faulty image recognition software, mass surveillance by the government, and loss of agency, their concern has come to fruition. Houser, *supra* note 24, at 476–82.

[151] *See generally* DENNIS D. HIRSCH ET AL., BUSINESS DATA ETHICS: EMERGING TRENDS IN THE GOVERNANCE OF ADVANCED ANALYTICS AND AI (2021), https://ssrn.com/abstract=3828239.

[152] *Id.* at 22–23 (explaining that U.S. data use law is tied to personally identifiable information, and would not expressly apply to de-identified information).

[153] Bob Evans, *Google Needs to Drop Its "Do No Evil" Thing*, FORBES (Sept. 2, 2011), https://www.forbes.com/sites/sap/2011/09/02/google-needs-to-drop-its-do-no-evil-thing/.

IV. NEXT-GEN DATA GOVERNANCE

Although the ability to collect and analyze data has resulted in valuable discoveries, it has also resulted in harm to those supplying the data–the data subjects. As such, a debate has unfolded over the proper way to protect data subjects without negatively impacting innovation. Current discussions lack actionable guidance for data governance improvement. In Part I, we analyzed the sources of data governance, including hard, de jure law, soft law, and institutional policies and procedures. Due to the extreme fragmentation within and among those sources, we determined that there is a lack of a uniform understanding of what data governance should entail, resulting in insufficient protection for data subjects. In Part II, we reviewed the lack of coordination between technologists and attorneys, the way that differing regulatory regimes and ideologies have led to a lack of uniformity, and how a lack of incentive to adopt data governance policies and procedures that protect data subjects have all contributed to the problem with considering data governance as data management. Although laws can serve to provide standards, the U.S. regime fails to meet even the minimum benchmark at the federal level. In addition, law cannot be developed quickly enough to address advances in technology and new uses for data. In Part III, we examined various models to determine if any could provide the necessary guidance. We quickly determined that notice and consent is an incomplete device, and most AI and ethical models are too general to serve as actionable data governance mechanisms. Despite its importance, data governance remains an under-researched field.[154] In the next section, we explore the potential for a data governance evolution from a paternalistic model to one that takes a stewardship approach.

A. Medical Code of Ethics

Medical ethics, particularly those around patient data and medical research, provides a great foundation for a data governance scheme. Furthermore, it is one of the earliest and most studied ethical canons on the planet.[155] Although many associate the Hippocratic oath with the statement "First do no harm," the actual oath emerged in 400 BCE as a series of pledges to behave ethically in the practice of medicine and did not include that phrase.[156] It includes promises to help, to not harm, to not

[154] MELANIE MCCAIGA & DAVAR REZANIA, A SCOPING REVIEW ON DATA GOVERNANCE 2 (2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3882450.

[155] Mittelstadt, *supra* note 137, at 2 ("This convergence of AI Ethics around principles of medical ethics is opportune, as it is historically the most prominent and well-studied approach to applied ethics.").

[156] Robert H. Schmerling, *First, Do No Harm*, HARV. HEALTH PUBL'G (Jun. 22, 2020), https://www.health.harvard.edu/blog/first-do-no-harm-201510138421.

engage in conduct in which they are not proficient, and to maintain confidentiality.[157] It further acknowledges that harms can arise from malpractice and the indiscriminate disclosure of highly private facts.[158]The first codification of this oath occurred in 1803 by Thomas Percival, an English physician.[159] When the American Medical Association was formed in 1947, it issued its first Code of Medical Ethics [the Code].[160] As the Code has evolved, it has expanded from physician conduct to addressing the expanded role of the medical profession, including medical research and the importance of patient privacy. The Code is also instructive as it has evolved over thousands of years. As such, it serves an important normative and practical function.

This sensible base ought to apply to data use. Health information is considered sacrosanct by many. As such, it is easy to understand how and why data governance and the ethical sharing of data is so important to the medical field and why it has always been a part of medical ethics. The Code has advantages over the previously discussed models like increased specificity, interpretive guidance, built-in accountability, and overarching concept of stewardship. It contains nine principles of medical ethics.[161] The Opinions relating to each of the principles furnish explanations and designate different levels of ethical obligations. For example, "must" means that the action is ethically required of the physician, while "should" indicates a best practice or recommendation. Although the Code is a set of guidelines, not law, the AMA recognizes that circumstances may require

[157] *Greek Medicine*, NATIONAL INSTITUTE OF HEALTH, https://www.nlm.nih.gov/hmd/greek/greek_oath.html (last updated Feb. 7, 2012). Some have even suggested a Hippocratic oath for data scientists. *See, e.g.*, Lucy C. Erickson et al., *It's Time for Data Ethics Conversations at Your Dinner Table*, BLOOMBERG (Mar. 23, 2018), https://www.bloomberg.com/company/stories/time-data-ethics-conversations-dinner-table/ ("One idea that has gained traction is the need for a 'Hippocratic Oath' for data scientists. Just as medical professionals pledge to 'do no harm,' individuals working with data should sign and abide by one or a set of pledges, manifestos, principles, or codes of conduct."); Tom Simonite, *Should Data Scientists Adhere to a Hippocratic Oath?*, WIRED (Feb. 9, 2018), https://www.wired.com/story/should-data-scientists-adhere-to-a-hippocratic-oath/ ("Microsoft released a 151-page book last month on the effects of artificial intelligence on society that argued 'it could make sense' to bind coders to a pledge like that taken by physicians to 'first do no harm.'").

[158] This is quite different from the Facebook (now Meta) motto, "Move fast, and break things."

[159] Sara Patuzzo et al., *Thomas Percival. Discussing the Foundation of Medical Ethics*, 89 ACTA BIOMED 343, 343 (2018).

[160] Robert M. Gellman, *Prescribing Privacy: The Uncertain Role of the Physician in the Protection of Patient Privacy*, 62 N.C. L. REV. 255, 267–68 (1984).

[161] *Code of Medical Ethics, Principles,*, AM. MED. ASS'N, https://code-medical-ethics.ama-assn.org/principles (last visited Nov. 19, 2023).

physicians to deviate. The more stringent the requirement in the Code, the stronger the justification needed to deviate. However, obligations indicated by the word "must" can only be violated in very rare circumstances. This flexibility is also needed for data governance.

Using the Code as a guide for data governance has been suggested by various scholars with good reason.[162] Medical ethics has evolved to consider patient care over the physician's interests. In the 1970s, an important ideological shift occurred due to the emergence of serious medical ethical issues, namely the need to remove decision-making from individual physicians. First, Associated Press reporter, Jean Heller, exposed the horrific Tuskegee Syphilis Experiment where black male patients suffering from syphilis were left untreated for decades for "research purposes" without their knowledge or consent. [163] Second, advancements in technology provided new treatment options, such as the mechanical ventilator, which allowed the harvesting of organs from patients without brain function. This signaled the need to remove decision-making from individual physicians and provide ethical guidance that could be scaled for the entire profession prompting a shift from a paternalistic model to one of stewardship. With respect to patient data, ethical rules involving confidentiality and the sharing of data arose out of this stewardship model with the physicians who possessed patient data being charged with the protection and proper handling of the data. By removing unguided fiat regularly practiced at the practitioner level to broad, science-based minimum standards is systematic as some might argue paternalism is also an ad hoc, situational decision-making that is standardized to remove situational bias of unaudited decisions with consensus developed by well-reviewed independent and objective judgments.

Similarly, decisions about the use of data should not be left to those collecting, using or sharing data, or creating new technologies. Without a robust legal framework, there is an urgent need for guidance. Companies must be able to point to some set of standards when making decisions about data use. The current model has produced real world

[162] *See generally*, Ali Abbas et al., *A Hippocratic Oath for Technologists*, *in* NEXT GENERATION ETHICS: ENGINEERING A BETTER SOCIETY 71 (Nov. 2019); Elaine Sedenberg & Anna Lauren Hoffmann, *Recovering the History of Informed Consent for Data Science and Internet Industry Research Ethics*, CORNELL UNIV. (Sept. 12, 2016 4:54 AM), https://arxiv.org/abs/1609.03266; Mittelstadt, *supra* note 137; Carissa Véliz, *Three Things Digital Ethics Can Learn from Medical Ethics*, 2 NATURE ELEC. 316 (2019).

[163] Jean Heller, *AP WAS THERE: Black Men Untreated in Tuskegee Syphilis Study*, AP NEWS (May 10, 2017), https://apnews.com/article/business-science-health-race-and-ethnicity-syphilis-e9dd07eaa4e74052878a68132cd3803a.

harms.[164] Similarly to the way that the Code shifted from paternalism to patient agency, the concept of data governance must be broadened to include the consideration of data subjects as well as expand the obligations of data collectors and users. In the following section, we explain how stewardship differs from management.

B. Stewardship

The traditional view of data governance as data management provides little protection for those whose data has been collected, analyzed, shared, and sold. The data being managed by organizations and governments *does not belong to them*, it is collected from data subjects.[165] In other areas of the law, when an entity has control over something that belongs to someone else, a relationship develops.[166] This relationship is one of stewardship. Under stewardship theory, "[a] steward is one who takes on the responsibility of *caring for* something on behalf of another person or group of people [emphasis added]."[167] Bailment, for example, is a form of stewardship where the bailee is responsible for the safe custody and transmission of goods entrusted to its care by the bailor.[168] We propose that data governance must expand to include the concept of stewardship.[169]

[164] Hemant Taneja, *The Era of "Move Fast and Break Things" Is Over*, HARV. BUS. REV. (Jan. 22, 2019), https://hbr.org/2019/01/the-era-of-move-fast-and-break-things-is-over (describing Facebook founder Mark Zuckerberg's "now famous" motto).

[165] *Compare,* Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject as Object*, 52 STAN. L. REV. 1373-1438 (2000) https://www.jstor.org/stable/1229517 (arguing for regulatory approach that pre-empts alienable property rights in PII) *with* Jeffrey Ritter and Anna Mayer, *Regulating Data As Property: A New Construct For Moving Forward*, 16 DUKE L. & TECH. REV. 220 (2017) (advocating property rights approach to PII) and Steven H. Hazel, *Personal Data as Property*, 70 SYR. L. REV. 1055 (2020) https://lawreview.syr.edu/wp-content/uploads/2020/12/1055-1113-Hazel.pdf.

[166] In a bailment relationship, for example, involving personal property held on behalf of another, the possessor (bailee) has a duty to take reasonable care of the goods on behalf of the owner (bailor). Danielle D'Onfro, *The New Bailments*, 97 WASH. L. REV. 97, 105 (2022). In a trustee-beneficiary relationship, the trustee assumes a fiduciary duty to protect the interests of the beneficiary with respect to the property held. Philip J. Ruce, *The Trustee and the Trust Protector: A Question of Fiduciary Power – Should a Trust Protector Be Held to a Fiduciary Standard?*, 59 DRAKE L. REV. 67, 83–84 (2010).

[167] KOJO MENYAH, *Stewardship Theory*, ENCYCLOPEDIA OF CORP. SOC. RESP. (Samuel O. Idowu et al. eds, 2013).

[168] DAVID MILLMAN, GOVERNANCE OF DISTRESSED FIRMS 20 (Edward Elgar ed., 2013).

[169] *See e.g.*, FIN. REPORTING COUNCIL, THE UK STEWARDSHIP CODE 2020 4 (2020) (defining stewardship in connection with managing assets as "the *responsible*

"Caring for" something indicates a higher standard than "managing" something. Adding the element of stewardship provides an enhanced obligation in the form of some type of duty owed to the data subject, while asset management merely involves a contractual obligation to the data user focusing on efficiency.[170] While both involve control, the beneficence of that control flows in opposite directions.

Without stewardship as a guiding principle, the needs of the data subjects are suppressed in favor of the wants of those collecting and using the data.[171] As explained by Astha Kapoor and Richard Whitt in their essay, *Nudging towards data equity: The role of stewardship and fiduciaries in the digital economy*, "treating data as a commodity resource magnifies the power and position of companies that hold the data surplus, while diminishing the agency of those whose lives are being harvested for profit."[172] Kapoor and Whitt suggest that granting data subjects greater agency in how their data is used can be readily accomplished through data stewardship.[173] Data stewardship is a concept with deep roots in the science and practice of data collection, sharing, and analysis and denotes a much broader approach than data management.

Ada Lovelace, the founder of scientific computing, predicted computing would eventually lead to big data analytics, stating, "A new, a vast, and a powerful language is developed for the future use of analysis, in which to wield its truths so that these may become of more speedy and

allocation, management and oversight of capital to create long-term value for clients and beneficiaries leading to sustainable benefits for the economy, the environment and society") (emphasis added). This definition was updated from the previous 2010 Code's focus on "[purposeful engagement] on strategy, performance and the management of risk" representing the evolution of stewardship. FIN. REPORTING COUNCIL, THE UK STEWARDSHIP CODE 2010 1 (July 2010).

[170] *See* Meeyeon Park, *Asset Management*, CORP. FIN. INST., https://corporatefinanceinstitute.com/resources/knowledge/finance/asset-management/ (last updated Feb. 19, 2023). *See also* MITTELSTADT, *supra* note 137, at 3 (noting that "the absence of a fiduciary relationship in AI means that users cannot trust that developers will act in their best interests when implementing ethical principles in practice").

[171] *See What Is Governance?*, GOVERNANCE INST. OF AUSTL., https://www.governanceinstitute.com.au/resources/what-is-governance/ (last visited Feb. 16, 2022) (advancing the definition of corporate governance as "a set of relationships between a company's management, its board, its shareholders and other *stakeholders*") (emphasis added).

[172] Astha Kapoor & Richard Whitt, *Nudging Towards Data Equity: The Role of Stewardship and Fiduciaries in the Digital Economy* 2–3 (Feb. 22, 2021), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3791845.

[173] *Id.* at 5.

accurate practical application for the purposes of mankind than the means hitherto in our possession have rendered possible."[174] The Ada Lovelace Institute, which advances the use of AI for the betterment of society, explains that similar to liberal constitutionalism, digital constitutionalism protects an individual's fundamental rights from intrusions by government and private companies.[175] As part of its *Future of Regulation* program, the Ada Lovelace Institute suggests a data stewardship approach for supporting "responsible and trustworthy data governance.[176] It goes on to define data stewardship as: "the responsible use, collection and management of data in a participatory and rights-preserving way."[177] Similarly, this article concludes that , next-generation data governance must incorporate the process by which responsibilities of stewardship are conceptualized and carried out.

Stewardship requires an acknowledgement that a duty to the data subject exists. This creates a tension in obligation which is why, for example, standards of care in bailment relationship vary on whose benefit the bailment is for. While some would argue that data users have a fiduciary obligation to put the interests of the data subjects above their own,[178] this is not the most likely model for a commercial enterprise.[179] There must be a balance between the needs of the organization and the needs of the data subject. A lengthy set of rules will not work for every organization, and vague aspirational principles are not actionable.

While governmental regulations, guidance, and industry standards all have a role to play, given the divergent operations of companies, it is

[174] Octavia Reeve, *Celebrating Ada Lovelace Day: What Ada Means to Us*, ADA LOVELACE INST. (Oct. 8, 2019), https://www.adalovelaceinstitute.org/blog/celebrating-ada-lovelace-day/.

[175] Ada Lovelace Institute, *How does digital constitutionalism reframe the discourse on rights and powers?* (Dec. 7, 2022), https://www.adalovelaceinstitute.org/blog/digital-constitutionalism-rights-powers/.

[176] Ada Lovelace Insitute, *The future of regulation*, https://www.adalovelaceinstitute.org/our-work/programmes/future-regulation/ (last visited Nov. 19, 2023).

[177] Ada Lovelace Institute, *Exploring Legal Mechanisms for Data Stewardship*, UK AI COUNCIL 23 (Mar. 2021), https://www.adalovelaceinstitute.org/wp-content/uploads/2021/03/Legal-mechanisms-for-data-stewardship_report_Ada_AI-Council-2.pdf.

[178] Jack M. Balkin, *The Fiduciary Model of Privacy*, 134 HARV. L. REV. F. 11 (2020) https://harvardlawreview.org/wp-content/uploads/2020/10/134-Harv.-L.-Rev.-F.-11.pdf; Andrew F. Tuch, *A General Defense of Information Fiduciaries*, 98 WASH. U. L. REV. 1897 (2021) https://openscholarship.wustl.edu/law_lawreview/vol98/iss6/12.

[179] Though not appropriate for commercial enterprises, it should be a more likely fit with governments.

ultimately up to each organization to adopt data governance principles. Firms should consider both the impact of its operations on data subjects and protect the company and upper-level management from liability due to a failure of data governance. This is especially true for public companies that will soon be subject to SEC rules on cybersecurity, risk management, strategy, governance, and incident disclosures.[180] While the federal government may be slow to move to an omnibus data protection model, states are rapidly expanding their requirements[181] which may provide data subjects with more options to hold companies liable for their failure to provide adequate data protection.[182] Companies would be well-served to prepare for these changes by acting now to incorporate Next-Gen Data Governance Principles. The following section briefly describes ten principles based on the Code. The attached Appendix provides examples of the principles as illustrations.

C. Ten Data Governance Principles

The next likely step is the maturation of data governance from data management to data stewardship. We offer the following ten principles on which organizations can model their own Data Governance strategy. The ten principles were built off key considerations from the Code.[183] See Appendix A.[184]

[180] Press Release, SEC, SEC Proposes Rules on Cybersecurity Risk Management, Strategy, Governance, and Incident Disclosure by Public Companies (Mar. 9, 2022) (on file with author), https://www.sec.gov/news/press-release/2022-39 (public companies will now need to disclosure their governance over cybersecurity and how the board oversees such risks. Disclosure fialures can result in significant liability for a corporation).

[181] *See* Anokhy Desai, *US State Privacy Legislation Tracker,* IAPPAPP (Oct. 7, 2022), https://iapp.org/resources/article/us-state-privacy-legislation-tracker/ (last visited Mar. 10, 2023) (showing that California, Colorado, Connecticut, Utah and Virginia have enacted data privacy laws and 21 other states including Minnesota, New Jersey, and Tennessee have active bills).

[182] *See* MASTODON, https://joinmastodon.org/ (last visited Dec. 2, 2022) (explaining how their social media site is decentralized and user data is not for sale).

[183] *See supra* Part IV.A.

[184] In the attached Appendix, the first column describes the principle, the second column refers to the Ethical Opinion or section of the Code. The third column provides examples of actual in-use cases demonstrating the principle. While not every example will be a fit for every organization, by providing actionable resources, the organization can customize the examples provided based on their specific operations.

IV. WHAT THE FUTURE HOLDS

While we obviously have no crystal ball that can predict the future development of GenAI technology over the next few years, there is no doubt that it will revolutionize many fields, not the least of which will be the legal and justice systems. Generating fake but believable text, audio, and video of ordinary people spouting lies, misinformation, or defamatory content, committing crimes, or breaking the law will become feasible for just about any person with a working computer. So, too, will anybody be able to generate competent pleadings, in a matter of minutes, with great benefit to access to justice coming alongside the risk of many more vexatious filings flooding court dockets. As a result of these technological developments, our current approaches to managing cases and evidence may need to change. The legal status of AI-generated art (in particular, with respect to copyright eligibility, copyright infringement, and trademark infringement and/or dilution) will need to be resolved. Judges themselves will have to sort through AI-generated pleadings and arguments, including perhaps even using an AI clerk to filter out or respond to junk claims or imaginary citations (if and when this becomes possible). Judges may eventually join the revolution, using new GenAI systems to help them decide their cases or draft their opinions more effectively and efficiently, after problems involving inaccuracy and bias are resolved. And one day, judges may even be replaced by AI,[185] giving new meaning to the phrase "having one's day in court."

APPENDIX A

Principle	**Code of Medical Ethics**	**Example**
1 – OVERSIGHT Boards must demonstrate the importance of data governance as part of their strategic planning and oversight obligations.	"To promote responsible innovation, health care institutions and the medical profession should: (m) Provide meaningful professional oversight of innovation in patient care."	**Board Committee**. Data governance will fall under either the Audit Committee or Corporate Governance Committee. Boards influence a firm's culture and should express the importance of data governance as a foundation for the remaining principles.[i]

[185] Tara Vazdani, *From Estonian AI judges to robot mediators in Canada, U.K.*, THE LAWYER'S DAILY, https://www.lexisnexis.ca/en-ca/ihc/2019-06/from-estonian-ai-judges-to-robot-mediators-in-canada-uk.page (last visited Nov. 10, 2023). Indeed, OpenAI's release of the research and code for its new text-to-3D model, Shap-E—while we were in the midst of writing this piece—may even allow judges to be printed at some point! *See* Avran Piltch, *OpenAI's Shap-E Model Makes 3D Objects From Text or Images*, TOM'S HARDWARE (May. 4, 2023), https://www.tomshardware.com/news/openai-shap-e-creates-3d-models.

	1.2.11 - *Ethically Sound Innovation in Medical Practice*	*See, e.g.*, *Carnegie Board-Level Guide* and *Board Checklist*.[ii]
2 – TRUST Firms must assure the public and data subjects that their data activities can be trusted. Firms must consider themselves stewards of the data collected, use sound judgment on behalf of the data subjects, and consider stakeholders in determining data use.	"The relationship between a patient and a physician is based on trust, which gives rise to physicians' ethical responsibility to place patients' welfare above the physician's own self-interest or obligations to others, to use sound medical judgment on patients' behalf, and to advocate for their patients' welfare." 1.1.1 – *Ethics of Patient-Physician Relationships* *See also* 1.1.3 – *Patient's Rights* 1.1.6 – *Quality* 1.1.8 – *Required Reporting of Adverse Events* 1.2.9 – *Use of Remote Sensing and Monitoring Devices* 1.2.11(c) – *Ethically Sound Innovation in Medical Practice* 2.1.3 – *Withholding Information from Patients* 8.6 *Promoting Patient Safety*	**Stewardship Models.** The Ada Lovelace Institute issued a report in 2021 explaining several legal mechanisms for data stewardship.[iii] *See, e.g.*, *Driver's Seat*, a data cooperative for Uber and Lyft drivers.[iv] **Data Trust**. A data trust can provide an intermediary between the data subject and data user with a fiduciary duty to protect the subject's data from breach and wrongful use.[v] *See, e.g., Virginia's Commonwealth Data Trust*[vi] and the *Brixham Data Trust*.[vii] **Encryption, Anonymization and Tokenization**. Where appropriate, data should be encrypted, anonymized, or tokenized.
3 - ACCURACY In order to avoid harm resulting from inaccurate data and to assure that the data will bring the most value to the firm, data sets must be balanced and representative. Additionally, algorithms must be designed by diverse teams, and predictions must be tested for accuracy.	Our AMA supports the systematic collection and utilization of physician feedback on administrative and support systems by health care organizations in efforts to reduce error and improve diagnostic accuracy. - Developing Physician Leadership in the implementation of Diagnostic Error Surveillance H-450.925 9.2 *Training in Data use* *See also* 9.4 *Method to Report and Correct Wrongful Data Hygiene*	**Quality data**. The effective and efficient use of data requires quality data which is easily accessible.[viii] Quality data also prevents harms to data subjects which may occur from inaccurate, unbalanced, non-representative data sets. *See e.g.*, the U.S. Agency for International Aid's *Data Quality Assessment* (DQA) instructions.[ix] **Accuracy testing**. As such, data sets must be continually tested for accuracy.[x] **Diversity**. Additionally, data sets must fairly represent members of society.[xi]
4 - CONSENT	"Informed consent to medical treatment is fundamental in both ethics and law. Patients have the right	**Informed consent**. Informed consent forms can provide the needed information

Companies must ensure that data subjects are aware that information about them is being collected and consent to its collection. In the absence of consent, data collectors must have one of the following in order to use the data: • Legitimate interest • Contractual necessity • Vital interest of the user • Legal obligation • Public interest	to receive information and ask questions about recommended treatments so that they can make well-considered decisions about care." 2.1.1 – *Ethics of Consent, Communication & Decision Making* *See also* 1.2.9 – *Use of Remote Sensing and Monitoring Devices* 2.2.1 – *Pediatric Decision Making* 7.1.2 – *Informed Consent in Research*	in order to initiate the collection of data from data subject. *See e.g*., Law Insider *Consent of the data subject* Sample Clauses.[xii] Article 13, GDPR, Information to be provided where personal data are collected from the data subject.[xiii]
5 - AGENCY Data subject agency allows data subjects to have a say in how their data is used. Firms must provide data subjects with the ability to determine the extent of the scope and use of their data.	"In general, patients are entitled to decide whether and to whom their personal health information is disclosed." 3.2.1 - *Confidentiality* 3.2.4 – *Access to Medical Records by Data Collection Companies*	**Agency** requires that there is a process for data subjects to exercise their rights.[xiv] *See e.g*., Stanford University *Consent* forms.[xv]
6 - PRIVACY Firms should conduct periodic Privacy Impact Assessments to determine why data is being collected (and its classification as personal data, sensitive data or non-human data) and how the data will be used, accessed, shared, safeguarded and stored to identify and mediate risks.	"Physicians must seek to protect patient privacy in all settings to the greatest extent possible." 3.1.1 – *Privacy in Health Care* *See also* 2.3.1 – *Electronic Communications with Patients* 3.3.2 – *Confidentiality & Electronic Medical Records* 7.3.7 – *Safeguards in the Use of DNA Databanks*	**Privacy Impact Assessment**: The purpose of a Privacy Impact Assessment (PIA) is to identify potential risks involving the collection, use, and sharing of persona data.[xvi] *See e.g*., Department of Homeland Security *Privacy Impact Assessment* template.[xvii]
7 - CONFIDENTIALITY & SECURITY Data must be protected from intrusions, breaches, and indiscriminate sharing. Not only must data be stored securely, and measures taken to guard against data breaches (security), but policies must be in place that	"Physicians in turn have an ethical obligation to preserve the confidentiality of information gathered in association with the care of the patient." 3.2.1 – *Confidentiality* *See also* 1.2.12 – *Ethical Practices in Telemedicine*	**Supply Chain Management**. While training will help promote data hygiene within an organization, third parties must also be managed. This could involve adding privacy and security audit requirements in third party agreements or requiring periodic compliance reports.[xix]

prevent the inadvertent or wrongful sharing of data (confidentiality).[xviii]	3.2.4 - *Access to Medical Records by Data Collection Companies* 3.3.2 – *Confidentiality & Electronic Medical Records* 3.3.3 - *Breach of Security in EMR*	**Data Security Guidance**. NIST can provide guidance on cybersecurity measures. *See e.g*., *Data Security*, National Cybersecurity Center of Excellence.[xx]
8 - RECORD MANAGEMENT Data Mapping is the process of tracking data held by firms from its source to its destination and will help firms: • Identify what personal data they hold, why it is held, and where it is held, • Assess any security or privacy risks to individuals, • Institute measures to mitigate those risks, • Provide for easy retrieval and transfer of data, and • Comply with their legal obligations.	"In keeping with the professional responsibility to safeguard the confidentiality of patients' personal information, physicians have an ethical obligation to manage medical records appropriately. This obligation encompasses not only managing the records of current patients, but also retaining old records against possible future need, and providing copies or transferring records to a third party as requested by the patient or the patient's authorized representative when the physician leaves a practice, sells his or her practice, retires, or dies." 3.3.1 – *Management of Medical Records* *See also* 3.3.2 – *Confidentiality & Electronic Medical Records*	**Data Mapping**: A data mapping firm that complies with the GDPR and/or the CCPA's requirements can automate data mapping for a firm. *See e.g*., Termly – *Complete Guide to Data Mapping*.[xxi]
9 - DATA REVIEW BOARD Firms must create a Data Review Board with "deep silo" expertise in information/data science, tech and securities law, behavioral sciences, corporate compliance, and risk assessment consisting of diverse members of society. The committee would be charged with creating policy and reviewing proposed data uses.	"Institutions have an obligation to oversee the design, conduct, and dissemination of research to ensure that scientific, ethical, and legal standards are upheld. Institutional review boards (IRBs) as well as individual investigators should ensure that each participant has been appropriately informed and has given voluntary consent." 7.1.1 – *Physician Involvement in Research* *See also* 7.1.3 – *Study Design & Sampling*	**Data Review Board**. Data governance cannot be siloed or left to a tech-focused C-level executive. University data governance committees can provide a model for firms provided they consist of multi-disciplinary teams from diverse members of society. *See e.g*., University of Wisconsin – Madison Data Governance Council[xxii] and the *Building Data and AI Ethics Committees* report by Northeastern University and Accenture.[xxiii] **Data Protection Impact Assessment (DPIA)**. Prior to the use of data, the Data Review Board would conduct a DPIA to

		ensure that the proposed use is compliant with these principles. *See e.g.*, UK Information Commission's Office - *Sample Data Protection Impact Assessment.*[xxiv]
10 - HUMAN IN THE LOOP Firms should ensure human oversight of any automated analysis. Human review by diverse members of society and periodic or sample testing is needed to assure accuracy and compliance with these Ten Data Governance Principles. Humans should also review proposed new innovations to make sure they align with these ten principles. Monitoring and quick response to inaccurate, unfair, or discriminatory results requires the ability to quickly identify problems, develop remediation solutions and install modifications.	Physicians who engage in biomedical or health research with human participants thus have an ethical obligation to ensure that any study with which they are involved: is consistent with the goals and values of the medical profession, is scientifically well-designed, minimizes risks to participants, safeguards confidentiality, does not have a disparate impact, and has been reviewed and approved by the oversight body. 7.1.3 – *Study Design & Sampling* *See also* 7.1.1 – *Physician Involvement in Research* 8.8 – *Required Reporting of Adverse Events* 9.4.1 – *Peer Review and Due Process* 9.4.2 – *Reporting Incompetent or Unethical Behavior by Colleagues*	**Chief Privacy Officer and Employee Training**. To ensure that the organization complies with these principles, the firm should conduct annual employee training on the policies and procedures and monitor compliance with the policies and procedures. This may be accomplished by the establishment of a Chief Privacy Officer with privacy certification or a legal/tech background to oversee compliance.[xxv] **Algorithmic auditing.** There must be a mechanism to discover unfair or discriminatory outcomes.[xxvi] **Data Ethics Hotline**. Firms must install a system to quickly identify and address instances of inaccurate, unfair, or discriminatory results. Not only would this help meet the board's governance duties, but it would also help maintain heigh levels of data integrity. One example would be to set up an anonymous reporting portal, where employees could note their observations without fear of retribution.[xxvii]

[i] Denise Lee Yohn, *Company Culture Is Everyone's Responsibility*, HARV. BUS. REV. (Feb. 8, 2021), https://hbr.org/2021/02/company-culture-is-everyones-responsibility (explaining how the Board of Directors "[g]uide[s]

the definition and development of the desired culture, ensuring that it aligns with business goals and meets the needs of all stakeholders").

[ii] *Board-Level Guide: Cybersecurity Leadership*, CARNEGIE ENDOWMENT FOR INT'L PEACE, https://ceipfiles.s3.amazonaws.com/pdf/FinCyber/English/1_Board_Guide.pdf (last visited Dec. 22, 2022); *Cybersecurity for Smaller Organizations, Board Checklist: Cybersecurity Leadership*, CARNEGIE ENDOWMENT FOR INT'L PEACE, https://ceipfiles.s3.amazonaws.com/pdf/FinCyber/English/1_Board_Checklist.pdf (last visited Dec. 22, 2022). Although these are specific to cybersecurity, they are a good starting point for data governance inquiries.

[iii] Ada Lovelace Institute, *supra* note 177.

[iv] DRIVER'S SEAT, https://driversseat.co/ (last visited Dec. 17, 2022).

[v] Houser & Bagby, *supra* note 18.

[vi] *Commonwealth Data Trust*, VA. OFF. OF DATA GOVERNANCE & ANALYTICS, https://www.odga.virginia.gov/commonwealth-data-trust/ (last visited Mar. 12, 2023) (providing multiple member and user agreements).

[vii] Natasha Nicholson & Pamela Charlick, *Placemaking with Data: A Data Trust to Serve a Coastal Community*, DATA TRUSTS INITIATIVE (Oct. 27, 2022), https://datatrusts.uk/blogs/data-and-place-creating-a-coastal-community-data-trust (incorporating data from a three kilometer radius from the town center of Brixham and fourteen square kilometers of the local marine environment); *see also* PROSPECT BRIXHAM, https://prospectbrixham.org/ (last visited Mar. 12, 2023) (presenting a place-based local data trust for the benefit of the people and the place, a coastal fishing village).

[viii] Christopher Roberts, *5 Reasons Why Data Accuracy Matters for Your Business*, MEDIUM (Oct. 26, 2019), https://chrisrob978.medium.com/5-reasons-why-data-accuracy-matters-for-your-business-b490d5e20bf1 (describing how data accuracy increases revenue, reduces costs, improves customer satisfaction, saves time, and boosts ROI).

[ix]*How-To Note: Conduct a Data Quality Assessment*, U.S. AGENCY FOR INT'L DEV. (Feb. 2021), https://usaidlearninglab.org/sites/default/files/resource/files/how-to_note_-_conduct_a_dqa-final2021.pdf.

[x] Ibarrera, *What is Data Accuracy, Why it Matters and How Companies Can Ensure They Have Accurate Data*, DATA LADDER (Sept. 25, 2020), https://dataladder.com/what-is-data-accuracy/.

[xi] Hessie Jones, *Artificial Intelligence Needs to Be Reset*, FORBES (Oct. 10, 2018, 2:49 AM), https://www.forbes.com/sites/cognitiveworld/2018/10/12/artificial-intelligence-needs-to-reset/?sh=7530684d386e; *see also* Kate Brodock, *Why We Desperately Need Women to Design AI*, FREECODECAMP (Aug. 4, 2017), https://www.freecodecamp.org/news/why-we-desperately-need-women-to-design-ai-72cb061051df/ [https://perma.cc/3S43-RHUX] (explaining the value of diversity in the development of AI).

[xii] *Data Privacy and Non-Disclosure Agreement*, LAW INSIDER, https://picpa.glueup.com/resources/protected/organization/959/event/24279/f1eda04f-040a-447a-b82c-bd5a8abfff63.pdf (last visited Mar. 12, 2023).

[xiii] Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), art. 13, 2016 O.J. (L 119) 1, 40–41.

[xiv] *Five Elements of Consent Under the GDPR*, PRIVACY POLICIES (Aug. 31, 2022), https://www.privacypolicies.com/blog/gdpr-consent-examples/#Five_Elements_Of_Consent_Under_The_Gdpr.

[xv] *Forms & Consent Templates*, STAN. RSCH. COMPLIANCE OFF., https://researchcompliance.stanford.edu/panels/hs/for-researchers/forms-templates (last visited Oct. 24, 2022).

[xvi] *See* Reuben Binns, *Data Protection Impact Assessments: A Meta-Regulatory Approach,* 7 INT'L. DATA PRIV. L. 22, 22–25 (2017) (explaining the origins and uses of PIAs).

[xvii] U.S. DEP'T OF HOMELAND SEC., PRIVACY IMPACT ASSESSMENT TEMPLATE.

[xviii] Confidentiality means limiting distribution of data to intended recipients and assuring its use is only for particular, authorized purposes. Information security means "protecting information and information systems from unauthorized access, use, disclosure, disruption, modification, or destruction in order to provide integrity . . . confidentiality . . . and availability." 44 U.S.C. § 3542(b)(1) (repealed 2014); 44 U.S.C. § 3552(b)(3) (2014).

[xix] Paul Mezzera, *Protecting the Modern Workforce Requires a New Approach to Third-Party Security*, VENTUREBEAT (June 25, 2022, 10:10 AM), https://venturebeat.com/datadecisionmakers/protecting-the-modern-workforce-requires-a-new-approach-to-third-party-security/.

[xx] *Data Security*, NAT'L CYBERSEC. CTR. OF EXCELLENCE, https://www.nccoe.nist.gov/data-security (last visited Oct. 24, 2022).

[xxi] Masha Komnenic, *Complete GDPR Data Mapping Guide*, TERMLY (May 12, 2022), https://termly.io/resources/articles/gdpr-data-mapping/.

[xxii] *UW–Madison Data Governance Council Charter*, UNIV. OF WIS.-MADISON, https://uwmadison.app.box.com/s/5um0d5jrt12brkf2z01otejb5obbwa0w (Oct. 22, 2019).

[xxiii] Ronald Sandler & John Basl, *Building Data and AI Ethics Committees*, ACCENTURE (2019), https://www.accenture.com/_acnmedia/pdf-107/accenture-ai-data-ethics-committee-report-executive-summary.pdf.

[xxiv] *Data Protection Impact Assessments*, INFO. COMM'R.'S OFF. (Oct. 14, 2022), https://ico.org.uk/media/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-impact-assessments-1-1.pdf.

[xxv] Scott Clark, *4 Ways a Chief Privacy Officer Can Help Your Company*, REWORKED (Dec. 1, 2020), https://www.reworked.co/information-management/4-ways-a-chief-privacy-officer-can-help-your-company/.

[xxvi] *See generally* Jon Kleinberg et al., *Algorithms as Discrimination Detectors*, 117 PROC. NAT'L ACAD. SCIS. 30096 (2019) (discussing how algorithms can make it easier to detect and prevent discrimination).

[xxvii] "Red flags" could then be brought to the board's attention for handling meeting its reporting and monitoring requirements. Patrick A. Lee, *Why Data Governance Should Be Part of Board Conversations*, KPMG (Nov. 2019), https://boardleadership.kpmg.us/relevant-topics/articles/2019/data-governance-part-of-board-conversations.html; Annette Weller-Collison et al., *Who Is Afraid of the DOJ? Why Companies Should Revisit Their Information Governance Program*, AM. BAR ASS'N (June 15, 2022), https://www.americanbar.org/groups/business_law/publications/blt/2022/06/info-gov-program/.