Washington and Lee Journal of Civil Rights and Social Justice

Volume 24 | Issue 2 Article 5

4-1-2018

Information and the Regulatory Landscape: A Growing Need to Reconsider Existing Legal Frameworks

Anjanette H. Raymond
Indiana University

Information and the Regulatory Landscape: A Growing Need to Reconsider Existing Legal Frameworks

Anjanette H. Raymond[*]

Table of Contents

* Associate Professor, Kelley School of Business, Indiana University; Director, Ostrom Workshop, Program on Data Management and Information Governance; Adjunct Associate Professor of Law, Maurer School of Law, Indiana University. This paper arose in the conversations that occurred at the Washington & Lee University School of Law Academic Roundtable entitled *Big Data: Understanding Algorithmic Power* (2017). Thank you to Margaret Hu and all roundtable participants for your valuable discussion, the Ostrom Workshop and the participants in the Colloquium and for the insight and editorial guidance of Steven Marino. All opinions are those of the author.

Introduction

Advanced artificial intelligence (AI) systems are already being used to enhance our lives and to transform the way businesses operate. Businesses across a broad spectrum of industries are exploring the potential gains offered by AI systems. In fact, the use of AI systems is already widespread in areas such as transport, finance, defense, social security, education, policing, public safety, and healthcare.[1] The recent explosion of machine learning technology is arguably a product of two things: "tremendous increases in computational power and enormous volumes of accumulated data."[2] Unsurprisingly, legal frameworks and industry-based governance regimes have failed to keep up with the

1. *See Emerging AI: 7 Industries Including Law, HR, Travel and Media Where AI IS Making An Impact*, CBINSIGHTS (2017), https://www.cbinsights .com/research/artificial-intelligence-emerging-industries/ (describing advances that are making AI crucial to modern industry and how various industries are using or are looking to use it) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

2. KINGSLEY ET AL., SLAUGHTER AND MAY & ASI DATA SCIENCE, SUPERHUMAN RESOURCES: RESPONSIBLE DEPLOYMENT OF AI IN BUSINESS 16 (2017), https://www.slaughterandmay.com /media/2536419/ai-white-paper-superhuman-resources.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice). The difference between the terms data and information are important, although often the terms are used almost interchangeably. *See What is the Difference Between Data and Information*, DQ GLOBAL (May 27, 2014), https://www.dqglobal.com/2014/05/27/what-is-the-difference-between-data-and-information/ (contrasting and defining data and information) (on file with the Washington & Lee Journal of Civil Rights & Social Justice). Data is raw, unorganized facts that need to be processed. *Id.* In general, data can be something simple and seemingly random and useless until it is organized. *See id.* (describing raw data as a "series of 1s and zeros that human would not be able to read"). In contrast, when data is processed, organized, structured or presented in a given context so as to make it useful, it is called information. *Id.*

newest AI. The existing gaps have led to industry attempting to fill the void, but these attempts are in their infancy and often fail to fully consider the various stakeholders impacted by the ubiquitous gathering and corresponding use of data.[3]

Consider the recent news splash concerning Google's advertising program, which is once again under fire for its use of highly secretive gathering, storing, and using of highly sensitive data.[4] According to the Electronic Privacy Information Center (EPIC) Google is gaining access to "highly sensitive information—the credit and debit card purchase records of the majority of U.S. consumers—without revealing how they got the information or giving consumers meaningful ways to opt out."[5] And, as is often the argument against the use of black box algorithms, EPIC asserts the "search giant is relying on a secretive technical method to protect the data."[6] Of course, this is not the first—nor the last—criticism of big businesses' use of black box algorithms. This Article seeks to further debates previously asserted by the author by examining the issue in light of the recent surge in attention algorithms are drawing, primarily because their use has grown exponentially.

---

3. *See, e.g.*, Kevin Petrasic et al., *Three Big Questions About AI in Financial Services*, WHITE & CASE (July 18, 2017), https://www.whitecase.com/publications/insight/ai-financial-services (noting that the stakeholders include consumers, technology companies, third-party data providers, and regulators, among others) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

4. *See, e.g.*, Hal Hodson, *Revealed: Google AI has access to huge haul of NHS patient data*, NEW SCIENTIST (Apr. 29, 2016), https://www.newscientist.com/article/2086454-revealed-google-ai-has-access-to-huge-haul-of-nhs-patient-data/ (discussing a document which revealed the previously unknown extent of private healthcare data acquired by DeepMind, Google's artificial intelligence company) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

5. Elizabeth Dwoskin & Craig Timberg, *Google's New Program To Track Shoppers Sparks a Federal Privacy Complaint*, WASHINGTON POST (July 30, 2017), https://www.washingtonpost.com/news/the-switch/wp/2017/07/30/googles-new-program-to-track-shoppers-sparks-a-federal-privacycomplaint/?utm_term=.f6932f73766b (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

6. *Id.*

I. The Growth of Advanced Artificial Intelligence Systems

Advanced AI systems are already being used to enhance our lives and to transform the way businesses operate. Businesses across a broad spectrum of industries are exploring the potential gains offered by AI systems. In fact, the use of AI systems is already widespread in areas such as transport, finance, defense, social security, education, policing, public safety, and healthcare.[7] Of course, AI is widely influencing our daily lives, as demonstrated by the widespread use by both Google and Facebook, to name but two.[8]

In many ways, people have become accustomed to the first level of the technological revolution which "has seen organizations automate repetitive, high volume, sometimes complex but typically rule-based (if X then Y) processes."[9] Yet, the recent explosion of machine learning technology is arguably a product of two things: "tremendous increases in computational power and enormous volumes of accumulated data,"[10] both of which are new occurrences. As a result, legal frameworks and industry-based governance regimes have failed to keep up with the newest AI. In fact, as will be explored in Section II, most legal frameworks are based in paper-based data gathering and/or basic automation and are thus frequently required to smash regulation of technology advancements into privacy and property based legal frameworks.[11] The existing gaps in legal regulation has led to industry attempting to fill the void, but these attempts are in their infancy and often fail to fully consider the various stakeholders impacted by the ubiquitous gathering and corresponding use of data.

---

7. Anand Rao et al., *Top 10 AI Technology Trends for 2018*, PWC NEXT IN TECH (Dec. 5, 2017), http://usblogs.pwc.com /emerging-technology/top-10-ai-tech-trends-for-2018/ (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

8. *See* Cade Metz, *Google, Facebook, and Microsoft are Remaking Themselves Around AI*, WIRED (Nov. 21, 2016), https://www.wired.com/2016/11/google-facebook-microsoft-remaking-around-ai/ (noting in addition that Amazon, Microsoft, and IBM "are also building cloud computing services specifically designed for artificial intelligence work") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

9. KINGSLEY ET AL., *supra* note 2, at 8.

10. *Id.* at 7.

11. *See infra* II. The Existing Landscape

II. The Existing Landscape

As will be explored in this Section, the existing landscape of information governance is haphazard, often limited by the sector, and designed without cohesiveness. Even more problematic are the limited legal contributions, often drawn from existing case law, and regulations that fail to fully consider the nuisances of ubiquitous information flows, which instead attempt to cram new issues into existing frameworks designed for paper and pencil, simple single-step, technology. This Section will orient the reader into the existing legal landscape and will demonstrate the limitations of the use of these approaches in the information centric world.

A. A Brief Privacy Primer

What we put out in the public eye we cannot expect to be private; but, "[o]ne who desires to live a life of partial seclusion has a right to choose the times, places, and manner in which and at which he will submit himself to the public gaze."[12]

Historically, this sentiment was captured repeatedly in case law, founded under constitutional protections from governmental intrusion,[13] some of the original torts[14] created to protect individuals from public gaze. Today, in the United States, there are a litany of protections afforded to specific types of information—often justified by the sensitivity of the information or the relationship that exists between the individual and the entity.[15] Some States have begun to more robustly protect

---

12. Pavesich v. New England Life Ins. Co., 50 S.E. 68, 70 (Ga. 1905).

13. *See generally* Daniel J. Solove, *A Brief History of Information Privacy Law*, *in* PROSKAUER ON PRIV., PLI 1–46 (2006) (explaining in detail the history of information privacy from the colonial era to now).

14. *See generally id.* (providing details and history on the tort of public disclosure of private facts, and the tort of false light, and the tort of breach of confidentiality which developed to protect disclosures of information in violation of trust within certain relationships).

15. *See generally* Family Educational Rights and Privacy Act, 20 U.S.C. § 1232g (1974) (protecting the privacy of student educational records); *see generally* Privacy Protection Act, 42 U.S.C. § 2000aa (1980) (protecting journalist and newsrooms from search by government officials); *see generally* Cable Communications Policy Act, 42 U.S.C. § 2000aa(a) (1984) (regulating cable

individuals' information gathered for particular uses, such as retailer-based information.[16]

In contrast to the U.S.-based sectoral approach, the vast majority of countries, including nearly every country in Europe, many in Latin America, and the Caribbean, Asia, and Africa, have adopted comprehensive data protection laws.[17] Arguably, an overwhelming number of countries agree on a list of basic legal protections, often based on Fair Information Practice,[18] the more comprehensive Organisation for Economic Co-operation and Development (OECD) Principles,[19] and the European Union Data

---

communications by federal, state, and local authorities); *see generally* Employee Polygraph Protection Act, 29 U.S.C. §§ 2001–2009 (1988) (prohibiting employers from using polygraph screening in most cases); *see generally* Electronic Communications Privacy Act, 18 U.S.C. § 2710(b) (1986) (extending government wiretap restrictions to include transmissions of electronic data); *see generally* Telephone Consumer Protection Act, 47 U.S.C. § 227 (1991) (restricting telephone solicitations, automatic dialing systems, and prerecorded voice messages); *see generally* Driver's Privacy Protection Act, 47 U.S.C. § 227(c)(5) (1994) (limiting disclosure of states' department of motor vehicles records).

16. *See generally* 2017 N.J. Laws 124. State laws governing the collection and use of personal information continue to proliferate. One of the latest state-based Acts—New Jersey—was signed on July 21, 2017, and restricts a merchant's ability to collect personal data of shoppers and share such data with third parties. *Id*. New Jersey's Personal Information Privacy and Protection Act further limits the retailer's ability to scan an identification card to a limited set of purposes—such as verifying the consumer's identity—and prohibits the retailer from sharing that data with a third party unless the retailer discloses its data-sharing practices to the consumer. *See* Cynthia Larose, *Retailers: Review Those Checkout Practices—Again*, MINTZ LEVIN (July 26, 2017), https://www.privacyandsecurity matters.com/tag/new-jersey-personal-information-privacy-and-protection-act/ (listing circumstances in which data can be collected and utilized) (on file with the Washington & Lee Journal of Civil Rights & Social Justice). Recent activity in California and Illinois law have followed New Jersey in enhancing privacy protections for individuals. *See State Laws Relating to Internet Privacy*, NAT'L CONF. ST. LEGISLATURES (June 20, 2017), http://www.ncsl.org/research/tele communications-and-information-technology/state-laws-related-to-internet-priv acy.aspx (providing state by state overview of internet privacy laws) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

17. *See* Graham Greenleaf, *Global Data Privacy Laws: 120 Countries National Data Privacy Laws, Including Indonesia and Turkey*, 145 PRIV. L. & BUS. INT'L REP. 10, 10, 13 (2017) (giving a detailed geographic distribution of countries engaged in data privacy law).

18. *See generally* Memorandum from Hugo Teufel III, Chief Privacy Officer, U.S. Dep't Homeland Sec. on Privacy Policy Guidance (Dec. 29, 2008), https:// www.dhs.gov/xlibrary/assets/privacy/privacy_policyguide_2008-01.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

19. *Privacy and Personal Data Protection*, OECD (Jan. 2007), https://www.

Protection Directive.[20] Many would agree, at a minimum, that the basic principles of data protection include:

- For all data collected there should be a stated purpose;
- Information collected from an individual cannot be disclosed to other organizations or individuals unless specifically authorized by law or by consent of the individual;
- Records kept on an individual should be accurate and up to date;
- There should be mechanisms for individuals to review data about themselves to ensure accuracy. This may include periodic reporting
- Data should be deleted when it is no longer needed for the stated purpose;
- Transmission of personal information to locations where "equivalent" personal data protection cannot be assured is prohibited, and;
- Some data is too sensitive to be collected, unless there are extreme circumstances that justify such collection (e.g., sexual orientation, religion).[21]

While these basic data protection laws are a valiant start to encouraging data stewardship, the laws fail to envision advanced AI and ubiquitous data gathering.[22] In many ways, privacy has

---

oecd.org/sti/ieconomy/37626097.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

20. Council Directive No. 95/46/EC of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data, art. 2(c), O.J. L 281/31 (1995).

21. Within this area may fall some of the currently labeled National Security exceptions to data collection. *See generally* USA PATRIOT Act of 2001, Pub. L. No. 107-56 (2001) ("Providing Appropriate Tools Required to Intercept Terror"); *see generally* The Homeland Security Act of 2002, 6 U.S.C. § 222 (2002) (establishing the Department of Homeland Security); *see generally* Real ID Act of 2005, H.R. 1268, Pub. L. No. 109-13 (2005) (improving security for drivers' licenses and personal identification cards).

22. *See* Richard Kemp, *Legal Aspect of Artificial Intelligence*, 22 No. 1 CYBERSPACE LAW. NL 2 (2017) (noting that "governments and policy makers around the world are just starting to grapple with what AI means for law and policy and the necessary technical and legal frameworks"); *see also* Sara Rosenbaum, *Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access*, HEALTH SERVS. RES. (Oct. 2010), https://

become the talisman of organized resistance to the encroachment of ubiquitous data gathering. It is a false idol as "private information" is a mischaracterization indicative of a much larger issue that looms on the horizon.

This is because much of the information that exists about a person is information that he or she freely share—thus, it is not, nor should it be, considered private. As such, the newest of data amalgamations create a dilemma for policy makers and privacy advocates, because people share a whole lot of information, which, in small, segregated, units, amounts to nothing. When combined, this information can amount to extremely accurate profiles of who we are as shoppers, as voters, as people, as neighbors, and as community members.[23]

This does not suggest that privacy is not important. It remains important in the traditional manner in which the concept has always existed, governments' intrusion upon areas of our lives that were never intended to be within public scrutiny[24] and areas of protections that should be afforded due to the sensitivity of the information,[25] both of which should continue be considered within the privacy rubric. That begs the question of how should we regulate data that is not, nor was ever intended to be, private? General information, such as age and gender can be gathered from a photograph or my Facebook page. How should we regulate the use of this data, for example, in the creation of a digital profile? As is discussed later, industry-based data stewardship and

---

www.ncbi.nlm.nih.gov/pmc/articles/PMC2965885/ ("Data stewardship is a concept with deep roots in the science and practice of data collection, sharing, and analysis. Reflecting the values of fair information practice, data stewardship denotes an approach to the management of data, particularly data that can identify individuals.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice). The concept of a data steward is intended to convey a fiduciary (or trust) level of responsibility toward the data. *Id*.

23. *See* Louise Matsakis, *This Is How Much Marketers Know About You Based On One Facebook Like*, VICE MOTHERBOARD (Oct. 4, 2017), http://www.motherboard.vice.com/en_us/article/434dpw/this-is-how-much-marketers-know-about-you-based-on-one-facebook-like (describing how easy it is for advertisers to target ads based on social media habits) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

24. *See* Solove, *supra* note 13, at 1–4 (describing the history of privacy as an American priority, dating back to the Revolutionary War).

25. *See id.* at 1–46 (noting the extensive control that private enterprises have over information technology resources).

compliance standards of impact assessments mitigate the consequences of this type of use. The ubiquitous nature of the gathering, storing, sharing, selling, and otherwise compiling information in the digitally connected, always "on" world, will make the governance of information "flows"[26] difficult to achieve.

B. Property and Proprietary Rights

Property as a defining legal structure for data is also troubling, as case law in this area further reinforces the belief that lawmakers simply do not understand new technology.

Databases have historically been treated as property within the law, as it is the storage system owned by an entity, and not the individuals' data that is contained within that storage system, that is entitled to legal protections. These original protections arise in intellectual property law, yet the protections have been retained—even expanded—as databases become more prevalent.

For example, prior to 1990, a directory of information was capable of receiving copyright protections under the Copyright Act of 1976.[27] Although circuits were split on the standard to be applied to determine the scope of protections, both approaches lead to the conclusion that a large majority of databases (then known as compilations) were in fact copyrightable.[28] This all changed in 1991

---

26. Information "flow" is a term of art used within a group of commentators to attempt to capture and explain the manner in which data and information moves through data systems. *See* MATT BISHOP, INTRODUCTION TO COMPUTER SECURITY 436 (2003) (discussing how information flow defines how data travels through a system). I have argued that narratives such as this are crafted to prevent a true understanding of information and to instead create an image of a river that people assume cannot be managed. In this author's opinion, this particular narrative feeds into the "magical thinking" narrative based in "black box," "clouds," and "mathematically based algorithms." The narrative, crafted by industry to (in this author's opinion) feeds the magic narrative—to encourage people to believe the vast majority of individuals cannot understand this technology and cannot control the river.

27. Illinois Bell Tel. Co. v. Haines & Co., 683 F. Supp. 1204, 1208 (N.D. Ill. 1988), *aff'd*, 905 F.2d 1081 (7th Cir. 1990), *vacated*, 499 U.S. 944 (holding that telephone book white pages meet the requirements for copyright protection).

28. *Compare* Feist Publ'gs, Inc. v. Rural Tel. Serv. Co., Inc., 499 U.S. 340, 352 (1991) (describing the sweat of the brow doctrine as granting copyright protection as "a reward for the hard work that went into compiling the work"), *with* Hutchinson Tel. Co. v. Fronteer Directory Co. of Minnesota, Inc., 770 F.2d 128, 131 (8th Cir. 1985) (holding that compilations must contain sufficient

when the U.S. Supreme Court determined in *Feist Publications v. Rural Telephone Service Co.*[29] that the telephone directory was in fact un-copyrightable.[30] In making this determination, the Court clarified that the sole basis for protection under U.S. copyright law is creative originality.[31]

The Court notes that "the vast majority of compilations will pass" the originality test.[32] The cases after *Feist* are informative of the Court's position. In fact, the subsequent cases suggest that the "originality" requirement is very easy to establish. For example, in *Key Publications, Inc. v. Chinatown Today Publishing Enterprises, Inc.*[33] the Second Circuit sustained the "copyrightability" of the yellow pages of a telephone directory, finding that the selection of entries in Key's directory was original.[34] In addition, the arrangement of the directory into categories was original when "viewed in the aggregate" because it "entailed the *de minimis* thought needed to withstand the originality requirement."[35]

While the scope of protection is very "thin,"[36] the ability to copyright is no longer the primary issue. Instead, the focus turns to the scope of protections. In the vast majority of post-*Feist* instances, the appellate cases have found wholesale takings from copyrightable compilations to be non-infringing.[37]

---

creativity in their selection, coordination or arrangement to render them "original works of authorship" entitled to copyright protection), *and* S. Bell Tel. & Tel. Co. v. Associated Tel. Directory Publishers, 756 F.2d 801, 809 (11th Cir. 1985) (stating that a compilation may be copyrighted "even where it merely consists of selection or arrangement of 'facts' which individually would not be copyrightable").

29. *See Feist*, 499 U.S. at 340 (holding that compilations of facts that lack a modicum of creativity are not eligible for copyright protection).

30. *Id.* at 362–64.

31. *See id.* at 358 (explaining that facts are never original, and the only claim on originality can be on the way facts are presented).

32. *Id.* at 359.

33. Key Publ'ns, Inc. v. Chinatown Today Publ'g Enter., Inc., 945 F.2d 509 (2d Cir. 1991) (holding that a published telephone directory for the Chinese-American community did not infringe the yellow page listings because of the lack of overlap between the directory and the yellow page business categories).

34. *See id.* at 513 (recognizing the choice of businesses based on the target audience as well as the author's personal knowledge of the businesses).

35. *Id.* at 514.

36. *Id.*

37. *See* Jane C. Ginsburg, *Copyright, Common Law and* Sui Generis *Protection of Databases in the U.S. and Abroad*, 66 U. CIN. L. REV. 151, 153 (1997)

In response to these legal protection limitations, businesses sought to alter the structure or content of their databases to incorporate greater creativity, thereby providing greater copyright protections. There is little getting around the inability to copyright facts. Thus, even in value-added databases, copyrights only protects information that is considered value-added.[38] To illustrate, West Publication has long held a database of case law.[39] While the case itself is not copyrightable, the case synopsis and indexing system are protected.[40]

Intellectual property rights also include protections for algorithms[41] under patent law. In 1994, the Court of Appeals for the Federal Circuit decided in *In re Alappat*[42] that an invention that had a novel software algorithm combined with a trivial physical step was eligible for patent protection.[43] While many may assume that this and subsequent cases opened the door for a large scale increase in the patenting of software, this has not generally been the case. A 2011 Berkman Center Research report found that "most software firms still do not patent, [and that] most software patents are obtained by a few large firms in the software industry

---

(discussing the shift in how courts viewed compilations following *Feist*).

38. *See* Baila H. Celedonia, *From Copyright to Copycat: Open Season on Data. The Supreme Court's Recent Decision on Directory Copyright Protection Will Affect Similar Works. Here's How to Protect Your Valuable—and Vulnerable—Property*, PUBLISHERS WKLY., Aug. 16, 1991, at 34 (recommending that compilers "consider enriching their publications in terms of subjective analysis of these facts," and attempt to incorporate "value-added subjective selection and arrangement" to make their products more likely to be protected under copyright).

39. *See generally Who We Are*, WEST ACAD., http://home.westacademic.com/legal-publisher (last visited Apr. 15, 2018) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

40. *See* Celedonia, *supra* note 38 (noting that a computer database arranges data through software).

41. *See* Tarleton Gillespie, *The Relevance of Algorithms*, http://www.tarletongillespie.org/essays/Gillespie%20-%20The%20Relevance%20of%20Algorithms.pdf (last visited Apr. 15, 2018), (recognizing that algorithms are encoded procedures for solving a problem by transforming input data into a desired output) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

42. See *In re* Alappat, 33 F.3d 1526, 1545 (Fed. Cir. 1994) (holding that a programmed computer is not patentable, but "a computer operating pursuant to software may represent patentable subject matter" and therefore a computer is an apparatus), *abrogated by In re* Bilski, 545 F.3d 943 (Fed. Cir. 2008).

43. *See id.* at 1542–44 (explaining the technicalities of the mathematical algorithm exception).

or in other industries . . . ."[44] In many ways, this outcome is somewhat unsurprising as the Courts struggled greatly with the issues, ultimately curtailing the availability of business method patents in the 2014 Supreme Court decision of *Alice Corp. v. CLS Bank Int'l*.[45] In fact, Jasper L. Tran notes, "[t]his Article estimates that—without accounting for selection bias—out of roughly 240,000 current patents in force related to computer-implemented inventions as of 2015, about 199,000 of those would likely be invalid patents under *Alice*, leaving about 41,000 valid patents."[46] Assuming the correctness of the estimate, the number is simply staggering. It is important to note, however, that patents remain valid until challenged and invalidated,[47] so the vast majority of these patents will remain in force[48] despite the potential for invalidation.

Moreover, one must consider the two research publications in tandem, the vast majority of these patents have likely been obtained by a "few large firms,"[49] and most will remain valid until invalidated through an expensive and time-consuming process that favors the large, financially robust, firm.[50] And, of course, one can imagine that large firms are in the best position to attack smaller, new market participants, who hold patents that have most likely been incorrectly granted. Consequently, it may be assumed that the existing structure is skewed heavily toward those that hold a larger share of the overall market.

---

44. James Bessen, *A Generation of Software Patents*, 18 B.U. J. SCI. & TECH. L. 241, 241 (2012).

45. *See* Alice Corp. v. CLS Bank Int'l, 134 S. Ct. 2347, 2351 (2014) (concluding generic computer implementation did not translate into a patent-eligible invention).

46. Jasper L. Tran, *Software Patents: A One-Year Review of* Alice v. CLS Bank, 97 J. PATENT & TRADEMARK OFF. SOC'Y 532, 532 (2015).

47. *See* 35 U.S.C. § 282 (2012) ("A patent shall be presumed valid . . . . The burden of establishing the invalidity of a patent or any claim thereof shall rest on the party asserting such invalidity.").

48. *See* Dennis Crouch, *What to do All These Invalid Patents?*, PATENTLYO (Aug. 28, 2014), https://patentlyo.com/patent/2014/08/these-invalid-patents.html (explaining the retroactivity of decisions determining that hundreds of thousands of issued patents lack eligible subject matter) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

49. Bessen, *supra* note 44, at 241.

50. *See id.* at 248 (recognizing the increased risks and costs for determining patent infringement).

Assuming that intellectual property law remains available—in a limited manner—to protect information, the area of law has long focused on the entity that created the original information in a database or algorithm.[51] While this approach may work in some areas, it is immediately obvious that individual information, shared publicly, should not be thought of as 'owned.' It is the creative or original compilation of data within the database that is protected, and the information inside the database that is used in the evidentiary process to demonstrate infringing activity.[52] Simply put, a business cannot own my information in its database, but the business is entitled to have protections that provide consequences to those that intrude upon their database and steal the information wholesale. However, one must wonder how long such protection will provide any comfort to the business. As more and more information is freely and openly shared, as processing power increases, and as technology is increasingly able to quickly and cheaply gather information from disparate locations and then compile it into a "new" database, it is hard to imagine database-based protections as a long-term solution. Like the phone book, databases are capable of minimal protections, but the information inside is "owned" by no one.[53]

The inability to resolve the issue of ownership has contributed to the rise of the control-based regulatory regime. Control—that is

---

51. *See* Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., 916 F.2d 718 (10th Cir. 1990), *cert. granted*, 498 U.S. 808 (1990) (finding a white pages telephone directory to be uncopyrightable and determining the sole basis for protection under U.S. copyright law is creative originality); *see also* U.S. COPYRIGHT OFFICE, REPORT ON LEGAL PROTECTION FOR DATABASES 21 (1997), https://www.copyright.gov/reports/db4.pdf ("In the wake of Feist, database producers were advised to increase the likelihood of copyright protection by incorporating a more subjective selection of facts or a more creative arrangement."); *see, e.g.*, Baila H. Celedonia, *From Copyright to Copycat: Open Season on Data?*, 74 PUB. WKLY. 34 (Aug. 16, 1991) (recommending that compilers "consider enriching their publications in terms of subjective analysis of the facts," and attempt to incorporate "value-added subjective selection and arrangement" to make their products more likely to be protected under copyright laws).

52. *See* Feist Publ'ns, Inc. v. Rural Telephone Service Co., Inc., 499 U.S. 340, 346 (1991) (describing originality as the "touchstone of copyright protection in directories and other fact-based works" and noting that "a compilation is copyrightable only to the extent that it features an original selection, coordination, or arrangement").

53. *Id.*

"what entity had control over the data/information at the time"[54] is growing as a conceptualization of obligations relating to the data/information. Entities within control of data/information have obligations relating to the protection of the information and similar responsibilities, including limitations on sharing with third parties.[55] While control is certainly useful to create obligations for the entity that has the data/information, the concept does not resolve the issue of ownership, which in a property-based regime certainly creates a more complete bundle of rights for all stakeholders.

Imagine property and control in another area of law, an automobile, for example. An individual can lend his/her car and the borrowing individual driving controls the vehicle. The borrowing driver has the responsibilities of driving the car, but he/she is not the owner. Despite the owner not being in control of the car, he/she still holds responsibilities as a car owner, even to those individuals on the road when the individual you lent the car to is driving. In terms of data, imagine if you are the owner of data and you rent out the data to another, you would still have responsibility as it relates to the data, such as ensuring the entity you lent the data to was actually capable of using the data safely. Although this may be difficult to imagine, control is a lesser and limited type of property based right, and full property rights are not being recognized in relation to data/information. Thus, while control can be used in limited circumstances, it does not resolve the issue of ownership and the corresponding rights and obligations that should exist for all stakeholders.

---

54. *See* INFO. COMM'RS OFFICE, DATA CONTROLLERS AND DATA PROCESSORS: WHAT THE DIFFERENCE IS AND WHAT THE GOVERNANCE IMPLICATIONS ARE 6, https://ico.org.uk/media/for-organisations/documents/1546/data-controllers-and-data-processors-dp-guidance.pdf (last visited Apr. 15, 2018) ("The data controller determines the purposes for which and the manner in which personal data is processed . . . . This means that the data controller exercises overall control over the 'why' and the 'how' of a data processing activity.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

55. *See* Commission Regulation 2016/679, art. 5, 2016 O.J. (L 119) 1 (obligating data controllers to ensure that personal data is "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures").

C. Consent and Data Compliance

In general, consent is a term best left to contract historians as the importance of consent as a legal doctrine has long been lost in the digital world.[56] In the United States, consent has been watered down to such a large extent that many scholars argue it is reduced to a mere check box in the privacy debate.[57]

As I have previously argued elsewhere, even when individuals are required to "give their consent," it is often part of a "must, rush, and trust" process. [58] That is, in the digital world, individuals *must* have the information, item, or webpage, they are in a *rush* to get what they want, and they *trust* the law to protect them.[59] Any affirmative click that stands in the way of this "must, rush, and trust" instinct is merely a formality, and thus, no longer serves the primary function that consent previously stood to fulfill.[60] Moreover, numerous research scientists have argued that even in the face of multiple click boxes and forced reading screens,

---

56. *See* Anjanette H. Raymond, *The Consumer as Sisyphus: Should We Be Happy with "Why Bother" Consent?*, 20 J. LEGAL STUD. BUS. 1, 2 (2017) (offering context for the place of consent in consumer purchasing).

57. *See id.* at 18–19 (noting that customers are largely helpless against "endless reams" of boilerplate terms they have no opportunity to negotiate or influence, and suggesting use "smart contracts" as one way to remedy this problem). That is not to write that this issue has been ignored by policy makers abroad, however, as in August 2017, the United Kingdom Department for Digital, Culture, Media and Sport announced its intentions to update the existing law. *See* DEP'T FOR DIG., CULTURE, MEDIA & SPORT, A NEW DATA PROTECTION BILL: OUR PLANNED REFORMS 2 (Aug. 17, 2017), https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/635900/2017-08-07_DP_Bill_-_Statement_of_Intent.pdf ("The Bill includes tougher rules on consent, rights to access, rights to move and rights to delete data. Enforcement will be enhanced, and the Information Commissioner given the right powers to ensure consumers are appropriately safeguarded.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

58. *See* Anjanette H. Raymond, *Yeah, But Did You See the Gorilla? Creating and Protecting an Informed Consumer in Cross-Border Online Dispute Resolution*, 19 HARV. NEGOT. L. REV. 129, 146 (Spring 2014) (noting that consumers save time by making presumptions about the terms presented to them, and knowing that they have little real choice in the matter, often speedily agree to contracts of adhesion).

59. *See id.* (stating that all consumer behavior of this sort relies on trust that the terms and conditions are not particularly egregious).

60. *See id.* at 144 (explaining that a consumer will act passively if he or she feels powerless).

individuals still fail to read terms, often fail to reflect upon terms they do not understand, and rarely refuse to click consent.[61] Thus, businesses are able to dictate vague, widely ambiguous, overly broad terminology that seeks to capture the individual's consent to any and all use of data, regardless of impact or intent, for now and forever.[62]

For example, in August 2017, Tesla, Inc. confirmed that the Tesla Model 3 has a driver-facing camera embedded into the rearview mirror of each vehicle.[63] Although the company insists the camera has not yet been activated, the company states it "will only become active after future software updates."[64] Of course, this would be an example of technology that is either covered under the existing vague terms of service or would be part of a 'accept new terms of service' consent that frequently accompanies software updates. This practice is so prevalent that popular Comedy Central program "Southpark" has prominently featured and discussed the issue for many years now.[65]

---

61. *See id.* at 143 (describing research that supports a consumer's increased propensity to provide consent when engaged in digital commerce).

62. *See* Dani Deahl, *Roombas have been busy mapping our homes, and now that data could be shared*, THE VERGE (July 24, 2017), https://www.theverge.com/platform/amp/2017/7/24/16021610/irobot-roomba-homa-map-data-sale (considering the case of the iRobot manufacturer, which after years of mapping private homes, changed TOS so that it may sell the information gathered about those homes) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

63. *See* Fred Lambert, *Tesla Model 3 is equipped with a driver-facing camera for Autopilot and Tesla Network*, ELECTREK (Aug. 1, 2017, 1:17 PM), https://electrek.co/2017/08/01/tesla-model-3-driver-facing-camera-autopilot-tesla-network/ (speculating why this camera has been placed in Tesla automobiles and how it will be used) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

64. *Id.*

65. *See* Jason D. O'Grady, *South Park parodies iTunes terms and conditions*, ZD NET (Apr. 28, 2011, 21:07 PM), http://www.zdnet.com/article/south-park-parodies-itunes-terms-and-conditions/ (accepting truth makes humor more accessible) (on file with the Washington & Lee Journal of Civil Rights & Social Justice); *see also* David Kravets, *TOS agreements require giving up first born—and users gladly consent*, ARS TECHNICA (July 12, 2016, 6:20 PM), https://arstechnica.com/tech-policy/2016/07/nobody-reads-tos-agreements-even-ones-that-demand-first-born-as-payment/ (describing a study demonstrating few people read terms of service agreements before offering consent) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

As noted by the Information Commissioners Office in the United Kingdom:

> This [consent] is seen as incompatible with big data analytics due to its experimental nature and its propensity to find new uses for data, and also because it may not fit contexts where data is observed rather than directly provided by data subjects.[66]

Of course, as technology designers take note of the legal limitations created by the simple binary model of consent, new processes may be developed that address concerns.[67] For example, it may be possible to have a "process of graduated consent, in which people can give consent or not to different uses of their data throughout their relationship with a service provider, rather than having a simple binary choice at the start."[68]

In addition, even if graduated consent is used, the sheer volume of data gathering and storage is an ongoing issue for organizations.[69] Data protection regulation in the European Union requires, for example, storing personal data securely,[70] keeping personal data up to date,[71] permitting data subjects to access their personal data,[72] complying with requests from data subjects for

---

66. INFO. COMM'RS OFFICE, BIG DATA, ARTIFICIAL INTELLIGENCE, MACHINE LEARNING AND DATA PROTECTION 30 (2017) [hereinafter ICO Big Data] https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

67. *See* Raymond, *supra* note 56, at 146 (suggesting the protection of consumers from clauses that allow the stronger party to use arbitration, where they have a heavy advantage).

68. *See* O'Grady, *supra* note 65 (accepting truth makes humor more accessible).

69. *See* Andrew Cave, *What Will We Do When The World's Data Hits 163 Zettabytes In 2025?*, Forbes (Apr. 13, 2017, 2:22 PM), https://www.forbes.com/sites/andrewcave/2017/04/13/what-will-we-do-when-the-worlds-data-hits-163-zettabytes-in-2025/#5bf09889349a (describing how a growing overabundance of data will require additional consideration by corporations) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

70. *See* Commission Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1 (setting out the rights of every data subject to obtain from the controller information regarding how and if their personal data is being used).

71. *See id.* at 36–37 (placing limitations permissible processing of personal information under Article 6).

72. *See id.* at 39–40 (mandating transparency and availability of personal data under Article 12 as rights).

their personal data to be deleted,[73] and actively deleting personal data once it is no longer required for the purpose for which it was collected,[74] to mention but a few ongoing data management issues. Organizations, therefore, must be active in their data management compliance as simple cyber security protections are no longer enough, nor is the creation of a simple single event data management policy.[75]

Yet, the regulation fails to appreciate the impossibility of such a data protection regime on business and the uselessness of it in protecting individuals and their information. Requesting data as an individual presupposes that the requesting party knows where the information is, what entity to request the information from, and that the data has not already been used by another entity.[76] In general, and the vast number of circumstances, you will not be able to untangle data from the information infrastructure. And, to keep individuals abreast of their information, any change of the use of the information requires new consent.[77] Non-stop email-based notifications and consent will not increase knowledge, and will only lead to information overload, resulting in individuals not glancing at the information but just merely clicking consent, or in the alternative, incentivizing businesses to include vague, over-encompassing, overly broad consent clauses which allow any and all information sharing.[78] Neither reality is one that society should support. Consent is irrelevant in the majority of situations in the digital world.

D. Transparency

Statistical models that accurately predict an outcome or classify an object have traditionally been transparent in their

73. *See id.* at 36–37 (outlining the "right to erasure" under Article 17).

74. *Id.*

75. *Id.*

76. *Id.* at 43.

77. *Id.*

78. *See* Schumpeter, *Too Much Information*, ECONOMIST (June 30, 2011), http://www.economist.com/node/18895468 (discussing the issues posed by data overload and suggesting that both individuals and corporations must work to combat the issue) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

reasoning, as scholars have been able to see their workings and interrogate their internal logic.[79] For some AI algorithms, this is much more difficult and, in some instances, in the current environment, is impossible.[80]

Of course, to date, the assumptions that support the operations have been rule based, in which computers are merely following instructions and any errors are attributed to the people.[81] Some AI will turn this process on its head, as the machine itself will find patterns and create rules.[82] In this way, it may just be that the machine is at fault. Moreover, as previously argued, automated processes and hidden black box worlds based in math and number crunching can still carry "an aura of objectivity and infallibility."[83] Thus, even when some transparency is possible, individuals tend to 'trust the numbers' despite evidence to the contrary.[84]

The inability to understand the inner workings of an algorithm should not be permitted to limit the ethical or legal decisions about the consequences of the actions taken by the algorithm. In this vein, the European Union is attempting to regulate the use of machine learning: for example, starting in 2018,

---

79. *See* Dave Weinberger, *Our Machines Now Have Knowledge We'll Never Understand*, WIRED (Apr. 18, 2017, 8:22 PM), https://www.wired.com/story/our-machines-now-have-knowledge-well-never-understand/ (describing the development of modern machine learning and the differences that machine created models have from ones that humans have created) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

80. *Id.*

81. *Id.*

82. *Id.*; *See also* Mark Wilson, *AI Is Inventing Languages Humans Can't Understand. Should We Stop It?*, CO.DESIGN (July 14, 2017), https://www.fastcodesign.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it? (discussing two Facebook artificial intelligences that created a shorthand code for communication before being reprogrammed) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

83. *See* Anjanette H. Raymond et al., *Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias*, Kelley School of Business, Research Paper No. 17-23, 7–13 (2017), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2921442 (discussing the black box and invisible bias surrounding algorithms, analytics, and machine learning) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

84. *See id.* at 11–13 (discussing the tendency to rely on machine learning even when the logic underlying the algorithms is problematic and not based on the scientific method).

EU citizens will be entitled to know how an EU institution arrived at a conclusion, even if machine learning and a black box were involved.[85] As University of Oxford researcher Bryce Goodman explains, the new data protection law entitled the General Data Protection Regulation (GDPR) is effectively a "right to an explanation" for decisions.[86] In fact, the law does more: it bans decisions "based solely on automated processing, including profiling, which produces an adverse legal effect concerning the data subject or significantly affects him or her."[87]

Unfortunately, the Data Protection Regulation will likely be viewed as a regulatory failure and an extreme over-reach that may just result in clever work arounds or the stifling of innovation.[88] For example, the use of the term "solely" will allow for compliance

---

85. *See* Bryce Goodman & Seth Flaxman, *European Union regulations on algorithmic decision-making and a "right to explanation"*, ICML WORKSHOP ON HUM. INTERPRETABILITY MACHINE LEARNING 1 (Aug. 31, 2016), https://arxiv.org/pdf/1606. 08813.pdf (summarizing "the potential impact that the European Union's new General Data Protection Regulation will have on the routine use of machine learning algorithms") (on file with the Washington & Lee Journal of Civil Rights & Social Justice). The use of "algorithmic regulation" may emerge as an option to fill in the gaps presented by static laws that cannot keep up with new advances in technology, as explained:

> Algorithmic regulation refers to decision-making systems that regulate a domain of activity in order to manage risk or alter behavior through continual computational generation of knowledge by systematically collecting data (in real time on a continuous basis) emitted directly from numerous dynamic components pertaining to the regulated environment in order to identify and, if necessary, automatically refine (or prompt refinement of) the system's operations to attain a pre-specified goal.

Karen Yeung, *Algorithmic Regulation: A Critical Interrogation*, KING'S C. LONDON L. SCH., Research Paper No. 2017-27, at 1 (2017) https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972505 (on file with the Washington & Lee Journal of Civil Rights & Social Justice). Please further note that this study maps the contours of several emerging debates surrounding algorithmic regulation, drawing upon insights from regulatory governance studies, legal critiques, surveillance studies and critical data studies to highlight various concerns about the legitimacy of algorithmic regulation. *Id.*

86. *See generally* Goodman & Flaxman, *supra* note 85, at 1 (summarizing the potential impact of the GDPR).

87. *Id.*

88. *See* Andrew D. Selbst & Julia Powles, *Meaningful information and the right to explanation*, 7 INT'L DATA PRIVACY L. 233, 233 (2017) ("The GDPR is an ambitious, complicated, contested law aimed at making Europe 'fit for the digital age.'").

so long as the decision was not based exclusively in advanced AI, for instance when a human is involved in even a rudimentary aspect of the process.[89] Additionally, the right to an explanation should not be viewed as a right to transparency of process.[90] In fact, it is a right to nothing more than a basic explanation of the decision process, such as the notifications that occur when you are rejected for credit.[91] Both of these serve as examples of the need to understand the basic inner working of the process and to seek to achieve a particular goal within the process structure. Otherwise, there is a risk of creating a set of rules that are likely to be easily subverted, and therefore meaningless.

Of course, if one conditions taking action on first achieving some kind of objective scientific understanding about what caused something to happen, there exists the risk of either obtaining a license not to take any action at all, or creating false narratives that seem to imply the existence of objective scientific understanding of something when none exists.

Currently, the inner workings of an advanced AI is almost impossible to understand, and no amount of transparency will remedy this predicament.[92] As a result, transparency should be abandoned, at least as the term is used within the context of process,[93] as it creates a false belief that human beings will

---

89. *See* Commission Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1 ("The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.").

90. *See* Goodman & Flaxman, *supra* note 85, at 1 (discussing the meaning of a "right to explanation" in the GDPR).

91. *See* Commission Regulation 2016/679, art. 13–15, 2016 O.J. (L 119) 1 (providing rights to "meaningful information" about the logic involved in automated decisions).

92. *See* Will Knight, *The Dark Secret at the Heart of AI*, MIT TECH. REV. (Apr. 11, 2017), https://www.technologyreview.com/s/604087/the-dark-secret-at-the-heart-of-ai/ (explaining that it may be impossible to explain AI processes because the computers have programmed themselves, and even those who build the technology cannot fully explain their process) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

93. That is not to write that all transparency is useless, it is not. Outcomes of processes should be transparent as they are what can be used to examine the effectiveness of the evidence-based problem solving that was deployed. *See* Nicole Laskowski, *Data transparency is the lifeblood of new information economy*, TECHTARGET (Feb. 2017), http://searchcio.techtarget.com/opinion/Data-transparency-is-the-lifeblood-of-new-information-economy (explaining the importance of

understand and interpret the inner working of AI algorithms, and allows an impossible objective to serve as a roadblock to the creation of regulatory frameworks. Instead, transparency of outcomes, with the ability to audit outputs[94] that are the product of the algorithm, should be part of the design process, a point that will be discussed in detail in Part D, Section 6 (below).

III. The Regulatory Future

The law and ethics surrounding new fields of innovation do not have to be handled in a substantially new manner, so long as policy makers do not succumb to the dazzle of the technology.[95] In fact, "[s]tates and societies are already equipped to navigate human error and typically have a range of escalating options for responding to the types of risks which could be associated with new products or services that are not responsibly deployed."[96] Consequently, some commentators argue that the best approach to regulation of new technology is to adopt a design model based upon the identified spectrum of risk.[97]

A. The Basics

In general, the following diagram sets out the areas of concern that, that must first be considered as the key divisions. These original delineations can then be expanded into more widespread

---

data transparency) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

94. *See* Anjanette H. Raymond, *A Meeting of The Minds: Online Dispute Resolution Regulations Should Be Opportunity Focused*, 16 U.C. DAVIS BUS. L. J. 189, 210–13 (2016) (discussing the importance of transparency as it relates to online dispute resolution).

95. *See* Bernd Carsten Stahl et al., *Ethics of Emerging Information and Communication Technologies: On the Implementation of Responsible Research and Innovation*, 44 SCI. & PUB. POL'Y 369, 369, 381 (2016) (discussing ethical considerations and responsible practices associated with handling certain information, especially with the emergence of new technologies).

96. KINGSLEY ET AL., *supra* note 2, at 16.

97. *See* RISK AND SOCIETY: THE INTERACTION OF SCIENCE, TECHNOLOGY AND PUBLIC POLICY 17–32 (Marvin Waterstone ed., 1992) (discussing risk analysis as a tool for policy decisions).

discussion within each area. It is essential to understand that this diagram does not truly capture information movement, it is simply an illustration to initiate conversation.

As the diagram captures, there are multiple points to consider in the overall governance structure. First, data, as discrete individual units, can be analyzed based on the data and only the data.[98] The diagram does not fully illustrate the multiple sources, both public and private, from which data arises—nor does it capture the ebb and flow of data into and out of the system.[99] Yet, in the most rudimentary sense, at some point, data exists and this data must be subjected to the same rigorous review that any data within social sciences is subjected.[100] Issues such as (1) where did the data come from; (2) does it capture the community we seek to capture; (3) is it verifiable, accurate, and robust; and similar

98. *See* THOMAS J. SANTNER & DIANE E. DUFFY, THE STATISTICAL ANALYSIS OF DISCRETE DATA 2 (Stephen Finberg et al. eds., 1989) (discussing the differences between discrete and continuous data).

99. *See* Bernard Marr, *Big Data: 33 Brilliant And Free Data Sources Anyone Can Use*, FORBES (Feb. 12, 2016, 2:42 AM), https://www.forbes.com/sites/bernardmarr/2016/02/12/big-data-35-brilliant-and-free-data-sources-for-2016/#19ca978db54d (discussing sources of data including public databases and data collected by private companies) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

100. *See generally* DAVID DE VAUS, ANALYZING SOCIAL SCIENCE DATA: 50 KEY PROBLEMS IN DATA ANALYSIS (2002) (explaining methods of analyzing social sciences data).

issues.[101] Second, the final box is the 'information' box and can be thought of as a single unit as well. The output from the transformative process should also be subjected to rigorous scrutiny.[102] Issues such as, is the outcome capable of replication; can it be generalized; is it biased or discriminatory on its face; and similar issues.[103]

It is the transformative process, the middle box, that is semi-unexplored to date. For example, the transformative process can demand that entities and policy makers think about: (1) What is allowed/acceptable to go into the transformative process? (2) What happens to the data during the transformative process, such as (a) what assumptions are allowable; (b) what automations should be allowed to occur; (c) what holes (or missing data) is allowed to be filled and how should those holes be filled; and (d) What biases are allowed, or otherwise permissible? (3) Does what is actually spit out match our expectations? Is it problematic on its face? (to name but a few).[104]

Fundamentally, assuming that the main points of discussion presented into the above diagram are correct, the question that remains is how to create a governance structure that appreciates the risks at the various points, considers the impacts of the risks, and attempts to reduce the impact of identified risks through mitigation. The following sections break out these issue in greater detail.

---

101. *See* FED. R. EVID. 702 (highlighting the questions to ask to determine whether expert testimony is reliable).

102. *See* UNIV. OF CAL. MUSEUM OF PALEONTOLOGY, BERKELEY, & REGENTS OF THE UNIV. OF CAL., THE SOCIAL SIDE OF SCIENCE: A HUMAN AND COMMUNITY ENDEAVOR 6–9 (2013), https://undsci.berkeley.edu/lessons/pdfs/social_side_of_science.pdf (explaining the importance of strict scrutiny of data in the scientific community) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

103. *See* DE VAUS, *supra* note 100, at 147–48, 152–53 (discussing factors to determine whether the data is reliable).

104. *See* FED. R. EVID. 702 (highlighting the questions to ask to determine whether expert testimony is reliable).

B. Spectrum of Risk Model

Risk management is the process of identifying risk, assessing risk, and taking steps to reduce risk to an acceptable level.[105] Traditionally used in the information technology industry, the cornerstone of the process is a risk impact assessment in which the entity determines the probabilities and consequences of risk events if the event were to occur.[106] The results are then used to prioritize risks in terms of their importance.[107] These rankings allow project's management to strategize resource allocation and to work to mitigate "high probability/high consequence risk events."[108] Within the field, enterprise environments are the emerging area.[109]

> Enterprise environments (e.g., the Internet) offer users ubiquitous, cross-boundary access to wide varieties of services, applications, and information repositories. Enterprise systems engineering is an emerging discipline. It encompasses and extends "traditional" systems engineering to create and evolve "webs" of systems and systems-of-systems that operate in a network-centric way to deliver capabilities via services, data,

---

105. GARY STONEBURNER ET AL., NAT'L INST. OF STANDARDS & TECH., RISK MANAGEMENT GUIDE FOR INFORMATION TECHNOLOGY SYSTEM 1 (2002), https://www.hhs.gov/sites/default/files/ocr/privacy/hipaa/administrative/securityrule/nist800-30.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice); *see also Risk Management Approach and Plan*, MITRE, https://www.mitre.org/publications/systems-engineering-guide/ acquisition-systems-engineering/risk-management/risk-management-approach-and-plan (last visited Apr. 15, 2018) [hereinafter *Risk Management Approach & Plan*] (providing a definition of risk management) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

106. *See Risk Management Approach & Plan*, *supra* note 105 (identifying risk impact assessments as a key step in the risk management processes).

107. *See Risk Impact Assessment and Prioritization,* MITRE, https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management/risk-impact-assessment-and-prioritization (last visited Apr. 15, 2018) [hereinafter *Risk Impact Assessment & Prioritization*] (explaining the use of prioritization and at what step of the risk impact assessment process it occurs) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

108. *See id.* (identifying the benefits associated with prioritizing risk).

109. *See Risk Management Approach & Plan*, *supra* note 105 (discussing the emergence of risk management in enterprises, and specifically in engineering field programs).

> and applications through an interconnected network of information and communications technologies.[110]

And, while enterprise risk assessment is a growing standard to be used by entities when considering the risk of interruption or failures in the delivery of a system/program—the principles that stand behind the risk assessment can be generalized to produce an overarching information governance framework.[111]

In the information governance risk assessment, an example of a preferred approach could start with "risk" being based on the risk of information being used in a manner that is detrimental to an individual or entity.[112] Thus, the process would seek to identify data and /or information and then consider the impact of the use of data/information upon various stakeholders.[113] Assuming the question is not one of mere legal compliance, one can imagine a risk management matrix that will document the following items:

1. Entity Identification, Data/Information in Question, and Intended Use
2. Risk and Consequences
3. Probability—probability of the risk occurring
4. Impact what is the impact on the various stakeholders if the risk should occur
5. Priority—based on impact and the probability of occurrence

---

110. *See id.* (explaining enterprise environments and their aid in the advancement of traditional engineering systems).

111. *See IT Governance*, MITRE, https://www.mitre.org/publications/systems-engineering-guide/enterprise-engineering/enterprise-planning-and-management/it-governance (last visited Apr. 15, 2018) (explaining the use of prioritization and at what step of the risk impact assessment process it occurs) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

112. *See What Constitutes Information Governance?*, INFOGOV BASICS, https://www.infogovbasics.com/what-is-infogov/what-constitutes-information-governance (last visited Apr. 15, 2018) (defining information governance) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

113. *See id.* (demonstrating that an information governance risk assessment is used to determine the severity in which an individual or entity is impacted in the case of irresponsible handling of that individual or entity's information).

6. Mitigation Response—a brief overview of mitigation steps to eliminate or reduce the risk.[114]

These questions can then guide the development of global information governance regime.[115] For example, consider the use of highly sensitive information, such as a significant medical diagnosis.[116] The risk of this information being gathered is that it could be used to discriminate in a variety of ways.[117] The first question is to 'classify' the information (or data).[118] To do this we first ask, what entity is seeking to gather/use this information?[119] How do they intend to use the information?[120] We then consider, will this use have a high impact on the individual (or other stakeholders)?[121] Moving on we consider what is the probability of

114. *See Risk Impact Assessment & Prioritization*, *supra* note 107 (outlining a risk management matrix and including questions to guide the development of global information governance).

115. *See* BERTRAND G. RAMCHARAN, INTERNATIONAL PEACE CONFERENCES 15 (2015) ("Global governance [is] the complex of formal and informal institutions, mechanisms, relationships, and processes between and among states, markets, citizens and organizations, both inter- and non-governmental, through which collective interests on the global plane are articulated, Duties, obligations and privileges are established, and differences are mediated through educated professionals.").

116. *See* Ajit Appari & M. Eric Johnson, *Information Security and Privacy in Healthcare: Current State of Research*, 16 (Aug. 2008), www.ists.dartmouth.edu /library/416.pdf (noting the highly sensitive nature of information contained in medical records) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

117. *See Risk Impact Assessment & Prioritization*, *supra* note 107 (describing risks associated with unauthorized information gathering).

118. *See Overview*, MITRE, www.mitre.org/capabilities /systems-engineering /overview (last visited Apr. 15, 2018) (outlining the best steps to take and questions to guide the development of global information governance regime in light of risk and sensitive information) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

119. *See IT Governance*, *supra* note 111 (explaining how to identify the gatherer of the information or data).

120. *See id.* (identifying a gatherer's specific intent to use certain information or data as relevant to the risk assessment).

121. *See id.* (describing the impact of the gatherer's intended use of information or data).

the risk occurring?[122] Are there mitigation steps that could be put in place to lessen the risk or impact?[123]

After this information gathering process is complete we can classify the data/information.[124] In the example of medical diagnosis in the hands of an insurer, one can see this information should be classified as critical (or highly sensitive).[125] This information deserves the highest protections and thus should be subject to legal protections and prohibitions on the ability to share the information beyond a one-time, single use.[126] This information should not be gathered unless absolutely necessary, should not be stored for any longer than absolutely necessary, and should not be used in a manner that could negatively impact the individual or others.[127] This type of data could also be subject to mandatory mitigation measures,[128] for example if the diagnosis was essential for billing purposes the information could be required to be coded in generalized terms (long term illness versus asthma).[129]

As can be surmised from the brief description of the risk model, the model can accommodate various concerns that arise in the lifecycle of data and the information web.[130] For example, data that is deemed critical (that is, data that has a high risk of harm if lost), will fall within the highest range of considerations and legal

---

122. *See id.* (identifying the probability of risk as a step in the overall information governance risk assessment process).

123. *See id.* (explaining that final steps of the information governance risk assessment process require an inquiry into whether mitigating steps would decrease the overall risk).

124. *See id.* (specifying classifications involved in the risk assessment process).

125. *See* Appari & Johnson, *supra* note 116, at 6 (referencing insurers and their relevant interests in obtaining medical information).

126. *See id.* at 3 (outlining the privacy regulations surrounding health information and the state of information security research).

127. *See id.* (explaining the threats gathering and storing health information pose to privacy).

128. *See id.* at 5 (discussing how effective mitigating measures vary based on the intended use of information).

129. *See id.* (describing mitigating measures when the intended use of information involves economic value).

130. *See Risk Impact Assessment & Prioritization*, *supra* note 107 ("Risk cutes across the life cycle of systems engineering, and MITRE SEs should be prepared to address the risk throughout.").

protections.[131] As such, critical data will need to be protected from loss (via breach, hack, accidental loss, etc.) and third party use through a regulatory structure.[132] Non-critical data (data that is widely available or is otherwise part of the public knowledge base) can be protected, if through industry-created standards.[133] The model can also accommodate the current concerns that arise in the use of data, such as the completion of data to create a digital profile.[134] The model can thus accommodate the use and thereby the impact of the data use within the risk framework.[135] For example, digital profile creation that is then used to create and customize game player experience, could be governed by industry standards as the risk of impact upon the individual or society is low.[136] Digital profile use in the insurance industry or in policing might be deemed high risk as these digital profiles may be used as mechanisms of discrimination or price inflation and, as such, this type of data use would be highly regulated.[137]

---

131. *See* Appari & Johnson, *supra* note 116 (noting legal considerations and regulations currently in place that attempt to protect information of all degrees of sensitivity).

132. *See id.* (recognizing the need for further advancements of the regulatory scheme involving the protection of critical data and potential breaches or third party use of that data).

133. *See* Woodrow Hartzog & Daniel J. Solove, *The Scope and Potential of FTC Data Protection*, 83 GEO. WASH. L. REV. 2230, 2233 (2015) (discussing how data is protected through industry standards).

134. *See* Annum Munir, *Learn How to Create a Complete Customer Profile Using Data (in 7 Steps)*, LOCALYTICS (June 3, 2015), http://info.localytics.com/blog/learn-how-to-create-a-complete-customer-profile-using-data-in-7-steps (explaining how to use data to create a customer profile) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

135. *See Risk Impact Assessment & Prioritization*, *supra* note 107 (discussing risk assessment model).

136. *See* Jukka Vahlo et al., *Digital Game Dynamics Preferences and Player Types*, 22 J. COMPUTER-MEDIATED COMM. 88, 88 (providing "new knowledge on how players' gaming preferences and different types of digital games can be analyzed within a single research framework based on activity theoretical considerations and game design concepts").

137. ACCENTUREDIGITAL, DIGITAL POLICING POWERED BY ANALYTICS: ACTIONABLE PUBLIC SAFETY INSIGHTS 2, 4–8 (2016), https://www.accenture.com/t20170412T030029Z__w__/us-en/_acnmedia/PDF-34/Accenture-Digital-Policing-Powered-by-Analytics.pdf#zoom=50 (discussing the use of analytics to assist police officers) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

In addition, the enterprise risk assessment can be used on an individual (or more local) level as well.[138] Locally, a business may use the model, for example to determine the use of a particular algorithm in a specific situation.[139]

C. Industry Key Recommendations

As one would imagine, we are most likely a long way from the creation of a framework to govern the regulation of the transformative process.[140] As such, industry should consider what steps it can and should take to begin the process of discovering the important aspects of the emerging governance.

One of the better examples of grassroots, industry-based technology regulation exists in the area of Bluetooth.[141] The Bluetooth Special Interest Group (SIG) is the body that oversees the development of Bluetooth standards and the licensing of the Bluetooth technologies and trademarks to manufacturers.[142] The SIG is a not-for-profit, non-stock corporation which does not make, manufacture or sell Bluetooth enabled products.[143] Any company incorporating Bluetooth wireless technology into products, using the technology to offer goods and services or simply re-branding a product with Bluetooth technology, must become a member of the

---

138. *See* KAREN HARDY, MANAGING RISK IN GOVERNMENT: AN INTRODUCTION TO ENTERPRISE RISK MANAGEMENT 9 (2d ed. 2010), https://enterrasolutions.com/media/docs/2013/09/RiskinGovernment.pdf (highlighting the benefits of the risk assessment framework when applied to at the individual level) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

139. *See id.* at 10 ("Understanding and managing risk is essential for any organization, public or private. In the private sector, risk management is a widely accepted practice designed to control risks that could lead to a business failure if not properly managed.").

140. *See Risk Management Approach & Plan, supra* note 105 (noting that at the enterprise level, governance and complexity risks become more apparent).

141. *See Bluetooth SIG*, REVOLVY, https://www.revolvy.com/main/index.php?s=Bluetooth%20SIG&item_type=topic (last visited Apr. 15, 2018) (describing the organizational structure of the Bluetooth group and identifying the Bluetooth group as an important body with significant oversight or development standards in the industry) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

142. *Id.*

143. *Id.*

Bluetooth SIG.[144] As such, Bluetooth SIG is able to create a membership of users that must complete the qualification and declaration process for their Bluetooth enabled product(s) to demonstrate and declare compliance to the Membership Agreements.[145]

I have written before about the ability of a technology developer to create a "captured" participation group and to insist that participants comply with standards to remain part of the user group.[146] One classic example is Apple or Facebook, who both create log in barriers to entry and remove members who fail to comply with platform rules. Industries that use the "captured" participants model are at an advantage to others because membership is revocable.[147] It is however, these industries that also have an incentive to monitor behavior and as such, may just be those with the greatest knowledge in community expectations going forward.[148]

In terms of the general industry guidance, it is possible to create user agreements that place restrictions upon data and insist upon compliance with existing community standards and laws.[149] And, although not discussed below, it is one of the easiest and possibly one of the more effective means to influence community and industry standards.[150]

That does not mean that industry must start from the beginning, some guidance already exists.[151] To this end,

---

144. *Id.*

145. *Id.*

146. *See* Anjanette Raymond, *The Dilemma of Private Justice Systems: Big Data Sources, the Cloud and Predictive Analytics*, 35 Nw. J. INT'L L. & BUS. 1A, 3A (2015) (discussing the binding impacts that big data and the continued advancements in technology have on consumers and users).

147. *See id.* (describing information gathering techniques employed by certain technology developers, which had negative effects on participants).

148. *See id.* at 5A (noting the incentive industries collecting information have to monitor gathering techniques and protect the community accordingly).

149. *See* Appari & Johnson, *supra* note 116, at 4–5 (outlining current regulations and laws affecting information gathering).

150. *See* Joshua Fairfield, *The Cost of Consent: Optimal Standardization in the Law of Contract*, 58 Emory L. J. 1401, 1455 (2009) (discussing the evolution of contracts and industry standards).

151. *See* Appari & Johnson, *supra* note 116, at 10–11 (noting the general industry guidance currently in place and the effectiveness and ineffectiveness of that guidance).

commentators have suggested the following key recommendations for any industry to consider as a first step in creating a framework for information governance.[152]

1. Incorporate Impact Assessment

According to Information Commissioners Office in the United Kingdom, privacy impact assessments (PIAs) "are a tool which can help organizations identify the most effective way to comply with their data protection obligations *and meet individuals' expectations of privacy*."[153] While the author's hesitation has already been written about, with the continued use of the term 'privacy,' it is important to note the full breadth of its use within this context and within this particular assessment, privacy includes physical and informational privacy,[154] including:

> [T]he ability of a person to control, edit, manage and delete information about themselves and to decide how and to what extent such information is communicated to others. Intrusion can come in the form of collection of excessive personal information, disclosure of personal information without consent and misuse of such information.[155]

This is a much broader conceptualization of privacy than exists in the United States and is thus a good reminder to always consider definitions and cultural influence in the development and commentary on policy.[156] The British conceptualization of the impact assessment is to "minimizing . . . the risk of harm through

---

152. *See generally* SUSAN MEAKIN, INFORMATION GOVERNANCE STRATEGIC MANAGEMENT FRAMEWORK 2016–2018 (2016), http://www.rdash.nhs.uk/wp-content/uploads/2014/06/IG-Strategic-Framework-v2-20162018.pdf (discussing how to create a framework for information governance) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

153. INFO. COMM'RS OFF., CONDUCTING PRIVACY IMPACT ASSESSMENTS CODE OF PRACTICE 4 (2014) [hereinafter ICO] (emphasis added) https://ico.org.uk/media/for organisations/documents/1595/pia-code-of-practice.pdf (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

154. *See id.* (noting the expanded breadth of the term privacy in the context of impact assessments).

155. *Id.* at 6.

156. *See* PRIVACY IN AMERICA: INTERDISCIPLINARY PERSPECTIVES 248 (William Aspray & Philip Doty eds., 2011) (discussing privacy in the United States).

use or misuse of personal information."[157] While the entire coverage of the Code is too large for inclusion in this paper, the Code provides numerous explanations, flow charts, formative questions, and explanations to assist organizations in conducting the assessment and all provided with an eye toward compliance with the Data Protection Act.[158] Most relevant for the paper is the key consideration that "[a]s part of the PIA process organizations should describe how information is collected, stored, used and deleted. They should explain what information is used, what it is used for and who will have access to it."[159]

Consider a simple example of an impact assessment, that is not necessarily privacy focused, but is nevertheless an example of how a broad impact assessment can measure impact risks based upon the impact of the particular use on multiple stakeholders. Consider the transformative process (middle box) mentioned above. If the data that is entered allows gender to be a data point within the transformative process, the designer should be expected to consider if the use of gender is impactful to the overall output. If so, and the impact is a positive one, then gender may be appropriate under the circumstances. But, if gender introduces negative impacts, such as the introduction of discrimination, without the need for such an introduction—then gender should be eliminated from inclusion.

Now, consider a real-world example: for a very long time the medical community assumed, despite years of research, that heart attacks presented in the same manner across individuals.[160] It turns out the research was incomplete.[161] It seems that gender,

---

157. ICO, *supra* note 153, at 7.

158. *See id.* (noting that neither compliance with the Code nor completion of the impact assessment is a requirement of the Data Protection Act).

159. *Id.* at 19.

160. *See* Colleen Story, *Symptoms of a Heart Attack*, HEATHLINE (Apr. 10, 2012), http://www.healthline.com/health/heart-disease/heart-attack-symptoms#symptoms-in-men3 (noting that symptoms of heart attacks show up in different ways and depend on a number of different factors) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

161. *See* Vidhi Doshi, *Why Doctors Still Misunderstand Heart Disease in Women*, ATLANTIC (Oct. 26, 2015), https://www.theatlantic.com/health/archive/2015/10/heart-disease-women/412495/ (discussing the misdiagnosis of heart attacks in women because studies of men's symptoms were used to diagnose women and women's symptoms were not studied until the '90s) (on file with the

when considering heart attack presenting symptoms, is a material factor.[162] The exclusion of gender from the conversation created an unreliable and very dangerous generalization.[163] In this instance, gender as data should have been considered within the conversation.[164] Thus, gender should have been considered. But of course, there are times when gender is an unnecessary consideration and may in fact do nothing but capture societies biases within the transformative process.[165] Turns out, despite what you may believe, women are not worse drivers than men when "worse drivers" are measured in terms of automobile accidents.[166] If designers are allowed or machine learning processes to capture this poor assumption, then designers need to mitigate for that negative impact.[167]

Despite what is a clear need, as demonstrated above, in most instances the existing frameworks do not envision codes of conduct or best practices for the deployment of algorithms that are built upon machine learning, and thus continual data gathering that is then used to create information.[168] This is the area that should be

---

Washington & Lee Journal of Civil Rights & Social Justice).

162. *See id.* (discussing different heart attack symptoms in men and women).

163. *See id.* (highlighting the death of Nancy Larko, who died from a misdiagnosed heart attack).

164. *See id.* (discussing the misdiagnosis of heart attacks in women because studies focused on the symptoms in men).

165. *See* Steven Pinker & Elizabeth S. Spelke, *The Science of Gender and Science Pinker Vs. Spelke: A Debate*, EDGE (May 16, 2005), https://www.edge.org/event/the-science-of-gender-and-science-pinker-vs-spelke-a-debate (highlighting that there is a difference between bias and actual genetic differences in men and women) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

166. *See* John Stossel, *Are Women Worse Drivers Than Men?*, ABC NEWS (May 7, 2017), http://abc news.go.com/2020/story?id=3148281&page=1 (explaining that it is a myth that women are worse drivers than men) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

167. *See* Tom Simonite, *Machines Taught by Photos Learn a Sexist View of Women*, WIRED (Aug. 21, 2017), https://www.wired.com/story/machines-taught-by-photos-learn-a-sexist-view-of-women/ (explaining how machine-learning software amplified the unintentional biases of its programmers) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

168. *See* Xavier Amatriain, *The Best Practices For Training Machine Learning Models*, FORBES (May 10, 2017), https://www.forbes.com/sites/quora/2017/05/10/the-best-practices-for-training-machine-learning-models/#47b0f71d7de8 (describing the best practices for training machine-learning software, including adding more data) (on file with the Washington & Lee Journal of Civil

added into consideration (although, admittedly this area of use would not yet be covered by the DPA).[169] Nevertheless, the considerations of impact can still stand as guiding principles within the wider use of an impact assessment. For example, currently, machine learning translation and the google translate algorithm are all the rage as an ever growing number of languages are capable of translation.[170] Although speech recognition and translation are vastly better than they have ever been,[171] they still have a long way to go, especially when attempting to translate languages that lack standard, predictable rules and patterns.[172] English is incredibly difficult for numerous reasons,[173] as anyone attempting to sort through colloquial phrasing and double use of words can attest to, everything is contextual.

One issue currently being considered amongst data scientists is the manner to measure "better" translations, which commentators argue ask researchers to consider the differences in translations produced from different models and to determine which translation more closely resembles the sentence, including meaning and phrasing.[174] Turns out, even poorly translated

---

Rights & Social Justice).

169. *See generally* Data Protection Act 1998, c.29 (Eng.) (protecting "personal data" defined as any data that can be used to identify an individual and proscribing certain uses of that data).

170. *See* Luba Belokon, *Machine Learning Translation and the Google Translate Algorithm*, DATA SCI. CENT. (Aug. 1, 2017), http://www.datascience central.com/profiles/blogs/machine-learning-translation-and-the-google-translat e-algorithm (providing a great explanation of machine learning translation and the Google Translate algorithm) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

171. *See Language: Finding a Voice*, ECONOMIST (May 1, 2017), http://www. economist.com/technology-quarterly/2017-05-01/language ("Computers have got much better at translation, voice recognition and speech synthesis, . . . they still don't understand the meaning of language.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

172. *See id.* (praising the advancements society made in speech recognition and language translation thus far but recognizing that languages which lack predictable patterns pose more difficult tasks and require further progress be made).

173. *See id.* (identifying English as a language which poses difficulties in certain advancements in language recognitions and translations).

174. *See* Daniil Korbut, *Machine Learning Translation and the Google Translate Algorithm: The basic principles of machine translation engines*, STATS & BOTS (Aug. 1, 2017), https://blog.statsbot.co/machine-learning-translation-96f0 ed8f19e4 (highlighting how Google Translate scans millions of documents in order

sentences can lead to someone understanding the main idea of the sentence, but the phrasing and language patterns add context.[175]

Why might all of this matter, especially in light of impact assessments? Imagine driving down a road and the driving assistant wishes for you to "turn left on Higgins." In order to translate this, an algorithm will use one of several possible translator algorithms and provide instructions that are most likely close to the direction that would have been given in English, but, as everyone who drives with one of these driving assistants knows, sometimes the program simply does not get it right.[176] For an impact assessment to be useful, we have to be able to anticipate the impact on these mistakes upon the user.[177]

Fortunately, the spoken word, in situations of driving, can often be corroborated or verified by visual cues, such as road signs, and available turning lanes, and thus, a poor translation may have less impact in these situations.[178] The absence of visual cues—or other sources of data that can be used to verify, support and provide context can in fact cause errors in the interpretation, communication or translation. For example, consider a situation in which police officers use translation algorithms to communicate with someone speaking another language in a real world setting. The potential absence of visual cues coupled with the intensity of a stressful event the individual is experiencing, may drastically

---

to provide the best translation) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

175. *See Translation is Not About Words. It's About What the Words are About.*, KEVIN HENDZEL (Dec. 14, 2012), http://www.kevinhendzel.com/translation-is-not-about-words-its-about-what-the-words-are-about/ (discussing how context matters more than the words) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

176. *See* Jeffrey Weiss, *Why Your Trusty GPS Sometimes Fails You*, CNN (Apr. 22, 2011), http://www.cnn.com/2011/TRAVEL/04/22/travel.gps.troubles/index.html (explaining how the GPS system occasionally fails to guide drivers) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

177. *See* Dan Bilefsky, *GPS Mix-Up Brings Wrong Turn, and Celebrity, to an American in Iceland*, N.Y. TIMES (Feb. 4, 2016), https://www.nytimes.com/2016/02/05/world/europe/iceland-american-tourist-gps.html (detailing a comedic scenario lived by a driver who made a minor error when inputting an address on his GPS) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

178. *See* Weiss, *supra* note 176 (discussing how drivers should not over-trust their GPS systems when there are alternative sources of direction).

alter the situation for all involved. The police officer may assume the translation is of assistance, while the individual may receive an oddly worded, out of context translation. The translation may not help the situation and may have a drastically negative impact upon the individual who cannot understand the police officer's questions.[179] Thus, a simple measure of the program is not enough as the same exact program will have drastically different impacts based upon the particular use of the instrument.[180]

2. Insist Upon Privacy by Design

While the intent of this Article is not to discuss cyber-security, this Paper would be incomplete without consideration of Privacy by Design within the industry recommendation section. Privacy by Design tends to encompass some of the principles that could be added into the impact-based risk assessment.[181] As such, a brief overview is necessary.

Privacy by Design, also known as "data protection by design and by default," is the design approach that will soon become a legal requirement in the E.U.[182] Within the design approach, the use of technology-based supports protect privacy.[183] For example, data controllers will be obliged to take "appropriate technical and organizational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed,"[184] such as: anonymization techniques, security measures to prevent data misuse, data minimization measures, purpose limitation and data segregation measures, and

---

179. *See* VA. CODE § 18.2-464 (2017) (treating failure to obey lawful order of a conservator of the peace as a misdemeanor).

180. *Compare* Bilefsky, *supra* note 177 (resulting in an interesting story), *with* VA. CODE § 18.2-464 (resulting in punishment).

181. *See* Commission Regulation 2016/679, art. 5, 2016 O.J. (L 119) 1 (examining the various ways by which certain personal data should be protected).

182. ICO Big Data, *supra* note 66, at 72.

183. *See* Commission Regulation 2016/679, 2016 O.J. (L 119) 1 ("In order to be able to demonstrate compliance with this Regulation, the controller should adopt internal policies and implement measures which meet in particular the principles of data protection by design and data protection by default.").

184. *Id.*

restrictions on meta data.[185] In terms of the broader discussion, Privacy by Design is an integral part of the conversation because the design process is an essential mitigation consideration in the risk assessment.[186] Yet, it should be noted that the design process in this instance focuses more on security and less on the transformative process.[187]

3. Eliminate Emotionally Crafted Narratives and Bad Data Science

USA Today authors Jefferson Graham and Laura Schulte described what some commentators consider a disturbing new use of technology that has Wisconsin workers voluntarily embedded with microchips to facilitate payment of purchases from workplace vending machines.[188] Graham and Schulte explained the sentiment of supervisor of the company, Three Square Market President Patrick McMullan, who captured the emotionally charged, illogical concerns:

> The chip is not a tracker nor does it have GPS in it, so the boss can't track your movements, company officials say. Still, to those who worry about Big Brother having more control over our lives, Three Square Market President Patrick McMullan says you should, "take your cell phone and throw it away."[189]

---

185. *See id.* (analyzing the different methods of protecting personal data).

186. *See* GIUSEPPE D' ACQUISTO ET AL., PRIVACY BY DESIGN IN BIG DATA: AN OVERVIEW OF PRIVACY ENHANCING TECHNOLOGIES IN THE ERA OF BIG DATA ANALYTICS 37 (European Union Agency for Network and Information Security ed., 2015), https://www.enisa.europa.eu/publications/big-data-protection/at_download/fullReport (noting that risk assessments require "significant investments in specialized training data sets and fast-evolving machine learning techniques which need to be further developed and considered") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

187. *See* Commission Regulation 2016/679, art. 22, 2016 O.J. (L 119) 1 (implementing methods of protecting privacy).

188. *See* Jefferson Graham & Laura Schulte, *Wisconsin Workers Embedded with Microchips*, USA TODAY (Aug. 1, 2017 2:16 PM), https://www.usatoday.com/story/tech/talkingtech/2017/08/01/wisconsin-employees-got-embedded-chips/529198001/ (describing the use of microchips embedded in workers in Wisconsin) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

189. *Id.*

Of course, the comparison of what cell phones track misses the point of the work place privacy issues and fails to appreciate the fallacy employed to outright reject a valid concern, especially in light of employees' inferior bargaining position.[190] Slippery slopes and similar fallacies are frequently employed to reject outright otherwise valid concerns.[191] Industry insiders must challenge this type of "you have already lost all your ability to complain" arguments. These and other arguments serve to stifle discussion and are used in fear mongering narratives that simply must stop.[192]

Moreover, data scientists must push back on an industry that insists upon no adherence to science and the guiding principle of the discipline, such as verification and informed consent.[193] This author previously wrote about the litany of "projects" undertaken with the full knowledge and expectation that individuals will be impacted yet, that impact is not considered, controlled, or monitored for harmful effects.[194] Facebook—and its various "adjustments"—have caused outrage amongst commentators and social scientists[195] because the impact upon individuals, in at least some cases, is irrefutable.[196] Data scientists and others must be held to the same standards of any social scientist, such as informed consent and human subject approval processes.[197]

---

190. *See id.* ("Three Square Market employees say they were having the chip installed to be part of the larger team, and help develop the technology.").

191. *See generally* CARL COHEN ET AL., INTRODUCTION TO LOGIC 109 (14th ed. 2016) (explaining slippery slope among other logical fallacies).

192. See *id.* at 110 (noting the trouble caused by logical fallacies in discussions).

193. *See* Raymond et al., *supra* note 83, at 12 (regarding the importance of fundamental scientific principles such as informed consent).

194. *See id.* at 2 (focusing on the impact of data on society and the individual).

195. *See* Vindu Goel, *Facebook Tinkers with Users' Emotions in News Feed Experiment, Stirring Outcry*, N.Y. TIMES (June 29, 2014), https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0 (describing Facebook's experimental psychological study examining how emotions can spread on social media) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

196. *See* Raymond et al., *supra* note 83, at 10 (exploring "the manner in which some of the most common pernicious impacts of invisible biases and misperceptions become engrained in algorithmic processes, with potentially devastating consequences").

197. *See id.* at 36 (arguing that data scientists and algorithms ought to be

Finally, the industry must stop using narratives that misrepresent or overgeneralize the issues or the actual technology being deployed.[198] Narratives are generally understood as storytelling "wherein the choice of what data to plot, and how, is tailored to the message the authors want to deliver."[199] While research generally supports the use of narratives to explain scientific information to a non-scientist,[200] the field also explored the ethical aspects of the use of narratives to persuade, instead of inform.[201] The distinction between the two becomes especially relevant in light of research that reveals that "narratives can also perpetuate misinformation and inaccuracies about science or about scientists themselves."[202] In fact, research into science based narratives suggests that because narratives are not subject to the same truth requirements as logical-scientific communications, the message is not easily countered.[203] In fact, accepted narratives are trusted so much that individuals rarely allow evidence to contradict the narrative.[204] Instead, evidence is altered to fit their narratives.[205]

---

held to similar accountability standards as social scientists).

198. *See* Yarden Katz, *Against Storytelling of Scientific Results*, 10 NATURE METHODS 1045, 1045 (2013) (arguing that storytelling is misplaced in science).

199. *See id.* (explaining various aspects of science-based narrative).

200. *See* Michael F. Dahlstrom, *Using Narratives and Storytelling to Communicate Science with Nonexpert Audiences*, 111 PNAS 13614, 13615 (2014), http://www.pnas.org/content/pnas/111/Supplement_4/13614.full.pdf (describing benefits of using narrative in scientific explanation) (on file with the Washington & Lee Journal of Civil Rights & Social Justice); *see also* Baruch Fischhoff & Dietram A. Scheufele, *The Science of Science Communication II*, 111 PNAS 13583, 13583 (2014), http://www.pnas.org/content/pnas/111/Supplement_4/13583.full.pdf (touching on moral implications of narration and storytelling for scientific expression) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

201. *See* Michael F. Dahlstrom & Shirley S. Ho, *Ethical Considerations of Using Narrative to Communicate Science*, 34 SCI. COMM. 592, 617 (2012) (debating the ethical elements in use of narration for scientific explanation).

202. *Id.* at 616.

203. *See* JEROME BRUNER, ACTUAL MINDS, POSSIBLE WORLDS 222 (1986) (exploring the implications of using narration to explain science).

204. *See generally* Katherine McComas & James Shanahan, *Telling Stories About Global Climate Change—Measuring the Impact of Narratives on Issue Cycles*, 26 COMM. RES. 30 (1999) (analyzing the impact of science-based narratives employed by the media on attention to climate change).

205. *See generally id.*

While some may argue that the current technology-based conversations should not be considered within the realm of science, one needs only to examine the arguments presented by the industry to appreciate the need for technology-based narratives is justified for the exact same reason as science-based narratives. Simply put, without the narrative no one will understand. While this justification is likely true, the purpose of the narrative is obliterated in the explanation. Science seeks to use narratives to inform and engage the larger community,[206] while the technology industry often utilizes narratives to further hide the issue.[207] The industry itself must begin to take control of these narratives and insist that narratives serve a supported purpose—that is to inform and engage—and should be rejected if the purpose is to oversimplify or misrepresent the technology being described.

4. Create Auditable Machine Learning Algorithms

While "auditing" an algorithm might sound like an impossible process, (and in the case of machine learning, unstructured data—based algorithms is most likely impossible) auditing can occur if we use a historically defined process within the fields of psychology and law.[208] Amongst other aspects of mental health treatment possibilities, cognitive behavioral therapy uses outwardly visible inappropriate or undesirable behaviors such as indicators of

---

206. *See* Melanie C. Green, *Narratives and Cancer Communication*, 56 J. COMM. S163, S183 (2006) (researching the public comprehension of scientific facts presented in narrative form); *see also* STEPHEN P. NORRIS ET AL., A THEORETICAL FRAMEWORK FOR NARRATIVE EXPLANATION IN SCIENCE 535–63 (2005) (explaining the usefulness of narrative for scientists with an audience of nonscientists); *see also* Lucy Avraamidou & Johnathan Osborne, *The Role of Narrative in Communicating Science*, 31 INT'L J. SCI. EDUC. 1683, 1707 (2009) (discussing the positive and negative impact of science-based narrative); *see generally* NAT'L SCI. BD., *Science and Technology: Public Attitudes and Understanding*, *in* SCIENCE AND ENGINEERING INDICATORS 2014, at 7–30 (2014) (evaluating public understanding of certain scientific concepts as well as the general attitude toward those concepts).

207. *See* NORRIS ET AL., *supra* note 206, at 545 (noting how narrative can be manipulated to alter perception).

208. *See* AARON T. BECK, COGNITIVE THERAPY OF DEPRESSION 8 (1979) (reviewing the historical and philosophical principles on which cognitive therapy relies).

maladaptive thinking in individuals.[209] In general, inappropriate or undesired behaviors are evidence of maladaptive thinking.[210] Assisting individuals in identifying undesired behaviors and examining the maladaptive thinking that leads to the repetition of those behaviors, is often a first approach to treatment in this field.[211] In a similar manner, outcomes—the information produced—can be examined for maladaptive machine logic, bad data, or other contaminants in the outcome.[212] It is the examination of outputs that leads to the need for closer scrutiny.[213]

While this process may sound daunting, it is, in fact, a process already used, although usually under well controlled circumstances.[214] For example, the now infamous Google search debacles.[215] In June of 2016, the Google search engine revealed underlying bias in its search algorithm when a search for "three white teenagers" turned up pictures of happy young white people, but the search of "three black teenagers" produced images of young black people in mug shots.[216] Google also came under fire in July 2015 when its photo app autonomously labeled a pair of black friends as animals.[217] The engineer in charge of the photo app

---

209. *See id.* at 4 (noting the variety of behavioral therapy methods used to treat depression).

210. *See id.* at 77 (describing treatment for maladaptive thinking).

211. *See id.* at 67 (explaining process by which cognitive behavioral therapists lead patients to resolving undesired behaviors by first examining the consequences of their actions).

212. *See* Philip Adler et al., *Auditing Black-Box Models for Indirect Influence*, 54 KNOWLEDGE & INFO. SYS. 1, 2 (Nov. 30, 2016) (addressing the use of output examination for determining solutions to undesired machine logic).

213. *See id.* at 9 (discussing the impact of data output during a black-box audit).

214. *See id.* (examining the use of black-box audits for correcting machine logic).

215. *See* Ben Guarino, *Google Faulted for Racial Bias in Image Search Results for Black Teenagers*, WASH. POST (June 10, 2016), https://www.washingtonpost.com/news/morning-mix/wp/2016/06/10/google-faulted-for-racial-bias-in-image-search-results-for-black-teenagers/?utm_term=.588b1434e31e (describing the algorithmic process that led to misidentification in image searches) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

216. *See id.* (detailing the problems google encountered in handling algorithms for image identification).

217. *See* Jeff Guo, *Why Google's Nightmare AI is Putting Demon Puppies Everywhere*, WASH. POST (July 8, 2015), https://www.washingtonpost.com/blogs/govbeat/wp/2015/07/08/why-googles-nightmare-ai-is-putting-demon-puppies-

believed the underlying program was fine, but the data used during training of the algorithm was "faulty" intimating that Google may not have appropriately trained the AI.[218] Of course, the ability to view undesired outputs such as this is the first step in identifying the underlying issue that may have led to the output.[219] In this manner, algorithms are auditable.

In fact, this author previously argued that some AI based algorithms, such as those in highly impactful areas that deprive individuals of constitutionally protected rights, should be subject to audits.[220] For example, there is wide concern about the use of AI based pattern recognition that is currently occurring in policing, probation, and parole.[221] Systems such as this are highly impactful on personal liberty and should thus be subject to the most stringent of review.[222] As such, it is possible that the "perfect" data set could be developed.[223] This data set could then be fed into the system to ensure the outputs were not out of line with expected outcomes.[224] If the outcomes did not correspond with expected outcomes, one could presume that the AI "learned" a bias it must

---

everywhere/?tid=a_inl (explaining how AI technologies perceive the world by image identification and recreation) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

218. *See id.* (recounting the speculations made by the programmer who discovered the issue).

219. *See id.* (noting that scientists can study the network and work with the outcomes to develop better training of AI).

220. *See* Raymond et al., *supra* note 83, at 36 ("[A]lgorithms used within an area of potential discriminatory impact, regardless of the source of the information being used, must be held to social and statistical science accountability standards. This requires three things: (1) accountability, (2) auditability, and (3) replication.").

221. *See id.* (advocating for strict review of AI affecting individuals and society).

222. *See id.* (promoting high standards for AI review).

223. *See id.* at 9 (noting that if data collection is limited or flawed, the machine's output will also be limited and flawed). Perfect here is an intentionally nebulous term, primarily because the perfect data set would need to be developed based on the specific parameters of the algorithm in development. *Id.* This is, of course, done to combat the "garbage in, garbage out" adage. *Id.* It is, however, the introduction of another potential area of bias. *See* Adler et al., *supra* note 212, at 1 (detailing the process for weeding out unwanted logical patterns).

224. *See* Adler et al., *supra* note 212, at 1 (examining the use of output for training machine logic).

address.[225] Although this process is probably difficult to imagine, it is, in fact, a process used amongst the industry to "train" systems and to build in corrective measures.[226]

Moreover, algorithms may be able to audit other algorithms.[227] While this may seem like HAL as the mechanism to turn off HAL Jr.,[228] it is in fact, a genuine possibility. [229] According to UK, "the 2016 International Conference on Data Mining showed a technique for algorithmic auditing that was evidenced as being effective at identifying discrete factors that influence the decisions made by algorithms."[230] While in the U.S., consultant companies are already being set up that specialize in providing algorithmic auditing services to their clients.[231]

And, while it is the case that it may be difficult to explain the inner workings of AI, processes are being developed every day that bring us closer to this potential reality.[232] For example:

> Other methods are being developed in natural language generation (NLG) to output text that explains why or how a decision was reached. Imagine if your favorite machine learning

---

225. *Id.*

226. *See* IFEOMA AJUNWA ET AL., HIRING BY ALGORITHM: PREDICTING AND PREVENTING DISPARATE IMPACT 17 (2016) ("Corporations and organizations bear a legal duty to correct algorithmic bias; and this duty is not mitigated by a lack of intent to discriminate or even a lack of awareness that an algorithm is producing biased results.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

227. *See id.* at 25 (acknowledging the concern "that mandating self-audits of hiring decisions that have already been delegated to a computer program negates the efficiency gains of automation in business").

228. *See* Raymond et al., *supra* note 83, at 1 (discussing the movie *2001: A Space Odyssey* and HAL (Heuristically programmed Algorithmic Computer) 9000, a computer that is artificial intelligence and the primary computer on the spaceship).

229. *See* Adler et al., *supra* note 212, at 1 (presenting a technique for auditing black-box models).

230. ICO Big Data, *supra* note 66, at 87.

231. *See generally* Cathy O'Neil, *It's the Age of the Algorithm and We Have Arrived Unprepared*, ORCCA, http://www.oneilrisk.com/ (providing example of one such firm) (last visited Apr. 16, 2018) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

232. *See generally* Nicholas Diakopoulos, *Accountability in Algorithmic Decision Making*, 59 COMM. ACM 56 (2016) (discussing the need for greater accountability for the use of algorithms in government and industry).

> library, say scikit-learn, could explain in a sentence why a particular input case was classified the way it was.[233]

While explanations for the more advanced AI based systems may be years away, today we should be entitled to expect the ability to audit algorithms when the algorithm either uses sensitive/critical data or when the impact of the outcomes is significant.[234]

D. Recommendations for Policy Makers

Policy makers also face a daunting task when considering issues surrounding the transformative process.[235] The first issue that must be considered is the overall framework that should be deployed within the area.[236] And of course, that requires a true conversation to occur to develop the areas upon which the framework will be built.[237] This section seeks to begin that conversation, by returning to the legal frameworks and industry recommendations discussed above and considering these within the creation of the framework.[238]

1. Reject the Privacy Narrative

Reject the narrative surrounding privacy as the interest that we seek to protect. As discussed above, privacy law originally envisioned governmental intrusion.[239] Today, the privacy conceptualization has expanded to include information that we

---

233. *Id.* at 61.

234. *See* Adler et al., *supra* note 212, at 10 ("[I]n privacy-preserving data mining, we do not trust the user of the results of classification with sensitive information from the input. In out setting, the 'sensitive' information must be hidden form the classifier itself.").

235. *See id.* (explaining the agreements and disagreements different approaches can elicit).

236. *See id.* (articulating the mathematical and computational frameworks for evaluating black-box models).

237. *See id.* (dissecting the agreements and disagreements different approaches can elicit).

238. *See id.* at 1 (discussing the legal issues that can arise from relying on algorithms).

239. *See* Solove, *supra* note 13, §§ 1–4, 1–7 (examining the historical role privacy law has played regarding governmental intrusion).

seek to shield from prying eyes.[240] In general, this information is considered sensitive—for a variety of reasons—and as such should be allowed to remain private.[241] However, if information is private, should one not be expected to protect it as important information that he/she values keeping private and should he/she not be allowed to, therefore, shield it from everyone, but for incredibly limited exceptions?

Individuals seeking to protect information as private find the task difficult to achieve.[242] For example, in January 2017 DuckDuckGo conducted *A Study on Private Browsing: Consumer Usage, Knowledge, and Thoughts*.[243] Almost unsurprisingly, the results clearly show a true confusion among users about what exactly privacy entails in an online environment.[244] According to the study, "46% of Americans have used Private Browsing" while those using it reported the "number one reason people use Private Browsing is 'Embarrassing Searches.'"[245] However, "76% of Americans who use Private Browsing cannot accurately identify the privacy benefit it provides."[246] In fact, "41.0 ±2.5% believe that Private Browsing Prevents websites from tracking me" and "39.1 ±2.6% believe that Private Browsing Prevents ads from tracking me."[247] And, do not assume this is a generational gap, as in general the misconceptions about Private Browsing are consistent.[248] Of course, when individuals are told about the true protections

240. *See* Larose, *supra* note 16 (explaining New Jersey's Personal Information Privacy and Protection Act and the legislative attempts to shield consumer information by using privacy laws).

241. *See* Dwoskin & Timberg, *supra* note 5 (describing some categories of sensitive information, such as debit and credit purchases of consumers).

242. *See generally* DUCK DUCK GO, A STUDY ON PRIVATE BROWSING: CONSUMER USAGE, KNOWLEDGE, AND THOUGHTS 4 (2017), https://duckduckgo.com/download/Private_Browsing.pdf (writing about how people view private browsing based on findings from a survey conducted with a random sampling of 5,710 Americans) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

243. *See id.* (discussing the confusion many consumers have regarding privacy).

244. *See id.* at 17 ("Of the 75.8 ±2.2% of people who have misconceptions about private browsing protections, 65.9 ±2.4% feel 'Surprised', 'Misled', 'Confused' or 'Vulnerable' upon learning about its real protections.").

245. *See id.* at 9–10 (reasoning why consumers use private browsing).

246. *Id.* at 13.

247. *Id.*

248. *See id.* at 14 (explaining the top three misconceptions).

provided from private browsing "65.9 ±2.4% feel 'Surprised', 'Misled', 'Confused' or 'Vulnerable.'"[249] Private and privacy do not mean what people think they mean—and thus, should be abandoned as safe and secure—the private information narrative is misleading and misunderstood.[250]

Individuals often have little choice in sharing information that everyone considers private, as this type of information is often used as a means to authenticate identity in the online world.[251] As such, much of my private information is shared, by myself, multiple times a day, with little choice in the matter, despite a true desire to protect it as private. Consider one's birthday, place of birth, and mother's maiden name—all of this information is used as security questions to establish identity to reconnect to certain accounts. Under the concept of privacy, how can one argue for this information to remain private when it is so readily share it over and over online every day? And that, of course, ignores the individuals who share this information and more on easily searched websites such as Facebook and Instagram.

Finally, consider one of the most heralded authentication devices: the social security number.[252] While most people assume a social security number is random—or otherwise sequential—based on some arbitrary system, in fact social security numbers are based upon simple bits of data.[253] Data that is often widely available and used often by individuals—even published on Facebook.[254] That means your social security number can be

---

249. *See id.* at 23 (discussing the large amount of people who incorrectly identify the protection private browsing provides).

250. *See id.* at 17 (exploring how private browsing actually does contrary to what consumers think it does).

251. *See* Brian Krebs, *Researchers: Social Security Numbers can be Guessed*, WASH. POST (July 6, 2009, 6:05 PM), http://www.washingtonpost.com/wp-dyn/content/article/2009/07/06/AR2009070602955.html (discussing how social security numbers were never meant to be used for identification purposes, but the private sector has largely ignored this) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

252. *See id.* (stating that if Social Security Numbers can be predicted from public data, they offer little protection for the consumer).

253. *See id.* ("Social Security number's first three digits . . . is issued according to the Zip code of the mailing address provided in the application form. The fourth and fifth digits . . . often remain constant over several years for a given region. The last four digits are assigned sequentially.").

254. *See id.* ("Records of an individual's state and date of birth can be obtained

guessed, with shocking accuracy, from nothing more than publicly available information, information that individuals regularly and publicly display.[255] According to Alessandro Acquisti, assistant professor of information technology and public policy at Carnegie Mellon University, "our work shows that Social Security numbers are compromised as authentication devices, because they are predictable from public data."[256] In fact, researchers found that it is possible to guess many—if not all—of the nine digits in an individual's Social Security number using publicly available information, as many could be guessed by simply knowing a person's birth data.[257] This is but one example of publicly available data being used to discover private data. And, it is a great example of the looming debate; how can one ever argue for the need for privacy, when the individual has publicly shared the information? All that has occurred is that an entity has gathered public information and compiled it to create an incredibly accurate profile, one that can easily lead to what many thinks of as private information.[258] Privacy must be abandoned as the guiding principle, as we publicly share too much information to realistically ever argue that information is private.[259]

Instead, U.S. citizens should use terminology such as sensitive and secure information to alleviate the confusion and reduce the potential to fall into the trap of using existing privacy laws as guidance.[260] Surely a social security number, if it is to continue to be used as an authentication device, should be labeled as sensitive

---

from a variety of sources, including voter registration lists and commercial databases. What's more, many people now self-publish this information as part of their personal profiles on blogs and social networking sites.").

255. *See id.* ("Researchers have found that it is possible to guess many—if not all—of the nine digits in an individual's Social Security number using publicly available information").

256. *Id.*

257. *See id.* ("Many numbers could be guessed at by simply knowing a person's birth data, the researchers from Carnegie Mellon University said.").

258. *See id.* ("'We can't pretend anymore that SSNs can be kept secret,' said Peter Swire, a law professor at Ohio State University and chief counselor for privacy during the Clinton administration.").

259. *See id.* (discussing the need for new approaches to identification due to the saturation of personal information on social media).

260. *See* Dwoskin & Timberg, *supra* note 5 (showing that these terms are already being used to a certain extent).

and thus has a high need of securing the information, regardless of how it was obtained.

Moreover, the terminology allows individuals to focus on the important aspect of the conversation; this information is not private, it is important to protect.[261] Thus, individuals have a responsibility to protect the information and gain the right to not disclose the information unless it is information that is important to the entity requesting the information.[262]

2. Reject Property as the Guiding Law

Issues of ownership of data—in both a personal and business context—have become an important consideration in terms of governance.[263] Despite a well-established history of property law, there is surprising uncertainty in the law regarding the ownership of information contained in records, on forms, and in other data repositories.[264] In general, the law regards records holding data as property owned by their creators, as was examined briefly above. But the real question is whether the data is owned at all.[265]

There are strong arguments that information cannot be owned, a problem that can be expected to intensify as paper records give way to a freely moving information on an electronic highway.[266] As Professor Mark Hall has observed when

---

261. *See id.* (illustrating this fact is the prevalence of consumer data online, such as credit and debit purchase information).

262. *See id.* (highlighting this problem is Google's ability to collect "a trove of highly sensitive information . . . without revealing how they got the information or giving consumers meaningful ways to opt out").

263. *See* Mark A. Hall, *Property, Privacy, and the Pursuit of Integrated Electronic Medical Records*, WAKE FOREST UNIV. L. STUD., Paper No. 1334963, at 2 (2009), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1334963 ("All parties are looking to the law to define the ownership, control, and commercialization potential of medical information. How these issues are resolved ultimately will determine which information network models are economically viable and what form they will take.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

264. *See id.* at 12 (explaining the various viewpoints regarding the ownership of medical records).

265. *See id.* (examining the legal uncertainty regarding who owns the information contained in medical records).

266. *See id.* at 11 ("The law's uncertainty over ownership and control of medical information is widely regarded as a major barrier to effective networking

considering medical records "[o]wnership was never much in doubt in an age of paper-based records"[267] because the paper record containing the information was owned by its creator.[268] The electronic information age has ushered in an era in which the content of information can be "digitized and freed from any particular storage medium."[269] In the paper-based world one could think of ownership and control as often found in tandem and certainly the relinquishing of control did not equate to the relinquishing of ownership (think of your car).[270] In the digital world, this is no longer the case as it is difficult—or nearly impossible—to identify the owner, and being legally in control of information is often a status to be avoided as it incurs heightened responsibilities.[271]

Unfortunately, the absence of clearly defined property rights within the area likely has significant consequences to governance and property-based rights.[272] As argued many times before, uncertain legal positions allow parties to attempt to simply to stake a claim.[273] In the paper-based world, claiming space forces a negotiation toward a contractual settlement that determines respective rights.[274]

Of course, the outcomes are only binding on the immediate parties, and the process of negotiation is an expensive barrier.[275]

---

of EMRs and policy analysts consider the legal status of medical information to be a critical question at or near the top of issues needing resolution.").

267. *Id.* at 1.

268. *See id.* (discussing the confusion that digitizing records has created).

269. *Id.*

270. *See id.* at 10 (describing the trustee model that places the patient-controlled record into the control of an "infomediary" or "records bank").

271. *See id.* at 14 (writing about how divided control over records make it difficult to control overall even though the collective benefits may be worth the costs).

272. *See id.* at 17 (discussing the challenges of digitizing medical records and the tragedy of the anti-commons).

273. *See id.* at 9 ("If legal positions are uncertain, parties can still attempt simply to stake a claim, forcing a negotiation toward a contractual settlement that determines respective rights, but such contracts bind only the immediate parties, and the process of negotiation is an expensive barrier.").

274. *See id.* ("Ownership was never much in doubt in an age of paper-based records, but now that information content can be easily digitized and freed from any particular storage medium, confusion reigns.").

275. *See id.* (discussing the parties bound by negotiations).

As such, it is argued that property rights must be clearly established so that the respective parties know their legal default positions.[276] In the digital world, this position is fraught with difficulty as the negotiations never truly materialize.[277] Instead, large scale technology creators control the market and limit entry to those that agree to their conceptualization of the property rights control.[278] Thus, even if property rights were established, the existence of consent-based membership would allow the powerful to reallocate the ownership rights as a term of entry into the system.[279]

One also must consider the problems created—or the clever work arounds—that currently exist in overcoming both the ownership and control dilemmas.[280] In terms of ownership, if an individual is allowed to own his or her digital profile and the information it contains, at what point is the record merely data and incapable of ownership? Consider a birthdate in a database. When it is one person's birthdate such that it is attached to identifying information, it has corresponding ownership rights, but when it is placed within a large database of anonymized data, it is no longer capable of ownership.[281] It is simply a number.

Maybe more importantly, owned data allows individuals to place limits on its use, which may fail to fill society's need to use data in a beneficial manner.[282] Consider smart cities, natural disasters, and medical emergencies. Should individuals be allowed to own data that prevents social benefit? Should they be allowed to

---

276. *See* Richard S. Murphy, *Property Rights in Personal Information: An Economic Defense of Privacy*, 84 GEO. L. J. 2381, 2387, 2403, 2395 (1996) (applying Coase's theorem to property rights in personal information).

277. *See* Hall, *supra* note 263, at 9–10 ("[U]ncertain legal rights over valuable property can spark a land grab that hoards rather than develops productive assets.").

278. *See id.* at 10 ("Once one party stakes its ownership claim, then so must all the other competing parties, for fear of being trumped.").

279. *See id.* at 12 (explaining the expensive barrier that negotiating creates).

280. *See id.* at 29 (describing the network externalities created by fragmentation which prevent an individual actor from achieving the majority of the social benefits of I-EMRs).

281. *See id.* (highlighting the need for stronger legal protections for anonymized information as the individual's claim to that information ceases to exist).

282. *See id.* (discussing how medical research and public health monitoring would increase public goods).

profit from it? Should the entity that gathers it be allowed to profit from it if the individual cannot? Of course, that is not to mention the obvious issue with data needing to be located and subjected to a particular property regime, a topic for another day.

It is a lengthy discussion that prompted many scholars and policy makers to abandon the ownership discussion as it is fraught with difficulties and, while intriguing as academic debate, it stands in the way of the creation of a governance regime.[283] As such, many now consider control to be the guiding paradigm that ignores the ownership issue and instead places responsibility upon the entity that controls the data.[284] Time will tell if this basic shift in the discussion allows the discussions to advance. The author fears by ignoring the true issue, we as a society will remain beholden to those that craft overly broad and wide sweeping permissions through the use of ubiquitous online contracts.[285]

3. Reject Solutions Designed in a Paper Based World

This Article does not suggest that the paper-based world should not be a large consideration in the creation of policy. In fact, the impact of data loss created in the paper world remains a constant concern.[286] For example, in August of 2017,health insurer Aetna allegedly "revealed 12,000 patients' HIV statuses by sending letters with a giant envelope 'window' that exposed confidential information."[287] Data loss occurs in the paper-based world as well—and often—and thus, it must be considered within the scope of any governance frameworks.[288] As noted by the Executive

283. *See id.* at 12 (discussing the problems of applying the principles of property and ownership to patient medical information in the U.S.).

284. *See id.* (explaining the efforts to entrench patient control over medical records wherever they are in the medical pipeline).

285. *See id.* (examining the formation of a contractual network to manage the sharing of medical records).

286. *See* Mia De Graaf, *Aetna Revealed 12,000 Patients' HIV Statuses by Sending Letters with Giant Envelope "Window" that Exposed Confidential Information*, DAILY MAIL (Aug. 24, 2017), http://www.dailymail.co.uk/health/article-4820680/Aetna-sued-revealing-scores-patients-HIV-statuses.html#ixzz4qhvTpHOs (describing accidental information disclosure using paper envelopes) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

287. *Id.*

288. *See id.* (discussing data loss from paper envelopes).

Director of the AIDS Law Project of Pennsylvania Ronda B. Goldfein, "Aetna letters casual disclosure of a person's HIV status or use of HIV medication is far more than a technical violation of the law . 'It creates a tangible risk of violence, discrimination and other trauma.'"[289] As Executive Director Goldfein notes, it is the impact of such disclosure upon the individual that has the potential to cause the greatest damage.[290] The paper-based world is an important consideration, but deploying informed consent based in a copious amount of information being delivered with the use of click wrap agreements to authenticate consent are historical paper-based solutions that no longer stand the test of time in the digital world.[291]

This author has written numerous times about the ability of technology to greatly reduce or completely eliminate the problems associated with the "must, rush, and trust" based consent world.[292] Simply put, no one—literally no one—pays attention to the terms presented to us.[293] In fact, very few people even skim the information presented.[294] And, the vast majority of individuals have become apathetic in the situation as they believe reading terms are a waste of time as they have no bargaining power and no real choice but to click and move on.[295] Moreover, presenting drastically one-sided terms is not even a barrier to consent as

---

289. *Id.*

290. *See id.* ("Sally Friedman, Legal Director of the Legal Action Center in New York City . . . said: 'Aetna's privacy violation devastated people whose neighbors and family learned their intimate health information. They also were shocked that their health insurer would utterly disregard their privacy rights.'").

291. *See* Raymond, *supra* note 56, at 129–71 (discussing the problems with online contracts determining consumers' consent).

292. *See id.* (explaining the consumer's need to quickly accept terms when purchasing online and the importance of trust); *see generally* Raymond, *supra* note 56, at 1, 2 (exploring the history and importance of consent in the digital age).

293. *See id.*, at 170 ("These behavioral understandings demonstrate that consumers simply do not read contract terms, even arbitration clauses, regardless of presentation style or emphasis.").

294. *See id.* ("Online consumers are a 'must have now, in a rush' community, do not read long contract clauses, prefer not to scroll though text passively, do not use online resources to understand terms, and generally feel powerless in the face of contract adhesion.").

295. *See id.* (discussing the power imbalance inherent with online contracts and suggesting new protections).

individuals believe that in many circumstances being a member of the community demands participation in the digitally connected world, such as Facebook. Individuals feel they have no real choice but to agree to the terms created and presented to them.

Of course, inroads are being made to improve the bargaining power of individuals. Smart contracts, information presented to explain terms, and mechanisms designed to encourage (or even force) individuals to read and assent to particular highly important terms are all being deployed. Yet, little incentive exists for many digital providers to deploy the technology as the ubiquitous presentation of click-based terms has created a system where individuals have already signed away their rights and are now completely apathetic to the process.

4. Embrace Being Uncomfortable

It should be noted, the deployment of advanced AI—even if done perfectly—can lead to uncomfortable patterns emerging. Many of these discoveries will be impossible to ignore once revealed. For example, Stanford University social psychologist and 2014 MacArthur fellow Jennifer Lynn Eberhardt is using machine intelligence to develop speech recognition and transcript analysis software for policing.[296] Her team is examining transcripts from traffic stops to recognize patterns of racial disparity, and the team's results are not necessarily unexpected but are nonetheless quite troubling.[297] Her team discovered that "officers were more likely to ask questions of black drivers, less likely to state a reason for pulling them over, and less likely to use respectful language."[298] This outcome should make readers uncomfortable, and we will all have to become accustomed to that uncomfortableness because advanced AI often reveals patterns our human consciousness seeks to ignore.

---

296. Laureen Murrow, *Cop Talk, the Sound of Bias*, WIRED (Sept. 6, 2017), http://www.wiredco.uk/article/police-racism-speech-linguistics-social-psychology (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

297. *Id.*

298. *Id.*

Of course, this is just the tip of the iceberg when it comes to these types of disparity.[299] Consider the research conducted by Cornell researchers Jeffrey J. Rachlinski and Andrew J. Wistrich examining judicial decision making. Highlighting the findings, they summarize:

> A wide range of experimental and field studies reveal that several extra-legal factors influence judicial decision making. Demographic characteristics of judges and litigants affect judges' decisions. Judges also rely heavily on intuitive reasoning in deciding cases, making them vulnerable to the use of mental shortcuts that can lead to mistakes. Furthermore, judges sometimes rely on facts outside the record and rule more favorably towards litigants who are more sympathetic or with whom they share demographic characteristics. On the whole, judges are excellent decision makers, and sometimes resist common errors of judgment that influence ordinary adults. The weight of the evidence, however, suggests that judges are vulnerable to systematic deviations from the ideal of judicial impartiality.[300]

While some of these finding are summaries and complications of prior research, the outcomes suggest the prevalence of the issue is more widespread than ever believed. Also, analytics applied to large swaths of large scale data has led to the re-examination of prior smaller scale research.

Consider one of the newest uses of data and algorithms: decisions relating to the granting of bail in the justice system. In the United States, the bail system, enshrined in the Bill of Rights, is meant to ensure that all defendants have an opportunity to remain free until convicted of a crime.[301] Concerns surrounding the

---

299. *See* Raymond, *supra* note 146, at 3A–5A (discussing the debate about the need for data regulation that focuses on the impact of the use of the data); *see also* Geoffrey Mohan, *Stanford's Jennifer Eberhardt Wins MacArthur "Genius" Grant*, L.A. TIMES (Sept. 18, 2014), http://www.latimes.com/science/la-sci-jennifer-eberhardt-genius-20140917-story.html (explaining Eberhardt's "efforts to examine how subtle, ingrained racial biases influence not just how we view people, but the objects of our daily world") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

300. Jeffrey J. Rachlinski & Andrew J. Wistrich, *Judging the Judiciary by the Numbers: Empirical Research on Judges*, ANN. REV. L. SOC. SCI. 1 (June 2, 2017), https://ssrn.com/abstract =2979342 (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

301. *See generally How Courts Work*, A.B.A., https://www.americanbar.org/groups/public_education/resources/law_related_education_network/how_courts_

defendant's willingness to return to court created a system in which defendant's pay bail, which is cash that is retained by the court should the defendant not attend trial. While this system may read as a reasonable one, it is widely criticized as favoring those with money,[302] as defendants without cash are unable to pay bail and thus languish, sometimes for extended periods of time in jail awaiting trial. In response to the bail dilemma, many States[303] have looked to data and advanced algorithms to gauge risk and hence, make bail recommendations.[304]

As Jon Schuppe of NBC News notes:

> Modern algorithms promise to objectively weigh whether someone will behave a certain way. But they fall short in one key aspect: they can never reflect the mystery and uncertainty of everyday life.[305]

Consider New Jersey's Public Safety Assessment algorithm, which uses a variety of weighted factors to produce a number that purportedly reflect the individuals risk of skipping court and committing a new crime.[306] The risk numbers appear in real time on the judge's computer screen and is then used to make a bail determination. Should the attorneys disagree with the assessment, they challenge the outcome and a further hearing is set in which they "must persuade a judge to override the algorithm's recommendation."[307] This process reflects a widely-

---

work/bail.html (last visited Apr. 16, 2018) (explaining how bail does not serve as a punishment and is returned after the defendant's trial) (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

302. *See* Jon Schuppe, *Post Bail: America's Justice System Runs on the Exchange of Money for Freedom. Some Say That's Unfair. But Can Data Fix It?*, NBC NEWS (Aug. 22, 2017), https://www.nbcnews.com/specials/bail-reform ("At the same time, the wealthy can buy their way out of pretrial detention on just about any offense, including murder.") (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

303. *See id.* (showing that three states to date have widely adopted such system, while another eleven have some use).

304. *See id.* (describing this approach as "what the new vision of American justice looks like").

305. *Id.*

306. *See id.* (noting New Jersey's use of an algorithm to predict whether a defendant will to return to court for trial or be arrested again).

307. *Id.*

held belief—as noted by NBC New Reporter Schuppe: "Algorithms need humans—flaws and all—to oversee them."[308]

Notably, the number of people being held before trial in New Jersey has dropped by nearly a third compared to last year.[309] Thus, commentators argue the system is working.[310] Unfortunately, the system also likely reflects a level of uncomfortableness. Simply put, the system allows for the overriding of the algorithm, which is the introduction of the prior bias that the algorithm was designed to eliminate, and it just reframes the reason given for the hearing.[311]

As can be seen, the patterns revealed or verified lead to a level of uncomfortableness in many societal institutions, even those with the power to inflict dire consequences. The reaction society has to these pattern revelations will define us as a community for decades to come.

Finally, discomfort must be considered in the context of the narratives that are crafted by the tech industry. There are many reasons to craft narratives that are simplistic, accessible and engaging, especially in the face of what can be a scary ununderstood world such as black box technology. Citizens are becoming worried about the Internet of Things (IoT) worried about their bank details being stolen, worried about weak passcodes, worried about where some of their "private" information is ending up. Narratives of the "flow" of information and creation of penalties and regulation make society feel safe, allow individuals to trust these systems, thereby encouraging continued use.

Yet, many of these narratives are based in less than accurate understand on the systems, and are thus, nothing more than rubbish. While narratives to assist individuals in feeling safe online is likely socially beneficial, we must not allow policy to be based on rubbish narratives. Policy makers have to become comfortable with being uncomfortable and then seek to become more comfortable. More comfortable with the way systems work,

---

308. *Id.*

309. *See id.* (discussing the impact of New Jersey's data systems).

310. *See id.* (describing a New Jersey official's statement that although individual decisions can turn out to be wrong, it is hard to argue with the results).

311. *See id.* ("The ultimate decision lies with the judge, but only after hearing arguments from defense lawyers and prosecutors, who must ask that someone be held before trial.").

more comfortable with technology, more comfortable with the way that individuals interact with technology, more comfortable with the fact that many do not understand technology at all. If policy makers do not embrace being uncomfortable in all of its impacts, the policy they create will continue to be out of line with the realities of the cyber world. Failure to draft and implement multi-layered policy that accurately reflects the cyber world will cause lasting negative impacts well beyond the current generation.

5. Embrace the Spectrum of Risk Analysis

As was discussed above, the use of risk analysis, on both the large scale and local level, will allow the governance regime to focus on the true concerns and to attempt to mitigate those concerns in the most efficient manner. Adoption of the Risk Assessment creates a guiding set of standards to be considered in each setting and each level of regulation.

6. Embrace Outcome Based-Impact Assessment

Reject the overused term of "transparency" and insist upon policy that focuses upon the outcomes of the process, including audits or outcomes and impact assessment. Consider the recent Consumer Financial Protection Bureau action in which two American Express banking subsidiaries were found to have discriminated against "consumers in Puerto Rico, the U.S. Virgin Islands, and other U.S. territories by providing them with credit and charge card terms that were inferior to those available in the fifty states."[312] In fact, some discrimination was predicated on the fact that some customers had Spanish-language preferences.[313] In

312. *CFPB and American Express Reach Resolution to Address Discriminatory Card Terms in Puerto Rico and U.S. Territories*, CONSUMER FIN. PROTECTION BUREAU (Aug. 23, 2017), https://www.consumerfinance.gov/about-us/newsroom/cfpb-and-american-express-reach-resolution-address-discriminatory-card-terms-puerto-rico-and-us-territories/ (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

313. *See id.* (noting that American Express discriminated against such customers in Puerto Rico by providing them with less favorable financial products and services).

this instance, the company discovered the unintentional discrimination when conducting an internal review of the various cards the company offered. The company "determined that certain cards issued in those markets through its international business did not uniformly have the same terms, conditions and features as the cards the company offered in the Continental United States."[314] The company discovered the error by noticing unexpected discrepancies amongst the lending environments (the outcomes) and sought to compare discreet, identifiable, data points, amongst the various groups.[315] The outcomes drew attention to underlying data revealing the source of the policy based discriminatory policies.

Outcomes, when monitored and audited, can reveal a great deal about underlying issues, regardless of the transparency of the process that occurs. In this instance, as a discriminatory practice, the outcomes had a significant impact as well as "more than 200,000 consumers were harmed" to the tune of "approximately $95 million."[316]

Impact assessments, that is the impact upon various stakeholders if the information is revealed (regardless of cause or source), is soon to be the only real means of considering harm as monetary loss will increasingly be difficult to demonstrate. The application of such considerations upon the use of data can be almost obvious when considered in light of some increasingly common activities of industry and states. For example, it is with increasing regularity that geolocation data is being used for a variety of purposes in our daily lives.[317] Geolocation data is generally defined as would be "non-content information" which is often "generated or derived from, in whole or in part," the operation

---

314. *American Express Resolves Regulatory Review of Products in U.S. Territories That Varied from Its Continental U.S. Offerings*, BUSINESSWIRE (Aug. 23, 2017), https://www.businesswire.com/news/home/20170823005813/en/American-Express-Resolves-Regulatory-Review-Products-U.S. (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

315. *See generally id.*

316. *Id.*

317. *See* CAMILLO GENTILE ET AL., GEOLOCATION TECHNIQUES: PRINCIPLES AND APPLICATIONS, at v (2012) ("While geolocation is a relatively new topic, in the multidisciplinary area of electrical, mechanical, and industrial engineering, it has grown very rapidly in the last decade due to the tremendous impact it is having on our everyday lives.").

of a mobile device.[318] The impact of the use of this data is the potential to "'infer' precise location of the (mobile) device."[319] In response to growing use of geolocation data that can infer location, Illinois General Assembly passed the Geolocation Privacy Protection Act seeking to limit the collection, use, retention, or disclosure of precise geolocation data from a mobile device without a person's prior express and written consent.[320] The argument, returning to various recommendations made within the paper, is that because the impact upon the data generator (or provider) can be significant, the legislative process needs to create regulation to prevent or reduce the impact. Although the Illinois law, the first of the kind in the nation, will likely result in new issues arising in the area of privacy and demand industry response, the process of consideration and response is most likely appropriate.

Conclusion

Data management and information governance are growing areas of ethical discussion. While much has been done in terms of data management, especially in the area of security, little has been done in terms of the regulation of transformative processes. This paper attempts to create a framework that could be used to identify global areas of concern, yet would be nimble enough to allow industry, and even individual businesses to consider policy creation.

Key to the creation of policy is the rejection of common—often overstretched—legal concepts such as 'privacy' and property-based rights. Moreover, the use of consent and transparency—while appropriate in some situations—must be carefully considered in light of the proliferation of information available to individuals. Instead, policy makers and designers should begin to incorporate

---

318. Edward R. McNicholas et al., *Illinois Becomes the First State to Pass a Geolocation Privacy Protection Bill*, DATA MATTERS: SIDLEY CYBERSECURITY, PRIVACY, DATA PROTECTION, INTERNET L. & POL'Y BLOG—SIDLEY AUSTIN LLP (July 5, 2017), http://datamatters.sidley.com/illinois-becomes-first-state-pass-geolocation-privacy-protection-bill/ (on file with the Washington & Lee Journal of Civil Rights & Social Justice).

319. *Id.*

320. *See id.* (noting the serious implications of this law on the proliferations of Internet of Things devices).

a risk analysis that allows designers to consider risk of data loss or misuse and analyzes the issues through the potential harm it causes individuals. This framework allows individuals to be protected in the event of loss of harmful information, places responsibility on the individual to protect sensitive information and asks that other information be deemed as requiring less robust protections.

Ultimately, more conversations about data, data loss, and data misuse must begin to take place, among all stakeholders, and these discussions must lead to the creation and use of a framework to assist designers in identifying information worthy of robust protection.