INFORMATION AND THE REGULATORY LANDSCAPE: A GROWING NEED TO RECONSIDER EXISTING LEGAL FRAMEWORKS

Anjanette H. Raymond[*]

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 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.

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. 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

---

[*] Associate Professor, Department of Business Law and Ethics, Indiana University, Kelley School of Business; Adjunct Associate Professor of Law, Maurer School of Law, Indiana University; Visiting Fellow in International Commercial Law, Centre for Commercial Law Studies, Queen Mary, University of London. This paper arose in the conversations that occurred at the Washington and Lee University School of Law Academic Roundtable entitled *Big Data: Understanding Algorithmic Power* (2017) and will be published in the *Washington & Lee Journal of Civil Rights and Social Justice*. Thank you to Margaret Hu and all roundtable participants for your valuable discussion. All opinions are those of the author.

[1]

[2] Slaughter and May White Paper at 7. The difference between the terms data and information are important, although often the terms are used almost interchangeably. Data is "Data is raw, unorganized facts that need to be processed. Data can be something simple and seemingly random and useless until it is organized." In contrast, "when data is processed, organized, structured or presented in a given context so as to make it useful, it is called information."

majority of U.S. consumers -- without revealing how they got the information or giving consumers meaningful ways to opt out."[3] 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." Of course, this is not the first- nor the last- criticism of big businesses use of black box algorithms. This paper seeks to further debates previously asserted by the author[4] by examining the issue in light of the recent surge in attention algorithms are drawing, primarily because their use has grown exponentially.

Table of Contents

[3] Elizabeth Dwoskin and Craig Timberg, *Google's New Program To Track Shoppers Sparks A Federal Privacy Complaint*, The 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-privacy-complaint/?utm_term=.f6932f73766b Complaint available at https://epic.org/privacy/ftc/google/EPIC-FTC-Google-Purchase-Tracking-Complaint.pdf

[4] Dwoskin, *Google's New Program To Track Shoppers*

I. THE GROWTH OF ADVANCED ARTIFICIAL INTELLIGENCE SYSTEMS

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.[5] And, of course is widely influencing our daily lives, as demonstrated by the widespread use by both Google and Facebook, to name but two.

In many ways, people have become accustomed to the first level of the "technological revolution in which organizations automated repetitive, high volume, sometimes complex but typically rule-based ("if X then Y") processes."[6] 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."[7] 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 XX, most legal frameworks are based in the 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. 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.

II. THE EXISTING LANDSCAPE

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

---

5

6 Slaughter… at 8.

7 Slaughter and May White Paper at 7

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 then expect to be private; however:

> One 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.[8]

Historically, this sentiment was captured again and again in case law, founded under constitutional protections from governmental intrusion,[9] to some of the original torts[10] created to protect individuals from public gaze. Today, in the United States, there is a litany of protections afforded to specific types of information – often times justified on the sensitivity of the information, or the relationship that exists between the individual and the entity.[11] And some States have begun to more robustly protect individuals' information gathered for particular uses, such as retailer based information.[12]

[8] Pavesich v. New Eng. Life Ins. Co., 50 S.E. 68 (Ga. 1905).

[9] Solove, Daniel J., A Brief History of Information Privacy Law. Proskauer On Privacy, PLI, 2016; GWU Law School Public Law Research Paper No. 215. Available at SSRN: https://ssrn.com/abstract=914271

[10] Such as the tort of public disclosure of private facts, and the tort of false light, and the tort of breach of confidentiality developed to protect disclosures of information in violation of trust within certain relationships. See id.

[11] Such as the Family Educational Rights and Privacy Act of 1974, (Pub. L. No. 93-380, 88 Stat. 484 (codified at 20 U.S.C. § 1232g).), the Privacy Protection Act of 1980 (Pub. L. No. 96-440, 94 Stat. 1879, codified at 42 U.S.C. § 2000aa.), the Cable Communications Policy Act of 1984 (42 U.S.C. § 2000aa(a)), the Employee Polygraph Protection Act of 1988, (Pub. L. No. 100-618, codified at 29 U.S.C. §§ 2001–09.), Electronic Communications Privacy Act of 1986 (18 U.S.C. § 2710(b).), the Telephone Consumer Protection Act of 1991 (Pub. L. No. 102-243, 105 Stat. 2394 (codified at 47 U.S.C. § 227).Driver's Privacy Protection Act of 1994 (47 U.S.C. § 227(c)(5)) to name but a few of the many sectorial approaches to privacy law in the U.S.

[12] State laws governing the collection and use of personal information continue to proliferate. One of the latest State based Act- 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. 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. New Jersey's Personal Information Privacy and Protection Act. And recent activity in California and Illinois has followed New Jersey in enhancing privacy based of individuals.

In contrast to the U.S. based sectoral approach, the vast majority of countries have adopted comprehensive data protection laws including nearly every country in Europe and many in Latin America and the Caribbean, Asia, and Africa.[13] It is however, arguably the case, that regardless of the source of protection, an overwhelming number of countries agree on a list of basic legal protections that are often based on Fair Information Practice and the more comprehensive Organisation for Economic Co-operation and Development (OECD) Principles and the European Union Data Protection Directive. 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 by 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 them, 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
- Some data is too sensitive to be collected, unless there are extreme circumstances that justify such gathering (e.g., sexual orientation, religion)[14]

While these basic data protection laws are a valiant start to encouraging data stewardship[15]- the laws fail to envision advanced AI and ubiquitous data gathering. In many ways, privacy has become the talisman of organized resistance to the encroachment of ubiquitous data gathering. It is however, a

[13] *See* Graham Greenleaf, *Global Data Privacy Laws: 89 Countries, and Accelerating*, Privacy Laws & Business International Report, Issue 115, Special Supplement, February 2012)

[14] Within this area may fall some of the currently labeled National Security exceptions to data collection, such as the USA PATRIOT Act of 2001, The Homeland Security Act of 2002, (6 U.S.C. § 222.) and the Real ID Act of 2005 (H.R. 1268, Pub. L. No. 109-13 (2005).)

[15] 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. The concept of a data steward is intended to convey a fiduciary (or trust) level of responsibility toward the data. See Sara Rosenbaum, Data Governance and Stewardship: Designing Data Stewardship Entities and Advancing Data Access, Health Serv. Res. 2010 Oct; 45(5 Pt 2): 1442–1455 (Oct 2010) https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2965885/

false idol as 'private information' is mischaracterization indicative of a much larger issue that looms on the horizon.

You see, much of the information that exists about me is information I freely share- thus, it is not, nor should it be, considered private. And 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 amount to nothing, but when combined can lead to extremely accurate profiles of who we are as shoppers, as voters, as people, as neighbors, and as community members.[16]

That is not to write that privacy is not important. It remains important in the traditional manner that the concept has always existed- governments intrusion upon areas of our lives that were never intended to be within public scrutiny[17] and areas of protections that should be afforded due to the sensitivity of the information-[18] both of which should continue be considered within the privacy rubric. However, how should we regulate data that is not- nor was it ever intended to be – private. Information such as my age, my gender, in general- information that 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 will be discussed later, industry based data stewardship and compliance standards of impact assessments mitigate the consequences of this type of use. However, 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'[19] difficult to achieve.

B. Property and Proprietary Rights

[16]

[17] See discussion, supra note xx-xx and in text

[18] See discussion, supra note xx-xx and in text

[19] Information 'flow' it 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. 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 authors 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 authors opinion) feed the magic narrative- to encourage people to believe the vast majority of individuals cannot understand this technology and cannot control the river. (One should admit- I have been criticized for being this obviously anti-business- it is a matter for another paper, however.)

Property as a defining legal structure for data is also troubling as the 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- that is, it is the storage system- owned by an entity- that is entitled to legal protections and not the individuals data that is contained within the data base. 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. 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.[20] This all changed in 1991 when the U.S. Supreme Court determined in *Feist Publications v. Rural Telephone Service Co*.[21] that the telephone directory was in fact uncopyrightable. In making this determination the court clarified that the sole basis for protection under U.S. copyright law is creative originality.[22]

The Court notes, however, that "the vast majority of compilations will pass" the originality test.[23] The cases after *Feist* are informative of the Court position, in fact, the cases subsequent suggest that the 'originality' requirement is in fact very easy to establish. For example, in *Key Publications, Inc. v. Chinatown Today Publishing Enterprises Inc*.[24] the Second Circuit sustained the copyrightability of the yellow pages of a telephone directory. The court found that the selection of entries in Key's directory was original.[25] In addition, the arrangement of the directory into categories was- when "viewed in the aggregate" original, because it "entailed

[20] While some circuits applied the 'sweat of the brow' doctrine, [see, e.g., Illinois Bell Tel. Co. v. Haines & Co., 683 F. Supp. 1204 (N.D. Ill. 1988), aff'd, 905 F.2d 20 1081 (7th Cir. 1990), vacated and remanded 499 U.S. 944 (1991); Rural Tel. Serv. Co. v. Feist Publications, Inc., 916 F.2d 718 (10th Cir. 1990)] others instead required that compilations contain sufficient creativity in their "selection, coordination or arrangement" to render them "original works of authorship" entitled to copyright protection. See, e.g., Hutchinson Tel. Co. v. Fronteer Directory Co. of Minnesota, 770 F.2d 128 (8th Cir. 1985); 22 Southern Bell Tel. and Tel. Co. v. Associated Tel. Directory Publishers, 756 F.2d 801 (11th Cir. 1985).

[21] 499 U.S. 340 (1991).

[22]

[23] Feist, at 359.

[24] 945 F.2d 509 (2d Cir. 1991).

[25] Key at 513

the de minimis thought needed to withstand the originality requirement."[26] It is the case, however, that the scope of protection is in fact very 'thin.'[27] As a result, the ability to copyright is no longer an issue, instead the focus turns to the scope of protections. In the vast majority of instances, post- *Feist* most of the appellate cases have found wholesale takings from copyrightable compilations to be non-infringing.[28]

Of course, 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. However, there is little getting around the inability to copyright facts- thus, even in value added databases, copyright only protects information that is considered value added.[29] For example, West Publication has long held a data base of case law- while the case itself is not copyrightable, the case synopsis and indexing system are protected.[30]

Intellectual property rights also include protections for algorithms,[31] generally under patent law. In 1994, the Court of Appeals for the Federal Circuit decided *In re Alappat* that an invention that had a novel software algorithm combined with a trivial physical step was eligible for patent protection.[32] While many may assume this case- and those after- opened the door for a large scale increase in the patenting of software, in fact this has not generally been the case. A 2011 Berkman Center Research report set out to examine the changes in the patenting behavior of the software industry. The researchers 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 or in other industries. . . " [33] In many ways, this outcome is somewhat

[26] Key at 514

[27] Key Publications, 945 F.2d at 514

[28] Jane C. Ginsburg, Copyright, Common Law and Sui Generis Protection of Databases in the U.S. and Abroad, U.CIN.L.REV. (1997).

[29] Baila H. Celedonia, From Copyright to Copycat: Open Season on Data?, PUB.WKLY.,74 Aug. 16, 1991, at 34 (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)

[30]

[31] algorithms are encoded procedures for solving a problem by transforming input data into a desired output. Gillespie, T. (2013) 'The Relevance of Algorithms.' In Gillespie, T, Boczkowski, P & Foot, K. (eds.) Media Technologies: Essays on Communication, Materiality and Society, Cambridge. MIT Press: MA Oddly, even this term lacks clear definition. See Id.

[32] In re Alappat, 33 F.3d 1526, 1542-44 (Fed. Cir. 1994).

[33] James Bessen, A Generation Of Software Patents, Boston University School of Law

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.*[34]. In fact, Jasper L. Tran notes:

> This 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. [35]

It is important to note, patents remain valid until challenged and invalidated,[36] so the vast majority of these patents will remain in force. [37] Moreover, one must consider the two research publications in tandem- consider the vast majority of these patents have likely been obtained by a 'few large firms' and that most will remain valid- until invalidated through an expensive and time consuming process- a process that one can imagine favors the large, financially robust, firm. And, of course, one can also 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.

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 database or algorithm. While this approach may work in some areas, it is immediately obvious that individual information, shared publically- should not be thought of as 'owned.' It is the data base that is protected, and the information inside the database that is used in the evidentiary process to demonstrate infringing activity. Simply put, a business cannot own my information in their database, the business is entitled to have protections that provide consequences to those that intrude upon their database and steal the information wholesale. But, one must wonder how long such a protection will provide any comfort to the business. As more and more information is freely- and openly –shared, as processing power increases, and

---

Working Paper No. 11-31 (June 21, 2011) Berkman Center Research Publication No. 2011-04 available at http://www.bu.edu/law/faculty/scholarship/workingpapers/2011.html

[34] 134 S. Ct. 2347 (2014).

[35] Tran, Jasper L., Two Years After Alice v. CLS Bank (June 20, 2016). Journal of the Patent and Trademark Office Society, Vol. 98, p. 354, 2016. Available at SSRN: https://ssrn.com/abstract=2798992

[36] 35 U.S.C. § 282 (2012). The Patent Act of 1952, Pub. L. No. 82-593, ch. 950, 66 Stat. 792, 812, codified the presumption of validity.

[37] Dennis Crouch, What to do About All These Invalid Patents?, PATENTLYO (Aug. 28, 2014), https://patentlyo.com/patent/2014/08/these-invalid-patents.html

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.

The inability to resolve the issue of ownership has contributed to the rise of the control based regulatory regime. Control- that is 'what entity had control over the data/information at the time' 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 the sharing with third parties. 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- your automobile. You can lend out your car and the individual driving the car is in control. The driver has responsibilities of driving the car- but he/she is not the owner. Despite you not being in control of the car you still have 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 is 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 too 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.

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.[38] In the

[38] Anjanette H. Raymond, The Consumer As Sisyphus: Should We Be Happy With 'Why Bother' Consent?, JOURNAL OF LEGAL STUDIES IN BUSINESS, Vol. 20, 1-26 (2017) (2015 SEALSB Best paper Award). That is not to write that this issue has been ignored by policy makers, however. On August 7, 2017, the United Kingdom Department for Digital, Culture, Media and Sport announced in its report A New Data Protection Bill: Our Planned Reforms announced its intentions to update the existing law. The Department

United States, consent has been watered down to such a large extent that many scholars argue it is a mere check box in the privacy debate.

As I have previously argued, even when individuals are required to 'give their consent' it is often part of a process which I call 'must, rush, and trust.'[39] In the digital world, individuals must have the information/item/or webpage, they are in a rush to get it, and the trust the law to protect them.[40] Any affirmative click that stand 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 fulfil.[41] Moreover, numerous research scientists have argued that even in the face of multiple click boxes and forced reading screens, individuals still fail to read terms, often fail to reflect upon terms they lack understanding of, and rarely refuse to click consent.[42] Thus, businesses are able to dictate vague, widely ambiguous, overly broad terminology that seeks to capture the individuals consent to any and all use of data- regardless of impact or intent, for now and forever.[43] For example, on August 3, 2017 Tesla confirmed that the Tesla Model 3 has a driver-facing camera embedded into the rearview mirror of each vehicle.[44] Although the company insists the camera has not yet been activated, the company states 'will only become active after future software updates." 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

writes: writes: "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."

[39] Anjanette H. Raymond, Yeah, But Did You See the Gorilla? Creating and Protecting an 'Informed' Consumer In Cross Border Online Dispute Resolution, 19 HARVARD NEGOTIATION LAW REVIEW 129, 129-171 (Spring 2014).

[40] Raymond, Yeah, But Did You See, supra note XX

[41] Id

[42] Id.

[43] Consider the case of the IRobot manufacturer which after years of mapping your home, has changed TOS so that it may sell the information gathered about your home. 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 iRobot was apparently caught off guard by the immediate and loud push back and has since responded with assurances that nothing has been done to date- or is officially- in the works. Brian Heater, iRobot's CEO defends Roomba home mapping as privacy concerns arise, Tech Crunch (July 25, 2017) https://techcrunch.com/2017/07/25/irobots-ceo-defends-roomba-home-mapping-as-privacy-concerns-arise/

[44] Fred Lambert, Tesla Model 3 is equipped with a driver-facing camera for Autopilot and Tesla Network, Electrek (Aug. 1, 2017) https://electrek.co/2017/08/01/tesla-model-3-driver-facing-camera-autopilot-tesla-network/

that popular Comedy Central program Southpark has featured the issue for many years now.[45]

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.[46]

Of course, as technology designers take note of the legal limitations created by the simple binary model of consent, new process may be developed that address concerns.[47] 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."[48]

In addition, even if graduated consent is used, the sheer volume of data gathering and storage is an ongoing issue for organizations. Data protection regulation in the European Union requires, for example, storing personal data securely,[49] keeping personal data up to date,[50] permitting data subjects to access their personal data,[51] complying with requests from data subjects for their personal data to be deleted,[52] and actively deleting personal data once it is no longer required for the purpose for which it was collected,[53] 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.

---

[45] Jason D. O'Grady, South Park parodies iTunes terms and conditions, ZD Net (April 28, 2011). Of course, truth makes humor more accessible, see David Kravets, TOS agreements require giving up first born—and users gladly consent, Ars Technica (2016)

[46] Information Commissioners Office (UK), Big Data, Artificial Intelligence, Machine Learning And Data Protection (2017) https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

[47] Anjanette H. Raymond, Yeah, But Did You See the Gorilla? Creating and Protecting an 'Informed' Consumer In Cross Border Online Dispute Resolution, 19 HARVARD NEGOTIATION LAW REVIEW 129, 129-171 (Spring 2014).

[48] ICO (2017)

[49] Supra note, XX

[50] Supra note, XX

[51] Supra note, XX

[52] Supra note, XX

[53] Supra note, XX

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. As an individual, should I wish to access my data with an organization, I can of course do this, but this truly misses the point as it is a rare case that data exists in this manner. It also presupposes that I know where the information is at- what entity to request the information from, and that the data has not already been used by another entity to create information about me. 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. Non-stop email based notifications and consent will not increase knowledge, it is information overload resulting in individuals not glancing at the information but just merely clicking consent or in the alternative, incentivizes business to include vague, over-encompassing, overly broad consent clauses which allow any and all information sharing. Neither reality is one that we should support, consent is irrelevant in the majority of situations in the digital world.

D. Transparency

There is one aspect of machine learning that stands out as a distinct challenge for policy makers and corporate governance. Statistical models that accurately predict an outcome or classify an object have traditionally been transparent in their reasoning. We have been able to see their workings and interrogate their internal logic. For some AI algorithms, this is much more difficult and in some instances, in the current environment, is impossible.[54]

Of course, to date, the assumptions that underlie the operations have been rule based, in which computers are merely following instructions and any errors are attributed to the people. However, some AI will stand this process on its head, as the machine itself will find patterns and create rules.[55] In this way, it may just be that the machine is at fault. Moreover, as I have previously argued at length, automated processes – and hidden black box worlds based in math and number crunching- can still carry 'an aura of

[54] Dave Weinberger (2017) Alien Knowledge. MINE

[55] See id. And may even create its own language to communicate amongst AI based technology. See Mark Wilson, AI Is Inventing Languages Humans Can't Understand. Should We Stop It?, Fast Code Design, (July 14, 2017) https://www.fastcodesign.com/90132632/ai-is-inventing-its-own-perfect-languages-should-we-let-it?utm_content=bufferc5387&utm_medium=social&utm_source=facebook.com&utm_campaign=buffer They did in fact shut it down, see id.

objectivity and infallibility.'[56] Thus, even when some transparency is possible, individuals tend to 'trust the numbers' despite evidence to the contrary.

However, the failure to understand the inner workings of an algorithm should not be allowed 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, EU citizens will be entitled to know how an EU institution arrived at a conclusion--even if machine learning and a black box was involved.[57] 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"[58] for decisions.[59] In fact, the law does more; it also 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."[60]

Unfortunately, the Data Protection Regulation will likely be viewed as a regulatory failure and an extreme over-reach that may just result in cleaver work arounds or the stifling of innovation. For example, the use of the term 'solely' will allow for compliance so long as the decision was not solely based in advanced AI, for instance when a human is involved in even a rudimentary aspect of the process. Additionally, the right to an explanation should not be viewed as a right to transparency of process. In fact, it is a right to nothing more than a basic explanation of the decision process, such as the

[56] Anjanette H. Raymond, Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias, (with Emma Arrington Stone Young and Scott J. Shackelford), 15 NORTHWESTERN JOURNAL OF TECHNOLOGY & INTELLECTUAL Property_ (forthcoming 2018).

[57] Bryce Goodman and Seth Flaxman, *EU Regulations On Algorithmic Decision-Making And A 'Right To Explanation,'* 2016 ICML Workshop on Human Interpretability in Machine Learning (WHI 2016), New York, *available at* https://arxiv.org/abs/1606.08813 . Do note, the phrase '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. See Yeung, Karen, Algorithmic Regulation: A Critical Interrogation (May 23, 2017). TLI Think! Paper 62/2017; Regulation & Governance, Forthcoming; King's College London Law School Research Paper No. 2017-27. Available at SSRN: https://ssrn.com/abstract=2972505

[58] *Goodman, EU Regulation, supra note X*

[59] *Id.*

[60] *Id.*

notifications that occur when you are rejected for credit. Both of these are serve as examples of needing to understand the basic inner working of the process and to seek to achieve a particular goal within the process structure- or risk creating a set of rules that are likely to be easily subverted and meaningless.

Of course, if you condition taking action on first achieving some kind of objective scientific understanding about what caused something to happen, you run the risk of either giving yourself a license not to take any action at all, or to create false narratives that seem to imply the existence of objective scientific understanding of something when none exists.

At the current time, understanding the inner workings of an advanced AI is almost impossible to achieve and no amount of transparency will remedy this predicament. As a result, transparency should be abandoned – at least as the term is used within the context of process,[61] as it creates a false belief that human beings will understand and interpret the inner working of AI algorithm, 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[62] that are the product of the algorithm should be part of the design process, a point that will be discussed in detail in XXX.

III. THE REGULATORY FUTURE

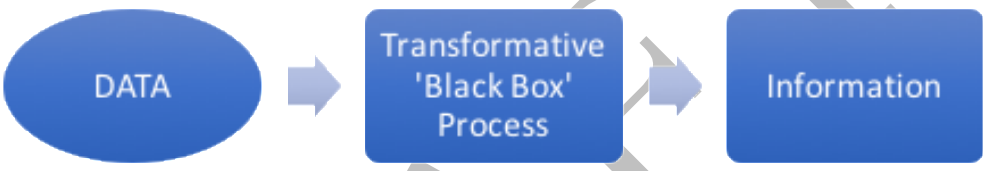
The law and ethics surrounding new fields of innovation do not have to be handled in a substantially new manner, so long as policy makers do not succumb to the dazzle of the technology. In fact, States and societies are already equipped to navigate human error and typically have a range of escalating options for responding to the types of risks which could be associated with new products or services that are not responsibly deployed. 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.

[61] 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.

[62] Anjanette H. Raymond, A Meeting Of The Minds: Online Dispute Resolution Regulations Should Be Opportunity Focused, U.C. DAVIS BUSINESS LAW JOURNAL, Vol 16(2) (2016).

A. The Basics

In general, the following diagram sets out the questions- or the push points- of concern that, in this authors opinion- must first be considered as the key divisions, these original delineations can then be expanded into more widespread discussion within each area. It is key to understand however; this diagram does not truly capture information movement- it is an example to begin discussions.



As the diagram attempts to capture, 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. The diagram does not fully capture 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. 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. 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 issues. Second, the final box- that is the 'information' box, can be thought of as a single unit as well. The output from the transformative process should also be subjected to rigorous scrutiny. Issues such as, is the outcome capable of replication; can it be generalized; is it biased or discriminatory on its face; and similar issues.

However, it is the transformative process – that 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; (d) What biases are allowed, or otherwise permissible? (3) What is actually spit out- does it match our expectations? Is it problematic on its face? (to name but a few)

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.

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.[63] 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.[64] The results are then used to prioritize risks in terms of their importance, these ranking allow project's management to strategize resource allocation and to work to mitigate "high probability/high consequence risk events." [65] Within the field, enterprise environments are the emerging area.

> 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, and applications through an interconnected network of information and communications technologies.[66]

[63] National Institute of Standards and Technology, July 2002, Risk Management Guide for Information Technology System, Special Publication 800-30, p. 1.

[64] MITRE https://www.mitre.org/publications/systems-engineering-guide/acquisition-systems-engineering/risk-management/risk-management-approach-and-plan

[65] MITRE

[66] MITRE

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.[67]

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. 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. 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

6. Mitigation Response - a brief overview of mitigation steps to eliminate or reduce the risk.

These questions can then guide the development of global information governance regime.[68] For example, consider the use of highly sensitive information, such as a significant medical diagnosis. The risk of this information being gathered is that it could be used to discriminate- in a variety of ways. The first question is to 'classify' the information (or data). To do

[67] MITRE

[68] Global governance may be defined as "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.

this we first ask- what entity is seeking to gather/use this information? How do they intend to use the information? We then consider, will this use have a high impact on the individual (or other stakeholders)? Moving on we consider- what is the probability of the risk occurring? Are there mitigation steps that could be put in place to lessen the risk or impact?

After this information gathering process is complete we can classify the data/information. 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). This information deserves the highest protections- thus should be subject to legal protections and prohibitions on the ability to share the information beyond a one time, single use. 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. This type of data could also be subject to mandatory mitigation measures, 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).

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. 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 protections. 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. However, non-critical data (that is data that is widely available or is otherwise part of the public knowledge base) can be protected- if at all- through industry create standards. 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. The model can thus, accommodate the use- and thereby the impact- of the data use within the risk framework. 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. However, 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.

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

C. Industry Key Recommendations

As one would imagine, we are most likely a long way from the creation of a framework to govern that regulation of the transformative process. 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. 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. The SIG is a not-for-profit, non-stock corporation which does not make, manufacture or sell Bluetooth enabled products. 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 Bluetooth SIG. 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.

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.[69] 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- as membership is revocable. 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.

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. 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.

That does not mean, however, that industry must start from the beginning, some guidance already exists. To this end, commentators have suggested the following key recommendations for any industry to consider as a first step in

69

creating a framework for information governance.

1. Incorporate Impact Assessment

According to Information Commissioners Office in the United Kingdom, "privacy impact assessments are a tool which can help organisations identify the most effective way to comply with their data protection obligations ***and*** *meet individuals' expectations of privacy.*"[70] While I have already written about the authors hesitation with the continued use of the term 'privacy' it is important to note the full breadth of its use within this context- within this particular assessment, privacy includes physical and informational privacy, including:

> the 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.[71]

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. The British conceptualization of the impact assessment it to "minimise . . . the risk of harm through use or misuse of personal information."[72] 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- all provided with an eye toward compliance with the Data Protection Act.[73] Most relevant for the paper is a key consideration:

> As part of the PIA process organisations should describe (to the identified internal and external stakeholders who share wide areas of expertise and interests in the area) how

[70] Information Commissioners Office (ICO) (U.K.) Conducting Privacy Impact Assessments, Code of Practice, 2014022, Version:1.0 (2014) https://ico.org.uk/media/for-organisations/documents/1595/pia-code-of-practice.pdf

[71] Information Commissioners Office, Conducting Privacy Impact Assessments, at 6.

[72] ICO at 8

[73] Neither compliance with the Code nor completion of the impact assessment is a requirement of the Data Protroaction Act.

> 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.[74]

Let's take a simple example of an impact assessment that is not necessarily privacy focused, but is none-the-less, 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. However, 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. 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.[75] Turns out- the research was incomplete. Seems gender- when considering heart attack presenting symptoms- are in fact very different. The exclusion of gender from the conversation, created an unreliable- and very dangerous- generalization. In this instance, gender as data should have been considered within the conversation. Thus, gender should have been considered. But of course, there are time when gender is an unnecessary consideration and may in fact, do nothing but capture societies biases within the transformative process. Turns out, despite what you may believe, women are not worse drivers than men- when 'worse drivers' are measured in terms of automobile accidents.[76] If designers are allowed- or machine learning processes capture - this poor assumption, then designers need to mitigate for that negative impact.

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. And, it is this area which I suggest should be added into consideration (although, admittedly this area of use would not yet be covered by the DPA). None-the-less, the considerations of impact can still stand as guiding principles within

---

[74] At 22

[75] http://www.healthline.com/health/heart-disease/heart-attack-symptoms#symptoms-in-men3

[76] http://abcnews.go.com/2020/story?id=3148281&page=1

the wider use of an impact assessment. For example, currently machine learning translation and the google translate algorithm is all the rage as more and more languages are capable of translation.[77] Although, speech recognition and translation are vastly better than it has ever been,[78] it still has a long way to go, especially when attempting to translate languages that lack standard, predictable rules and patterns.[79] English, is incredibly hard for numerous reasons as anyone attempting to sort through all the colloquial phrasing and double use of words can attest to- everything is context.

One issue that is 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 consider which one more closely resembles the sentence, including meaning and phrasing. Turns out, even poorly translated sentences can lead to someone understanding the main idea of the sentence (turn left of Huggins) but the phrasing and language patterns add context.

Why might all of this matter, especially in light of impact assessments? Imagine you are driving down a road and the driving assistant wishes for you to 'turn left on Higgins'- 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. For an impact assessment to be useful, we have to be able to anticipate the impact on these mistakes upon the user. Fortunatly, the spoken word – in situations of driving- can often be corroborated or verified by visual ques, such as road signs, and available turning lanes, and thus, a poor translation may have less impact in these situations. However, if translation algorithms are used by police officers in the field- the absence of visual ques and the intensity felt by the individual may in fact have a drastic impact upon the individual who cannot understand the police officer's questions. Thus, the same exact program will have drastically different impacts based upon the particular use of the instrument.

[77] For a much better explanation, see Luba Belokon, Machine Learning Translation and the Google Translate Algorithm, Data Science Central (Aug. 1, 2017) http://www.datasciencecentral.com/profiles/blogs/machine-learning-translation-and-the-google-translate-algorithm

[78] Language: Finding A Voice, The Economist, (May 1, 2017) http://www.economist.com/technology-quarterly/2017-05-01/language

[79] Finding a Voice

2. Insist Upon Privacy by Design

While this paper is not intending to discuss cyber-security, the paper would be incomplete if privacy by design was not considered within the industry recommendation section. This is primarily because the privacy by design tends to encompass some of the principles that could be added into the impact based risk assessment. 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.[80] Within the design approach, privacy is protected through the use of technology based supports. For example, data controllers will be obliged to take "appropriate technical and organisational measures for ensuring that, by default, only personal data which are necessary for each specific purpose of the processing are processed,"[81] such as: anonymization techniques, security measures to prevent data misuse, data minimization measures, purpose limitation and data segregation measures, and restrictions on meta data.[82] In terms of the broader discussion, privacy by design is an integral part of the conversation because the design process is an essential consideration mitigation considerations in the risk assessment.[83] Yet, it should be noted- the design process in this instance focuses more of security and less on the transformative process.

3. Eliminate Emotionally Crafted Narratives and Bad Data Science

USA Today authors Jefferson Graham and Laura Schulte, describe what some commentators consider a disturbing new use of technology that has Wisconsin workers voluntarily being embedded with microchips to facilitate payment of purchases from workplace vending machines. The sentiment of supervisor of the company Three Square Market President Patrick McMullan captures the emotionally charged, bad logic based concerns:

---

80 https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf

81 GDPR Article 25

82

83 D'Acquisito, Giuseppe et al, Privacy By Design In Big Data. An Overview Of Privacy Enhancing Technologies In The Era Of Big Data Analytics, ENISA, December 2015. https://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/big-data-protection

> 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."[84]

Of course, the comparison of what cell phones track misses the point of the work place privacy issues and fails to appreciate the fallacy that it employs to reject outright a valid concern, especially in light of employees inferior bargaining position. Slippery slopes and similar fallacies are frequently employed to reject outright otherwise valid concerns. 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 narrative that simply must stop.

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. I have previously written about the litany of 'projects' that are undertaken with the full knowledge and expectation that individuals will be impacted- yet, that impact is not considered, controlled, or monitored for harmful effects.[85] Facebook- and its various 'adjustments' have caused outrage amongst commentators and social scientists[86] – as the impact upon individuals, in at least some cases, is irrefutable. Data scientists- and others- must be held to the same standards of any social scientist.

Finally, the industry must stop using narratives that misrepresent or overgeneralize the issues or the actual technology being deployed. Narratives are generally understood storytelling, "wherein the choice of what data to plot, and how, is tailored to the message the authors want to deliver."[87] While research generally supports the use of narratives to explain scientific

---

[84] Jefferson Graham and Laura Schulte, Wisconsin workers embedded with microchips, USA TODAY (Aug. 1, 2017) https://www.usatoday.com/story/tech/talkingtech/2017/08/01/wisconsin-employees-got-embedded-chips/529198001/

[85] Anjanette Raymond, Emma Arrington Stone Young and Scott J. Shackelford Building a Better HAL 9000: Algorithms, the Market, and the Need to Prevent the Engraining of Bias, 15 Northwestern Journal Of Technology & Intellectual Property, (forthcoming 2018).

[86] https://www.nytimes.com/2014/06/30/technology/facebook-tinkers-with-users-emotions-in-news-feed-experiment-stirring-outcry.html?_r=0

[87] Katz Y, (2013) Against storytelling of scientific results. Nat Methods 10(11):1045.

information to a non-scientist,[88] the field has also explored the ethical aspects of the use of narratives to persuade, instead of inform.[89] The distinction between the two becomes especially relevant in light of the research that reveals that "narratives can also perpetuate misinformation and inaccuracies about science or about scientists themselves."[90] In fact, research into science based narratives argues that because narratives are not subject to the same truth requirements as logical-scientific communications- the message is not easily countered. [91]In fact, accepted narratives are trusted so much that individuals rarely allow evidence to contradict the narrative.[92] Instead, evidence is altered to fit their narratives.

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 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,[93] the technology industry often utilizes narratives to further hide the issue. 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 are rejected should the purpose be to over simply or misrepresent the technology being described.

[88] Michael F. Dahlstrom, Using narratives and storytelling to communicate science with nonexpert audiences, Arthur M. Sackler Colloquium of the National Academy of Sciences, "The Science of Science Communication II," held September 23–25, 2013, at the National Academy of Sciences in Washington, DC. http://www.pnas.org/content/111/Supplement_4/13614.full

[89] Dahlstrom MF, Ho SS (2012) Ethical considerations of using narrative to communicate science. Sci Commun 34(5):592–617 (2012).

[90] Barriga CA, Shapiro MA, Fernandez ML (2010) Science information in fictional movies: Effects of context and gender. Sci Commun 32(1):3–24.(2010)

[91] Bruner J,(1986) Actual Minds, Possible Worlds (Harvard Univ Press, Cambridge, MA), p 222. (1986)

[92] McComas K, Shanahan J, (1999) Telling stories about global climate change—Measuring the impact of narratives on issue cycles. Communic Res 26(1):30–57 (1999)

[93] Green MC (2006) Narratives and cancer communication. J Commun 56(Suppl 1):S163–S183; National Science Board (2012) Science and technology: Public attitudes and understanding. Science and Engineering Indicators 2012 (National Science Foundation, Arlington, VA); Norris SP, Guilbert SM, Smith ML, Hakimelahi S, Phillips LM (2005) A theoretical framework for narrative explanation in science. Sci Educ 89(4):535–563; Avraamidou L, Osborne J (2009) The role of narrative in communicating science. Int J Sci Educ 31(12):1683–1707.

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 in fact occur, if we use a historically defined process within the fields of psychology and law. Amongst other aspects of mental health treatment possibilities, cognitive behavioral therapy uses outwardly visible inappropriate or undesirable behaviors as indicators of maladaptive thinking in individuals. In general, inappropriate or undesired behaviors are evidence of maladaptive thinking- assisting individuals in xx undesired behaviors and examining the maladaptive thinking that leads to the repetition of those behaviors, is often a first approach to treatment in the area. In a similar manner, outcomes- the information that is produced - can be examined for maladaptive machine logic, bad data, or other contaminants in the outcome. It is the examination of outputs that leads to the need for closer scrutiny.

While this process may sound daunting, it is in fact a process that is already used- although usually under well controlled circumstance. For example, the now infamous Google search debacles. 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.[94] Google also came under fire in July 2015 when its photo app autonomously labeled a pair of black friends as animals.[95] In this instance, however, the engineer in charge of the program 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. [96] 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. In this manner, algorithms are auditable.

In fact, I have 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. For example, there is wide

[94] Ben Guarino, Google faulted for racial bias in image search results for black teenagers, The Washington 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

[95] https://www.washingtonpost.com/blogs/govbeat/wp/2015/07/08/why-googles-nightmare-ai-is-putting-demon-puppies-everywhere/?tid=a_inl

[96] https://www.washingtonpost.com/blogs/govbeat/wp/2015/07/08/why-googles-nightmare-ai-is-putting-demon-puppies-everywhere/?tid=a_inl

concern about the use of AI based pattern recognition that is currently occurring in policing and probation/parole. Systems such as this are highly impactful on personal liberty and should thus, be subject to the most stringent of review. As such, the 'perfect' data[97] set could be developed- this data set could then be fed into the system to ensure the outputs were not out of line with expected outcomes. If the outcomes did not correspond with expected outcomes, one could presume that the AI had 'learned' a bias that must be addressed. Although this process is probably difficult to imagine, it is in fact a process that is used amongst the industry to 'train' systems and to build in corrective measures.[98]

Moreover, algorithms may be able to audit other algorithms. While this may seem like HAL as the mechanism to turn off HAL Jr.[99] it is in fact, a genuine possibility. [100] According to UK, "at 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."[101] While in the U.S. in the US, consultant companies are already being set up that specialize in providing algorithmic auditing services to their clients.[102]

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. 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 library, say scikit-learn, could explain in a sentence why a particular input case was classified the way it was.[103]

---

[97] Intentionally nebulas term, primarily because the perfect data set would need to be developed based on the specific parameters of the algorithm in development. This is of course done to combat the 'garbage in- garbage out' adage. It is however, the introduction of another potential area of bias.

[98] Ajunwa, Ifeoma et al. Hiring by algorithm; predicting and preventing disparate impact. SSRN, 10 March 2016. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2746078

[99] HAL 2001 A Space Odyssey….

[100] Adler, Philip et al. Auditing Black-box Models for Indirect Influence. IEEE International Conference on Data Mining. December 2016 reported in ICO report

[101] ICO report

[102] For example, O'Neil Risk Consulting and Algorithmic Auditing. http://www.oneilrisk.com/

[103] Diakopoulos, Nicholas. Accountability in algorithmic decision making. Communications of the ACM 59, no. 2 (2016): 56-62

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.

D. Recommendations for Policy Makers

Policy makers also face a daunting task when considering issues surrounding the transformative process. The first issue that must be considered is the overall framework that should be deployed within the area. And of course, that requires a true conversation to occur to develop the areas upon which the framework will be built. 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.

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. Today, the privacy conceptualization has expanded to include information that we seek to shield from prying eyes. In general, this information is considered sensitive- for a variety of reasons, and as such should be allowed to remain private. However, if information is private, shouldn't I be expected to protect it as important information that I value keeping private and shouldn't I be allowed to therefore, shield it from everyone, but for incredibly limited exception.

Yet, even individuals seeking to protect information as private find the task difficult to achieve. For example, in January 2017 DuckDuckGo conducted A Study on Private Browsing: Consumer Usage, Knowledge, and Thoughts.[104] Almost unsurprisingly, the results clearly show a true confusion among users about what exactly privacy entails in an online environment. According to the study, "46% of Americans have used Private Browsing," while those using it report the "number one reason people use Private Browsing is "Embarrassing Searches."[105] However, "76% of Americans who use Private Browsing cannot accurately identify the privacy benefits it provides." In fact, "41.0 ±2.5% believe that Private Browsing Prevents websites from tracking me" and "39.1 ±2.6% believe that Private Browsing

[104] https://duckduckgo.com/download/Private_Browsing.pdf
[105] DuckDuckGo

Prevents ads from tracking me."[106] And, do not assume this is a generational gap, as in general the misconceptions about Private Browsing are consistent. "However, younger audiences are more likely to believe a search engine couldn't see their searches in Private Browsing mode, whereas older audiences are more likely to believe that Private Browsing mode would protect their IP address from being seen." [107] Of course, when told about the true protections provided from private browsing "65.9 ±2.4% feel "Surprised", "Misled", "Confused" or "Vulnerable.""[108] Private and privacy does not mean what people think it means- and thus, should be abandoned as the safe and secure- i.e. private information narrative is misleading and misunderstood.

Moreover, 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. 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 my birth day, my place of birth, and my 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 private- how can I argue for this information being private when I so readily share it over and over online every day. And that of course, ignores the individuals who share this information and more on the easily search websites- such as Facebook, and Instagram.

Finally, consider one of the most heralded authentication devices- the social security number. 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. Data that is often widely available and used often by individuals- even published on Facebook. That means your social security number can be guessed- with shocking accuracy- from nothing more than publically available information, information that you yourself probably display in public. 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 if they are predictable from public data."[109] In fact, researchers found that it is

[106] DuckDuckGo 14
[107] DuckDuckGo 15
[108] DuckDuckGo 23
[109] Brian Krebs, Researchers: Social Security Numbers Can Be Guessed, Washington Post (July 6, 2009) http://www.washingtonpost.com/wp-

possible to guess many -- if not all -- of the nine digits in an individual's Social Security number using publicly available information, many could be guessed at by simply knowing a person's birth data.[110] 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 publically 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 thing of as private information. Privacy must be abandoned as the guiding principle- as we publically share to much information to realistically ever argue that information is private.

Instead, I suggest we use terminology such as sensitive and secure information to elevate the confusion and reduce the potential to fall into the trap of using existing privacy laws as guidance. Surely a social security number, if it is to continue to be used as an authentication device, should be labeled as sensitive 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. 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.

2. Reject Property as the Guiding Law

Issues of ownership of data- in both a personal and business context- has become an important consideration in terms of governance. 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. 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.

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 electronic highway. As Professor Mark Hall has observed when considering medical records, "Ownership was never much in doubt in an age

dyn/content/article/2009/07/06/AR2009070602955.html

[110] Krebs, SS Numbers can be guessed, supra note XX

of paper-based records"[111] because the paper record containing the information was owned by its creator.[112] However, 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.''[113] 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!). In the digital world, this is no longer the case as it is difficult – or near impossible- to identify the owner and, being legally in control of information is often a status to be avoided as it incurs heightened responsibilities.

Unfortunately, the absence of clearly defined property rights within the area likely has significant consequences to governance and property based rights. As argued many times before, uncertain legal positions allow parties to attempt to "simply to stake a claim." In the paper based world, claiming space forces a negotiation toward a contractual settlement that determines respective rights. Of course, the outcomes are only binding on the immediate parties, and the process of negotiation is an expensive barrier. As such, it is argued that property rights must be clearly established so that the respective parties know their legal default positions.[114] In the digital world, this position is fraught with difficulty as the negotiations never truly materializes. Instead, large scale technology creators control the market and limit entry to those that agree to their conceptualization of the property rights control. 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.

One also must consider the problems created- or the cleaver work arounds, that currently exist in overcoming both the ownership and control dilemmas. In terms of ownership, if an individual is allowed to 'own' their digital profile and the information it contains, at what point is the record merely data and nor capable of ownership. Consider a birthdate in a database- when it is *my* 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- it is simply a

[111] Mark A. Hall, Property, Privacy, and the Pursuit of Integrated Electronic Medical Records. Wake Forest University Legal Studies # 1334963 (2009) SSRN [accessed on August 15, 2017]. Available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1334963

[112] ABOVE-

[113] Hall, supra note X at 1

[114] 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);

number.

Maybe more importantly, 'owned' data allows individuals to place limits on its use- which may fail to fill societies need to use data in beneficial manner. Consider smart cities, natural disasters, and medical emergencies, should individuals be allowed to own data that prevents social benefit- should they be allowed to 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- to be subjected to a particular property regime, which is a matter for another day.

It is a lengthy discussion that prompted many scholars and policy makers to abandoned 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 creation. As such, many now consider control to the be guiding paradigm that ignores the 'ownership' issue and instead places responsibility upon the entity that controls the data. Time will tell if this basic shift in the discussion allows the discussions to advance> The author fear 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.

3. Reject Solutions Designed in a Paper Based World

That is not to write that the paper based world should not be a large consideration in the creation of policy, in fact the impacts of data loss created in the paper world remain a constant concern. For example, in August of 2017 it is alleged that health insurer Aetna "revealed 12,000 patients' HIV statuses by sending letters with giant envelope 'window' that exposed confidential information."[115] 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. However, as noted by the Executive 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. . ." As Executive Director Goldfein notes, it is the impact of such disclosure upon the individual that has the potential

[115] Mia De Graaf, Aetna revealed 12,000 patients' HIV statuses by sending letters with giant envelope 'window' that exposed confidential information,, Daily Mail (August 24, 2017) http://www.dailymail.co.uk/health/article-4820680/Aetna-sued-revealing-scores-patients-HIV-statuses.html#ixzz4qhvTpHOs

to cause the greatest damage. The paper based world is an important consideration, but deploying 'informed consent' based in 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.

I have 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.[116] Simply put, no one- and I literally mean no one, pays attention to the terms presented to us,[117] in fact, very few people even skim the information presented.[118] 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.[119] Moreover, presenting drastically one sided terms are not even a barrier to consent as 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 they 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

[116] Anjanette H. Raymond, Yeah, But Did You See the Gorilla? Creating and Protecting an 'Informed' Consumer In Cross Border Online Dispute Resolution, 19 HARVARD NEGOTIATION LAW REVIEW 129, 129-171 (Spring 2014); The Consumer As Sisyphus: Should We Be Happy With 'Why Bother' Consent?, JOURNAL OF LEGAL STUDIES IN BUSINESS, Vol. 20, 1-26 (2017);

[117] Raymond, Gorilla

[118] Raymond, Gorilla

[119] Raymond, Gorilla

analysis software for policing. Her team is examining transcripts from traffic stops to recognize patterns of racial disparity, and the teams' results are not necessarily unexpected but are none-the-less quiet troubling. 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."[120] 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.

Of course, this is just the tip of the iceberg when it comes to these types of disparity.[121] 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.[122]

And, 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. And, analytics applied to large swaths of large scale data has led to the re-examination of prior smaller scale research.

---

[120] Laureen Murrow, Cop Talk, The Sound of Bias, WIRED (Aug 2017)

[121] See HAL See also, Geoffrey Mohan, Stanford's Jennifer Eberhardt wins MacArthur 'genius' grant, Los Angeles Times (Sept. 18, 2014) http://www.latimes.com/science/la-sci-jennifer-eberhardt-genius-20140917-story.html

[122] Rachlinski , Jeffrey J. and Wistrich, Andrew J., Judging the Judiciary by the Numbers: Empirical Research on Judges (June 2, 2017). Annu. Rev. Law Soc. Sci. 2017. 13:X--X, doi: 10.1146/annurev-lawsocsci-110615-085032 ; Cornell Legal Studies Research Paper No. 17-32. Available at SSRN: https://ssrn.com/abstract=2979342

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. However, concerns surrounding the 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[123]- 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[124] have looked to data and advanced algorithms to gauge risk- and hence, make bail recommendations. [125]

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.

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.[126] 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."[127] This process reflects a widely-held belief- as noted by NBC New Reporter Schuppe: "Algorithms need humans — flaws and all — to oversee them."[128]

Notably, the number of people being held before trial in New Jersey has

[123] 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

[124] Schuppe, Post Bail. Three States to date have widely adopted such system- while another 11 have some use. Schuppe, Post Bail

[125] Schuppe, Post Bail

[126] Schuppe, Post Bail

[127] Schuppe, Post Bail

[128] Schuppe, Post Bail

dropped by nearly a third compared to last year.[129] Thus, commentators argue the system is working.[130] 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- it just reframes the reason given for the hearing.

As can be seen, the patterns revealed- or verified- lead to 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, uncomfortable must be considered in the context of the narratives that are crafted by 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 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 'flow' of information and creation of penalties and regulation make society feel safe, allow individuals to trust, encourage 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, 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 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.

[129] Schuppe, Post Bail
[130] Schuppe, Post Bail

Adoption of the Risk Assessment creates a guiding set of standards to be considered in each setting and each level of regulation.

{Author note: add more??}

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 50 states."[131] In fact, some discrimination was predicated on the fact that some customers had Spanish-language preferences.[132] In this instance, the company discovered the unintentional discrimination when conducting an internal review of amongst 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."[133] 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. The outcomes drew attention to underlying data revealed 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."[134]

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

[131] Consumer Financial Protection Bureau, CFPB and American Express Reach Resolution to Address Discriminatory Card Terms in Puerto Rico and U.S. Territories CFPB Newsroom, (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/

[132] CFPB Newsroom, supra note XX

[133] CFPB Newsroom, supra note XX

[134] CFPB Newsroom, supra note XX

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. 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 of a mobile device. The impact of the use of this data is the potential to "infer" precise location of the (mobile) device." 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.[135] 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.

---

[135] Edward R. McNicholas, Colleen Theresa Brown and Stephen McInerney, Illinois Becomes the First State to Pass a Geolocation Privacy Protection Bill, Data Matters: Sidley Cybersecurity, Privacy, Data Protection, Internet Law and Policy Blog - Sidley Austin LLP (August 2, 2017) http://datamatters.sidley.com/illinois-becomes-first-state-pass-geolocation-privacy-protection-bill/