A Survey of Data Ethics: Problems New and Old

Michael Skirpan
University of Colorado Boulder

I. Introduction

There should be nothing controversial about the suggestion that computer science and modern society are being reconfigured by new relationships with data. Machine learning, cloud computing, and network scalability are applied engineering areas undergoing accelerated expansion as data abundance and hardware innovations continue to open up possibilities and avail new insights. The "Fourth Paradigm" of science has been named.[1] Any human interacting with an online system is now entangled in a constant process of being tracked, modeled, and solicited information.[2]

But these changes did not magically spawn out of a computational ether. Much like any technological development, applied engineering and research have been informing and responding to one another along a historical chain that brought about our Data Era. As far back as 1967, researchers at the RAND Corporation created the Relational Data File as a way to use a computer for "the logical analysis of large collections of factual data" (Levien and Maron 1967). Within the realm of HCI theory, there exists a rich history of discussion on topics such as contextual privacy, situated/desituated action, and grammars of action[3]--the presence and capture of data being a substantive anchor in each conversation.

What has changed is that data-driven systems have recently been deployed like wildfire due to the growing availability of data-intensive infrastructure such as Apache's Hadoop and HDFS, Amazon EC2, and now Google's Tensor Flow. Within the past decade and a half innovations in computer virtualization, NoSQL databases, and GPU computing have pushed us toward new capacities for distributed, machine-intelligent systems. However, with any new technological and social arrangement must come a fresh set of risks, harms, and ethical norms. Some of these have been pre-empted; others we have only come to recognize in practice. For

[1] The fourth paradigm was coined as a way to describe the shift in methodology happening in modern science. Beyond hypothesis-driven experimentation, scientific discovery can now be steered by data acquisition and analysis. That is, we can capture data without having any clue of what we are looking for to only later avail hidden information within the dataset. For a longer treatment of this see: Hey, T., Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: data-intensive scientific discovery* (Vol. 1). Redmond, WA: Microsoft research.

[2] As of 2014, the online advertising economy was at $120 billion. The fuel of this economy is personal data mined by online trackers, which is bought and sold by third-parties that may or may not have any interest in the original context. See more in: "Getting to Know You." *The Economist*, September 13, 2014. http://www.economist.com/news/special-report/21615871-everything-people-do-online-avidly-followed-advertisers-and-third-party.

[3] Look to the work of Helen Nissenbaum, Jonathan Grudin, and Phil Agre, respectively for crucial treatments of these topics.

instance, I take it AOL did not realize how easy it would be to de-anonymize their users when they released a search query dataset back in 2006. On the issue, The New York Times quoted the executive director of the Electronic Privacy Information Center saying, “the unintended consequences of all that data being compiled, stored and cross-linked...” are “a ticking privacy time bomb” (Barbaro and Zeller 2006).

It is this unforeseeable expansion in our notion of *what’s possible* that has recently catalyzed researchers, engineers, and the government. Using the Association of Computing Machinery’s (ACM) online digital library search engine, we find a growth in publications with the keyword “big data” going from 9 in the year 2011 to 223 in 2015. The Obama Administration began a National Big Data Research and Development Initiative in 2013. Writers at The Harvard Business Review have even called it early and determined the data scientist as the “sexiest job of the 21st century” (Davenport and Patil 2012).

Within the context of these new innovations, we must continue to ask ourselves, “what could go wrong?” Some of the discovered challenges fall into areas researchers, engineers, and legislators have dealt with for some time. The Belmont Report provides guidelines for IRB panels that oversee human subject research in our academic institutions. The US Government has designed regulations through acts like HIPAA and FERPA to protect individuals against the negligent treatment of sensitive information in health and educational contexts. In tandem, computer scientists have built their own standards such as RSA, SSL, and SHA-3 to ensure our ability to provide a secure online experience. What has only recently become clear is that our prior solutions to concerns such as privacy and consent cannot fix all the problems Big Data has exposed.

What are data ethics?

The long history of theorizing and debating about privacy and human rights in the context of computing falls under the larger category of “applied ethics.” Applied ethicists are domain experts that project more abstract concepts such as *rights*, *ownership*, and *duty* into more specific subject matter. In what follows, a survey of applied ethical questions will be considered that are specific to matters concerning data and data-driven technologies. For our purposes “Big Data” will not be taken up as a particular quantity of data; rather, as a paradigmatic shift in the role data has in science, engineering systems, and everyday life. In this view “data ethics” is an applied field encompassing the problems that arise alongside the accumulation of massive quantities of data and the corresponding changes to scientific and engineering practice. These are questions pertinent to engineers, HCI designers and theorists, lawyers, and users as we witness and become enmeshed in this explosion of data-driven technologies. Domain experts have hardly had the time to reassess the relevant laws, social norms, and practitioner guidelines being challenged by Big Data.

We now have risk assessment and predictive policing systems in place around the US that transpose our available data into a numeric score that courts can use to alter a defendant’s treatment (Angwin et al. 2016). Trends like this put the heat on subject-matter experts and practitioners to quickly and soberly bring our attention to questions--new and old--regarding the ethical use of data in both engineering and society. Already, across the many domains touched by Big Data innovations, we are unearthing imminent ethical challenges in need of answers.

In HCI and social science research, we are rediscovering ethical problems concerning collecting, storing, and sharing human-subject data. When using social media data, how does one get consent from a subject who does not even know they are being researched? When sharing data, how can we be sure a dataset is sufficiently anonymized given what information is now publicly available? Is informed consent relevant to strictly online research? Do industry data collectors need IRB approval to study their own data?

For computer scientists and engineers, we are constantly making choices with data that will shape other people. Is it fair to train a model using a particular dataset? Or will that dataset favor a particular group of people? When data mining, what metrics should be captured and how should they be classified? What risks come in deploying systems whose models are uninterpretable to the engineers who design them?

Lawyers and legislators further have to answer these questions for themselves in order to ensure proper protections and avenues of recourse are in place. Can an algorithm enact illegal bias or discrimination on protected classes of people? Who is liable when a computer model makes a decision that unjustly harms someone? Do users have any rights to own or control the data they produce?

And finally we have to ask fundamental, philosophical and social questions about what we want life to be like. Can we trust the most powerful technical actors to be benevolent with our data? Is there any danger in our relationship with proprietary search engines when looking for factual information? At what granularity and frequency do we care to be notified about changes to privacy policies in the online systems we use?

Looking Across Domains

The following survey will frame and elaborate on the numerous issues being complicated and raised by data-driven technologies. The questions outlined above span a number of expert domains and express concerns that intersect with and depend on one another. However, in effort to cogently summarize such a broad topic, the remainder of this survey is structured into four distinguishing categories: 1) Privacy, 2) Discrimination, 3) Consent and User Attitudes, and 4) Algorithmic Impact.

The opening section on privacy will frame canonical questions about privacy in the context of Big Data. Having a longer history than other data-relevant topics, we'll see how older solutions to privacy questions have now proven to be insufficient and how those inadequacies bring light to new challenges. Establishing concepts such as "personally identifiable information," "anonymization," and "personalization" will in turn provide an important foundation for further elaborations.

From there, the focus will move to discrimination. Before the relevancy of discrimination to computing can be understood, an overview of the specific regulations that set the precedent for what legally counts as "discrimination" will be covered. The discussion will then move to contemporary engineering solutions that use machine learning techniques to train models for software applications. Any model that is *trained* requires data, thus we must look at how data may bias and affect the results obtained by our algorithms. Understanding bias, we can continue to see how often machine learning is a process overtly leveraged to discriminate and grapple with when discrimination is ethical or not.

Establishing a backdrop of foundational issues in privacy and discrimination will allow for a more lucid examination of questions directly concerning the user. The following section will introduce new problems surrounding consent involving terms of service, privacy policies, and human-subjects research. Grounding these concerns in present day controversies, attention will be paid to particular recent cases which have fomented critical discussion about the limitations of consent. In light of these complicated cases, we'll also look at the attitudes and concerns understood from the perspective of the users coming out of contemporary HCI and social science research.

Moving up a level of abstraction, we will lastly encounter the topic of algorithmic impact. In this section, the nature of how algorithms modify and re-habituate our patterns of life will be discussed. After characterizing the vectors through which algorithms act on humanity, the remainder of the discussion will look at the broad consequences data-driven systems can have on society.

Finally the essay will end by summarizing what's currently being done to mitigate these challenges and what future avenues we might pursue to improve our ethical situation.

II. Privacy: A Right and a Preference

Most notable among the list of issues raised by all the new data around us is privacy. By no means are privacy concerns new phenomena created by Big Data; rather, there is a long historical precedent governing the rights individuals have to privacy. In the legal sphere, the right to privacy interplays directly with a protection from harm. On the other hand, in HCI literature and our lives, privacy is a nebulous concept rife with normative judgments and cultural differences defining where we draw lines between the public and private spheres of life and in which contexts those lines hold. For instance, for one person the history of their love life may be a private concern, restricted to closest friends and never put on Facebook; whereas, another person may disclose their timeline of love publicly without any issue. The primary distinction which causes debate is between privacy as a right and as a preference. Specifically interpreting, "when has one's *privacy rights* been violated?"

A naive interpretation of privacy rights versus preferences in today's online era would be to assume that there is a categorical divide between sensitive and insensitive data. Sensitive data are those special numbers and characters that we do not share to anyone but our most trusted confidants such as doctors, lawyers, and spouses. These should only be passed around by confidential means and those who obtain them should be heavily regulated. Outside of those special pieces of information, everything else is personal preference: share what you want about your life, but if it's online, presume its public.

Assuming we could conjure some standard for what's sensitive and what's not, we still face a number of challenges. First, what if enough insensitive data allows us to infer sensitive data? If I have access to your purchase history, might I be able to tease out attributes of your medical record? Or what about cases like the Netflix Prize Dataset where a dataset believed to be anonymized--ie, uniquely identifying data removed--was then de-anonymized using outside data sources, allowing for the recovery of sensitive information (Narayanan and Shmatikov, 2006)? Both of these holes in our naive privacy view--the threats of inference and de-anonymization--are in fact real, tangible threats brought about by amassed data sources and new analysis techniques.[4]

What this naive treatment hopefully shows is that basic assumptions made about our *ability* to protect privacy no longer hold. In the absence of these formerly-trusted techniques and protections, computer scientists and lawyers must search for new ideas. Let's begin unveiling these issues by first looking backwards at how privacy has been handled historically and slowly move forward into an understanding of how it is today's privacy scene has changed.

Legal Protections of Privacy

For lawyers, the provenance of information privacy law starts with an article, *The Right to Privacy*, by Samuel Warren and Louis Brandeis in 1890 (Ohm 2010). The two lawyers argued that the rise of tabloid journalism brought about the need for privacy torts. That is, courts should allow plaintiffs to seek legal redress over the harms being done to victims unjustly exposed by tabloids. It took nearly seventy years, but finally the call for torts came in 1960 when William Prosser established four privacy torts that are still widely recognized in the US (Prosser 1960). What these amount to is legal precedent that one cannot intrude on someone's private affairs, publicly embarrass another over private information, conjure a false image about someone using publicity, or appropriate the identity of someone for external advantage. These torts operate on the notion of *harm* that can be done with information.

Had things stayed this way, computer scientists would have little concern since these violations presume some defendant intentionally enacted a privacy harm. Common privacy harms place no burden on people who actually collect or store information, rather guide judicial determination when someone calls attention to or actively exposes information for malicious purpose. Essentially, "no harm, no foul" is the doctrine when it comes to torts. Though as the government began to use computers for record keeping, a transition began from focusing on *harm* and to *prevention*.

The new world of digital storage erupted a new set of questions about what information can be stored, for how long, where should it be kept, and who is authorized to use the computer? Over a decade of back-and-forth finally led to a the "Fair Information Principles" (FIPS) and the legal ratification of those principles through the The Privacy Act of 1974. FIPS established many of the norms we find common today when collecting and storing personal information. Requirements such as *notice and consent*, *individual's right of access*, and *individual's right of amendment* along with corresponding enforcement and penalties became

[4] Consider Latanya Sweeney's analysis that 87% of US Citizens can be uniquely identified with just zip code, gender, and date of birth: Sweeney, Latanya, "Simple Demographics Often Identify People Uniquely" (Data Privacy Working Paper 3, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2000), http://dataprivacylab.org/projects/identifiability/paper1.pdf.

legally instilled (United States Department of Justice 2015). A more fundamental change created in this policy was Congress embracing, “a wholly data-centric approach, the PII approach, to protecting privacy” (Ohm 2010). Embedded in our regulatory approach to privacy was now an assumption that we could evaluate the risk associated with different data categories and precisely identify which fields in a database must be regulated.

Moving ahead in time, FIPS created a number of tactics to ensure these risky data entries would be appropriately handled. These tactics structured the foundation for now-commonplace anonymization approaches to sensitive data. Specifically, the 1996 enactment of the Health Insurance Portability and Accountability Act (HIPAA), which regulated the management of health records, designated a “de-identification of health information” standard (Ohm 2010). Subsequent refinements such as converting birth dates to years and reducing zip codes to three digits were made standard, allowing doctors to share information without infringing on their patient’s privacy. For academics this should all seem familiar since these standards were further adopted under the Family Educational Rights and Privacy Act (FERPA) and are now widespread across research as a risk mitigation tactic to storing personal data. We will later return to touch upon other aspects of FIPS, laying the groundwork for the privacy policies used by online platforms.

While small adjustments have been made such as the zip code and birth year changes mentioned above, the PII framework remains standing against the trampling stampede of Big Data innovations. As the acceleration of available data is growing and the locations where it’s being collected more ubiquitous, the primary protection of our data amounts to specific columns of a CSV file being removed or obfuscated. Given this backdrop of data regulation, we can insert it into the context of HCI to grasp how these protections fit into the views espoused by the community of interaction researchers.

Contextual Privacy and HCI

Unlike lawyers who concern themselves with liability of risk when handling data, HCI researchers emphasize and explore varying dimensions of how humans might interact with data representations through computer systems. Thus, when it comes to privacy, major questions involve “what context does the user believe a particular action to be happening in?” and “how is that context communicated via the structure of the interaction?” An example HCI concern involving privacy may be the use of automated notification systems. A pop-up notification on a screen may raise privacy violations for the person receiving the message, the person sending the message, or both. Importantly, the context matters since the pop-up is fully private if alone at a desk, but may cause professional damage when giving a public talk.

Paul Dourish describes the two primary HCI concerns of context as:

> The first is mutual relationship between physical form and activity; how we can design computationally-enhanced devices and how their form as much as their interactive ability affects likely patterns of action and interaction...The second concern...is how computation can be made sensitive and responsive to the setting in which it is harnessed and used. (Dourish 2004)

If our goal is to design systems that cooperate with a user's needs and enhances their abilities while performing some contextual action, we must be aware of the norms carried in a particular context of action. In terms of privacy, we are *required* to design interactions that safeguard users from harm (ie, we must follow FERPA) and *should* design with contextual norms and expectations in mind (ie, we should hide notification content when a computer is in presentation mode). This means health monitoring, social network messaging, and email clients all face separate challenges to protecting privacy.

The problem of interpreting context is very challenging in and of itself. How are designers supposed to know exactly when and how a user will attempt to use their system? How can we adopt privacy standards that actually fit an uncertain context? Germane to these questions, Palen and Dourish have come up with a list of "boundaries" designers can use as guides in considering how a person might manage their privacy in a particular context (Palen and Dourish 2003):

1. Disclosure: what information may be disclosed through an action and under what circumstances?
2. Identity: how is identity displayed and maintained for each party during a technology-mediated interaction?
3. Temporality: how may the action be interpreted across the past, present, and future?

Leveraging the these considerations, HCI designers are meant to configure interactions that offer control and limit risk when confronted by these boundaries. Helen Nissenbaum applies similar points from a legal vantage. Arguing that universal privacy principles will cause too much disagreement, she ends up at a similar conclusion to Palen and Dourish: no principle is universal since cultures, opinions, and circumstances vary (Nissenbaum, 2004). Or in the original HCI terms, boundaries are fluid across contexts. With this, she comes up with a tenet she calls "contextual integrity." Meaning privacy policies should seek, "compatibility with presiding norms of information appropriateness and distribution" (Nissenbaum 2004). Nissenbaum attempts to combine our legal history of privacy with the insights of HCI researchers, suggesting a reading of the law that equates *privacy harm* with *violation of reasonable expectations of context*. Implementing such a regime would require its own systematic (and debatable) categorization of contexts and agreed-upon standards. While these standards may bring an improvement to our current situation, by no means should we expect them to settle the issue given the wide disparities that exist culturally and individually (Wang, et al. 2011).

Unfortunately for users of computer systems, no matter how much work goes forward pursuing contextual privacy by design, as it stands today, most information is immediately captured and prepared for uses not necessarily pertaining to the context of creation. The boundaries Palen and Dourish wish to interpret no longer take on knowable forms as disclosure and temporality remain uncertain at the moment of interaction. What we will go on to see is that neither PII protection and anonymization nor contextual privacy considerations are easily tractable in the world of Big Data.

Ubiquitous Data Capture

As things stand in 2016, data is being collected on consumers asynchronously, autonomously, and constantly. Of course, studying consumers is not a new phenomenon in business. Retailers have long shared information about who their customers are through subscribers lists and purchase histories (Larson 1994). What is different now is that data exchange is not isolated to conscious, isolated transactions. While typing this sentence, drafting this essay in a Google Document, Google is storing each revision made to the document. Most modern websites take part in behavioral tracking, social media networks convert our relationships into massive datasets, and mobile phones offer location-based services to add fine-grained geospatial details into the capture of online actions.

Consider a group of researchers' findings while testing for the kinds and quantities of data shared through popular Android and iOS mobile applications. Of 110 apps, 73% of Android apps shared personal information such as email addresses with third parties and 47% of iOS apps shared geo-coordinates with third parties (Zang, et al. 2015). Do all these asynchronous updates really correspond with the expectations of the context? Beyond the growth of data capture, the size of a payload has grown as well. Latanya Sweeney details the augmentation of person-specific information at a number of common locations such as the grocery store, during a hospital visit, and at birth . A particularly enlightening passage:

> ...a consumer in 1983 could purchase items from a supermarket and the only recorded evidence left behind were roughly an inventory debit and a record of the total amount purchased and the amount of tax paid. There was no knowledge necessarily of the identity of the consumer or of the consumer's personal habits and behaviors in terms of goods typically purchased and the times and days of the consumer's shopping experiences...Nowadays consumer transactions can be stored and analyzed, and by doing so, information about each consumer's lifestyle, behavior, beliefs and habits can usually be revealed. (Sweeney 2001)

She calls the new policies technologists are adopting with data "collect more," "collect specifically," and "collect it if you can." Let's remind ourselves that these remarks by Sweeney are focused on *explicit* data transactions. Moments where an individual actually recognizes that an information exchange is occurring. Online machinations for behavior tracking abide by the same collection logic, yet often the person involved has no idea what kind of data is leaking out from them. A description of an online advertising interaction might look like:

> When a user navigates to a publisher who contracts with an ad network, the ad server simultaneously transmits an ad, looks-up the ad network's cookie in the user's browser, and logs certain information about that user's activity in a database. (Barocas and Nissenbaum 2009)

What this means is that online moments always have the potential to be more than they seem. Whether it's Facebook's "Like Button" lingering around on a page or an imperceivable pixel tag[5],

[5] These are 1x1 pixel images embedded in pages specifically to act as a tracker. Websites usually load pictures with an HTML <img> element that has a particular *src* attribute defining a URL where the image lives. Pixel tags exploit the automated call to this URL to identify your browser to a third-party tracker purchasing user data from the first-party website being visited.

the moment your browser parses an HTML file asynchronous callbacks are being sent to observing parties all across the internet.

Why is it that online services are so interested in squeezing every last bit of information out of us? Often it is their business model: they offer free content, they get paid by serving you ads. Advertisers only want to pay if the ads actually target a user likely to click, making behavioral profiling crucial to revenue generation. Other times the data is used to personalize your experience on the website. Collecting information about you may allow for better recommendations directly to you or make more robust inferences about people associated with you in your social graph. Most important to engineers, data is a powerful and necessary tool for training new models using machine learning or AI techniques. The more and better data you have, the higher likelihood your model will be useful (and profitable).

Establishing that data is being collected more frequently and in larger quantities does not immediately explain why these practices are harmful to privacy. If the data is handled correctly, why should we believe any harm is enacted? Personalization can improve computer interactions, so why not try to make better online experiences? What we now move on to discuss is how vast quantities of data makes de-anonymization and re-identification a much easier task and how personalization comes with a disclosure trade-off riddled with harmful potentials.

De-contextualized Action, Identification, and Personalization

Before we address these new privacy issues, let's recap. Earlier we discussed how the legal framework for privacy protection focuses on PII where specific information that could cause harm must be anonymized or left out before sharing. Further, we saw that in HCI privacy is considered an issue of context where we look to protect norms involved in particular uses of technology.

The first thing to note pertains to context. Looking to HCI theorists Jonathan Grudin and Phil Agre, we find early warnings that contexts presented to users are slowly slipping away from their actions. As Grudin has discussed, the moment that information is put online, the context of "here and now" fades away (Grudin 2001). The possibilities of how that information may be used or interpreted goes far beyond what can be imagined in that moment. Phil Agre, on the other hand, discusses how we reconstitute systems into particular "grammars of action" to support the desired data capture from the system (Agre 2001). He presciently warns us that actions will likely take on looser couplings with their built environments due to the separate concerns of data capture and human activity. Merging these two ideas with the above discussion about online data tracking, we see that the practices of Big Data has completed the arc of dislodging our data from context. Not only do our online actions go far beyond their original intentions as Grudin predicted, but online environments are now built to allow for data capture that is not congruent with the presented context. Bringing back Paul Dourish's first concern of context in HCI regarding the mutual relationship between form and activity, we should recognize that Agre's arguments had already pierced through the heart of his framework. The need for data capture is hegemonic on the logic of systems design. Respecting norms of context has been relegated to superficial concern of appearance. Unless we somehow reel in the now-common practices of capture used by online platforms, a serious shadow of doubt is cast on the ability for contextual privacy to save us from violations possible in the Big Data Era.

Putting theoretical notions aside, let's take a step back to our legal protections. One may ask, "Even if droves of data are being collected, the really private information that could cause negative impact, that's secure, right?" Or, "Even if a bunch of decontextual data is piling up on servers somewhere, isn't it kept anonymous?" While the prima facie answer to these questions is "yes", this doesn't quite salvage our privacy concerns.

We first must recognize that the anonymity criteria involves practices like deleting or obfuscating data entries that contain social security numbers, last names, home addresses, dates of birth, etc. What makes these entries important is that they allow for *unique identification* of a person to whom they relate. While it's very easy to identify someone if you already have their name or address, these are not the only pieces of information that uniquely identify a person. This fact leads us to a practice called *re-identification attacks*:

> These attacks depend on a variety of methods: overlaying an anonymized dataset with a separate dataset that includes identifying information, looking for areas of overlap (commonly described as a linkage attack) or performing a sequence of queries on an anonymized dataset that allow the attacker to deduce that a specific person must be in the dataset because only one person has all of the queried attributes (differencing attack). (Barocas and Nissenbaum 2014)

Computer science researchers such as Arvind Narayanan have proven and formalized methods for de-anonymization of large datasets (Narayanan 2008). What this amounts to is any anonymized dataset having the potential to be de-anonymized by an attacker who has the right complementary dataset. And with the vast quantities of data anyone can get through the Facebook or Twitter APIs, it does not take long to legally and easily amass plenty of information to perform such an attack. In *Broken Promises of Privacy,* Paul Ohm argues that we have to transition out of a regulatory scheme that relies on PII as a central component (Ohm 2010). Looking at the same evidence, he shows us that anonymization is no longer a way to protect citizens from privacy harms. A telling anecdote comes from Barocas and Nissenbaum who say an engineer at Google claimed, "We don't want the name. The name is noise" (Barocas and Nissenbaum 2014). And this claim should come as no surprise. Names are not all that unique; much less so than your browsing history.

Volumes of information can pragmatically be considered volumes of inferences. And it's the inferences that matter. Because it is *very hard* to predict what will later be inferred from any body of information, the constant accruing of online data only exacerbates privacy concerns. The power of inference goes beyond violations of privacy. A data set of Twitter Tweets, once conversations about daily life, may later become a tool for discovering rates of heart disease.[6] It may appease researchers to say these findings are advancements to science and humanity, but that may not satisfy folks on the ground if insurance premiums go up in the locations determined to be at higher risk. As we will discuss at the end of this essay, the worrisome ability to de-anonymize datasets is leading to lots of new research around differential privacy,

[6] Researchers have actually shown validity in this measurement: https://www.sciencedaily.com/releases/2015/01/150121114913.htm

probabilistic programming, and client-side information storage. However, before moving on, there is one last privacy concern that big data uncovers even if you do not actually have the data: personalization.

The act of personalizing experience does not necessitate actually owning data on the specific user. Instead it could mean using contextual information to perform algorithmic methods like collaborative filtering or Bayesian classification such that you merely tailor content to the user. In many circumstances, personalization can lead to very positive experiences for end-users. Though, it is crucial to note that it also comes with certain risks. Recommendations often imply something about past behavior. Further, categorizing groups of people into clusters or groupings reveals information through association. These practices can often lead to moments of embarrassment and have the ability to compromise privacy. Take for example Facebook's "Beacon" advertising program. The program allowed activities such as purchasing a product or adding a product to a wish list to be published to users' friends' feeds. After plenty of negative media attention and a class action lawsuit, the program had to be shut down (Toch et al. 2012). And these stories do not end here, Google's "Buzz" and Facebook's follow-up ad program "Instant Personalization" both had to be shut down due to privacy complaints. What this means for HCI designers is not that personalization cannot happen, but that it's direly important to consider potential information disclosure and inference as a consequence of a personalized interaction.

What the above has shown is that our new data-driven society has complicated previously trusted notions of privacy. Both protected information and contextual integrity are challenged by Big Data practices. These concerns have many implications for the future design and implementation of computer systems. They also mean there is likely to be new legislation upcoming to try and minimize potential harm as these threats grow. We will discuss some of the more promising potentials at the end of this essay. For now, questions left dangling are how and when can we share data and retain privacy protection for our users? How might we impose limitations or make more transparent how data capture operates in certain contexts? What controls should a user have on future transmissions of their personal data?

III. Discrimination: Reorienting the Problem to the Machine

Much like a human who learns from experience, algorithms for learning may inherit unwanted bias coloring future decisions. One of the primary reasons for engineers to deploy data mining operations is to build statistical models for discrimination. A primary use of machine learning is to codify rationality into a model that can meaningfully distinguish between users and identify the features that make them statistically similar or different. Statistics is a science of distributions and data mining is a tactic to characterize a population distribution and make inferences based on prior information. So what happens when training data contains hidden bias based on gender, race, or class? Or what do we do when a model's recommendations differ across lines of identity?

To elucidate this point, let's roughly sketch a sample machine learning application.[7] Assuming we are using some training data, and thus are performing a *supervised learning* task, the goal "is to learn a mapping from inputs *x* to outputs *y* given a labeled set of input-outputs *D*" (Murphy 2012). D is our training set, the data we are using to "teach" our system about its task. Now for each input, *x*, we have a number of dimensions, or *features*, which are the attributes of the data (e.g., height, weight, eye color, age) we believe correspond to our output. For the learning procedure we will need to define a *cost function* and a *target variable*. The target variable is the output we want our system to learn about and the cost function is how we evaluate how right or wrong the system is while training it. A possible application could be a system designed to identify good job candidates for a company. In this case, our input values would be gathered from an applicant's CV (e.g., age, university, prior employers, years of employment) and the system would output a score between 0 and 1 (i.e., 0 is a poor-fit candidate and 1 is a well-fit candidate). The learning process will involve us using past hiring decisions and training our model to correctly identify the past candidates that ended up being good employees. Seems like a great way to cull out a few good applications from the flood that applies for a desirable position.

In a field like engineering, dominated by white men, an immediate problem we may run into is that the model seems to only rank white men as well-fit candidates. This is a serious issue because in the US we have equal opportunity employment laws that protect certain classes of people from being discriminated on the basis of their identity.[8] However, unlike the former days of an HR representative or a recruiter manually filing through applications, now an approximated mathematical function or a network of numerical arrays has computed a decision. And it is here we see how data has become a new issue for civil rights in our society. Machines are now in a position to enact the same kind of illegal and unethical decisions humans might make.

Legal Protections

Ahead of enumerating the different ways machines can discriminate, let's briefly cover some background on legal protections we have in the US. A summary of the laws applicable to machine discrimination was covered by the FTC in their recent report *Big Data: A Tool for Inclusion or Exclusion* (FTC 2016). They listed:

1. Fair Credit Reporting Act (FCRA)
2. Equal Opportunity Laws such as the Equal Credit Opportunity Act (ECOA), Title VII of the Civil Rights Act (Title VII), the Americans with Disabilities Act (ADA), the Age Discrimination in Employment Act (ADEA), the Fair Housing Act (FHA), and the Genetic Information Nondiscrimination Act (GINA).
3. Section 5 of the Federal Trade Commision Act (Section 5)

[7] For a more detailed treatment of machine learning procedures, see my article *How Do Neural Networks Learn*: http://blog.fastforwardlabs.com/post/129793362663/how-do-neural-networks-learn

[8] Not to mention the personal, humane ethics of striving for equality across all kinds of people.

These laws were put in place due to a long history of civil rights movements that have attempted to equal the playing field in terms of opportunity and treatment in the United States. The FCRA applies to reporting agencies that compile and sell consumer reports. These reports are used as a basis for determinations regarding credit, employment, insurance, housing, etc. FCRA ensures reasonable procedures are in place that maximize accuracy of the reports and allow customers to access their own information and correct errors. Already two data aggregators *Spokeo* and *Instant Checkmate* have been successfully prosecuted under FCRA for discriminatory data practices that lack FCRA compliance (FTC 2016).

Across equal opportunity laws, we find explicit protections for discrimination on the basis of race, color, sex or gender, religion, age, disability status, national origin, marital status, and genetic information. Using Title VII as an example, we find liability for discrimination applied under two primary strains: disparate treatment and disparate impact (Barocas and Selbst 2014). Disparate treatment is either proven intent to discriminate or evidence of formal disparate treatment of similarly situated people along lines of identity. Disparate impact applies to policies that prima facie appear neutral, but in practice have an adverse impact on a protected class of individuals. These two approaches to litigation under Title VII have historically proven to be difficult.[9] Case histories have slowly moved the burden of proof onto the plaintiff. Explicit intent to discriminate must be shown or, in the case of proven discrimination, if the discriminatory policy is seen as a “business necessity” the plaintiff must show an alternative policy exists that achieves the same results sans discrimination. In light of big data systems, another layer of complexity has been added since discriminatory action is dislocated from any human and can be codified in an algorithm.

Finally, Section 5 is in place to prohibit unfair or deceptive acts or practices in or affecting commerce.[10] Section 5 is a blanket regulation applicable across all market sectors and applies to most companies. This is a consumer protection that prevents companies from making statements, designing ad campaigns, or omitting information that may mislead a consumer acting reasonably. More than protecting against predatory practices, a further protection offered by Section 5 is the sale of consumer data to customers that a company knows or has reason to believe will use the data for fraudulent purposes. The FTC has already prosecuted multiple companies that collect consumer data under this principle.[11]

The scope of these protections amounts to a serious and encompassing regulatory system to protect citizens from unfair discrimination. Knowing these will help us investigate how data and common data practices have brought about new ethical challenges to discrimination.

Data as a Tool for Discrimination

[9] Look to *Wards Cove Packing Co. v. Antonio* following which acceptable business justifications for discriminatory practice loosened and burdens of proof moved to the plaintiff to first show that the practice in questions was not necessary.

[10] 15 U.S.C. § 45(a)(1) (2012).

[11] Cf. *FTC v. Sequoia One, LLC* and *United States v. ChoicePoint, Inc* for example cases won under Section 5.

ProPublica has recently begun a spotlight investigation under the title “Machine Bias.” One of their central stories follows the the use of a risk assessment software called COMPAS (Angwin et al. 2016). The software is supposed to help determine whether a defendant is eligible for probation or treatment programs; however, they uncover several examples of judges using these scores in their sentencing decisions. In several cases, black defendants with lesser charges from prior offenses received higher risk scores than their white counterparts charged with the same crime. After hearing a few of these stories, the reporters did an analysis of judicial decisions in a county actively using the software, Broward County, Florida. They found that from the population of criminals who did not commit a subsequent offense 44.9% of African Americans were still labeled high risk in comparison to 23.5% among whites. And of those that were labeled low risk they found 28% of African Americans did re-offend against 47.7% of whites.

What makes this account even more frustrating is that COMPAS is a proprietary software meaning they do not have to disclose their methods. The potential for harm speaks for itself. There’s an uninterpretable algorithm that uses undisclosed information from people’s background to decide how likely they are to commit crimes and it just so happens it tends to more often deem people of color higher risk. Multiplying this concern is that it’s very likely the engineers who designed this system had no malintent. Given the inequality in prosecuting and incarcerating African American’s in America, almost any training methodology using historical data is likely to find correlations between being colored and being a criminal. And this is why former US Attorney General Eric Holder (United States Department of Justice 2014), the FTC (FTC 2016), and legal scholars (Harcourt 2008) are all warning us of the potential harms data discrimination may have on our justice system.

In *Big Data’s Disparate Impact*, Solon Barocas and Andrew Selbst compile a list of engineering choices that may lead, in practice, to a discriminatory system. Though not exhaustive, this list is an important start for engineers or lawyers needing to audit data practices (Barocas and Selbst 2014):

1. *Training data* that either over- or under-represents some class of individuals.
2. Certain ways of *labeling data* can inadvertently or advertently cause your classification basis to be discriminatory.
3. *How you collect your data* may skew toward certain populations by making assumptions about technological access or preferring particular behaviors or locations.
4. *Feature selection* while determining the input variables for your system may insert bias by limiting what information the algorithm receives.
5. *Proxy variables* that highly correspond to identity (e.g., school district with race) may end up approximating protected categories while not explicitly used.
6. *Masking* can happen where someone who intends to discriminate can be removed from exposure by choosing a method that is likely to be bias.

Taking this list seriously, there are potential pitfalls throughout the entire design process for any machine learning system that will make decisions of substantive consequence. What makes this a particularly hard problem is the double-edged sword inherent in any attempt to regulate these actions. Imposing a liability on engineers would disincentivise exploring the field of machine learning and make systems design a quagmire given how difficult it is to interpret these risks. At the same time, being lax about this problem could make hard-fought civil rights protections toothless whenever algorithms are involved in decision-making.

As intelligent systems are deployed more widely, the problem of identifying discrimination becomes more recalcitrant. Discrimination has already reared its ugly head in several other types of online interactions. Through a blackbox analysis with Google's ad system researchers at Carnegie Mellon found that being female makes it less likely to be served ads for high paying jobs (Datta et al. 2015). In the same vein, Latanya Sweeney further found Google's AdSense to be much more likely to serve ads suggestive of arrest records and criminal history when performing searches on traditionally African American names (Sweeney 2013). Advertisements may not be deemed as important as financial, criminal, and employment decisions, but they often act as access points to services and carry along cultural information that is taken up by society.

At the moment, many other research efforts are beginning to assess the extent of discrimination in other important areas. Voter microtargeting has already burst onto the scene as a contentious issue (Barocas 2012; Hillygus and Shields 2008; Nielsen 2012; Reston 2012). Already well established is that political campaigns are data machines that do everything in their power to get fine-grain information about their targeted demographics. Whether it's polling phone calls or online ads, parties and politicians work hard to change the minds of voters by any means necessary. An interesting project underway out of the University of Wisconsin-Madison is *Project Data*.[12] They install browser plugins that keeps track of political ads targeting you in hopes of getting a better understanding of how persuasion ads are being delivered, who pays for them, and who receives them.

Price discrimination has also put into question the nature of consumer rights (Alessandro 2006). Having a lot of information about your customers means you can make inferences about how much value a person places on an item at a given time. The demand for such knowledge is so high that Google even has a patent on a particular method for dynamic pricing.[13] Perhaps useful for companies, this practice can be hugely unfair for consumers. Already price discrimination has been observed, but it is still difficult to say on what grounds the discrimination is occurring (Mikians et al. 2012). Since certain kinds of price discrimination are legal, such as adapting prices to current demands like commonly found purchasing airline tickets, there is a significant burden on the consumer to determine when and if unfair discrimination is happening. Critically, we still are searching for what exactly one's rights are once the unfair discrimination is suspected.

This threat to basic civil and consumer rights has caused certain advocates like Kate Crawford to demand a right to procedural data due process (Crawford and Schultz 2013). As opposed to regulating on categories of personal data, due process would give a positive right to recourse given certain determinations made by algorithms. Any adjudicative process where Big Data plays a role in determining attributes or categories of the individual would be open to be

[12] http://www.eyesonelections.com

[13] https://www.google.com/patents/US20080154798

contested in court. A right like this would be a huge step forward in ensuring individuals like the ones described in the beginning of the section at least had an opportunity to question the decision made by an algorithm. Big Data is on the advent of mediating our insurance premiums, mortgage rates, job and college decisions, and police interactions, yet we do not have defined modes of legal recourse for someone suspicious that they have been harmed by an algorithm. Issues like this seem destined for major legal determinations in our near future. As outlined above, there's a long history of government intervention on issues of discrimination. Therefore, it is simply a matter of time until a precedent concerning algorithms is set. Engineers joining in with the likes of Crawford to come up with best practices and regulatory suggestions will dampen the likelihood of that decision coming out of ignorance and thus having lasting negative consequences for everyone.

Summarizing questions to be answered: Should certain data be restricted from use within particular systems? What kind of transparency standards could allow citizens and juries to make sense of fair and unfair discrimination? Can data scientists come up with techniques to audit datasets and algorithms?

IV. Online Research, Consent, and User Attitudes

Ethical principles for human subject research first became widely accepted following the Nuremberg War Crime Trials. The Nuremberg Code established standards to judge the work done by physicians and scientists on humans in concentration camps during World War II. Subsequently, these rules would begin framing a long-term effort to protect participants involved in research. Similar efforts have been made to shape norms protecting consumers. Such is the case with Section 5 (discussed above), the Gramm-Leach-Bliley Act which obliges companies that offer financial products to explain their information-sharing practices, and the FTC's self-regulation principles for online behavioral advertising (Federal Trade Commission 2009). Central to both research ethics and consumer protections are the concepts of *notice* and *consent*. Specifically, 'notice' pushes organizations who use or collect information to explain their practices so that consumers can 'consent' by making a meaningful choice of whether or not to participate (ie, opt-in/out protocols).

The norms around notice and consent are manifest in the form of IRB protocol for researchers and Terms of Service agreements and Privacy Policies for commercial actors. However, as Big Data practices continue to enter into more aspects of life and society these standards have come under scrutiny. As research using online data continues to grow, questions around consent have started to surface. What does it mean to do research on human subjects who do not know they are being researched? How do we effectively communicate opt-in and opt-out choices to massive online subject pools? Is it even fair to use data that was never meant for research purposes?

In the corporate world there are fewer standards which has led to public outrage regarding some recent work proving that private companies are experimenting on their customers without notice. Moreover, Privacy Policies and Terms of Service are often nothing more than click-through agreements completed without the consumer batting an eyelash. And so the question emerges, “As consumer data collection becomes a bigger part of online life, what does it look like to provide meaningful consent to consumers?” Perhaps even more concerning is how do we actually relay the information in a way that users understand the consequences of their consent?

This section will be dedicated to understanding the regulations and practices currently at work regarding consent, how Big Data practices are challenging those norms, and what we have learned so far about how users actually feel about these systems.

The Belmont Report and the Problem of Online Research

In the aftermath of horrible injustices committed by researchers on vulnerable communities such as in the Tuskegee Syphilis Study, necessary provisions were created to grant protections to participants of research studies. The Belmont Report (1979) initially laid out the guiding ethical principles for human-subject research and Common Rule legislation in 1991 explicitly regulated them through Title 45 of the Code of Federal Regulations (US Department of Health and Human Services 2009). The results of these government interventions have been crystalised in the form of the Institutional Review Board (IRB). These boards, in place within academic institutions across the country, operate as oversight committees to review, approve, and require modifications to research activities applicable under the Common Rule. The ethical principles stated in the Belmont Report and applied by IRB committees are “respect for persons,” “beneficence,” and “justice.” Important to our discussion is how the IRB has traditionally interpreted “respect for persons” through means of *informed consent*.

Before the advent of large-scale online data collection tactics, informed consent was fairly straight forward. Not to say disagreements and difficulties did not emerge, but mostly informed consent was about being sure to clearly assess and state the possible risks of taking part in a study and then communicating those to your subjects. The result is usually a lengthy form that each subject reads and signs giving the researcher legal rights to perform the intervention and keep the resulting data. In light of new availability of public data, however, this traditional method of consent is proving to be insufficient.

Nowadays more people are participating in online spaces such as social media networks which offer relatively easy ways to collect large amounts of data through APIs. This novel infrastructure is accelerating possibilities of online research through social media, online forums, trace ethnographies, text mining, and activity monitoring (Vitak 2016). Meanwhile, as online research efforts are ramping up, concerns of the new risks coming to light from Big Data are beginning to stimulate fresh ethical conversations about how to adapt new principles (Vitak 2016; Shilton 2012; Zimmer 2010). Informed consent with online research has largely gone out the window due to practical difficulties and a lacking imagination of possible risks. That is, most online research is deemed exempt by the IRB. When collecting information from thousands or even millions of Twitter accounts, how does one meaningfully communicate research intentions? Unlike highly controlled experiments in lab settings, online subjects enter into research on an asymmetric footing. They are there to socialize, learn, play games, and generally do the activities of their daily lives while researchers often only enter after the fact with separate interests unknown to the participants. The first headline case of online research gone wrong involved a 2008 study called *Tastes, Ties, and Time*. Within days of the project's first data release, which included information scraped from 1700 Facebook profiles, properties of the dataset were identified (Zimmer 2010). As days went by more aspects of the dataset were de-anonymized until finally the researchers had to retract the data.

Pertinent to the prior discussion about the robustness of anonymization, what these researchers learned was the extreme difficulty in removing all unique identifiers from a dataset. Since the time of this mishap, work has been done to get a handle on where the Belmont Report fails to engage with extant problems in online research. Jessica Vitak (2016) has begun to study the ethical norms held by people in different disciplines regarding online research. She found lacking shared principles across research communities and a general belief from academics that they were held to higher standards than industry researchers.[14] Further, she has pointed out that we are in dire need for new creative means of transparency and encouraged continued ethical conversations between online researchers.

While this academic discussion continues, industry has begun to take advantage of their large datasets. In 2014, Facebook released their results on the now-famous Contagion Study (Kramer et al. 2014). Modifying users' social feeds, where users receive updates of recent activities across their friend network, Facebook researchers wanted to see if positive and negative emotions were contagious. By selectively showing users messages with more positive or negative sentiment, they studied subsequent posts to see if the content invoked similar emotions in the user. With no IRB board, Facebook deemed the study ethical on their own terms. And, given that they own the data, there was little that could have been done to stop them. After the public erupted in anger, OKCupid came out in solidarity with Facebook admitting that they too do experiments on their users (Rudder 2014). It turns out, behind the scenes, the popular dating site was conducting a number of experiments including suggesting people as matches who actually were not (based on their algorithm).

Straying from weighing in on whether the studies were actually harmful, the point is clear that our networked interactions are being toyed with by industry researchers who have no real oversight. The sheer acknowledgement of these studies forces us to reevaluate the place of the IRB and regulatory systems applicable to human-subject research. Now that online software products are sites of monitored and stored behavioral action, where do we attempt to draw the line on what situations require regulatory compliance? Here we must return to our notions of *notice* and *consent* as the bastions of ethical choice. The current solution to this problem depends on those lengthy terms of service agreements we all have to click through before registering for any online platform. And this brings us to our next topic: notice and consent enacted through online terms of service and privacy policies.

Privacy Policies, Terms of Services, and What Users Actually Think

[14] Although, interestingly, the one exceptional discipline where researchers did not express this difference between academic and industry standards was in computer science.

Whether in implicit or explicit terms, the moment one takes part in a networked interaction that interaction is governed by the policies of the hosting web domain or software owner. We call these delineations *Terms of Service* (TOS) and *Privacy Policies*.[15] Practices vary around where and when explicit consent is required, but it is common to require a user's conscious consent during registration for an online service or at the moment of installation for software. TOS outline the specific rights a user has in relation to the service and the special details of how the service is meant to operate. Sometimes privacy policies are included in a TOS, but they are usually separate declarations of the entity's information practices. The primary distinction here being the handling of information (privacy policy) versus the general rights of use (TOS).

Returning to a legal framework from our privacy deliberations, FIPS also sets practice guidelines for what must be provided to the user to fulfill notice and consent. Namely, FIPS demands users be, "given notice, that is to say informed who is collecting, what is being collected, how information is being used and shared, and whether information collection is voluntary or required" (Barocas and Nissenbaum 2014). There is no mention of specific rights that must be offered nor standardized language that must be upheld. Currently we are in the wild west of online contracts. Most regulators and practitioners keep their fingers crossed for a day where plain language and easy-to-understand policies are standardized.

Critical to this hope, Barocas and Nissenbaum (2014) discuss "The Transparency Paradox." They claim that simplicity and clarity is a trade-off with fidelity. Each piece of software or web platform has its own complexity that requires certain edge cases and caveats. This makes it extremely hard to settle on a single set of terms or practices that universally apply. Pushing us further, Barocas and Nissenbaum argue that the very idea of notice and consent is infinitely difficult to track in the online world. Developers are constantly trying out new features, business deals emerge, contracts expire, and all the while, the terms the user signed do not even apply to the third-party services shared with by the original provider. It's as if I tell you a secret, then have you sign a contract saying you can only tell Billy and Cindy today, since I know they are trustworthy, but don't bother with whom Billy and Cindy might tell nor if you decide to tell other people a few weeks from now. The chain of who touches someone's data and what they do with it is indeterminate and unpredictable.

Adding complexity to the issue, even if we could come up with some legal standards, there is mounting evidence that users don't understand these contracts, and when they do, they don't actually want what they are being offered. A number of researchers have shown that copyright implications of TOS are misunderstood (Fiesler 2016), privacy policies are too complicated for common users to read and digest (Luger et al. 2013; Jensen and Potts 2004), and many users do not even want the forms of personalization offered using their data (Anton, et al. 2010; Turow et al. 2005).

One study found that only 11% respondents could understand a text description of an opt-out cookie, a common mode of control offered to users against tracking (McDonald and Cranor 2010). The same study found only 20% of participants wanting the "benefits" of targeted

[15] It is also commonplace to have a *Data Policy* alongside or in place of a *Privacy Policy*, but for the purposes of this essay both types of notice will be discussed under the term *Privacy Policy*.

advertising. Google researchers submitted a study to CHI showing that users do not believe they are receiving extra benefits by allowing their data to be shared to third parties (Bilogrevic and Ortlieb 2016). A mere 23% of nearly a thousand participants in their study believed they received benefit if the first-party company uses their data and a dismal 6% thought it benefited them if the data was shared. Another researcher at Pomona College (Andrejevic 2014) found 59% of respondents to a survey about web tracking believed websites collect too much data. The same respondents widely agreed upon support for stronger do-not-track options (92%), a requirement to delete personal data (96%), and real-time notifications of tracking taking place (95%).

The contrasting facts of a) users not understanding what they are consenting to and b) when surveyed they often do not want what's being offered, is concerning. Our only real resolution of this seeming paradox is to say users should quit using the same services. While this pragmatic mindset functions logically, some researchers argue people feel powerless given the social pressures and lacking alternatives (Andrejevic 2014). The felt powerlessness seems realistic given that many privacy advocates have pushed for better do-not-track policies with minimal success. Of many legislative attempts[16] only California Assembly Bill AB 370 has gone through, which merely requires websites and services to disclose how they respond to a do not track signal (and only in the State of California).

It's unclear whether notice and consent can hold on as our ethical solution to online research and contractual user agreements. As has been suggested several times in this essay, data's use value is often unknown at the time of collection. This complicates the very idea of a one-time verification of consent when its ideas and commitments are ephemeral. For HCI researchers, this leads to a major challenge upcoming of how to improve user knowledge of what actions mean in networked systems. Legal experts, on the other hand, will need to ultimately decide what rights a person has while online. With these solutions still in limbo two central questions to focus on are: "What does it mean for someone to be informed enough to give meaningful consent?" and "How do we manage consent over longer time periods?"

V. Algorithmic Impact

Perhaps the most common way which a human is affected by data is through interactions with an algorithm. In a previous section, the problem of discrimination was taken up as a potential harm that can result from the determinations of an algorithm. Discrimination is a specific issue that fits snugly into a more comprehensive legal framework built long before our current techno-social lives. What we now take up is the multifaceted ways in which algorithms restructure aspects of our life and have long-term consequences on our society. Though the broad topic of algorithms could stretch wide, we will focus in specifically on the category of algorithms which are directly constructed from and adapted to data. Questions about when to use a greedy algorithm or dynamic programming solution are not of concern. Instead, we may

[16] The best summary of all Do Not Track Legislation attempts has been well documented by Wikipedia: https://en.wikipedia.org/wiki/Do_Not_Track_legislation

wonder whether there should be concern about the hidden and proprietary nature of Google's search algorithm? Or how might the use of PageRank as opposed to EdgeRank influence epistemology?

One way to describe the algorithms of interest would be "public relevance algorithms" (Gillespie 2014). These are computations "whose inputs are composed of our personal and collective activities, expressions, and preferences." Crucially, we must refocus our thinking from algorithms as technical means with a fixed purpose to society in favor of artifacts with their own morality and ideology. A simple example is to consider a search algorithm that preferentially indexes websites based on the number of known links to a page as opposed to the quantity of string matches within the page. The first assumes that if there are more links to a page, it is more important; whereas the second assumes that the page that uses your exact phrasing is likely to be more important to you. Widespread use of either will have tangible consequences to what information people access and therefore to what people know and believe. What this means is that we should not simply consider "the search engine" as an isolated phenomenon that solves a single category of problems. Rather, we should consider aspects such as how a particular choice of what data to use, what relationships between data are privileged, and what transparency mechanisms are offered to determine the ethical standing of a particular algorithm.

In this section, brief ethical considerations will be introduced to highlight the many locations data-driven algorithms are making their marks on us as individuals and communities and on our society as a whole.

Calculated Publics

A term beginning to be used by researchers such as Kate Crawford and Tarleton Gillespie is "calculated publics." As opposed to "networked publics" which are communities assembled by new mediums of networked interaction such as online forums and video games, "calculated publics" are those communities implied by groupings and suggestions made by algorithms (Gillespie 2014). Kate Crawford (2016) meditates on the meaning of searching for a book and being suggested a series of other books under the heading: "Customers who bought this also bought." There already is a semblance of a book-buying public that is constructed from something like The New York Times Best Sellers List. But unlike a bestsellers list, whose members we intuitively understand, what is the relationship between being derived between the people buying books on Amazon? Should we be reading the other books those like us read? Is anyone paying to make sure their book is connected to the book just bought?

Perhaps more impactful is a category of profile visibility users are able to choose from on Facebook: friends of friends. This choice raises questions of trust within your social network. Are my friends' friends likely to be good people? What kind of people do I accept as a friend on Facebook and do those people use the same discretion as me? Gillespie (2014) brings up a more concerning set of human relationships implied by Google's search algorithm when he searches the term "she invented" and Google asks "Did you mean *he invented*?" While Google did not hardcode that into their system, their learning algorithm was able to learn our culture's misogynist tendencies and is now imposing them back on us in a search recommendation. Similarly, advertising constructs its own communities through grouping people by taste and interest. As people wake up every morning and go online to read the news, many people originating on different websites will be suggested the same article or presented with the same click-bait headline. Using a payment website like Venmo will automatically begin showing you people's recent financial transactions who they believe are related to you.

Certain algorithms like collaborative filters or markov chains inherently associate people by saying "other people who have attributes similar to you liked this" or "other people in the same position as you did this next." Again, there is nothing wrong with these algorithms, but as our communities, friendships, and public groupings are shaped by them, we may ask questions about their appropriateness. Should algorithms that associate people use filters that ignore negative stereotypes? If we start accepting recommendations as objective, may we become vulnerable to unwelcome advertising embedded within? What choice should people have to disassociate from an algorithmic grouping?

Knowledge and the Structure of Information

Networked access to information has largely been celebrated as a triumph of the internet. Websites such as Wikipedia have removed major financial and physical barriers to knowledge being disseminated equally. While these advances should be seen in high regard, we must also ask whether there are risks associated with converting our investigative and reasoning skills into digital means. Moreover, as data stacks up, how are we meant to interpret all of it? How can data be manipulated to make arguments seem valid without actually being so?

The first example to consider comes from a recent study by the American Institute for Behavioral Research in Technology. They were interested in finding out to what extent biasing search results could change the opinions people formed. Further, they wanted to know if people would even notice that they were interacting with a biased search engine. After studying people in different global locations during real election cycles, they found across the board 20% of undecided voters were influenced by biased results with certain demographics being even more vulnerable (Epstein and Robertson 2015). Perhaps even more concerning, the presence of bias was entirely undetectable by their subjects. They have coined the term "search engine manipulation effect" (SEME) to describe the phenomenon. Considering realities such as the creation of The Groundwork, a Google-backed data operative working for the Clinton Campaign, results like these should warrant some pause.[17] Moreover, this should engender concerns about the consequences of the well-known "search bubble" or "search filter" problem. Now that most search engines look at your past searches, location, and profile information to determine what you are likely to want, it is becoming easier to only see the arguments and opinions you already hold.

Another recent happening that places into question how we trust data is the Volkswagon data scandal.[18] The EPA found that Volkswagen was manipulating data during emissions tests.

[17] Many news organizations have discussed this in a number of lights, but for an example see: http://www.bloomberg.com/politics/articles/2016-05-19/clinton-bets-on-tech-strategy-to-defeat-trump

[18] For a good overview of what happened and what it all means see: https://www.theguardian.com/business/2015/sep/22/volkswagen-scandal-q-and-a-emissions-scandal

By using physical characteristics of the car to automatically determine when the car was undergoing a test, the car could activate different emissions controls than what would be found on the road. Thus a basic data collection practice used for environmental protection has proven to be falsifiable so long as it's possible to figure out when it's happening. With auditing practices as a widely accepted method for oversight, this scandal should bring into question how it is we can validly audit any system using algorithms to tailor itself to environmental variables.

As if making sense of scientific findings was not hard enough already, now with the advent of Big Data, the possibility of *p-hacking* has become of heightened concerned (Aschwanden and King 2015). In traditional scientific investigation, one tests the validity of their hypothesis by conducting an experiment and seeing whether or not their intervention had measurable effect. A commonly used metric to determine if the treatment group was significantly different than the control group is a *p-value*. However, now with an easy accumulation of massive data sets, it has become possible to simply search for correlations that have significance then slap an explanation on it. Inversely, if a scientist has a theory they *want* to prove, they can collect lots of data and simply choose the subsection of it that contains the correlation best fit to their theory.

Some questions raised by these trends are: How do we assess the validity of a data source when it's far too large or complex for a human to manually go through? Should there be public standards on what can go into a search engine? Should we regulate relationships between online information providers and institutions such as political campaigns and research groups?

Inclusivity and Exclusivity

As already discussed, algorithms require some discerning choices in regard to what data matters to it and what does not. Once put into the context of different systems, we see that these choices often amount to ideological commitments. For instance, in the context of predictive policing, should my political views be relevant to a risk assessment of my criminal likelihood? Choices like this must be made in almost all data-driven systems. Take for instance a controversy that arose with Amazon a few years ago. Attempting to be family friendly, Amazon does not index adult books in their sales rankings and recommendations. This seemingly understandable practice led to quite the political statement when Amazon decided to categorize all LGBT-related books as adult and overnight removed their presence in sales rankings and recommendations (Gillespie 2014). More recently, there was a conspiracy theory floating around that Google was manipulating their autocomplete feature to hide scandals by presidential candidate Hillary Clinton.[19] However, Google responded that there was no scandal, but they simply do not allow offensive or disparaging autocompletions.[20] While the conspiracy theory is wrong, it is interesting to reflect on how the choice to remove certain terms effectively shifted the appearance of objectivity. It avails that Google had determined they have license to decide for us and our families which lexical relationships are fair, kind, just, etc.

[19] This video was the catalyst: https://www.youtube.com/watch?v=PFxFRqNmXKg&feature=youtu.be

[20] http://www.washingtontimes.com/news/2016/jun/10/google-denies-burying-bad-hillary-clinton-stories/

“Shadow bodies” has become a useful term to express the relationship our physical bodies has to the person described by our data traces (Gillespie 2014). Personalized features and anticipatory mechanisms embedded in systems cause software providers to encourage us to share as much as possible. Salient to this practice is what information is deemed relevant to a human profile and what is not. Keeping track of someone’s search queries and email tendencies may tell a very different story than knowing their actual favorite books and hobbies. The former may have traces of the latter, but could also be misleading. These choices of “what matters” are being made autonomously such that the consequences may be out of even the developers’ control. This was the case when Google recently found their image tagging software labeling African Americans as “gorillas.”[21] Unable to determine what features their image AI system had codified as important, they simply removed the tag from the system.

Concerns to reflect on due to these trends are: Should users be able to decide what categories of information are collected on them and for what purposes? How transparent do content providers need to be when making choices about what to show and what not to show? When algorithmic services hurt or embarrass someone, should anyone be held accountable?

Data as Power

In her provocative essay, *Can Algorithms Be Agnostic?* (2016) Kate Crawford raises a higher-order question: “How does the device or system generate a means for authority and/or power?” She claims that disagreement and dialectics are the heart of a healthy democracy. The concern is that as we replace systems of human deliberation with systems of algorithmic determination, will we lose critical sites for political struggle? The advent of data-driven systems may prove very useful when lost in a new city, but could they be doing us a disservice when they choose not to show us information on the protest happening 30 minutes from our doorstep? In relationship to these questions, ethicist Lucas Introna has argued for what he calls “disclosive ethics” (Introna 2007). Essentially he asks us to keep account of the moral implications of pragmatic and technical decisions, “at the level of code, algorithms, and the like--through to social practices, and ultimately to the production of particular social orders, rather than others.” His argument starts with a recognition that algorithms do not always disclose their presence or intentions to us. He distinguishes between transparent (e.g., a garage door opener) and opaque technologies (e.g., embedded face recognition software) to expound the need to keep track of technological deployments in our environment that otherwise may be difficult to see. Disclosive ethics essentially amounts to having genealogies of technological choices and their consequences such that we can maintain a transparent view of “how we got here” when we run into a particular problem.

Crucial to this essay’s discussion of data and the above ethical dilemmas is the recognition that *data is power*. Choosing to give a certain company or institution your data is an act of power changing hands. Data is the fodder of science, it’s the brick and mortar of machine

[21] http://blogs.wsj.com/digits/2015/07/01/google-mistakenly-tags-black-people-as-gorillas-showing-limits-of-algorithms/

learning models, and it has the ability to come back into our lives in complex and undetectable ways. In the mode of disclosive ethics, we can see from the history of companies who own or control a particular avenue of data capture, those companies *will* have a competitive advantage in the future market of related algorithms and systems. Taking this line of thinking a step further, the visions of the designers and engineers who take ownership over our data have the power to shape the future of our society. Someone may argue that anyone with a sharp vision and some skills may be able to do this. The rebuttal to this is to consider a problem like speech recognition: no matter how innovative an interface or algorithm is, it will be an uphill battle to compete with companies like Apple, Amazon, or Google who already own all the best training data.

It is in this spirit of reconsidering our rights and the rights of data, this essay moves ahead to its conclusion. The final section will discuss some of the promising avenues of research and engineering that are attempting to tame the chimeric animal that is Big Data.

VI. The Future of (Data) Ethics

The story painted by the above deliberations may make the future of Big Data appear grim. Portents that privacy as we know it may be dead, the reinvention of identity based discrimination, and the adaptive swarm of internet algorithms out for your attention--taken in isolation, look bleak. On the brighter side, these same methods may lead to major advances in medical diagnostics, easier and fairer exposure for small businesses and artists, and a reduction in bureaucratic overhead on time-consuming, perfunctory tasks. The future remains unknown. What is taken to be the case in this essay is that there is a partnered relationship between knowledge and control. The more we understand what is happening in this rapidly changing technological landscape, the better potential we can harness it for human betterment rather than wake up to a world for which no one feels they signed up. As engineers and designers conjure up new visions of how our world could be, those possibilities must be evaluated with careful, critical eyes. We may soon come across the day where a machine decides whether a bomb drops, a car swerves into oncoming traffic, or whether you should get that line of credit to start the business of your dreams. As the interrelationship between human and machines deepen, it’s crucial we forewarn ourselves of what risks are being borne by our users and society. Perhaps even more so when the innards of those machine “minds” are built out of bits of information we emit through our daily lives. Looking at the issue through this lens, the machines are mere approximations of us. The questions then become which approximation, whose machine, and for which purposes.

Before we conclude this wide examination of concerns raised by the Big Data Society, let’s briefly see what progress has already been made in thinking through these issues.

Privacy

As we explore ways to recover privacy protections being attenuated by new accessibility and capabilities of Big Data, several engineers have already pitched promising solutions. One solution to the privacy conundrum comes under the name of “differential privacy.” Proponents frame the problem as such: some trusted party wants to hold onto a dataset filled with sensitive information, but want to allow others to access statistical and global information about the data. Providing real aggregated statistics may lead to de-anonymization attacks and thus, it comes into question whether the data should be shared or stored at all. Differential privacy claims there is no need to use, or even have, the raw data in order to provide statistical information to third-parties. The proposal is to use select randomization algorithms that take a dataset as input and outputs a new dataset that retains the same statistical properties but differs in terms of single elements (Dwork and Roth 2013). In this way datasets can retain much of their practicality while not openly exposing information about the people being represented.

For engineers who want to create personalized systems, but do not want to bear the risks of owning abundances of user data, client-side storage may be an option. The idea is to store all the relevant data on the user’s system and only call it into RAM when a personalized feature requires it (Toch et al. 2012). This may not solve problems for systems where the goal is to train a model, but it could be used in the case of features that simply want to evaluate past queries or locations while not storing them externally. It also would allow the user to control the persistence of the information by having the choice to delete the memory associated with the application. In practice, engineers could still capture, but not persist data externally, allowing them to feed data to their models for training, but then only be able to evaluate their models upon user interaction.

A separate standard to consider is Latanya Sweeney’s (2002) *k-anonymity*. Her technique is to remove information until there are *k* entries that cannot be distinguished. In the case of a medical database, you could imagine if all entries regarding heart disease were reduced to patient’s sex and primary symptom, there would be a lot of overlap across the dataset. Sweeney adopts this idea as a method for analyzing when enough removal or obfuscation has occurred. What number *k* is may change with the size of the dataset, the context of the data release, or other factors.

Consent

Deriving a guarantee for informed consent may be the most challenging issue of all since it is predicated on the idea of a universal standard of knowledge for all users. However, a few recent research efforts have shown interesting results. One idea was to allow privacy policies to be socially annotated so that users who are more expert could share knowledge and elucidate concerns to privacy policies (Balestra et al. 2016). In a first exploration of the idea, researchers found users reporting a higher level of comfort after reading through annotated policies; though not necessarily a higher degree of competency with the ideas afterwards. Though still in its early-stages, we could imagine a space opening up where users can share information and debate relevant questions publicly in relation to privacy policies. Having these annotations could also be helpful in the event of a future court injunction that requires some historical evidence or in the event of a privacy policy changing and users’ wanting to compare.

Another group out of Carnegie Mellon has introduced the idea of the “privacy nutrition label” (Kelley et al. 2010). The essence of the idea is to transplant the standards the FDA put on food labels into the realm of privacy policies. Users would see standard evaluations of areas that may be concerning such as copyright and licenses, third-party sharing, and anonymity. While the idea is attractive, we as a society still have not determined the requisite language and standards to make it useful.

Legislation and Oversight

Slightly ahead of America, the EU has already begun implementing laws to protect users’ information online. The first landmark attempt has been the infamous “Right to be Forgotten.”[22] What this amounts to is a positive right for users to request that data be removed from the public internet and unlinked from search results. Requests are handled with discretion around how problematic the information is, whether removing it would adequately address the concern, and how much time has passed. By no means is this law perfect given this can both help and hurt: it may be bad that powerful people can effectively erase the internet’s memory of their misdeeds. Regardless, this is a nice experiment to see whether such an effort mitigates certain harms done to people by slanderous or embarrassing information being taken offline.

Separately, the EU has very recently passed legislation slated to go into effect in 2018, which gives citizens a “right to explanation” (Goodman and Flaxman 2016). This idea resembles Kate Crawford’s priorly-discussed notion of “data due diligence.” The law both places limitations on how autonomous or algorithmic systems can make decisions that “significantly affect” users and gives users the right to request an explanation of how the decision was made. It’s not clear yet how this will work in practice, but the idea definitely progresses us in terms of accountability under the threat of unjust discrimination by algorithms.

Most recently, a consortium of researchers from Harvard, MIT, and the University of Zurich have released an article outlining *Elements of a New Ethical Framework for Big Data Research* (Vayena et al. 2016). Many of their ideas have been touched upon in various places above, but it’s worth offering an overview since it may be the most comprehensive account out there of a way forward. Possibilities the authors propose are:

1. Universal Coverage: put all research, both industry and academic, under the same oversight regulation and create new participant-led boards to handle the overload this would put on IRBs.
2. Conceptual Clarity: enforce specific language to standardize terms such as “privacy,” “security,” “sensitivity,” etc and declare a formal method for revising these terms as technology changes.
3. Risk-Benefit Assessments: advise or enforce internal systematic risk assessment and train outside parties who can conduct reviews and create guidelines for review processes.

[22] The EU has released this very useful factsheet about the relevant case history and implications: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

4. Standardize Procedural and Technological Solutions: adopt a list of approved standards for technical and algorithmic implementations, privacy policies, etc so researchers have mandates on what's acceptable practice.
5. Tailored Oversight: journals and communities can establish required reviews that must be passed before publishing or granting. Tiered access to data can be mandated based on the sensitivity of the particular context.
6. Multistakeholder Processes: Setup boards, consortiums, and other conglomerations of people in expert domains and put them together with people who represent privacy and research interests to set recommendations or perform oversight.

Naturally none of these solutions offer a panacea for how to fix our ethical challenges with data. In fact, many of them would be quite hard to implement and oversee without risks to over-complication or nepotistic determinations. However, the suggestions create a starting point for people who are seeking solutions to these problems and may provide useful for future legislators responsible for making major decisions.

Where We Go Now

A lot of ground has been covered throughout this survey. Hopefully it provides both macro- and micro-viewpoints into the challenges our new Big Data Society has brought us. As a closing remark, I would like to offer up a few questions that should be seen as central to future research and work in this area.

1. What rights should an individual have to the data they produce?
    a. What would a meaningful opt-out choice look like?
    b. Should there be an auditable trail of who data has been shared with?
    c. Under what circumstances should data records be deleted?
2. Are their categories of decisions unfit for algorithmic aid?
    a. Should weaponized AI ever be allowed?
    b. Should humans ever be entirely out of the loop for decisions of significant impact?
    c. In areas with historical injustice, is it possible to train unbiased algorithms?
3. How do we communicate risks to users and the public?
    a. How do we communicate about data and technology to the Public's ability to make meaningful choices with technology?
    b. How do we communicate the long-term implications of a user sharing their data?
    c. Is there a fair way to inform the user about uncertainty?
4. What does a fair risk analysis look like for data-driven algorithms and systems?
    a. Should risk assessments be public for all consumer-related technologies?
    b. How do we think about risks that may be long term or not yet applicable?
    c. How do we bring users into the process of assessment?

Considering the nature of ethical discourse, we should never expect that these questions will be perfectly answered. The best we can hope for is an establishment of agreeable norms that create a relationship of trust between an individual offering their data and the entity taking ownership. With the growing ubiquity and commonality of data-driven systems, it seems likely these “data ethics” will slowly become indiscernible from the more general category of applied ethics. Hopefully with the growing need will come a growing interest by researchers, programmers, designers, and lawyers.

Bibliography

Acquisti, Alessandro. "Price Discrimination, Privacy Technologies, and User Acceptance." *CHI Workshop on Personalization and Privacy*. 2006.

Ananny, M. “Toward an Ethics of Algorithms: Convening, Observation, Probability, and Timeliness.” *Science, Technology & Human Values* 41.1 (2016): 93–117.

Angwin, Julia, et al. "Machine Bias: There’s software used across the country to predict future criminals. And it’s biased against blacks." *ProPublica, May* 23 (2016). https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing.

Andrejevic, Mark. “Big Data, Big Questions: The Big Data Divide.” *International Journal of Communication 8.0* (2014): 17.

Antón, Annie I., Julia B. Earp, and Jessica D. Young. "How internet users' privacy concerns have evolved since 2002." *IEEE Security & Privacy* 8.1 (2010): 21-27.

Aschwanden, C., and R. King. "Science isn’t broken." *August* 11 (2015): 2015. https://fivethirtyeight.com/features/science-isnt-broken/

Balestra, Martina, et al. "The Effect of Exposure to Social Annotation on Online Informed Consent Beliefs and Behavior." *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016.

Barbaro, Michael, and Tom Zeller Jr. “A Face Is Exposed for AOL Searcher No. 4417749.” The New York Times, August 9, 2006. http://www.nytimes.com/2006/08/09/technology/09aol.html.

Barocas, Solon. “The Price of Precision: Voter Microtargeting and Its Potential Harms to the Democratic Process.” *Proceedings of the First Edition Workshop on Politics, Elections and Data*. ACM, 2012. 31–36.

Barocas, Solon, and Helen Nissenbaum. "Big data's end run around anonymity and consent." *Privacy, big data, and the public good: Frameworks for Engagement* (2014): 44-75.

Crawford, K. "Can an Algorithm Be Agonistic? Ten Scenes from Life in Calculated Publics." *Science, Technology & Human Values* 41.1 (2016): 77–92.

Crawford, Kate, and Jason Schultz. "Big data and due process: Toward a framework to redress predictive privacy harms." *BCL Rev.* 55 (2014): 93.

Datta, Amit, Michael Carl Tschantz, and Anupam Datta. "Automated Experiments on Ad Privacy Settings." *Proceedings on Privacy Enhancing Technologies* 2015.1 (2015): 92–112. DeGruyter.

Davenport, Thomas H., and D. J. Patil. "Data Scientist: The Sexiest Job of the 21st Century." Harvard Business Review, October 1, 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century.

Dourish, Paul. "What We Talk about When We Talk about Context." *Personal and ubiquitous computing* 8.1 (2004): 19–30.

Dwork, Cynthia, and Aaron Roth. "The Algorithmic Foundations of Differential Privacy." *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2013): 211–407.

Federal Trade Commision. "Federal Trade Commission Staff Report: Self-Regulatory Principles For Online Behavioral Advertising" (2009) Retrieved from the Federal Trade Commission's Website: http://www1.ftc.gov/os/2009/02/P085400behavadreport.pdf

Fiesler, Casey, Cliff Lampe, and Amy S. Bruckman. "Reality and perception of copyright terms of service for online content creation." *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016.

Gillespie, Tarleton. "The Relevance of Algorithms." *Media technologies: Essays on communication, materiality, and society* (2014): 167.

Goodman, Bryce, and Seth Flaxman. "EU regulations on algorithmic decision-making and a" right to explanation"." *arXiv preprint arXiv:1606.08813* (2016).

Grudin, Jonathan. "Desituating action: Digital representation of context."Human-Computer Interaction 16.2 (2001): 269-286.

Harcourt, Bernard E. Against prediction: Profiling, policing, and punishing in an actuarial age. University of Chicago Press, 2008.

Hillygus, D. Sunshine, and Todd G. Shields. *The persuadable voter: Wedge issues in presidential campaigns*. Princeton University Press, 2014.

Introna, Lucas D. “Maintaining the Reversibility of Foldings: Making the Ethics (politics) of Information Technology Visible.” *Ethics and Information Technology* 9.1 (2007): 11–25.

Jensen, Carlos, and Colin Potts. "Privacy policies as decision-making tools: an evaluation of online privacy notices." *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. ACM, 2004.

Kelley, Patrick Gage et al. “Standardizing Privacy Notices: An Online Study of the Nutrition Label Approach.” Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2010. 1573–1582.

Kramer, Adam DI, Jamie E. Guillory, and Jeffrey T. Hancock. “Experimental Evidence of Massive-Scale Emotional Contagion through Social Networks.” *PNAS* 111.29 (2014): 10779.

Larson, Erik. The naked consumer: How our private lives become public commodities. H. Holt, 1992.

Levien, R. E., & Maron, M. E. (1967). A computer system for inference execution and data retrieval. Communications of the ACM, 10(11) (November 1967), 715–721.

Luger, Ewa, Stuart Moran, and Tom Rodden. "Consent for all: revealing the hidden complexity of terms and conditions." *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2013.

McDonald, Aleecia, and Lorrie Faith Cranor. "Beliefs and behaviors: Internet users' understanding of behavioral advertising." TPRC, 2010.

Mikians, Jakub, et al. "Detecting price and search discrimination on the internet." *Proceedings of the 11th ACM Workshop on Hot Topics in Networks*. ACM, 2012.

Murphy, Kevin P. Machine learning: a probabilistic perspective. MIT press, 2012.

Naranyanan, Arvind, Joanna Huey, and Edward Felten. “A Precautionary Approach to Big Data Privacy.” In Data Protection on the Move, 357–85. Springer Netherlands, 2016. http://randomwalker.info/publications/precautionary.pdf.

Nielsen, Rasmus Kleis. *Ground wars: Personalized communication in political campaigns*. Princeton University Press, 2012.

Prosser, Privacy. "48CALIF." L. REV 383 (1960): 389-92.

Reston, Maeve. “Voter Data Crucial to Romney's Victory.” *Los Angeles Times*. Jan 10, 2012. http://articles.latimes.com/2012/jan/10/nation/la-na-romney-analysis-20120111

Rudder, Christian. “We Experiment On Human Beings!” *OkTrends*, July 28, 2014. https://blog.okcupid.com/index.php/we-experiment-on-human-beings/

Shilton, Katie. “Participatory Personal Data: An Emerging Research Challenge for the Information Sciences.” *Journal of the American Society for Information Science and Technology* 63.10 (2012): 1905–1915.

Sweeney, Latanya. “Discrimination in Online Ad Delivery.” *Queue* 2013: 10.

Sweeney, Latanya. "k-anonymity: A model for protecting privacy."International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10.05 (2002): 557-570.

Sweeney, Latanya. "Simple demographics often identify people uniquely." *Health (San Francisco)* 671 (2000): 1-34.

US Department of Health and Human Services. "Code of federal regulations. Title 45 Public welfare. Department of Health and Human Services. Part 46: Protection of human subjects." (2009).

United States Department of Justice. “Attorney General Eric Holder Speaks at the National Association of Criminal Defense Lawyers 57th Annual Meeting and 13th State Criminal Justice Network Conference” (2014) Retrieved from the Department of Justice’s Website: https://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-association-criminal-defense-lawyers-57th

United States Department of Justice. “Overview of the Privacy Act of 1973” (2015) Retrieved from the Department of Justice’s Website: https://www.justice.gov/opcl/overview-privacy-act-1974-2015-edition

Vitak, Jessica, Katie Shilton, and Zahra Ashktorab. "Beyond the Belmont Principles: Ethical challenges, practices, and beliefs in the online data research community." *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 2016.

Vayena, Effy, et al. "Elements of a New Ethical Framework for Big Data Research." *Washington and Lee Law Review Online* 72.3 (2016): 420.

Wang, Yang, Gregory Norice, and Lorrie Faith Cranor. "Who is concerned about what? A study of American, Chinese and Indian users' privacy concerns on social network sites." *International Conference on Trust and Trustworthy Computing*. Springer Berlin Heidelberg, 2011.

Zang, Jinyan, et al. "Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps." *Proceeding of Technology Science* (2015).

Zimmer, Michael. "'But the Data Is Already Public': On the Ethics of Research in Facebook." Ethics and Information Technology 12.4 (2010): 313–325.