Communications of the Association for Information Systems

Volume 45 Article 26

12-2019

Addressing the Growing Need for Algorithmic Transparency

Hugh J. Watson
University of Georgia, hwatson@uga.edu

Conner Nations
University of Georgia

Recommended Citation

Watson, H. J., & Nations, C. (2019). Addressing the Growing Need for Algorithmic Transparency. Communications of the Association for Information Systems, 45, pp-pp. https://doi.org/10.17705/1CAIS.04526

Communications of the Association for Information Systems

Tutorial DOI: 10.17705/1CAIS.04526 ISSN: 1529-3181

Addressing the Growing Need for Algorithmic Transparency

Hugh J. Watson
Department of MIS
University of Georgia
hwatson@uga.edu

Conner Nations
Department of MIS
University of Georgia

Abstract:

Today, many organizations use personal data and algorithms for ads, recommendations, and decisions. However, some have expressed concern that this use negatively impacts individual privacy and poses a risk to individuals and society. In response, many have called for greater algorithmic transparency; that is, for organizations to be more public and open about their use of personal data and algorithms. To better understand algorithmic transparency, we reviewed the literature and interviewed 10 experts. We identified the factors that influence algorithmic transparency, the Association for Computing Machinery's principles for ensuring that one uses personal data and algorithms fairly, and recommendations for company best practices. We also speculate about how personal data and algorithms may be used in the future and suggest research opportunities.

Keywords: Algorithmic Transparency, Algorithms, Analytics, Personal Data, Models, Privacy, Best Practices.

This manuscript underwent editorial review. It was received 02/07/2019 and was with the authors for 2 months for 2 revisions. Young Jin Lee served as Associate Editor.

1 The Impacts of Personal Data and Algorithms

Algorithms have come to increasingly impact individuals, organizations, and society (Pasquale, 2015). They drive personal recommendations that influence what we buy, which movies we watch, and the social media networks we form. Other algorithms impact decisions about our lives, such as whether we get a car loan, receive an interview for a job, and even receive prison sentence rather than probation. A *Wall Street Journal* article suggests that an algorithm may be your next boss and will schedule vacations, determine work teams, and assign day-to-day tasks (Schechner, 2017).

Personal recommendations have a privacy/convenience tradeoff. When they can use more personal information, recommendations become more customized and helpful, but they compromise privacy. Companies such as Facebook, Google, and Twitter have a fundamental business model in which they apply algorithms to personal data, create targeted ads, and generate income. People vary in their reactions and acceptance of this tradeoff: some welcome the recommendations, while others dislike the intrusion into their privacy. On a societal level, this business model can cause concern since organizations can use the same personalization technology to create a dystopian society where they manipulate "the truth" and beliefs through targeted messages and newsfeeds (Tafekci, 2017).

A nuanced topic, people's willingness to share personal information includes considerations such as individual's attitudes about privacy, the benefits they receive, and how others will use the information. To illustrate, most people willingly share information so that an app can serve its intended, primary purpose but do not want the same information shared with unknown groups and companies for unknown purposes.

Algorithms that drive decisions about people's lives can facilitate decision making but potentially pose a risk. In her influential book, *Weapons of Math Destruction*, Cathy O'Neil argues that algorithms can increase inequality and even threaten democracy (O'Neil, 2016). Algorithms can amplify structural discrimination, produce errors that deny services to individuals, and influence election results (Diakoulus & Friedler, 2016). Researchers have increasingly recognized that the public should be concerned about the societal risks that algorithms pose and hold the companies that deploy them accountable (Smith, Patel & Munoz, 2016). Companies need to be especially careful when automating decisions that affect people's lives due to the potential negative consequences that errors can have.

The growing use of algorithms has led to an increased interest in algorithmic transparency (also referred to as algorithmic accountability); that is, organizations being more public and open about how they use algorithms. The interest pertains to more than just the algorithms themselves and includes the data that organizations collect, use, and share; how they develop and deploy algorithms; and the consequences of using them. Gartner has identified digital ethics and privacy as an issue that will have significant disruptive potential over the next five years (Hill, 2018).

While a need for increased algorithmic transparency exists, there are reasonable limits as to what information companies should be required to share. For example, someone who has been denied a loan should be entitled to an explanation and be given the opportunity to appeal the decision, but the company should not be required to compromise competitive and intellectual property rights (e.g., trade secrets) by fully disclosing the algorithmic details. On the other hand, the government should have the ability to determine that an algorithm does not violate the law, such as by using protected class data (e.g., race, gender). Also, the government should not always need to provide details about some of their algorithms, such as algorithms it uses for screening tax returns for possible audit and preventing terrorism or drug trafficking.

In order to better understand the issues involved with algorithmic transparency, we reviewed the academic, professional, and popular literature and interviewed 10 experts (i.e., managers, professionals, and consultants) (see Appendix). The experts actively participate in companies' predictive modeling initiatives and regularly deal with transparency issues. Based on these sources, we introduce the "creepiness scale" to help explain people's reactions to the use of personal data and algorithms, explore the factors and issues that affect algorithmic transparency, discuss the Association for Computing Machinery's (ACM) recommended algorithmic transparency principles, recommend company best practices, speculate about possible future algorithmic transparency scenarios, and suggest research opportunities. This tutorial should help students and faculty better understand algorithmic transparency and business and analytics managers and professionals who deal with the need for algorithmic transparency in their companies.

2 The Creepiness Scale

One can think about people's reactions to recommendations based on personal data and algorithms as varying along a creepiness scale that ranges from "this is helpful" to "this is creepy" to "this is so wrong" (see Figure 1). For example, many people find the product recommendations on Amazon helpful. In contrast, having a Gmail conversation with a friend about a trip to Las Vegas that results in a pop-up ad for a Vegas hotel might strike those same people as creepy (at least the first time it occurs). At the extreme end, many would view a message to a friend about newly diagnosed breast cancer that generates hospice care ads as inappropriate and disturbing. Table 1 provides additional examples of algorithm-driven recommendations and decisions that span the creepiness continuum.

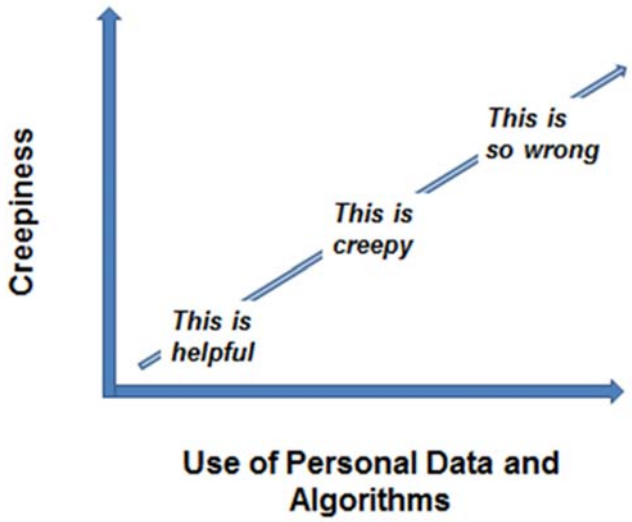

Figure 1. The Creepiness Scale for the Use of Personal Data and Algorithms

Table 1. Possible Reactions to the Use of Personal Data and Algorithms

This is helpful	• Movie suggestions on Netflix • Traffic information details from Google Maps before leaving for work • Recommendations and discounts for nearby restaurants from Yelp • Google's Home Advisor asking if you would like a reservation after you ask about a restaurant's operating hours • LinkedIn matching job recruiters and applicants
This is creepy	• Seeing someone you just met at a professional meeting suggested as "people you may know" on Facebook • Google telling you how long it will take to get to a destination without you saying where you are going • Ads from Instagram based on how you use your phone's microphone • Researching sickness symptoms and then seeing advertisements for specialists • Google Photos' ability to pull up every picture of you
This is so wrong	• Facebook's ability to influence your world view through news feeds • Screening job applicants based on analyzing their smile • Visiting a hospital emergency room and receiving ads from personal injury lawyers • Receiving an ad from a reseller of engagement rings after changing your relationship status to "single" from "engaged" on Facebook • Health insurance decisions based in part on your Facebook friends list • Receiving ads for writing papers and taking tests after changing your profile status to "study abroad in the US" on the popular Chinese website Weibo

While an appealing and intuitive concept, creepiness is hard to define, includes several factors, and varies with the individual. It also relates to privacy and ethics. In general, individuals perceive behavior as creepy when it varies from the norm in a potentially threatening way (Smith, 2016). For example, some people

perceive clowns as creepy (and scary). In the digital world, people can find a company's using their personal data and algorithms creepy (and possibly upsetting and harmful) when they do not anticipate that the companies will do so. To illustrate, an app called Girls Around Me that provided personal information and location data on girls in close proximity to the app's users once existed (Tene & Polonetsky, 2014). Though not illegal, people found it sufficiently creepy that the major social media companies ultimately removed it from their APIs, which lead to the app's demise.

The creepiness axis (y-axis) includes two major factors: 1) an individual's reactions to others' using their personal information in a way that potentially invades their privacy and 2) an individual's reactions to others' using algorithms that results in actions or decisions that they potentially perceive as harmful. To illustrate the former, consider how ads from retailers can follow users wherever they go after they view an item online. To illustrate the latter, consider an algorithm that recommends a prison sentence rather than probation based on factors such as whether the person's father was ever arrested, whether any friends were ever arrested, and how long the person has lived at their current address.

What people perceive as creepy depends on the individual. For example, people may accept something they once found creepy after the surprise factors goes away. The way people react to how companies use their personal data and algorithms normally relates to how much control people have over that use. For example, when people opt in to allow a company to use their personal data, such as with Facebook, they will be less likely to reject that use compared to when they have no control over its use (e.g., when a company shares data with another company without the user's knowledge). Also, a person may not mind sharing some personal information, such as where they work, but not want to share other information, such as their current location. Some personal information carries greater risk; for example, the disclosure that a person has a serious medical condition may affect employment and insurance decisions.

People view some ways to use personal data and algorithms as creepy because they invade their privacy. A multifaceted concept, privacy has received much research attention over the years, and its conceptualizations have evolved over time (especially recently due to advances in technology such as personalization platforms that target ads and newsfeeds).

One cannot easily define privacy. However, researchers often consider it something that one can lose, diminish, intrude on, invade, violate, or breach (Kennesaw State University, 2019). What one perceives as private depends on the content, usage context, and the individual (Levy, 2019). For example, people commonly give doctors access to their medical records but not to their employer. No one has access to a person's credit file unless they provide permission. Criminal records are publicly available.

Theories tend to view privacy based on either 1) non-intrusion, 2) non-interference, and 3) control over/restricting access to one's personal information (KSU, 2019). Non-intrusion (also referred to as accessibility privacy) refers to being left alone and free from government intrusion. Non-interference (decisional privacy) focuses on freedom from interference in making decisions. Control over/restricting access (informational privacy) refers to control over the flow of personal information, which includes its transfer, sharing, and exchange. Accessibility and informational privacy become most salient when considering online privacy and creepiness.

People view some applications as potentially creepy because an algorithm at least partially makes the decision. In a sense, the algorithm becomes a decision maker that users may not perceive as ethical and fair. Some applications use considerable personal information, such as the previously mentioned algorithms for generating prison sentence recommendations, while others, such as pricing algorithms for airlines, rental cars, and hotel rooms, use little or no personal data.

People want other individuals, organizations, and governments to treat them in an ethical manner. Companies should follow behavioral standards (e.g., decisions and actions) that comply with ethical principles, such as do right and good actions, minimize harm, respect autonomy, and protect rights (Chonko, 2019). "Decisions of the heart" (see Section 5.9) require especially close scrutiny due to their potential impact on people's lives.

3 The Factors

Multiple factors influence algorithmic transparency (see Figure 2). Depending on how companies address these factors, they can find themselves anywhere along the creepiness scale. With care, companies can benefit from and avoid the penalties associated with using personal data and algorithms inappropriately.

The recommended best practices that we present later can help ensure that companies do not move too far toward the "this is creepy" and "this is so wrong" portion of the scale.

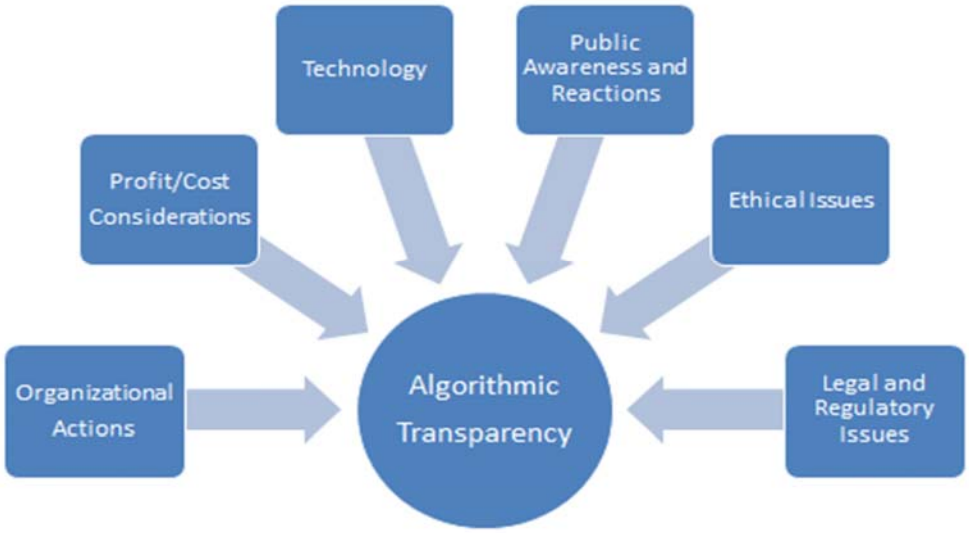

Figure 2. Factors that Affect Algorithmic Transparency

3.1 Organizational Actions

Advances in collecting, storing, and analyzing big data, dramatic increases in computational power, enhanced data-analysis techniques (which neural networks exemplify), and real business opportunities have motivated and enabled companies to more widely use algorithms in their advertising and decision making. They can now better personalize and target ads. They can also now automate decisions, which essentially creates a new, external decision-making entity (Arthur, 2017).

3.2 Profit/Cost Considerations

The use of personal data and algorithms has many potential benefits. With personal recommendations, companies have opportunities for improved customer service and increased revenues and profit. Further, algorithmic-driven decision making allows companies to make faster, less expensive, and potentially better decisions. Companies that use personal data and algorithms in an ethically sound and legally compliant way have a position for long-term success.

On the cost side, developing, maintaining, updating, deploying, and operating algorithmic systems comes with costs. Further, companies can face penalties for failing to comply with laws and regulations. In 2015, the Federal Communications Commission fined AT&T US$25 million for various data breaches that exposed the personal information of nearly 280,000 customers (Kastrenakes, 2015). More recently, Equifax agreed to pay up to US$700 million for its "failure to take reasonable steps to secure its network" in its 2017 data breach (Whittaker, 2019). Other costs associated with not securing personal data can be substantial. In March, 2018, Facebook's stock lost over US$100 billion following news coverage about a data breach involving Cambridge Analytica (Shen, 2018).

Less explicit costs include the ones that arise from using algorithms in ways that damage a company's brand, such as how Target used algorithms to predict pregnant women. The blowback began when an angry father complained to a Minneapolis store manager that his 16-year-old daughter had received pregnancy-related coupons and that the coupons promoted teen pregnancy. Publications such as *The New York Times* and *Forbes* reported the story, and, while Target had conducted legal activities, they struck some people as creepy and damaged Target's brand. It did not matter that the father later learned that his daughter was indeed pregnant.

3.3 Technology

Companies have become good at capturing and analyzing personal data. Many have also excelled at creating or accessing personalization architectures that use personal data and algorithms to generate highly targeted recommendations and ads. Companies can also turn to data brokers such as Alteryx or LexisNexis to provide data and platforms for sourcing, blending, and analyzing data ingested from multiple sources. Sources have estimated that as many as 4,000 firms have made a US$21B business of buying and selling people's personal information (Washington Post Editorial Board, 2018). One expert we interviewed observed that: "In some companies today that collect, combine, and sell personal data (data brokers), there is the attitude that the public will never know the company's role in the ads the public sees".

Technology also helps companies and individuals disseminate "fake news" (also called "disinformation") and false, inaccurate, or misleading product promotions (Lomas, 2018). Bot-based amplification can increase the popularity of news stories and, combined with paid product influencers, stimulate demand for promoted products.

New technologies have expanded the kinds of data one can capture, analyze, and use. Amazon Echo and Google Home permanently store what people say. Facial Action Coding Systems visually capture individuals' facial movements (i.e., microexpressions) and analyze how they feel (e., anger, joy). These and other technologies have increasingly appeared in various areas such as security, healthcare, and marketing and have both exciting but also potentially troubling consequences (Randall, 2011).

3.4 Public Awareness and Reactions

The public currently seems to vaguely understand and have mixed concerns about what personal data organizations collect, analyze, and use. People typically perceive ads as an irritant or the cost of receiving a free service (which they are). However, the growing use of personal data and the large number of security breaches at companies such as Yahoo and Equifax have focused the public's attention on personal data. Further, the U.S. Government's probe into Russia's involvement in the 2016 presidential election has contributed to the US public's awareness. In a 2018 study, the Pew Research Center found that just over half (51%) of the public believes that technology companies should be regulated more than they are now (Smith, 2018).

People's concerns increase dramatically when an algorithmic decision negatively affects their life in a personal way. An application does not even have to use personal data to create adverse publicity and reactions. In September, 2017, some airlines faced criticism for dramatically raising fares for people in south Florida who were trying to flee before Hurricane Irma made land (Mindock & Calder, 2017).

One can use personal data and algorithms in subtle ways that many people perceive as unethical. For example, consider companies that share location data. People typically do not mind revealing their location to an app such as Yelp! or TripAdvisor if it recommends highly rated nearby restaurants. However, people may dislike companies that give their location to a third-party data broker who sells ads to other companies (Washington Post Editorial Board, 2018). Further, users may dislike how Google can track Android users' location by collecting the addresses of nearby cell towers even when they have turned location services on their phone off. In behalf of companies, Google can use this information to increase advertising revenue by delivering ads to people who have visited certain locations (Nakashima, 2018).

Decisions to avoid building applications in the "this is so wrong" category should not be difficult to make, but others involve more challenge. For example, as we move to driverless cars, should a car's occupants or pedestrians have higher safety priority in a crisis situation? As we develop more expensive, potentially life-saving medical treatments, what factors (e.g., age, wealth) should algorithms that screen candidates for treatment include?

3.5 Legal and Regulatory Issues

Most industries in the US have laws that regulate how organizations collect, store, and release data, such as the Health Insurance Portability and Accountability Act (HIPPA) in healthcare; the Third Basel Accord, known as BASEL III, in financial services; the Fair Credit Act (FCA) in credit reporting; and the Family Educational Rights and Privacy Act (FERPA) in education. Companies in these industries usually have much familiarity with operating according to these laws and regulations. But, as one expert we interviewed

observed: “When it comes to online marketing, social media, and the IoT, at this time there is limited federal or state legislation that affects the use of algorithms”.

Legislation about personal data and algorithms has begun to emerge. Over 50 countries have recently enacted data privacy laws. The European Union’s General Data Protection Regulation (GDPR) went into effect on 25 May, 2018, and its 99 articles (i.e., regulations) apply to all companies that process European Union (EU) resident data (EUGDPR, 2016). It replaced national laws and the data directive that existed for 20 years and provided only recommendations for handling personal data. Any company that either operates in the EU or collects European citizens’ personal data now needs to comply with the GDPR even if the company does not have physical operations (e.g., a data center) there. As such, the GDPR affects any company that conducts any business in the EU and interacts with European residents anywhere in the world. The GDPR will likely influence non-European consumers’ expectations, and they will likely press for similar regulatory protections. Bilateral data sharing agreements between the US and EU will also likely foster similar laws to GDPR in the US.

Companies that do not comply with the GDPR can receive fines up to four percent of global revenues or €20 million (approximately US$24.5 million)—whichever is greater—for the most serious violations, although, as one expert we interviewed said: “No one knows how quickly the regulators are going to act or how far they will go in enforcing the regulations, but it could be a game changer”. While regulators may initially have more interest in helping companies comply rather than leavening fines, the French data regulator CNIL fined Google €50 million (approximately $57 million US) on 21 January, 2019, because its data-collection policies violated the GDPR (Kaiser, 2019).

With the GDPR, authorizations to collect and use personal data must be specific and unambiguous, and organizations must use data for a specific, well-understood business purpose. People must opt in for a company to collect their data. Companies cannot use an “opt in” mechanism such as asking consumers to agree to pre-ticked boxes; rather, consumers must “opt in” by clicking on an unchecked box marked “I agree” (Mathews & Bowman, 2018).

In some circumstances, people also have the right to have their data deleted (the so-called “right to be forgotten”), such as when a company no longer needs it for the purpose it collected it for, if one withdraws consent, or if the company unlawfully processed it (Burgess, 2018). The GDPR expands the scope of personal data to include any data that one can link or attribute to an individual, such as through an IP address. People must be alerted to compromises of their personal data within 72 hours (compare that to six weeks for Equifax and over a year for Über). While exceptions exist, the law entitles European citizens to explanations about automated decisions, to challenge such decisions, and to opt out. One expert we interviewed predicted that: “The dominant change from GDPR will be the requirement for the purposeful use of personal data—companies will no longer collect data for one reason and then use it for something else”.

In the US, the California Consumer Privacy Act (CCPA) went into effect on 1 January, 2020 (Mathews & Bowman, 2018). The CCPA applies to all larger companies that do business in California (even if they do not physically operate there) that collect and control California residents’ personal information. The CCPA has similarities to GDPR, such as applying to companies outside physical borders, but also has differences, such as more focus on consumer privacy rights and company-required disclosures to consumers. Among its most salient features include a requirement that companies have a link on their website titled “do not sell my personal information” and that people can request companies to delete their personal information (Spirion, 2018).

Local governments across the US have also begun to pass data-related laws. At the end of 2017, New York City passed a bill that created a task force to provide recommendations for how data collected by the city’s automated decision systems can be shared with the public and how agencies may address instances of people being harmed by those systems (Bernard, 2017). The American Civil Liberties Union (ACLU) supported the bill and believes that algorithmic source code should be publicly available to help identify and remedy flaws and biases.

After years of claiming they could self-regulate, technology companies now generally better support and participate in developing federal privacy laws and regulations. They have developed this new posture at least partially due to a fear that state and local governments will enact consumer-protection laws, such as CCPA, that restrict them more than what the U.S. Government might pass at the federal level. They also fear that the various laws will create a patchwork of regulations that make conducting business challenging (Allan, 2018; Guliani, 2018).

4 Algorithmic Transparency Principles

The Association for Computing Machinery (ACM) in the US and Europe, working both separately and together, codified seven principles for ensuring that organizations use personal data and algorithms fairly (Garfinkel, Matthews, Shapiro & Smith, 2017). We list these principles in Table 2 and discuss them in the context of today's algorithmic transparency environment. The principles provide a foundation for recommending effective company policies and practices.

Table 2. ACM Algorithmic Accountability Principles (Garfinkel et al., 2017)

Principle	Description
Awareness	Ensuring that individuals recognize what personal data organizations collect, analyze, use, and share and the extent to which they automate decisions
Access and redress	Investigating and correcting erroneous decisions
Accountability	Ensuring that companies and the people who develop and use algorithms are responsible for their actions
Explanation	Communicating the algorithm's logic in human terms
Data provenance	Knowing the data sources used and their trustworthiness
Auditability	Recording the processes followed in developing and using algorithms so they can be reviewed
Validation and testing	Ensuring that automated systems perform as intended

4.1 Principle 1: Awareness

Some companies (e.g., Facebook) generate tremendous revenues from business models that rely on using personal data and algorithms to make recommendations and target ads. In a typical scenario, users download an app to a smart device and quickly click through the permissions to gain access to it. The users give little thought to and have little concern over losing control over their personal data. The app has a long agreement and presents it in small font and legalese, and it gives the vague right to use and share personal data with business partners. In this scenario, users typically find it difficult to or cannot reject or limit how the app collects and uses their personal data (short of not using the app). In this scenario, the app's usefulness and convenience typically outweigh users' privacy concerns.

Most people do not fully recognize how extensively companies collect, analyze, use, and share personal data. Emerging technologies outpace individuals' and society's ability to keep track of events. Indeed, technologies (especially social media apps) can collect extensive amounts of personal data. For example, the popular dating app Tinder includes data from an individual's Facebook account to help find matches (Thompson, 2013). It looks at one's friends, one's common interests, one's college, who one has conversations with, and who one swipes "yes" and "no" to. It also monitors how one uses the app by looking at how often and for how long one has conversations to see if it should show similar users.

For high-profile, digital-native companies, using personal data can raise the public's ire when that use receives negative media attention. For example, in 2015, the music streaming company Spotify released a revised privacy policy that allowed it to track people's location, access their photos and contacts, collect and store information and posts from Facebook, and share this information with third-party partners. In response to criticism, CEO Daniel Etz felt compelled to acknowledge that the policy "caused a lot of confusion about what kind of information we access and what we do with it" (Goldman & King, 2015). Many companies collect and use personal data in similar ways to Spotify. While few companies currently experience such significant blowback on their use of personal data, that has begun to change.

4.2 Principle 2: Access and Redress

Algorithms can make mistakes due to poor development processes and the statistical uncertainty associated with some models (Diakopouols, 2014). Organizations should have mechanisms that allow individuals and groups to contest decisions and correct errors that algorithms make. Organizations should offer and make it easy for people to access information about who to contact and what processes to follow in order inquire about possible errors. As Cathy O'Neil says, when an algorithm has affected someone's life, "at the very least they should be asking how do you know this is legal? That it isn't discriminating?" (Chalabi, 2016).

4.3 Principle 3: Accountability

Individuals (especially IT professionals who understand how to build systems) can find it irritating to hear the computer has caused a problem when human error or the system's design likely caused it. In a similar way, organizations should be accountable if an algorithm makes a mistake or its use has unintended consequences. "The algorithm did it" (especially with life-altering decisions) does not represent an acceptable excuse (Diakoulus & Friedler, 2016).

Data scientists, managers, and companies should have responsibility for the kinds of systems they create, the design choices they make, and how they use systems. As one expert we interviewed said: "Data scientists should have the same responsibility as physicians—'first do no harm'". They should be responsible for the technical, legal, and ethical aspects of their work.

4.4 Principle 4: Explanation

Unless one knows the variables and algorithms that organizations and government agencies use to make decisions, one cannot know whether they engage in deceptive, discriminatory, illegal, or unethical practices (EPIC, 2018). People whom algorithms negatively impact should have the right to an explanation about actions and decisions related to them.

However, while one can easily understand some algorithms (e.g., with multiple regression analysis, one knows the independent variables and their relative importance; the same clarity exists with decision trees), one cannot easily understand some more advanced algorithms, especially the hierarchical neural networks associated with deep learning. Many organizations now use these deep learning models for predictive purposes in situations with many independent variables and large data sets. Once one inputs data for the independent and output variables into the neural network, each layer of the model feeds its analysis results to the next level in the hierarchy. In this analysis approach, one cannot easily identify the relationships among the various input variables to identify and explain a decision in human terms (in other words, they constitute "black boxes").

Sometimes, organizations make predictions based on multiple rather than a single model. This approach combines multiple models in order to develop better predictions than one can develop with a single model. For example, when predicting income, one model may be best for individuals with low to medium incomes while another may be best for individuals with higher incomes. Alternatively, one can combine predictions from multiple models to make a prediction (e.g., random forests). While such an arrangement may result in better predictions, it also increases the difficulty of explaining the logic behind a given recommendation or decision.

In addition to being able to explain an algorithm's results, organizations also need to be able to explain the processes they use in building, testing, updating, and using algorithms. This understanding ensures model confidence and credibility.

4.5 Principle 5: Data Provenance

Developers should be familiar with the data domain and recognize and be concerned about the accuracy and potential biases in the data they use when building and testing models. For example, prescreening bias represents a common, serious model-building problem. To illustrate, a model that decides who to admit to a university that one builds and tests using only the data from students who attended the university in the past will carry the biases embedded in previous admission decisions.

Developers can unintentionally introduce bias into the models they build. They can exclude variables and data that would provide better predictions. They can make analysis choices (e.g., the number of categories specified with k-means) that affect the models that they create. While models may appear to epitomize fairness and objectivity, they potentially contain biases in the data sources used and the model-building design choices made.

4.6 Principle 6: Auditability

Internal and external parties should be able to review how organizations develop, use, and maintain algorithms. By allowing others to monitor and critique algorithms, organizations can make better design choices and more rapidly change models and processes when necessary (Diakoulus & Friedler, 2016).

Organizations can attain various benefits from recording how they develop and use algorithms. They need to update models as company and environmental conditions change. Further, by documenting the data sources and algorithms they use, the model training and testing processes they employ, and how they implement and update models, they better facilitate the reusability of previous work.

4.7 Principle 7: Validation and Testing

While all algorithms should generate accurate predictions, organizations should hold decisions that affect people's lives to the highest standards. For example, an organization might consider a one percent lift in sales from a targeted ad a great success, whereas 90 percent accuracy from a medical diagnostic algorithm might not be good enough. Required algorithmic accuracy must be set on a case-by-case basis.

The most common approach to developing and testing models involves splitting up the data into training and testing data sets (often an 80/20 split). While many methods for making the splits exist, modelers use the same data for building and testing. In a sense, the data gets "worn out". For this reason, experienced modelers often hold out a third data set that they use only to validate the final model. An even better practice involves collecting new data and using it for final testing.

After organizations put an algorithm into production, they must monitor its performance. It is not always clear in advance how an algorithm will behave, and some of the possibilities are unfavorable (Dickey, 2017). Like all things in life, models age. Consequently, organizations must have systems to continuously ensure that models perform as intended. Organizations will likely need algorithms that monitor the performance of other algorithms to ensure that the algorithms that automate decision making are fair, perform as intended, and comply with laws and regulations (Angwin, 2016). Organizations should also put systems in place to monitor unfavorable reactions to their algorithms, such as news media reports and social media posts.

5 Algorithmic Transparency Best Practices

Multiple factors contribute to the growing interest and concerns about algorithmic transparency, such as the increasing frequency with which organizations use personal data and algorithms to generate ads, security breaches that focus attention on personal data's safety, and the GDPR's impacts and consequences. In the future, more companies will have to practice algorithmic transparency; consequently, companies should prepare for it now.

Based on reviewing the literature, conducting interviews with experts, and our experiences, we recommend the following best practices (see Sections 5.1 to 5.10). They cover business (e.g., strategic positioning), organizational (e.g., governance), development (e.g., model building), technology, and other issues and that help satisfy the ACM's algorithmic transparency principles.

5.1 Best Practice 1: Recognize that One Size Doesn't Fit All

Much of the conversation about algorithmic transparency involves technology companies such as Facebook, Google, and Twitter. These large and highly visible companies have business models that rely on personal data and algorithms to generate revenues. However, the importance of algorithmic transparency to a company depends on various factors, such as industry, competitive landscape, company goals, current algorithmic maturity, and the data and applications it uses. A company's positions relative to these factors can change, which can, in turn, change algorithmic transparency's importance.

Figure 3 shows the interaction between 1) the extent to which an organization uses personal data and algorithms (x-axis) and 2) its impact on individuals (y-axis). When both are low (the chill quadrant), algorithmic transparency does not pose an issue. Even when an organization uses personal data and algorithms at a high level, problems will likely not arise if the algorithms do not affect individuals much (the quiet waters quadrant). On the other hand, if organizations use personal data at a low level but the algorithms have a high impact on individuals, problems can arise (the dangerous road quadrant). Finally, the fireworks quadrant represents companies that use personal data and algorithms extensively for purposes with high impact. Next, we consider some examples that the experts we interviewed gave.

Figure 3. Algorithmic Impact Matrix

FirstData (a credit card processing company that Fiserv recently acquired) has a model to predict in-store and online credit card fraud that employs a neural network with over 150 predictor variables. Businesses that accept credit cards make up the model's customers. Though the application uses considerable personal data (e.g., previous purchase history) and advanced analytics, it causes no problems for the public. The largely invisible application (unless a false positive for fraud occurs) helps protect card holders. The network presumably negatively impacts only individuals who perpetrate fraud, so this usage falls into the quiet waters quadrant.

Next, consider State Farm. The company started its analytics journey with a focus on workflow applications to better enroll customers, manage claims, and provide excellent customer service. The data and analytics (e. g., statistical modeling) increased its operational efficiency and effectiveness and did not raise concerns from the public. Next, it used predictive models that identified, for example, which customers would most likely purchase particular policies.

State Farm currently explores analytics applications that use data sources and analytics methods that the public will more likely notice. For example, it employs camera drones to capture data on insured property and analyzes this data through advanced technology such as neural networks to fulfill claims faster and more accurately after natural disasters. Thus, this example falls into a higher exposure quadrant.

Companies should assess the ways they currently and plan to use personal data and algorithms and how such uses will impact individuals. With this assessment (and an awareness of the changing societal and legal environments), companies should be able to understand the importance of algorithmic transparency.

5.2 Best Practice 2: Analytics Leaders Should Take the Lead

Ideally, senior management should participate in all things analytics. More realistically, senior management should at least providee the vision and guidelines for analytics (Burkhardt, Hohn, & Wigley, 2019). Experience has shown, however, that senior management may provide only vague directives for analytics leaders. As one expert we interviewed said: "My charge was to figure out what could be done with customer data". Due to situations such as this one, the analytics team may need to take the lead in identifying specific analytic opportunities and addressing the need for algorithmic transparency.

Analytics leaders should identify and avoid practices and applications that expose the company to customer, societal, legal, and financial risk. For example, leaders may have to educate senior management about how models should be developed and used. One expert we interviewed described

how his organization created a propensity-to-buy model to use with existing customers. It proved successful and senior management wanted to use the same model with potential customers as well. The analytics leader had to explain that the analytics team could not fully transfer the model to potential customers and had to redevelop it using data from new customers.

Organizations also need to ensure that their algorithmic capabilities are not used in unanticipated ways that create problems (e.g., Russia's involvement in the 2016 presidential election). We have only just begun to fully appreciate the ways individuals and organizations can use social media platforms and the impact that such uses can have (Lomas, 2018).

Analytic teams should not simply build the best possible predictive model but more broadly consider their work's possible impacts and consequences. While one cannot anticipate all possible situations, analytic team leaders should think beyond current modeling efforts and ways to use technology. They should not wait until problems occur that put their company and senior management in difficult situations. This best practice helps satisfy the ACM's accountability principle.

5.3 Best Practice 3: Apply Business Sense when Using Personal Data and Algorithms

For every application that uses personal data and algorithms, organizations should carefully consider the business value and risk. As one expert we interviewed pointed out: "There should be a focus on how the customer is likely to react to a recommendation, ad, or an automated decision". Organizations should stay in the "this is helpful" portion of the creepiness scale and avoid the "this is so wrong" end. Analysts and data scientists who build the algorithms should be sensitive to this perspective. Further, organizations need to involve business leaders who have good business acumen and good relationships with customers in the process since they understand the business consequences of inappropriately using data and algorithms. This best practice helps satisfy the ACM's accountability principle.

5.4 Best Practice 4: Govern the Use of Personal Data, Algorithms, and Applications

Senior management should put people, committees, and processes in place to ensure employees develop and use algorithms in legal and ethical ways that benefit individuals, the organization, and society. Governance committee members should have a mix of business, legal, ethical, statistical, modeling, and systems engineering backgrounds, and a C-level executive such as a CIO, chief data officer, or director of data science should ideally lead them.

Analytics requires a combination of modeling, data management, and business domain knowledge, but, as privacy, ethical, and legal considerations have become more important, companies should assign a person or team to consider emerging regulations, concerns, and issues. Companies should have someone who always asks and answers the questions: "Is this fair to the consumer?" "Is this legal". One of the experts we interviewed provided a legal but possible troubling example: a gaming company purchasing the names and contact information of people who had received treatment for a gambling addiction and targeting them with ads and offers.

Good governance also involves explicit policies that undergo regular reviews and updates to ensure they comply with new analytics initiatives. The legal team should closely participate in deciding what data to use and store, who to share it with, how to analyze it, and how to use the output—preferably as members of the governance group or alternately as part of a final review and sign-off process. This best practice helps satisfy the ACM's accountability principle.

5.5 Best Practice 5: Be Transparent about the Use of Personal Data and Algorithms

While laws, regulations, and public pressure require greater transparency, companies also benefit from being more open and transparent. As one expert we interviewed said: "Companies should trust that people will make the correct decision to share their personal data when there is a logical reason and perceive that it is beneficial to do so".

One way to engender more trust involves requesting personal data (and explaining what will be done with it) at the time one needs it rather than asking for upfront blanket approval, such as what typically occurs when users initially use a new app on their smartphone. One can also capture the right to collect, store, and use personal data by employing a template or form that asks for permission. The form should identify

the various kinds of data the app collects, how the app collects it (e.g., microphone), how the organization will use it, whether the organization will share it, and, if so, with whom and for how long.

In the Spotify example we refer to in Section 4.1, the public poorly reacted to the way the company collected location data. One might ask why a person's location pertains to streaming music, and it turns out that the company did have a valid reason—it wanted to tailor music to what people did. If a person began to run (and, thus, changed location quickly), Spotify would stream more upbeat songs that matched the runner's pace. By simply explaining its reasons, Spotify might have prevented at least some of the media and public backlash. This best practice helps satisfy all of the ACM's principles.

5.6 Best Practice 6: Be Alert and Seek Technological Solutions

Many CMOs tend to collect and store as much personal data as possible in case it becomes useful at some point. This practice means that many companies have considerable personal data stored in various databases and systems that multiple applications use and that the companies employ for different purposes. Organizations should revisit this practice. Unless data has a well-defined purpose, organizations risk exposure and regulatory violations.

Companies need to understand and have control over this data. Companies that collect and use personal data will have to show that they have a person's consent to collect it, track what data they use, how they transform it, what analyses they perform on it, how they use it, and who has access to it (i.e., its lineage). A recent study found that many companies experience challenges in complying with the GDPR's data-discovery and data-mapping elements (ISACA, 2018). In the future, customers will be able to ask what data organizations store, check its accuracy, have corrections made, and have data deleted. In order to prepare for this new norm, companies should put processes and technology in place now even if they currently use customer data in a limited and relatively basic way; they will find the experience they gain from doing so valuable.

Having appropriate technology enables effective governance (see Section 5.4). One expert we interviewed observed that "the software (for governance) is currently under-developed, but this is likely to change once vendors see the market for new and enhanced offerings and customers expect their vendors' products to help meet regulatory requirements". Companies can consider some of the current software alternatives that can help (Laszlo, 2018).

Companies use master data management (MDM) to record and link their critical data to one file called a master file (Sabat, 2018). This file provides the best or "golden" records (e.g., customer address) for applications that need that data. Companies can use MDM systems to understand what customer data they store, where it resides, the relationships among customers (e.g., husband and wife), who has access to the data, how they make changes to the data, and how they match customers (i.e., identifying the same customer). While MDM can support GDPR and other privacy compliance requirements, it does not represent a complete solution. MDM systems do not typically include some data customer data, such as behavioral data.

However, other technologies can help. Companies cannot use some personal data, such as race and gender, for most applications due to federal laws and regulations. To ensure that they do not inadvertently use it, companies can implement access-control mechanisms through database technology. They can also provide similar controls for other data that has the potential to cause problems, such as ZIP codes (given they correlate highly with race). Data-mapping software can also help companies to automatically record what data they store, where it resides, and how it flows through various applications. Software can also record what personal data and algorithms they use. They can maintain all this information in a centralized hub for regulatory and other purposes.

As for algorithms, companies and researchers have begun to work on developing software that helps explain the logic behind algorithmic output, especially from deep learning models. Should they succeed, these efforts will help explain algorithmically driven decisions and show government regulators that models do not discriminate against protected classes. This best practice helps satisfy all of the ACM's principles.

5.7 Best Practice 7: Act on Europe's General Data Protection Regulation (GDPR) and Future Laws and Regulations

Though the GPDR has already taken effect, many companies in and outside Europe do not fully comply with it for various reasons (e.g., because they incorrectly think it does not apply to them or do not fully understand its requirements and consequences)(Addario, 2018).However, given the GDPR's extensive requirements, it will affect most companies.

Companies should take certain steps to prepare for potential changes to the regulatory landscape. For example, they should first identify all personal data that they and third parties use, where it comes from, where they store it, who has access to it, and whom they share it with. Companies should think about contingency plans if they can no longer use some personal data in predictive models. They should be sure that they have up-to-date privacy and communications policies and have systems for recording and managing customer consent. They should develop procedures to respond to customer requests for data access and deletion. They should implement and use technologies to support compliance (see Section 5.6). They should build systems to quickly communicate data breaches and develop processes for responding to and correcting errors in algorithmically driven decisions. Though some firms currently take a "silo" approach (e.g., Europe only) to comply with GDPR, the smart ones have already begun preparing for possible global implementation.

While responding to GDPR demands much from an organization's IT, HR, marketing, and legal departments, doing so will likely improve the organization's data management, data security, and customer confidence in how it handles personal data, which is good for business (Hrubey, 2018). Complying with all rules and regulations helps satisfy all of the ACM's principles.

5.8 Best Practice 8: Have People and Applications that Can Explain Analytical Output

When organizations use model ensembles and powerful machine-learning algorithms, one often cannot easily interpret and explain the outputs. When algorithms significantly affect lives, companies need to be able to communicate and explain whether a human or a machine made a decision, and—if the machine—the process and logic involved. Companies should make an appeal process for machine-made decisions available. Several experts we identified expressed the belief that the public currently has little interest in how models work other than mild curiosity about those that the large digital natives use, though this situation will likely change. Because executives often do not wish to implement models they do not understand, modelers do wind up spending considerable time explaining model logic to business sponsors who request and ultimately use them.

In response, organizations should create an "analytics interpreter" role to explain analytical output to management and external stakeholders (Watson, 2017). This role may be someone's primary job or an additional task that the business intelligence (BI) staff, business analysts, or data scientists perform. In either case, the person must understand the audience, the business domain, the opportunity or problem the model addresses, and must have excellent communication skills (i.e., writing, presentation, and storytelling).

Some companies have begun to develop methods to trace back through a model to explain its logic in a way that humans can understand. In particular, credit bureaus companies such as Equifax that experience close regulatory scrutiny require this capability. These methods provide "narrow AI" in that they pertain to a specific model/application domain. Broader, more generalizable AI involves much more challenge, but researchers have begun to research it (Datta, Sen, & Zick, 2018). This best practice helps meet all of the ACM principles but has a critical role in satisfying the explanation principle.

5.9 Best Practice 9: Be Careful When Automating Decisions of the Heart

Organizations have used algorithms such as economic order quantity (EOQ) and the simplex method of linear programming for many years to increase their operational efficiency. Organizations have recently begun using models to understand and predict human behavior using personal data. Even more recently, organizations have begun using models to make decisions for what we might call "decisions of the heart" (i.e., important decisions about people's lives that one can automate but that many people believe humans should make or at least review since they have empathy).

To illustrate, consider a person who has an accident and cannot work for several weeks. They fall behind in paying bills, which results in a decreased Fair, Isaac, and Company (FICO) score. Whereas an algorithm for granting a loan would probably not know about or consider the reason for the lower score, a human might consider it when making a credit-granting decision. Other examples include the length of prison sentences or whether one qualifies for government assistance.

In a significant legal development that limited algorithms' use in sentencing in the United States, the Wisconsin Supreme Court ruled that an algorithmically determined risk score could be "determinative" in deciding whether a defendant is jailed or put on probation but that the presentencing report submitted to the judge must include a caveat about the limits of the algorithm's accuracy. This case may suggest future limitations on using algorithms, especially those that involve decisions of the heart (Angwin, 2016).

Even though decisions of the heart appeal to humans' empathy, one needs to maintain perspective. People are imperfect and inconsistent decision makers, which can result in "unfair" decisions. They often have conscious or unconscious biases in their decision making. As one expert we interviewed said: "We all bring biases…and [there is] no way to avoid it except to use data that stands on its own and does not need to be interpreted". Algorithms may not be perfect, but one can test and monitor them for biases and provide consistent results with the same inputs. This best practice helps satisfy all of the ACM's principles.

5.10 Best Practice 10: Recognize that You Are Building New Systems, Ways of Doing Work, and Business Models

Many professions and industries (e.g., travel agents, stock brokers, and bank tellers) have seen varying degrees of disintermediation. The current business models that allow many companies to compete require new technologies (including analytics) to drive down costs, compete in the marketplace, and create new business opportunities.

Lemonade, a renters and home insurance company, provides an interesting example of the new competitive business landscape. The company appeals to tech-savvy millennials in several ways. It has no physical agencies: it does everything online or via mobile. It has no deductibles and one can insure anything. Individuals provide documentation electronically, so the company wastes no paper, and it donates some of its profits to customers' favorite charities. Most interesting from an analytics perspective, however, is that AI-based chatbots have largely replaced human insurance agents. One can converse with a chatbot to sign up for insurance, file a claim, or receive other forms of customer service. One has no need to interact with a human unless the chatbot does not know how to respond to a specific question or situation.

The Lemonade example illustrates that management needs to recognize that analytics changes both business models, work processes, and the use of models. These models and processes can include how the organization competes, what work gets done and how, the new operational systems that employ models, and how the organization maintains the systems.

Managers, data scientists, analysts, and IT professionals should take a "privacy by design" approach that places privacy at the forefront of considerations when creating new models, processes, and systems. This approach does not represent a new concept (its origins trace back to the 1970s), but the need for its implementation has intensified over the past few years (Heavin & Power, 2019).

There is a growing need for people, processes, and technology for integrating and maintaining analytic models in operational systems. This is a challenging task, especially when companies have thousands of models deployed. Models embedded in operational systems must be maintained and updated in a largely automated way or companies will face a choice between the consequences of out-of-date models or the expense of constant model and system rebuilding. DevOps thinking and approaches (perhaps better called AnalyticsOps in this context) which unifies software development and operations are very appropriate when it comes to building and implementing analytical solutions.

The recognition of the need for model management is not new. In the days of OR/MS and DSS, a significant difference was that the focus of DSS was on the system – data, models, and dialog/interface – rather than primarily on the model, which was the focus of OR/MS. The maintenance of models in DSS was though a model management system (Sprague and Watson, 1975). This best practice helps satisfy the auditability and validation and testing principles.

6 Possible Future Scenarios

We cannot predict with certainty the future of algorithmic transparency. However, we have little doubt that it will continue to increase in importance as individuals, society, and governments become more concerned and involved.

While "going off the grid" has an innate appeal for some people, most people will not likely choose to do so. Most people enjoy the "this is helpful" benefits they receive from sharing their personal information. Would many people choose to go back to paper maps and asking for directions rather than sharing their location with a navigational app such as Waze? Would many people give up the ability to find their former friends and classmates on Facebook? Probably not.

Individuals' willingness to share their personal information constitutes a nuanced topic and includes considerations such as their attitudes about privacy, the benefits they receive, and how companies will use the information. For example, most people willingly share information so that an app serves its intended, primary purpose but do not want the same information shared with unknown groups and companies for equally unknown purposes. In the United States, Senator Ron Wyden (D-Oregon) has proposed a way for the public to gain control over how companies share their personal information (Washington Post Editorial Board, 2018). People could enroll in a "do-not-track" registry that would prohibit companies from sharing most personal data collected.

The impacts of algorithmic transparency will vary with the firm. Companies in the dangerous road and fireworks quadrants will most likely feel the effects. The large digital natives receive the most media and legislative attention and are most likely to have to change their business models and practices in order to secure long-term viability (Rohnama, 2018).

To illustrate, Facebook, which has widely received criticism for how it uses personal data and targets ads and newsfeeds, will likely modify how it does business. For example, people might 1) pay a monthly or annual fee and Facebook would make limited or no use of their personal data or 2) people would continue to have free use but agree to the how the company uses their personal data. As the storm of criticism has continued to grow, more dramatic changes seem likely. Facebook Chief Executive Officer (CEO) Mark Zuckerberg has announced that Facebook would become a "privacy-focused communications platform" with end-to-end encryption for personal messages and business transactions. WhatsApp and Instagram (the latter two owned by Facebook) would use this encryption, and Facebook would build profit-making business services on these applications. The company has also discussed allowing people to decouple their Internet browsing history from their Facebook user profiles, which would limit Facebook's ability to target ads and, as a consequence, the company's profitability (Ovide, 2019).

6.1 Use Scenarios

People will likely gain more control over their personal data and how others use it. In one scenario, they may even "own" their personal data and only allow companies to use it if they perceive that they will obtain sufficient enough benefits from doing so. Giving individuals more control over their personal data can have interesting consequences in both small and large ways. In the former case, for example, users would allow companies to collect, analyze, and use their personal data as long they perceive it to benefit them. For example, Amazon has begun trying to revolutionize the grocery shopping experience with its first Amazon Go store, which opened in January, 2018, and is mulling over plans to open 3,000 stores by 2021 (Leswing, 2018). Customers enter the store and scan a QR code using the Amazon Go app on their phone. As they shop, they add items to their digital cart. When they are done, they simply walk out of the store. The system records everything in their cart, charges the cost of the groceries to the credit card they have stored on their Amazon account, and sends a receipt to their phone.

The personal data collected adds just a little to what Amazon already knows about its customers, including what consumers buy and their shopping patterns. The next logical step involves using this additional data to improve customer satisfaction and generate revenues. For example, Amazon may send in-store shopping recommendations to consumers' phone based on their previous shopping behavior and the items currently in their cart. It may also send pop-up electronic discount coupons as consumers move throughout the store. However, Amazon may also sell consumers' shopping purchases information to third parties, such as companies that supply grocery products.

A person's medical data provides an interesting, higher-impact example. If individuals integrated and stored their complete medical history in the cloud, they could release their data to physicians as needed.

This practice would lead to more complete, up-to-date information on which physicians could base medical diagnoses and treatments (not to mention avoiding having to repeatedly provide the same information with every doctor visit). One expert we interviewed pointed out that this scenario makes the heroic assumption that companies that provide medical system software would willingly work in a cooperative way despite little evidence they would.

If people's medical history included their DNA, they might release their medical data for research purposes to better understand, predict, and treat specific illnesses (Roca & Letouze, 2016). For example, by age 50, half the population has hypertension. Physicians commonly have patients try multiple combinations of drugs in order to find one "cocktail" that controls the high blood pressure with minimal side effects. The findings from medical research will likely use biomarkers that allow physicians (or eventually artificial intelligence applications) to develop more personalized patient treatment plans to more quickly and better control hypertension (and other diseases).

Of course, such a situation would require safeguards. The public has begun to increasingly recognize security breaches and the possible negative consequences of storing such sensitive information electronically. Companies would also have to provide assurances that they will not release the data for unauthorized purposes, such as an insurance company seeking a person's medical history in order to predict expensive future medical problems and deny coverage.

6.2 External Review

Many companies now either must or choose to have a third-party audit their financial statements, accounting practices, and internal controls. These audits satisfy regulatory requirements and provide credibility about a firm's financial health and practices. One expert we interviewed observed: "We may be on a path where companies have external audits of their use of personal data and algorithms". We can find support for this possibility in Uber's settlement with the Federal Trade Commission in the aftermath of its passenger and driver data security breach. The agreement calls for Uber to create a comprehensive privacy program and to submit reports from required third-party audits (Bensinger, 2017). Much like Sarbanes-Oxley Act (SOX) created new business opportunities for auditing firms, similar opportunities may arise due to the demands for greater algorithmic transparency.

7 Research Opportunities

Algorithmic transparency provides a varied, rich set of research opportunities. The opportunities range from societal to individual to business to technical in their focus. In this section, we consider some possibilities.

7.1 Use the Algorithmic Impact Matrix as a Framework to Better Understand how Companies Should Respond to Demands for Greater Algorithmic Transparency

The algorithmic impact matrix provides a useful framework for thinking about and researching the potential impact that using personal data and algorithms has on companies. Researchers could next conduct an analysis that places companies and industries in the matrix. For example, tech firms such as Facebook and Google obviously reside in the fireworks cell, but what about companies in transportation, gaming, and hospitality? Researchers could look at the location of industries today and where they will likely go in the future as analytics become increasingly important across all industries.

The best practices recommendations have a generic nature in that they provide guidance for all firms. That said, the recommendations have the most importance for firms that use personal data and algorithms in ways that have the greatest current visibility and impact on individuals (e.g., those in the dangerous road and fireworks cells). Future research could refine these recommendations based on which cell/industry a company resides in. For example, organizations in high-impact cells and regulated industries need to explain decisions based on analytical output the most.

7.2 Refine the Creepiness Construct

The creepiness scale represents a visually, intuitively appealing way to discuss the possible reactions that people have to how companies use their personal data and algorithms. Students in classes and executive development programs have received it well. However, it does not represent a carefully created, theory-

based, empirically tested construct. If researchers conduct further work using creepiness, they need theory-based, empirically validated instruments for measuring it.

Creepiness is a formative rather than a reflective construct because it results from a combination of possible variables or factors. While limited specific theory about creepiness as applied to personal data and algorithms exists, the literature does contain some potentially useful theories about privacy and ethics (Tene & Polonetsky, 2014; Kennesaw State University, 2019).

Researchers may also need to separate creepiness into creepiness associated with 1) how companies use personal data and 2) how companies use algorithms. Each has factors that seem to be more important to creepiness. For example, personal data use pertains more to privacy concerns, while some algorithms pertain more to ethical concerns.

7.3 Gain a more Nuanced Understanding of Privacy in the Digital World

Even though researchers have extensively researched privacy, we need more research on it. While such research could go in many possible directions, it would be especially interesting and helpful to better understand how feelings about privacy vary with age, gender, culture, and nationality. The initial research suggests significant differences, for example, with age groups and cultures (Kezer, Sevi, Cemalcilar, & Baruh, 2016). These finding could help companies follow practices that individuals do not perceive as "creepy" or "so wrong" with different groups.

The research may result in nuanced findings about privacy. For example, many often say that millennials lack concern about privacy as the amount of personal information they place on social media evidences. While they do post a lot, they also quickly block some people (e.g., ex-partners) from viewing it.

7.4 Design, Build, and Test Platforms that Meet the Requirements for Algorithmic Transparency

While some technologies that help companies achieve algorithmic transparency exist, they typically do not focus on transparency as a key consideration. For example, the GDPR did not exist until recently, and most products do not fully comply with it. This situation creates opportunities for commercial and open-source vendors and academics to develop platforms' logical and/or physical design in a way that supports algorithmic transparency. Recognizing this opportunity, researchers have recently developed a logical design for a blockchain-based system that complies with the GDPR and provides users with easy control over their personal data and the ability to sell the data for monetary compensation (Faber, Michelet, Weidmann, Mukkamala, & Vatrapu, 2019).

8 Conclusion

Hardly a week passes without news coverage about how companies have misused personal data and algorithms. Individuals should better recognize the business models that drive technology companies, understand why apps are free, know how companies use their personal data, be more careful about the permissions they give when installing an app, and learn how to protect themselves from unwanted use of personal data. Smart managers and analytics professionals should follow our recommendations to help ensure that their companies do enter the headlines due to poorly thought out and inappropriate personal data- management and algorithmic practices. Researchers should pursue some of the many research opportunities associated with algorithmic transparency.

References

Addario, E. (2018). GDPR will be a harsh wake-up call for most U.S. companies. *Information Management*. Retrieved from www.information-management.com/opinion/the-general-data-protection-regulation-will-be-a-harsh-wake-up-call-for-most-us-companies

Allan, D. (2018). California's new data privacy law could be a regulatory disaster. *Fortune*. Retrieved from http://fortune.com/2018/10/23/california-data-privacy-law-gdpr/

Angwin, J. (2016). Make algorithms accountable. *The New York Times*. Retrieved from www.nytimes.com/2016/08/01/opinion/make-algorithms-accountable.html

Arthur, W. B. (2017). Where is technology taking the economy? *McKinsey Quarterly*. Retrieved from www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/where-is-technology-taking-the-economy

Bensinger, G. (2017). Über settles with FTC over data-privacy protections. *The Wall Street Journal*. Retrieved from www.wsj.com/articles/uber-settles-with-ftc-over-data-privacy-protections-1502819449

Bernard, Z. (2017). The first bill to examine "algorithmic bias" in government agencies has just passed in New York City. *Business Insider*. Retrieved from www.businessinsider.com/algorithmic-bias-accountability-bill-passes-in-new-york-city-2017-12

Burgess, M. (2018). What is GDRP? The summary guide to GDPR compliance in the UK. *Wired*. Retrieved from https://www.wired.co.uk/article/what-is-gdpr-uk-eu-legislation-compliance-summary-fines-2018

Burkhardt, R., Hohn, N., & Wigley, C. (2019). Leading your organization to responsible AI. *McKinsey*. Retrieved from www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/leading-your-organization-to-responsible-ai

Chalabi, M. (2016). Weapons of math destruction: Cathy O'Neil adds up the damage of algorithms. *The Guardian*. Retrieved from https://www.theguardian.com/books/2016/oct/27/cathy-oneil-weapons-of-math-destruction-algorithms-big-data

Chonko, L. (2019). Ethical theories. *DSeF*. Retrieved from http://www.dsef.org/wp-content/uploads/2012/07/EthicalTheories.pdf

Datta, A., Sen, S., & Zick, Y. (2018). *Algorithmic transparency via quantitative input influence: Theory andexperiments with learning systems.* Retrieved from www.andrew.cmu.edu/user/danupam/datta-sen-zick-oakland16.pdf

Diakopoulos, N. (2014). Algorithmic accountability: Reporting on the investigation of black boxes. Tow *Center for Digital Journalism.* Retrieved from https://academiccommons.columbia.edu/doi/10.7916/D8ZK5TW2

Diakopoulos, N., & Friedler, S. (2016). How to hold algorithms accountable. *MIT Technology Review*. Retrieved from www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/

Dickey, M. R. (2017). Algorithmic accountability. *TechCrunch.* Retrieved from https://techcrunch.com/2017/04/30/algorithmic-accountability/

EPIC. (2018). Algorithmic transparency: End secret profiling. Retrieved from www.epic.org/algorithmic-transparency/

EUGDPR (2016). General data protection regulations. *Advisera.* Retrieved from https://advisera.com/eugdpracademy/gdpr/

Faber, B., Michelet, G., Weidmann, N., Mukkamala, R. R., & Vatrapu, R. (2019). BPDIMS: A blockchain-based personal data and identity management system. In *Proceedings of the Hawaii International Conference on System Sciences.*

Garfinkel, S., Matthews, J., Shapiro, S. S., & Smith, J. M. (2017). Toward algorithmic transparency and accountability. *Communications of the ACM*, *60*(9).

Goldman, D., & King, H. (2015). Spotify CEO apologizes for super-creepy new privacy policy. *CNNTech*. Retrieved from https://money.cnn.com/2015/08/21/technology/spotify-privacy/index.html

Guliani, N. S. (2018). Tech industry pushing for federal privacy protection—watch out. *The Washington Post*. Retrieved from www.miamiherald.com/opinion/op-ed/article219453010.html

Heavin, C. & Power, D.J. (2019). How should digital ethics impact analytics and data mining? *DSSResources*. Retrieved from http://dssresources.com/faq/index.php?action=artikel&id=454

Hill, P. (2018). Gartner: Top 10 strategic technology trends for 2019. *Forbes*. Retrieved from www.forbes.com/sites/peterhigh/2018/10/22/gartner-top-10-strategic-technology-trends-for-2019/#586bffa41423

Hrubey, P. S. (2018). Why GDPR is the best opportunity data managers ever had. *Information Management*. Retrieved from www.information-management.com/opinion/why-gdpr-is-the-best-opportunity-data-managers-ever-had

ISACA. (2018). GDPR: The end of the beginning. *ISACA Research Report*. Retrieved from https://m.isaca.org/Knowledge-Center/Documents/2018-GDPR-Readiness-Survey-Report.pdf

Kaiser, M. (2019). Google vs. GDPR: The ripple effect of the biggest data protection fine to date. *Techradar* Retrieved from www.techradar.com/news/google-vs-gdpr-the-ripple-effect-of-the-biggest-data-protection-fine-to-date

Kastrenakes, J. (2015). FCC fines AT&T $25 million for customer data thefts. *The Verge*. Retrieved from www.theverge.com/2015/4/8/8370515/att-fcc-settlement-data-thefts-25-million-fine

Kezer, M., Sevi, B., Cemalcilar, Z., & Baruh, L. (2016). Age differences in privacy attitudes, literacy and privacy management on Facebook. *Cyberpsychology.* Retrieved from https://cyberpsychology.eu/article/view/6182/5912

Kennesaw State University. (2019). *Theories of privacy* [presentation]. Retrieved from https://bit.ly/2P0BIHo

Laszlo, D. (2018). Data mapping: A key challenge in achieving GDPR compliance. *ISACA*. Retrieved from https://bit.ly/2Mv9SRU

Leswing, K. (2018). Amazon's grocery store of the future opens today: No cashiers, no registers, and no lines. *Business Insider*. Retrieved from www.businessinsider.com/amazon-go-grocery-store-future-photos-video-2016-12

Levy, E. (2019). Data privacy and protection fundamentals. In *Proceedings of the TDWI Conference.*

Lomas, N. (2018). Report calls for algorithmic transparency and education to fight fake news. *Techcrunch*. https://techcrunch.com/2018/03/12/report-calls-for-algorithmic-transparency-and-education-to-fight-fake-news/

Mathews, K. J., & Bowman, C. M. (2018). *The California Consumer Privacy Act of 2018.* Retrieved from https://privacylaw.proskauer.com/2018/07/articles/data-privacy-laws/the-california-consumer-privacy-act-of-2018/

Mindock, C., & Calder, S. (2017). Irma: Airline ticket prices surge up to 600% as incoming hurricane sparks mass evacuations. *Independent.* Retrieved from www.independent.co.uk/news/world/americas/irma-hurricane-flights-florida-tickets-prices-costs-hiked-gouging-delta-united-airlines-a7933471.html

Nakashima, R. (2018). APNewsBreak: Google clarifies tracking location-tracking policy. *AP News*. Retrieved from www.apnews.com/ef95c6a91eeb4d8e9dda9cad887bf211

O'Neil, C. (2016). *Weapons of math destruction*. New York, NY: Crown.

Ovide, S. (2019). Facebook feeling the pinch of privacy. *Bloomberg*. Retrieved from www.onlineathens.com/opinion/20190726/ovide-facebook-feeling-pinch-of-privacy

Pasquale, F. (2015). *The black box society.* Boston, MA: Harvard University Press.

Randall, K. (2011). Human lie detection: Paul Ekman decodes the faces of depression, terrorism, and joy. *Fast Company*. Retrieved from www.fastcompany.com/1800709/human-lie-detector-paul-ekman-decodes-faces-depression-terrorism-and-joy

Roca, T., & Letouze, E. (2016). Open algorithms: A new paradigm for using private data for social good. *Devex.* Retrieved from www.devex.com/news/open-algorithms-a-new-paradigm-for-using-private-data-for-social-good-88434

Rohnama, H. (2018). Algorithmic transparency is the next disruption for tech companies. *Readwrite*. Retrieved from https://readwrite.com/2018/05/01/algorithmic-transparency-is-the-next-disruption-for-tech-companies/

Sabat, T. (2018). Master data management: Answer to GDPR? *GRAKN.AI.* Retrieved from https://blog.grakn.ai/master-data-management-answer-to-gdpr-114598d8011

Schechner, S. (2017). Meet your new boss: An algorithm. *Wall Street Journal.* Retrieved from www.wsj.com/articles/meet-your-new-boss-an-algorithm-1512910800

Shen, L. (2018). Facebook stock is in the red for the year after the FTC confirms investigation. *Fortune*. Retrieved from http://fortune.com/2018/03/26/facebook-stock-ftc-investigation-cambridge-analytica/

Spirion. (2018). *Everything you need to know about the California Privacy Act of 2018.* Retrieved from www.spirion.com/blog/everything-you-need-to-know-about-the-california-consumer-privacy-act-of-2018/

Smith, A. (2018). Public attitudes toward technology companies. *Pew Research Center*. Retrieved from www.pewinternet.org/2018/06/28/public-attitudes-toward-technology-companies/

Smith, D. L. (2016). A theory of creepiness. *Aeon*. Retrieved from https://aeon.co/essays/what-makes-clowns-vampires-and-severed-hands-creepy

Smith, M., Patel, D. J., & Munoz, C. (2016). Big risks, big opportunities: The intersection of big data and civil rights. *The White House.* Retrieved from https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights

Sprague, R. H., & Watson, H. J. (1975). Model management in DSS. In *Proceedings of the 7th Annual Meeting of the American Institute for Decision Sciences*.

Tafekci, Z. (2017). We are building a dystopia just to make people click on ads. *TED*. Retrieved from www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads#t-15885.

Tene, O. & Polonetsky, J. (2014). A theory of creepy: Technology, privacy, and shifting social norms. *Yale Journal of Law and Technology*, *16*(1).

Thompson, C. (2013). What companies are doing with your intimate social data. *CNBC.* Retrieved from www.cnbc.com/2013/10/30/what-companies-are-doing-with-your-intimate-social-data.html

Washington Post Editorial Board. (2018). Yes, you are being followed. *StarNews Online.* Retrieved from www.starnewsonline.com/opinion/20181218/editorial-washington-post-yes-you-are--being-followed

Watson, H. J. (2017). Data visualization, data interpreters, and storytelling. *Business Intelligence Journal.* Retrieved from www.researchgate.net/publication/316605154_Data_Visualization_Data_Interpreters_and_Storytelling

Watson, H. J. (1973). Simulating human decision making. *Journal of Systems Management*, *26*(2), 24-27.

Whittaker, Z. (2019). FTC slaps Equifax with a fine up to $700M for 2017 data breach. *TechCrunch*. Retrieved from https://techcrunch.com/2019/07/22/equifax-fine-ftc/

Appendix: The Expert Interviews

We first conducted lightly scripted initial interviews with experts that covered various topics depending of their expertise and work experiences. They typically took an hour. We then sent the experts a draft of the manuscript and conducted a second round of interviews with them that lasted about half an hour to obtain their reactions and additional thoughts that we then incorporated into the tutorial. We recorded all interviews.

Table A1. Experts Interviewed

Expert	Title and Company
Michael Anton	Product Manager, Security & Fraud Solutions, FirstData/Fiserv
Helayna Bostian	President, Reworked
David Dunmire	Director, Product and Platform Development, AT&T
Carla Gentry	Digital Marketing Manager, Samtec
Lisa Loftis	Principal Consultant, CI Advisory Services—SAS Best Practices
Matt McGivern	Managing Director, Information Technology Consulting, Protiviti
Jake Meek	Manager, Assurance Practice, PwC
Bryan Samples	Systems Manager, Data Sciences, State Farm (at the time of the interview)
Stephen Smith	Research Director, Data Science, Eckerson Group
Anonymous	Manager of CRM Analytics. Large Retailer

About the Authors

Hugh J. Watson is a Professor of MIS and a holder of a C. Herman and Mary Virginia Terry Chair of Business Administration in the Terry College of Business at the University of Georgia. Hugh is a leading scholar and authority on business intelligence and analytics, having authored 24 books and over 200 scholarly journal articles. He helped develop the conceptual foundation for decision support systems in the 1970's, researched the development and implementation of executive information systems in the 1980's, and for the past twenty years has specialized in data warehousing, BI, and analytics. Hugh is a Fellow of the Association for Information Systems and The Data Warehousing Institute and is the Senior Editor of the *Business Intelligence Journal* and on the Advisory Board of the *Communications of the Association for Information Systems.* He is also the founder and a Fellow of the Teradata University Network, a free portal for faculty and students who want to learn about data warehousing, BI/DSS, analytics, and database.

Conner Nations is a recent graduate of the MIS program at the University of Georgia and is a business analyst at Randstad US in Atlanta, Georgia.