GUEST EDITORIAL

Reflections on the 2021 Impact Award: Why Privacy Still Matters

By: Heng Xu, Kogod School of Business, American University, Washington, DC, U.S.A.
Tamara Dinev, College of Business, Florida Atlantic University, Boca Raton, FL, U.S.A.

The purpose of this editorial is to reflect on our 2011 article, "Information Privacy Research: An Interdisciplinary Review," which was recognized with *MISQ's Impact Award for 2021*.[1] Our 2011 article starts with an observation that the "evolution of the concept of privacy in general—and information privacy in particular—follows the evolution of information technology itself" (Smith et al., 2011, p. 990). In the decade since the publication of our article, we have seen a revolution in the development and use of information technology, especially related to platforms, data, and artificial intelligence (AI) (Sejnowski, 2018). It is thus no surprise that the landscape of privacy—from what it means to how the relevant issues are addressed through social, legal, and technological means—all underwent seismic shifts during this time. In developing this editorial, we were delighted to find that most aspects of our 2011 article remain useful and relevant today. Our goal in this editorial is to acknowledge these enduring issues and, even more importantly, to explore what we could have done better and what new challenges to privacy scholarship deserve your attention.

We organize this editorial in three parts. First, we review the context within which we wrote the 2011 article, emphasizing the individual and societal concerns about privacy that were salient at the time, and how these concerns shaped the theoretical arguments and debates of the day. We explain how this historic context inspired the basic premise of the article, which was to elide the debate on "what privacy means" by focusing on examining people's privacy concerns as a measurable proxy of privacy and a central construct for empirical privacy research. By recounting the trajectory of privacy scholarship since the publication of our 2011 article, we offer conjectures as to why the article has had a sustained impact. We also expound on what we would have done differently.

Second, we ask how changes in IT since the publication of our 2011 article have modified the nature of people's privacy concerns and the conceptualization of privacy. In particular, we explain how advances in AI, especially deep learning (Goodfellow et al., 2016), have fundamentally shifted IT's reach and limits. By juxtaposing the major privacy outcries of the current decade with the decade prior, we suggest that the changing landscapes for privacy have pushed the locus and focus of people's privacy concerns beyond the flow of their data, into the uncertain, unknown, and potentially unknowable territory of what can be learned from data through technologies available today or in the future.

Third, we examine whether the issues and questions discussed in our 2011 article are still relevant today, or if new challenges have arisen. This examination gives rise to many perplexing research questions that we could not have foreseen in 2011. For example, can our extant theories speak to the different effects of privacy-enhancing tools available to individual users (e.g., incognito for private browsing) vs. privacy assurance techniques adopted by organizations (e.g., data anonymization)? How about the advanced computer-vision (Szegedy et al., 2016) and natural-language (Vaswani et al., 2017) models that threaten to turn all our online traces (e.g., texts, photos) into structured information about our socioeconomic status or purchase preferences? How about the interactions between those models and algorithms like generative adversarial networks (Goodfellow et al., 2014), which can create fake photos or speeches that look or sound just like us? We suggest that the emergence of these questions marks a transition point in privacy scholarship, when our research and theory could stand to gain by bringing in ideas and mechanisms from adjacent disciplines to help make our extant theories better aligned with the changing landscape of privacy regulation and industry practice.

[1] The award honors the paper published a decade earlier (2010-2012) that the selection committee deems to have: (1) the most significant and sustained scholarly impact, as shown by citations, by how it led to a change in thinking in the field, and by its prescience in identifying an important issue today; and 2) a real or potential impact beyond academia, especially through how it influences the way our field engages in an important real-world domain. As part of the award, the authors are invited to write a reflective editorial on the topic of the paper.

Back Then: The Concerns on Ubiquitous Data Collection

In the passages that follow, we first briefly review the historical context in terms of the major privacy-related incidents at the time, before explaining the genesis of our 2011 article given the state of privacy research back then.

Historical Context

In 2007, when we first started working on our 2011 article, privacy outcries were just beginning to resonate among internet users and shock many businesses that, by then, had grown comfortable with the idea that technological advances would eventually render the concept of privacy obsolete (Brin, 1998). One of the most widely reported privacy outcries in the era was the AOL search-log release incident. On August 4, 2006, AOL, with the noble intent of facilitating research on web search, released an excerpt of its users' web search query logs, covering about 650,000 users over a three-month period. Before releasing the data, AOL removed all attributes collected about a user (e.g., IP addresses, browser configurations) except a numeric ID and the search queries issued by the user, believing that doing so would eliminate all privacy concerns. Yet the release was met with an immediate uproar among AOL users for the sensitive information in the released queries, from names to phone numbers to addresses, allowing the identities of users to be inferred from the released data (e.g., User #4417749, Thelma Arnold from Lilburn, Georgia—Barbaro & Zeller, 2006). After TechCrunch published a scathing report on August 6 (Arrington, 2006), AOL took the dataset offline within a few hours, issued an apology the next day, and accepted the resignation of its chief technology officer on August 21, after "two weeks of intense criticism from privacy advocates" (Zeller, 2006).

Another outcry followed the publication of Sweeney (1997), which arguably shaped a major provision in the Health Insurance Portability and Accountability Act (HIPAA), the signature privacy legislation in the U.S. at the time. Sweeney (1997) linked two publicly available datasets—the voter registration list from Cambridge MA, and the medical data of the state employees of Massachusetts—to uniquely identify the medical profile of then-Governor Weld. The study's remarkable finding, which was cited as the motivation for broadening the definition of "identifier" in HIPAA (Federal Register, 2000), was that over 80% of Americans can be uniquely identified by a combination of gender, ZIP code, and birth date. Since none of the three attributes can, on its own, uniquely identify an individual, all three were retained in the medical dataset, which allowed Sweeney to identify Governor Weld based on the same data in the voter list.

In our view, both incidents speak to the root cause of privacy concerns at the time—i.e., the ease and scale of data collection cultivated by the explosive growth of Internet use. In both cases, privacy issues arose directly from the collection and sharing of an individual's personal data. For the AOL case, it was the release of search queries containing sensitive information. For the Sweeney study, it was the sharing of ZIP code, gender, and date of birth in combination. In either case, an individual could easily point to a certain part of the data flow as the culprit for the resulting privacy outcry. From this perspective, it should be no surprise that the prevailing mechanism for privacy protection at the time was to compose *rules*—from legislations to organizational policies—to govern flows of personal data (Beales & Muris, 2008).

Three Questions We Asked Back Then

To properly govern data flows, one needs to specify and then enforce rules. The enforcement part was amenable to technical solutions given the rich literature on access control (Sandhu & Samarati, 1994). The attention of many privacy scholars was therefore on the specification part, in particular the "ought" question of how these rules ought to be set in an ideal world. This inevitably triggers the normative question of "what privacy is." If we believe privacy is a human right (Milberg et al., 2000), then the rules should be set according to normative ethical theories. If we treat privacy as a commodity (Bennett, 1995), then the political economy of information markets should determine the rules. If we view privacy as an internalized norm, whether at the individual (Westin, 1967) or transaction (Margulis, 1977) level, then we need to study and understand the amorphous preferences of different individuals before customizing the rules. Given the importance of "what privacy is" to "who should set the rules," the normative notion of privacy was at the center of debates when we wrote our 2011 article.

While normative arguments about privacy are important, privacy scholars started realizing that debates about them are often rooted in deep-seated political and cultural beliefs, sometimes charged with emotions, and are rarely conclusive. There was a growing recognition that relentless debates on fundamental yet unresolvable questions like "what privacy is" or "who should

set the rules" might not be the best use of time when many empirical questions, both approachable and practically important, were left unanswered. This is why, in the 2011 article, we asked whether and how privacy scholars could actualize the potential of empirical research without necessarily taking a normative stand on the hotly debated ontological questions about privacy.

We did so in three steps. First, we asked how privacy has been defined in the literature. After reviewing the conceptualization of privacy as a right, commodity, state, or control, it became clear that there is no single empirical measure of privacy that could work across all these lenses because they attend to fundamentally different aspects of the privacy phenomena. A choice then naturally arose. We could develop parallel streams of empirical research, each adopting a different conceptualization of privacy and a different empirical measure it entails. While doing so allows privacy scholars to stay true to their normative beliefs, it also fragments privacy research and, as a result, exacerbates the already incendiary debate. Alternatively, instead of measuring "privacy" itself, we could seek a *proxy* that has generalizable value across normative definitions, so as to allow the discovery of empirical insights that are valuable regardless of one's belief of "what privacy is."

The second question we asked was whether there exists such a measurable proxy of privacy and, if so, what it is. To answer this question, we first noted an important condition such a proxy must satisfy. As discussed earlier, different normative concepts of privacy give primacy to different privacy-related constructs—e.g., privacy regulation is prominent when privacy is viewed as a right, while cost/benefit analysis is salient when privacy is considered a commodity. If a proxy of privacy were to have generalizable value across conceptualizations, it would have to be substantively linked to constructs that are prominent under different conceptualizations of privacy. In other words, to identify the proxy, we should seek a construct that is at the "center" of privacy-related constructs that have been studied. Based on a review of the literature, we identified *privacy concern* (e.g., beliefs, attitudes, perceptions) as this central construct. It is both measurable (e.g., Smith et al., 1996) and richly connected to many other privacy-related constructs, from privacy regulations to trust, from privacy awareness to people's behavioral reactions, as summarized in the Antecedents → Privacy Concerns → Outcomes (APCO) macromodel first proposed in our 2011 article and later extended in the enhanced APCO model (Dinev et al., 2015).

Once we substantiated privacy concerns as a proxy of privacy that is generalizable across normative concepts, the next step was to ask whether privacy scholars could build an overarching theory about privacy that is similarly generalizable. This was the third and final question of the 2011 article. We started by noting a key challenge facing the generalization of privacy theories. Just like how different conceptualizations of privacy give primacy to different privacy-related constructs, different *contexts* tend to make people ascribe to different normative notions about privacy (e.g., government surveillance typically triggers privacy-as-a-right views) and thus animate different privacy-related constructs and relationships. If we overlooked the importance of context in empirical research, we would either be confronted with inconsistent findings stemming from studies in different contexts, or have studies that essentially "fold" a multitude of contexts into one "average" finding that lacks precision in any specific context. Our review of the literature suggested that this context deficit was widely prevalent back then. Our answer to the final question was thus a call for building context-contingent theories that explicate the mechanisms through which contextual factors influence privacy-related constructs. To facilitate this development, we identified key contextual factors in need of further study, from the types of information to the technological applications involved.

What We Did Right

Our 2011 article was published at an opportune time when there was no end in sight for the normative debate about what privacy is, yet an urgent need for researchers and practitioners to understand why some forms of data collection and use seem largely acceptable to people while others would trigger privacy outcries. It was with this backdrop that we called for a turn from a normative to an empirical focus in privacy research. We argued that, by studying a measurable proxy with generalizable value across different privacy conceptualizations, researchers could unveil useful empirical insights without necessarily engaging with the normative question of what privacy is. Indeed, empirical studies on information privacy have flourished since the publication of our article, both within and beyond IS (see Acquisti et al., 2016; Acquisti et al., 2020; Baruh et al., 2017; and Popovic et al., 2017, for privacy literature reviews in economics, marketing, communication, and IS). Now looking back, we speculate that this call for a normative-to-empirical turn was the most important reason behind the impact of our article on privacy research.

Another reason our 2011 article has been highly cited might reside in the two foci we emphasize in our analysis of the literature: (1) level of analysis (i.e., individual, group, organizational, and societal), and (2) treatment of context. The article was published

during a time when universalism and "grand" theorizing gradually gave way to a more nuanced, situational, and contingent view of theorizing, both within IS and beyond, as highlighted by seminal work on multilevel analysis (Kozlowski & Klein; 2000; Burton-Jones & Gallivan, 2007) and context-contingent theorizing (Johns, 2006; Hong et al., 2014). This trend matched neatly with the growing realization among privacy scholars that privacy phenomena in general reflect multifaceted interests that could vary considerably from one level of analysis to another, from one situation to another, and from one individual to another (Acquisti et al., 2015; Nissenbaum, 2010). The two trends combined together promoted a flexible view of how privacy theories could function across situations. A decade later today, the privacy literature has become much richer on multilevel analysis (Bélanger & James, 2020) and context-contingent theorizing (Xu & Zhang, 2022a), echoing the recommendations outlined in our 2011 article.

What We Would Have Done Differently

With the benefit of hindsight, it became clear to us that we should have included a review of the technical development of privacy assurance techniques, a notable example of which was data anonymization, in our 2011 article. When we wrote the article, there was an abundance of ongoing research in computer science on how to design algorithms that "anonymize" a dataset of personal information by, on the one hand, preventing any record from being linked to an individual while, on the other hand, retaining the value of the dataset for purposes such as statistical analysis (cf. review by Fung et al., 2010). This stream of technical privacy research dates back to the 1980s, when techniques such as random-noise insertion (Beck, 1980; Denning & Schlorer, 1983) were developed to enable statistical analysis over a dataset without accessing individual data records. It gained renewed interest in the 2000s, perhaps thanks to HIPAA's requirement of data anonymization. Sweeney (2002), for example, developed algorithms to achieve *k*-anonymity, meaning that each record in the dataset should be made indistinguishable from at least $k - 1$ other records.

When attempting to recall why we excluded technical research on data privacy from our 2011 review, we found an old draft noting our collective belief (back then) that the technical notion of privacy was an oversimplification of people's privacy desires and needs in reality. As mentioned in the 2011 article, even though data anonymization "may enable privacy control," "many other avenues to such control also exist." We reasoned that, since privacy is so amorphous and fluid across situations, a rigid technical notion like *k*-anonymity may not be flexible enough to address people's dynamic privacy needs that arise. What we did not anticipate was the speed at which the developments of privacy assurance techniques gained traction in practice. In the European Union (EU), the European Commission (2014) specifically analyzed how different data anonymization algorithms could fit the requirements of the EU General Data Protection Regulation (GDPR). In the U.S., a variety of state agencies, like the Texas Department of State Health Services (2012), applied data anonymization before publishing their data. Differential privacy, as one of the most popular privacy assurance techniques, was adopted in the 2020 U.S. Census, triggering intense debates on how to balance privacy protection with the accuracy of census results (boyd & Sarathy, 2022). Almost every iOS or Android device today has differential privacy built into the operating system, protecting the privacy of users while allowing Apple or Google to collect data in support of functions such as autocomplete and autocorrect for text entry (Greenberg, 2016).

To us, a remarkable moment happened in 2016, when Apple decided to devote several minutes in the keynote of their widely publicized product-release event to explain the idea of differential privacy to a largely non-technical audience. This clearly spoke to how the rapid development of privacy assurance techniques had grown from an academic pursuit to become a quintessential part of the public discourse about privacy. Just like one cannot ignore the capability or limitation of cryptography when studying cybersecurity, privacy scholars today need to treat the rapid technical development of data protection as part of the privacy phenomena rather than a literature so distal that it can be safely ignored. This will become evident at the end of this article, after we review how the privacy phenomena have morphed during the past decade.

The Past Decade: A Sea Change in Information Privacy

Now looking back, we found the publication of our article to coincide with a turning point in IT history. If the two decades between 1990 and 2010 were marked by the rise of the internet and the resulting ease of web-based data collection, then the decade since 2010 would be best characterized by the rise of AI and the emerging possibilities in deriving value from the collected data. In this section, we first briefly review the deep-learning revolution that happened during the past decade, before explaining how this revolution led to a sea change in the social, legal, and technological landscape of privacy development.

Technological Backdrop: The Rise of AI

In the past decade, one of the most profound advancements in IT was the advent of deep learning (Goodfellow et al., 2016) and, along with it, the possibility that AI could revolutionize many aspects of work and life (Berente et al., 2021). A common myth about deep learning is that it represents a revolution in the algorithms used to learn patterns from data. While considerable improvements have been made to learning algorithms in recent years (e.g., Goodfellow et al., 2014), all the critical ingredients of deep learning, from the architecture of deep neural networks (Rosenblatt, 1958) to algorithms like back-propagation (Rumelhart et al., 1986), were invented decades ago. As Goodfellow et al., noted, "the learning algorithms reaching human performance on complex tasks today are nearly identical to the learning algorithms that struggled to solve toy problems in the 1980s" (2016, p.19). A key contributor to the resurrection of deep learning in the past decade was that these algorithms were finally provided with the huge amount of data they need to succeed. For example, the curation of ImageNet (Deng et al., 2009), a collection of more than 14 million images, is often credited as a key enabler for the deep-learning revolution in visual recognition. As famously stated by the creator of ImageNet, Fei-Fei Li, "big data would change the way machine learning works", as "data drives learning" (Parloff, 2016).

From this perspective, it was not fortuitous that the deep-learning revolution followed the third era of privacy evolution. After two decades of ubiquitous information collection, by the time our article was published in 2011, many organizations had already accumulated massive amounts of data about their customers, employees, etc. As estimated by McKinsey, "by 2009, nearly all sectors in the US economy had at least an average of 200 terabytes of stored data per company with more than 1,000 employees" (Manyika et al., 2011). Despite the massive amount of collected data, organizations were unable to fully capture the business value in data because their ability to do so was limited by the capacity of learning algorithms available back then (e.g., kernel machines—Cortes & Vapnik, 1995). Yet the data were not collected in vain, as it was exactly the availability of such "big data" that finally unleashed the power of deep learning and, in turn, allowed businesses to start unlocking the value in the data they collected. The unlocked value generated immense enthusiasm toward AI, and often surprised even the developers of the algorithms themselves (Lehman et al., 2020). Nonetheless, what is of value to businesses could very well be a minefield of privacy concerns for their customers and employees. As we discuss next, the "rise of AI" fundamentally altered the social, legal, and technological landscape of privacy development during the past decade.

Recent Privacy Outcries: from Data Flows to Inferred Knowledge

Today, it is rare to have a public discourse about privacy without mentioning the Cambridge Analytica scandal. Cambridge Analytica was a British consulting firm that harvested the personal data of about 87 million Facebook users through a personality test app "this is your digital life." The collected data were then used to learn the users' psychological profiles, so as to target potential swing voters through political ads for campaigns such as the U.S. presidential campaign of Donald Trump, the U.S. presidential primary campaign of Ted Cruz, etc. Cambridge Analytica's data practice was first reported in 2015 (Davies, 2015), before gaining widespread attention in 2018 after Trump's election (Rosenberg et al., 2018).

Another widely publicized privacy incident involved the startup company, Predictim, which offered a service that screens potential babysitters by using natural language processing (NLP) algorithms to scan their public social-media profiles for "risk indicators" such as aggressiveness, drug use, and harassment (Harwell, 2018). After an exposé on *Washington Post*'s front page, an outcry ensued, with "criticism centered around privacy" (Hosanagar, 2019). Within a week, Facebook, Twitter, and Instagram, effectively banned Predictim's access to social media data. Predictim ceased operation soon after.

When we juxtapose these recent privacy outcries with the ones discussed earlier, an intriguing difference emerges. Unlike the earlier privacy incidents, which centered around data flows, these recent incidents are more about knowledge that could be *learned* or *inferred* from collected data. This point was exemplified by a comment in the earliest report of Cambridge Analytica: "it's one thing for a marketer to try to predict if people like Coke or Pepsi," and "another thing for them to predict things that are much more central to our identity" (Davies, 2015). In other words, the recent privacy outcries were less about unauthorized access to personal data and profiles and more about how companies may *learn* from the collected data about people's personalities and intentions and then *manipulate* their decisions and activities. Privacy concerns were centered more on the knowledge drawn from the data—e.g., a prediction model, a scoring mechanism, etc.—rather than the data flow itself.

The Challenge Brought by the Rise of AI

When the target of privacy concerns shifts from *data* to *knowledge*, it brings into question whether we could continue the canonical practice of protecting privacy by regulating data flows, because doing so requires us to trace the *provenance* of information when and how data are transformed to knowledge. For example, in the earlier outcries, one could always point to a part of the data flow (e.g., certain variables like ZIP code, gender, and date of birth) as the culprit for privacy issues, implying that blocking that part of the data flow would effectively alleviate one's privacy concerns. If we hope to do the same in addressing knowledge-centered privacy concerns, then we would have to clearly trace how each piece of knowledge is learned from which parts of the data. Only by doing so would we be able to address a knowledge-centered concern by blocking parts of the data flow.

This is where the rise of AI brought a fundamental challenge to our privacy research. As the inference of knowledge from data is mostly done through learning algorithms today, the increasing complexity of these algorithms makes it nearly impossible for humans, even experts in AI, to explain which parts of the data are used by an algorithm to learn which part of the knowledge in its output (Rahwan et al., 2019). In other words, the provenance of information, which used to be clear (e.g., how AOL collects and then releases query logs), is now murky. The Cambridge Analytica scandal is a notable example as, even today, it remains unclear how their algorithm managed to learn the link between psychological profiles and voting tendencies (Hu, 2020), and the extent to which each user's Facebook profile was used in the learning process (Gibney, 2018).

Without a clear provenance from data to knowledge, a perplexing issue abounds for privacy protection. When an individual is concerned about learning certain knowledge that cannot be specifically tied to the data of that individual, is regulating data flows still a viable recourse for addressing privacy concerns? For example, the machine learning literature has repeatedly noted the possibility of inferring a more detailed profile of an individual based solely on data about people who "look like" the individual (e.g., Yang et al., 2016; Kang et al., 2019). In this case, if we are concerned about the profiles being inferred about us by an organization, yet no data actually flow from us to the organization, what shall we ask the organization to protect? Should the organization protect our "privacy" by limiting data flows from those people who happen to look like us (whom we likely do not even know)? What if those people who look like us actually wanted to share their data with the organization? As the link between data and knowledge becomes murkier, so does the link between privacy concerns and privacy protection. See Figure 1 for an illustration of the emerging *knowledge*-centered privacy concerns.

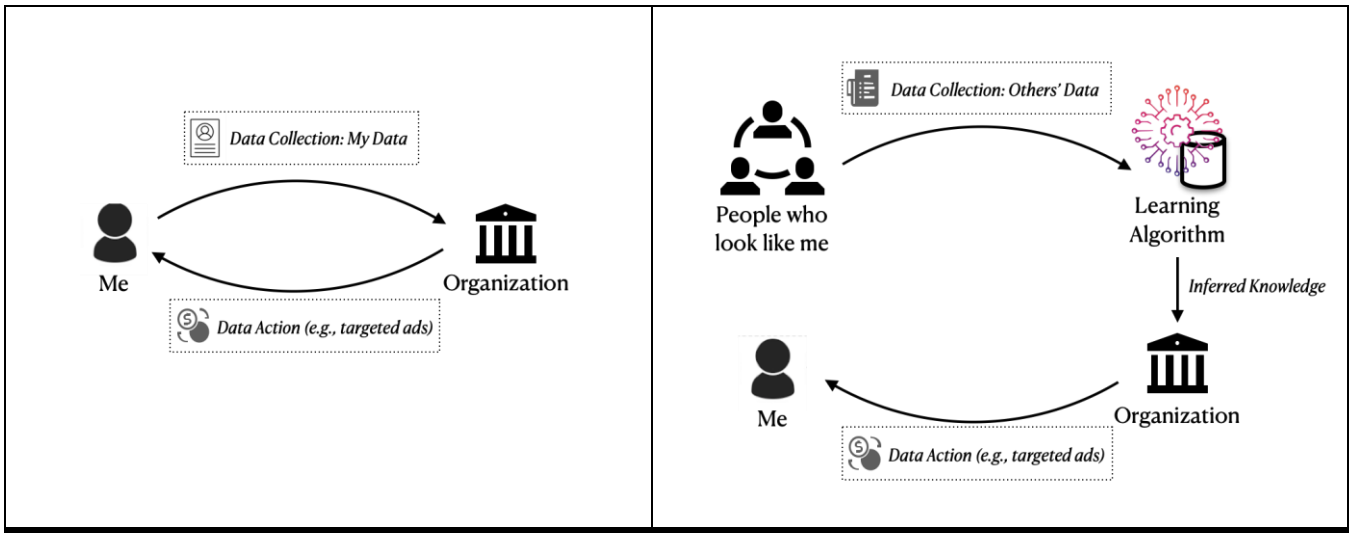

Figure 1. (a) Data-Centered Privacy Concerns (left), and (b) Knowledge-Centered Privacy Concerns (right)

Recent Legal and Technical Developments

The challenge brought by this perplexing issue is evident in the current legal landscape. Consider privacy laws like GDPR and the California Consumer Privacy Act (CCPA). Unlike HIPAA (e.g., its safe harbor provision), which focuses on regulating data flows, both GDPR and CCPA explicitly acknowledge consumers' privacy interests in preventing *inference*, which creates

knowledge "through deduction or reasoning rather than mere observation or collection from the data subject" (Wachter & Mittelstadt, 2019:515). Yet neither clarifies the extent to which consumers can limit the inference process and, critically, how to do so (Holder, 2020). This challenge was evident in the European Commission's (2014) review of potential mechanisms for privacy protection, which found that none of the seven reviewed could definitively eliminate the risk of inference.

The development of privacy assurance techniques centered around the exact same issue of *inference* over the past decade. Many data anonymization algorithms studied before, like the aforementioned *k*-anonymity, were found to be ineffective in preventing the inference of sensitive knowledge (European Commission, 2014). Further, it became evident that no technical solution can limit what an organization may learn through inference without first knowing what background knowledge is possessed by the organization (Kifer & Machanavajjhala, 2011). Since it is always possible for an organization to acquire additional background knowledge by joining different datasets in the future, this essentially implies that it is technically infeasible to limit learning through inference at the time of data collection.

With this backdrop, the vast majority of development in privacy assurance techniques turned to the notion of differential privacy (Dwork et al., 2006) in the past decade. Recognizing the infeasibility of limiting what an organization may learn through inference, the key idea of differential privacy is to limit what *additional* knowledge an organization may learn by virtue of having an individual's data shared with the organization. In other words, while differential privacy cannot prevent an organization from learning any specific knowledge, what it can do is to assure an individual that, whatever the organization learns, it can learn anyway even without access to the individual's data. One of the most attractive features of differential privacy is that this guarantee can be afforded regardless of what background knowledge the organization might possess. The question, on the other hand, is whether it resolves people's privacy concerns on the inferred knowledge, or simply limits the scope of such concerns by proclaiming that an individual cannot allege privacy interests in a piece of knowledge that can be learned without the individual's data.

Moving Forward: Two Questions for Future Research

As summarized in Table 1, the emphasis on privacy governance before 2010 was about regulating data flows. As we have shown, the center of the privacy phenomena has now morphed from the flow of personal data to the inference of knowledge. Going forward, we suggest that there are two important questions brought up by (1) the shift from data-centered privacy concerns to knowledge-centered privacy concerns, and (2) the rapid adoption of privacy assurance techniques in data protection.

Table 1. Evolution of Information Privacy (adapted from Smith et al., 2011)

Period	Characteristics of IT and privacy development	Emphasis on privacy governance
First era (1961 -1979)	The rise of information privacy as an explicit social, political, and legal issue. Early recognition of potential dark sides of the new technologies, formulation of the fair information practices (FIP) framework, and the establishment of government regulations.	***Back then:*** ***Focus:*** to point to a specific part of the data flow as the culprit for privacy issues ***Goal:*** regulating *data flows* to prevent data overcollection, unauthorized release, or misuse
Second era (1980 - 1989)	The rise of computer and network systems, and database capabilities. More and more government regulations designed to channel the new technologies into FIP.	
Third era (1990 - 2010)	The rise of the internet, social media, and the terrorist attack of 9/11/2001 dramatically changed the landscape of information exchange.	
Fourth era (2011 - present)	The rise of AI and machine learning (especially deep learning) shifted the locus of privacy concerns from the flow of personal information to how learning algorithms can infer knowledge from massive data collections.	***In recent times:*** ***Focus:*** to trace how each piece of knowledge is learned from which parts of the data ***Goal:*** regulating *inference* to prevent harmful impacts of learning algorithms on individuals' autonomy and agency

Question #1: Is My Privacy Still about My Private Information?

A core idea in our 2011 article was to turn the normative debate about *what privacy is* into a focus of empirically understanding people's privacy concerns. However, as the technological landscape has shifted over the last decade, so too have the focus and locus of privacy concerns. It is now common for privacy outcries to stem from people's concerns about the *knowledge* learned by an organization from large amounts of collected data. Unlike a specific data record, this knowledge does not "belong to" any particular individual. This fundamental change has profound implications for privacy research and theoretical development. Many of our core assumptions on data-centered privacy concerns and the associated research and theory may be outdated. This shift ostensibly challenges the relevance of the individualistic view of privacy concerns and invites the question of who should set the rules that govern the inference process. Ironically, this appears to bring us right back to the era of normative debates, as only when we settle "what privacy is" can we convince each other "who should set the rules."

While normative debates on inference are ongoing (Holder, 2020), we suggest that there are many new questions for IS researchers to tackle. To this end, we would like to reiterate our call in the 2011 article for a shift away from having the individual as the salient unit of analysis. When knowledge about an individual can be learned from the data of a population, protecting the privacy of an individual requires an understanding of the distributional configuration of collective privacy concerns across the population. As discussed earlier, the advances in AI have demonstrated the possibility for an organization to infer knowledge about us so long as *some other people* (who resemble us in certain aspects) are willing to share their data with the organization. To study whether an organization can actually do so in practice, and how accurate the inference might be, we cannot only examine an individual's concerns about the knowledge inferred from his or her own data. Instead, we need to examine whether other people are concerned about the same knowledge inference; which subpopulation may be more or less concerned than others; and whether there would be bias in the knowledge inferred by the organization from those who are willing to share their data. This last issue clearly also depends on the algorithm used for knowledge inference. In sum, privacy scholars may need to be attentive to both the distributional configuration of collective privacy concerns and the design of learning algorithms in order to launch a proper empirical examination of privacy concerns over the inference of knowledge.

Question #2: What is the Role of Privacy Assurance Techniques in Data Protection?

Over the past decade, rapid developments of privacy assurance techniques have gained traction in practice. Apple and Google now use differential privacy to anonymize our text entries on smartphones before collecting them to their servers. Government agencies also use anonymization techniques to process our records before releasing them to the public. The burst of innovation around privacy assurance techniques not only "point at the possibility of protecting individuals' privacy while allowing beneficial analytics to advance," but also "portend a world where privacy by design is possible without undermining the value of data" (Acquisti et al., 2022, p. 272).

However, everyday people by and large do not understand what these techniques do with their data and how these techniques protect their privacy (Cummings et al., 2021). Even experts do not agree on the societal impacts of these techniques, as evidenced by the controversies surrounding the use of differential privacy in the 2020 U.S. Census (boyd & Sarathy, 2022). Further, recent academic research has identified potential negative impacts of using privacy assurance techniques like differential privacy on the detection of health disparity (Xu & Zhang, 2022b), the fairness of resource allocation (Steed et al. 2022; Tran et al., 2021), or even the future of democratic representation (Kenny et al., 2021). Other studies have challenged the adequacy of how leaders in the high-tech industry implemented these techniques (Tang et al., 2017). With privacy assurance techniques becoming an increasingly prevalent part of data exchange in today's world, future research needs to treat them as such and attend to both the technical design of these techniques and their societal impacts. It is our belief that IS researchers are being presented with a golden opportunity to demonstrate the value of sociotechnical thinking in the next generation of privacy research. As an interdisciplinary field that combines behavioral, econometric, analytical, and computational methods, the IS research community is well-positioned to contribute to the discourse, both on the technical design side, by bringing in the behavioral understanding and societal expectation of privacy, and the human side, by unfolding the technological black box.

Final Remarks

When we started writing this editorial, we visited Google Trends to understand how the public's search interest in privacy has changed over the last decade. Figure 2 shows the trend from 2004 to 2022. Interestingly, the publication of our 2011 article coincided with a turning point when the interest in privacy stopped the declining trend since 2004 and started rising all the way to today. We feel that this chart is a fitting conclusion to this editorial, as it vividly shows how privacy is not a stale, frozen-in-time, concept but very much alive and morphing. The locus of people's privacy concerns today has fundamentally shifted from what they were a decade ago. As privacy scholars, the decision is ours on whether to drive in the rear-view mirror, or to confront the emerging privacy issues brought forth by the rise of AI, the rapid adoption of privacy assurance techniques, and, most importantly, the changing social, legal and technical landscape surrounding the evolution of privacy.

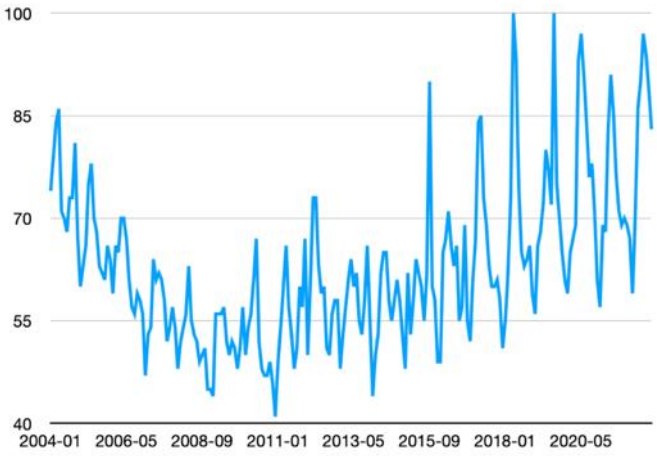

Figure 2. Change of Search Interest in Privacy (from Google Trends)

Acknowledgments

Developing a *theory and review* paper is harder than we thought and more rewarding than we could have ever imagined. Our original 2011 paper would not exist without the inspiration and contribution of our co-author H. Jeff Smith, who passed away in October 2019. Additionally, we are immensely grateful to the original editorial review team (Lynne Markus, Paul Pavlou, and three anonymous reviewers) for their support and constructive feedback on our paper. We sincerely thank Andrew Burton-Jones for his helpful comments on the editorial. We also owe an enormous debt of gratitude to the 2021 Impact Award selection committee (Andrew Burton-Jones, Arun Rai, Paulo Goes, Virpi Kristiina Tuunainen, and Galit Shmueli) for their confidence and their belief in our work.

Associate Editor Reflections—Paul Pavlou, University of Houston

When I served as AE for Smith et al. (2011), could I have predicted that it would go on to become a classic? I think the answer is probably yes! While the *MISQ* Impact Award did not exist at the time of the paper's publication, I think it had all the necessary ingredients to become a classic. The paper tackled an important problem, offered a new theoretical perspective, and opened an exciting research agenda for privacy. It was a pleasure to serve as AE on such an influential paper.

As the authors note in this editorial, the salience of privacy has skyrocketed since the original publication with the advent of advanced technologies, such as artificial intelligence (AI) and machine learning (ML). As the authors rightly point out, researchers and practitioners must deeply understand the changing social, legal, and technical landscape surrounding the evolution of privacy, especially due to the rise of AI/ML. As information systems (IS) researchers, we need to play our part in helping our societies to

create the privacy regulation needed to ensure the responsible and ethical use of new powerful technologies and effectively protect privacy in cyberspace.

Against this increasingly troubling backdrop of how new powerful technologies can infringe on any individual person's privacy, often without even using the individual's own personal data (given the ability to infer knowledge about the individual from other massive sources of data), the original article by Smith et al. (2011) still provides a useful paradigm to inform the study of privacy by identifying a set of key contextual factors, including types of information and the technological applications involved, to specify the mechanisms through which these contextual factors affect privacy-related constructs. In doing so, the Smith et al. article still sets the stage for future research to examine how inferred knowledge from other sources can have an impact on one's own individual privacy and how privacy assurance technologies may help protect people's privacy in the rise of AI and advanced AI/ML systems.

Like 10 years ago, we are at a crossroads again. We must work collectively to ensure that our societies will be safer and more protected 10 years from now, as technology is bound to proliferate in formidable ways in the next decade. Understanding the contextual factors and how they affect privacy-related constructs would be an important means to develop effective privacy assurance tools to combat the rapidly growing power of AI/ML systems to make inferences about people's privacy. Accordingly, it is important for IS researchers to prioritize the design and testing of privacy assurance techniques that would help protect individual privacy in the backdrop of the rise of AI.

Senior Editor Reflections—M. Lynne Markus, Bentley University

Reflecting on the 2021 Impact Award is particularly meaningful for me because I edited and accepted Smith et al.'s 2011 paper as the then-senior editor of *MISQ*'s Theory and Review Department. As Xu and Dinev's editorial makes clear, writing excellent theory and review articles offers exceptional rewards (citations, recognition for intellectual leadership, and a platform for a productive future career of empirical and theoretical research) in return for the very hard work involved.

Less frequently remarked is the very hard work involved in editing such articles. This gives me the opportunity to reflect on the rewards to our field (and its flagship journals like *MISQ*) for editing and publishing important theory and review papers. As Xu and Dinev note here, the privacy domain, like virtually every other aspect of information systems, has changed radically in the decade since their article was published. Despite the many things the authors "got right" and their many contributions to the IS literature and public policy debates, new questions about information privacy must now be asked, and new concepts and theories must be developed to answer them. This inevitable consequence of ongoing societal transformation does not reduce the value of a seminal theory and review paper like Smith et al. (2011). By contrast, it reveals its value even more clearly.

Smith et al.'s 2011 paper encapsulated the then-existing knowledge about the phenomenon and challenges of information privacy, and it offered a well-argued case for viewing this important concept in an innovative way. Only by rereading papers like this one, can we in our field see clearly how rapidly and fundamentally things have changed. Insights like this provide us with the motivation and the justification for periodically reviewing and refreshing our basic concepts, assumptions, and theories. Empirical findings about the sociotechnical world are unlikely to endure. But the accumulation of knowledge requires deep awareness of what we have learned and still need to know.

I congratulate the authors on their well-deserved award. And I congratulate the *MISQ* editorial board and award committee for inaugurating the Impact Award and selecting Smith et al.'s 2011 theory and review paper as its first recipient.

References

Acquisti, A., Brandimarte, L., & Hancock, J. (2022). How privacy's past may shape its future. *Science*, *375*(6578), 270-272.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*, *30*(4), 736-758.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015). Privacy and human behavior in the age of information. *Science*, *347*(6221), 509-514.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, *54*(2), 442-92.

Arrington, M. (2006). *AOL Proudly Releases Massive Amounts of Private Data*. TechCrunch. https://techcrunch.com/ 2006/08/06/aol

European Commission. (2014). *Article 29: Data Protection Working Party*. https://ec.europa.eu/justice/article-29/documentation/ opinion-recommendation/files/2014/wp216_en.pdf

Barbaro, M. and Zeller, T. (2006). A face is exposed for AOL Searcher No. 4417749. *New York Times*. https://www.nytimes.com/2006/08/09/technology/09aol.html

Baruh, L., Secinti, E., & Cemalcilar, Z. (2017). Online privacy concerns and privacy management: A meta-analytical review. *Journal of Communication*, 67(1), 26-53.

Beales III, J. H., & Muris, T. J. (2008). Choice or consequences: Protecting privacy in commercial information. *University of Chicago Law Review*, 75, 109-135.

Beck, L. L. (1980). A security mechanism for statistical database. *ACM Transactions on Database Systems, 5*(3), 316-338.

Bélanger, F., & James, T. L. (2020). A theory of multilevel information privacy management for the digital era. *Information Systems Research*, *31*(2), 510-536.

Bennett, C. J. (1995). *The political economy of privacy: A review of the literature.* Center for Social and Legal Research.

Berente, N., Gu, B., Recker, J., & Santhanam, R. (2021). Managing artificial intelligence. *MIS Quarterly*, *45*(3), 1433-1450.

boyd, d., & Sarathy, J. (2022). Differential perspectives: Epistemic disconnects surrounding the US Census Bureau's use of differential privacy. *Harvard Data Science Review, Special Issue 2*, https://doi.org/10.1162/99608f92.66882f0e

Brin, D. (1998). *The transparent society: Will technology force us to choose between privacy and freedom?* Addison-Wesley.

Burton-Jones, A., & Gallivan, M. J. (2007). Toward a deeper understanding of system usage in organizations: A multilevel perspective. *MIS Quarterly*, *31*(4), 657-679.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273-297.

Cummings, R., Kaptchuk, G., & Redmiles, E. M. (2021). "I need a better description": An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security* (pp. 3037-3052).

Davies, H. (2015). *Ted Cruz using firm that harvested data on millions of unwitting Facebook users*. The Guardian. https://www.theguardian.com/us-news/2015/dec/11/senator-ted-cruz-president-campaign-facebook-user-data

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 248-255).

Denning, D. E., & Schlorer, J. (1983). Inference controls for statistical databases. *Computer*, *16*(07), 69-82.

Dinev, T., McConnell, A., & Smith, H. J. (2015). Informing privacy research through information systems, psychology, and behavioral economics: Thinking outside the "APCO" box. *Information Systems Research*, *26*(4), 639-655.

Dwork, C., McSherry, F., Nissim, K., & Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In Theory of cryptography conference (pp. 265-284). Springer.

Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, *42*(4), 1-53.

Gibney, E. (2018). The scant science behind Cambridge Analytica's controversial marketing techniques. *Nature*, March 29, 2018.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... and Bengio, Y. (2014). Generative adversarial nets. In *NIPS Proceedings: Advances in Neural Information Processing Systems*

Greenberg, A. (2016). Apple's "differential privacy" is about collecting your data—but not your data. *Wired.* https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/

Harwell, D. (2018). Wanted: The "perfect babysitter." Must pass AI scan for respect and attitude. *Washington Post.* https://www.washingtonpost.com/technology/2018/11/16/wanted-perfect-babysitter-must-pass-ai-scan-respect-attitude/

Holder, A. E. (2020). What we don't know they know: What to do about inferences in European and California data protection law. *Berkeley Technology Law Journal, 35,* 1331-1364.

Hong, W., Chan, F. K., Thong, J. Y., Chasalow, L. C., & Dhillon, G. (2014). A framework and guidelines for context-specific theorizing in information systems research. *Information Systems Research, 25*(1), 111-136.

Hosanagar, K. (2019). AI Assistants or Digital Despots? *The Pennsylvania Gazette.* https://thepenngazette.com/ai-assistants-or-digital-despots/

Hu, M. (2020). Cambridge Analytica's black box. *Big Data and Society*, *7*(2), https://doi.org/10.1177/2053951720938091

Johns, G. (2006). The essential impact of context on organizational behavior. *Academy of Management Review*, *31*(2), 386-408.

Kang, S., Hwang, J., Lee, D., & Yu, H. (2019). Semi-supervised learning for cross-domain recommendation to cold-start users. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 1563-1572).

Kenny, C. T., Kuriwaki, S., McCartan, C., Rosenman, E. T., Simko, T., & Imai, K. (2021). The use of differential privacy for census data and its impact on redistricting: The case of the 2020 US Census. *Science Advances, 7*(41), Article eabk3283.

Kifer, D., & Machanavajjhala, A. (2011). No free lunch in data privacy. In *Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data* (pp. 193-204).

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein and S. W. J. Kozlowski (Eds.), *Multilevel theory, research, and methods in organizations: Foundations, extensions, and new directions* (pp. 3–90). Jossey-Bass.

Lehman, J., Clune, J., Misevic, D., Adami, C., Altenberg, L., Beaulieu, J., ... and Yosinski, J. (2020). The surprising creativity of digital evolution: A collection of anecdotes from the evolutionary computation and artificial life research communities. *Artificial Life*, *26*(2), 274-306.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. https://www.mckinsey.com/business-functions/ mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation

Margulis, S. T. (1977). Conceptions of privacy: Current status and next steps. *Journal of Social Issues*, *33*(3), 5-21.

Milberg, S. J., Smith, H. J., & Burke, S. J. (2000). Information privacy: Corporate management and national regulation. *Organization Science*, *11*(1), 35-57.

Neate, R. (2018). Over $119 bn wiped off Facebook's market cap after growth shock. The Guardian. https://www.theguardian.com/technology/2018/jul/26/facebook-market-cap-falls-109bn-dollars-after-growth-shock

Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.

Parloff, R. (2016). From 2016: Why deep learning is suddenly changing your life. *Fortune.* https://fortune.com/longform/ai-artificial-intelligence-deep-machine-learning/

Popovic, A., Thong, J. Y. L., & Wattal, S. (2017). Information Privacy. MIS Quarterly Research Curations. *MIS Quarterly*, https://www.misqresearchcurations.org/blog/2017/6/30/information-privacy

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., ... and Wellman, M. (2019). Machine behaviour. *Nature*, *568*(7753), 477-486.

Rosenberg, M., Confessore, N., & Cadwalladr, C. (2018). How Trump consultants exploited the Facebook data of millions. *The New York Times*. https://www.nytimes.com/2018/03/17/us/politics/cambridge-analytica-trump-campaign.html

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386-408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533-536.

Sandhu, R. S., & Samarati, P. (1994). Access control: Principle and practice. *IEEE Communications Magazine*, *32*(9), 40-48.

Sejnowski, T. J. (2018). *The deep learning revolution*. MIT Press.

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly*, *35*(4), 989-1015.

Smith, H. J., Milberg, S. J., & Burke, S. J. (1996). Information privacy: Measuring individuals' concerns about organizational practices. *MIS Quarterly*, *20*(2), 167-196.

Federal Register (2000). *Standards for privacy of individually identifiable health information* (65 Fed. Reg. 82462). https://www.federalregister.gov/documents/2000/12/28/00-32678/standards-for-privacy-of-individually-identifiable-health-information

Steed, R., Liu, T., Wu, Z. S., & Acquisti, A. (2022). Policy impacts of statistical uncertainty and privacy. *Science*, *377*(6609), 928-931.

Sweeney, L. (1997). Guaranteeing anonymity when sharing medical data, the Datafly System. In *Proceedings of the AMIA Annual Fall Symposium*.

Sweeney, L. (2002). *k*-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *10*(05), 557-570.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818-2826).

Tran, C., Fioretto, F., van Hentenryck, P., & Yao, Z. (2021). Decision making with differential privacy under a fairness lens. In *Proceedings of the International Joint Conferences on Artificial Intelligence Organization* (pp. 560-566).

Tang, J., Korolova, A., Bai, X., Wang, X., & Wang, X. (2017). *Privacy loss in Apple's implementation of differential privacy on MacOS 10.12.* Available at https://arxiv.org/abs/1709.02753

Texas Department of State Health Services (2012). *Texas hospital inpatient discharge public use data file*. https://www.dshs.state.tx.us/thcic/hospitals/UserManual1Q2012.pdf

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... and Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of big data and AI. *Columbia Business Law Review*, 494-620.

Westin, A. F. (1967). *Privacy and Freedom.* Atheneum.

Xu, H., & Zhang, N. (2022a). From contextualizing to context theorizing: Assessing context effects in privacy research. *Management Science*, *68*(10), 7383-7401.

Xu, H., & Zhang, N. (2022b). Implications of data anonymization on the statistical evidence of disparity. *Management Science*, *68*(4), 2600-2618.

Yang, Z., Cohen, W., & Salakhudinov, R. (2016). Revisiting semi-supervised learning with graph embeddings. *Proceedings of the International Conference on Machine Learning* (pp. 40-48).

Zeller, T. (2006). AOL technology chief quits after data release. *New York Times.* https://www.nytimes.com/2006/08/21/technology/21cnd-aol.html

About the Authors

Heng Xu is a professor of information technology and analytics in the Kogod School of Business at the American University in Washington, D.C., where she also serves as the director of the Kogod Cybersecurity Governance Center. Her current research focuses on information privacy, data ethics, responsible AI, and algorithmic fairness. Her scholarly work has appeared in premier outlets across different fields such as information systems, human-computer interaction, and psychology, including *Information Systems Research*, *Management Information Systems Quarterly*, *Management Science*, *Proceedings of the ACM Conference on Human Factors in Computing Systems*, *Psychological Methods*, and many others.

Tamara Dinev is a professor and chair of the Department of Information Technology and Operations Management (ITOM) in the College of Business at the Florida Atlantic University in Boca Raton, Florida. Her research focuses on individual/employee behavior and multicultural aspects regarding computer security and information privacy. She has published in premier journals in the field of Information Systems, including *Management Information Systems Quarterly*, *Information Systems Research*, *Decision Sciences*, *Journal of the Association for Information Systems*, *Journal of Strategic Information Systems*, *European Journal of Information Systems*, *Information Systems Journal*, *Communications of the ACM*, and many others.