A Multi-Disciplinary Perspective for Conducting Artificial Intelligence-enabled Privacy Analytics: Connecting Data, Algorithms, and Systems

SAGAR SAMTANI, Department of Operations and Decision Technologies, Indiana University, Indiana, USA
MURAT KANTARCIOGLU, Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, TX, USA
HSINCHUN CHEN, Department of Management Information Systems, University of Arizona, AZ, USA

Events such as Facebook-Cambridge Analytica scandal and data aggregation efforts by technology providers have illustrated how fragile modern society is to privacy violations. Internationally recognized entities such as the National Science Foundation (NSF) have indicated that Artificial Intelligence (AI)-enabled models, artifacts, and systems can efficiently and effectively sift through large quantities of data from legal documents, social media, Dark Web sites, and other sources to curb privacy violations. Yet considerable efforts are still required for understanding prevailing data sources, systematically developing AI-enabled privacy analytics to tackle emerging challenges, and deploying systems to address critical privacy needs. To this end, we provide an overview of prevailing data sources that can support AI-enabled privacy analytics; a multi-disciplinary research framework that connects data, algorithms, and systems to tackle emerging AI-enabled privacy analytics challenges such as entity resolution, privacy assistance systems, privacy risk modeling, and more; a summary of selected funding sources to support high-impact privacy analytics research; and an overview of prevailing conference and journal venues that can be leveraged to share and archive privacy analytics research. We conclude this paper with an introduction of the papers included in this special issue.

CCS Concepts: • **Security and privacy → Human and societal aspects of security and privacy**; **Social aspects of security and privacy**;

Additional Key Words and Phrases: Privacy, artificial intelligence, data, analytics, systems, theories

ACM Reference format:
Sagar Samtani, Murat Kantarcioglu, and Hsinchun Chen. 2021. A Multi-Disciplinary Perspective for Conducting Artificial Intelligence-enabled Privacy Analytics: Connecting Data, Algorithms, and Systems. *ACM Trans. Manage. Inf. Syst.* 12, 1, Article 1 (March 2021), 18 pages.
https://doi.org/10.1145/3447507

This material is based upon work supported by the National Science Foundation under Grant Numbers OAC-1917117 (CICI), CNS-1936370 (SaTC CORE), CNS-1850362 (CRII SaTC), and DGE-2038483 (SaTC-EDU).
Authors' addresses: S. Samtani (corresponding author), Department of Operations and Decision Technologies, Indiana University, 1275 E. 10th St., Bloomington, Indiana 47405, USA; email: ssamtani@iu.edu; M. Kantarcioglu, Erik Jonsson School of Engineering and Computer Science, University of Texas at Dallas, 800 W. Campbell Rd., Richardson, TX 75080, USA; email: muratk@utdallas.edu; H. Chen, Department of Management Information Systems, University of Arizona, 1130 E. Helen St., McClelland Hall 430, Tucson, AZ 85721, USA; email: hsinchun@arizona.edu.

2158-656X/2021/03-ART1
https://doi.org/10.1145/3447507

1 INTRODUCTION

Privacy has long been a critical topic in society. Historically, privacy has focused on providing individuals and groups the ability to selectively restrict themselves or their information from consumption by others. However, the rapid proliferation of information technology has made maintaining privacy in the digital age a non-trivial task. Events such as the Facebook–Cambridge Analytica scandal, cyber-attacks, and data aggregation efforts by major technology providers (e.g., Amazon, Facebook, Google, Apple, and Twitter) have illustrated how fragile modern society is to privacy violations. Recent regulations such as the European Union General Data Protection Regulation (GDPR) in 2018 and the California Consumer Privacy Act (CCPA) in 2020 accelerated the attention and focus on enhancing and maintaining privacy. However, there remains a critical need for innovative technical research and solutions to help enhance the privacy of citizens in modern society.

To date, substantial progress has been made in various aspects of privacy research, including privacy concerns [5, 11, 14, 59], privacy controls [12, 22, 26], privacy risk [28], and privacy preservation [31]. These research streams have largely been driven by social science perspectives, including behavioral, economics, psychology, cognitive science, and others. Systematic efforts have also been undertaken by dedicated groups, including the Usable Privacy Project at Carnegie Mellon University and Robust Analytics Lab at American University (formerly at Pennsylvania State University). Despite significant progress in these initiatives, globally recognized entities such as the National Science Foundation (NSF) and National Science and Technology Council have indicated that significant opportunities remain in developing Artificial Intelligence (AI)-enabled privacy analytics to manage, analyze, and derive insight out of large quantities of data for high impact privacy applications for individuals, companies, regulators, and other stakeholders [38, 40]. Despite its significant promise, AI-enabled privacy analytics remains in its nascency. Multi-disciplinary efforts are required to manage prevailing data sources, systematically develop AI-enabled privacy analytics to tackle emerging challenges, and deploy systems to address critical privacy needs. To this end, we aim to make the following contributions in this article:

(1) First, we summarize prevailing data sources pertaining to privacy, including legal documents, social media platforms, internet browsing data, the Dark Web, data breach services, people search engines, and biometrics scholars and practitioners can consider when developing AI-enabled privacy analytics.
(2) Second, we provide a novel framework for scholars and practitioners to consider when developing novel AI-enabled privacy analytics. This framework emphasizes a multi-disciplinary approach and illustrates how data, analytics, and systems can be developed to support high-impact privacy research. For each major component of the framework, we provide several examples of promising future research directions.
(3) Third, we provide a summary of prevailing grant funding opportunities that can help support systematic AI-enabled privacy analytics research, education, and transition to practice efforts. This summary aims to provide an up-to-date (as of 2021) summary of options that can be considered when building out selected programs.
(4) Finally, we provide an overview of prevailing conference venues and journal outlets suitable for presenting, publishing, and archiving selected AI-enabled privacy analytics research activities. These venues can serve as excellent mechanisms for ensuring the long-term viability and health of the discipline.

The remainder of this article is organized as follows. First, we provide a comprehensive overview of prevailing data sources for AI-enabled privacy analytics. Second, we present the integrated AI-enabled privacy research framework. Third, we summarize selected NSF funding opportunities

that can support AI-enabled privacy research. Fourth, we present the summary of conference and journal venues amenable to privacy analytics research. The final section presents the papers in this special issue and concludes this work.

2 PREVAILING DATA SOURCES FOR AI-ENABLED PRIVACY ANALYTICS

Recent privacy regulations such as GDPR and CCPA aim to strengthen citizen's control (e.g., opt-out, right to access, consent) over the data they collect, process, store, and share about their customers. Oftentimes, the data in question is Personally Identifiable Information (PII) that can directly or indirectly identify a natural person (i.e., data subject). However, many Social Behavioral and Economics (SBE) theories indicate that privacy regulations that emphasize increasing consumers' control over their PII may not achieve their desired affect due to the privacy paradox, wherein an individual shows significant concern about their privacy yet continues to engage in behaviors that disclose their personal information [3]. The privacy paradox's leading causes include biased perception, unawareness, and underestimating the probability of data breach [3, 5, 43, 61]. Increasing users' privacy awareness can help resolve the privacy paradox. The development of AI-enabled analytics can enable novel solutions (e.g., intelligent privacy assistants) that can be used to increase individuals' privacy risk awareness [2, 17]. To support the development of AI-enabled analytics, the collection and storage of promising PII data from a multitude of sources is necessary. Seven major categories of data can exist that can support AI-enabled privacy analytics: (1) Digital Legal Documents, (2) Social Media Platforms, (3) Internet Browsing Data, (4) Dark Web Sites, (5) Data Breach Services, (6) People Search Engines, and (7) Biometrics. We summarize each major category of data in Table 1. For each category of data, we provide a sample data source (e.g., platform), a description, and sample metadata and/or data relating to PII. Following the table summary we provide a detailed description (e.g., background, collection strategies) for each major category of data.

2.1 Legal Documents

Legal documents play a crucial role in helping citizens, businesses, and countries adhere to predefined rules of engagement and/or operations. Within the context of privacy, selected prevailing legal documents include privacy policies that summarize the ways a party (e.g., organization) collects and manages data, government records that provide public records (e.g., names, birthdates, address, and salary) about individuals and companies, terms of service that summarize agreements between users and providers for a particular service, and non-disclosure agreements that confirm confidentiality between two partners. Collection strategies for these for privacy policies, public government records, and terms of service often rely on developing custom crawlers to scrape data from the web. The data coverage may vary based on various government regulations (e.g., Florida's Sunshine Laws). With regards to non-disclosure agreements, these are often under the jurisdiction of the organizations for whom they were developed. Irrespective of data availability and collection mechanisms, nearly all legal documents contain primarily textual or categorical data.

2.2 Social Media Platforms

Social media platforms provide citizens an unprecedented ability to communicate and connect with others across the globe. Many platforms such as Facebook, Twitter, TikTok, Instagram, Flickr, Reddit, and others offer users the ability to share text, images, videos, and other content with their connections and beyond. However, the platform creators are often aggregating, curating, selling, and leveraging the data about their users to refine the recommendation algorithms and provide more powerful services. Many platforms also offer Application Programming Interfaces to systematically and programmatically access their user data to facilitate the development of novel

Table 1. Summary of Prevailing Data Sources to Support AI-enabled Privacy Analytics

Type	Sample Data Source	Description	Sample Metadata and/or Data Relating to PII
Legal Documents	Privacy Policies	Legal documents that discloses the ways a party gathers and manages a customer's data	Text, section headers
	Public Government Records	Publicly accessible records about individuals and companies	Names, birthdates, address, salary, arrests
	Terms of Service	Legal agreements between a user and a service provider	Date, text, names
	Non-disclosure agreements	Agreement between two partners that maintains confidentiality	Date, text, names
Social Media Platforms	Facebook	General purpose social networking site to create profiles and connect with users	Username, posts, likes, geotags
	Twitter	Microblogging site that allows users to share short tweets (280 characters)	Username, pictures, videos, posts
	TikTok	Platform for sharing short videos	Username, video, length
	Instagram	Photo and video sharing platform	Username, photos, videos, geotags
	Reddit	Social news aggregation and forums	Username, date, post
Internet Browsing Data	Browsing History	A log of browsing patterns	Timestamps, clickthrough rates
	Search History	Search log of queries via search engines	Timestamps, sites searched, search terms
Dark Web Sites	Hacker Forums	Discussion forums that enable sharing of content and knowledge	Date, author, forum text
	Carding Shops	Shops selling stolen credit cards	Bin, quantity, price
	Social SecurityNumber Shops	Shops selling stolen social security numbers	Quantity, price
	Paste Sites	Sites that allow the anonymous posting of large plain-text contents	Date, message
	DarkNet Marketplaces	Shops selling illicit and stolen goods	Price, title, quantity, seller
Data Breach Services	Have I Been Pwned	Site allowing users to check if their emails have been involved in breaches	Email search
	BreachAlarm	Service that scans the Internet to check if passwords are publicly accessible	Email search
	Dehashed	Multi-lingual search engine to look up breached emails, names, and passwords	Search via email, username, IP, name, address, phone number
	Have I Been Sold	Site that determines if an email address has been sold	Email search
	Inoitsu	Service that detects if an email address has been part of a breach	Email breach data classes, recent breaches

(Continued)

Table 1. Continued

Type	Sample Data Source	Description	Sample Metadata and/or Data Relating to PII
People Search Engines	Spokeo	Search engine aggregating online and offline personal information and physical addresses	Property, consumer, court records
	BeenVerified	Background check company	Name, phone, email, address, username
	MyLife	Service for determining the reputation of individuals	Court records, names, ages, addresses
	That's Them	Search engine for finding people anonymously	Address, phone, email, IP
	Xlek	Search engine for online data records from public sources	Marketing, property, vehicle, court records
	Whitepages	Background check service	Phone number, address
	AnyWho	Directory of people, places, and businesses	Name, zip code, age
Biometrics	Fingerprints	Physical impressions on the tip of fingers	Moisture, grease, ridges
	Eye movements	Movement of eyes	Rate, color, direction
	Photos	Visual artifact of a particular moment	Pixels, colors, and contours
	Videos	Electronic medium for recording and broadcasting visual media	Frames, rates, images
	Accelerometers	Measurement of the acceleration of movement(s)	Speed, velocity
	Voice	Sounds generated from vocal cords	Pitch, tone
	Mouse movements	How computer users move their mouse	Speed, directionality
	Pulse	The heart rate of a human	Pump rates

platforms and mash-ups for selected business tasks. While providing an unprecedented ability to drive forward the next generation of social computing, significant concerns have been raised by governing bodies within the US regarding the privacy of users on these platforms. These include the manner in which the data is stored, managed, and in some cases, sold to related organizations or data collection companies.

2.3 Internet Browsing Data

The Internet has provided an unparalleled ability for citizens to search for content relevant to their interests and needs. How users browse and search for content can illuminate key traits about them and ultimately reveal private information about users and user bases (often unbeknownst to the user). For example, a user's browsing history can reveal the sites they visit, rates, time spent, and others. Such information can be used to enhance recommendation algorithms. Similarly, search histories can provide logs of the searches users have made on a particular service (e.g., search engine). Often these data are difficult for the public to obtain. More commonly, these data are available to Internet Service Providers (e.g., Comcast and Xfinity) and companies that provide the Internet browsers (e.g., Firefox, Google Chrome, and Microsoft Edge), search engines (e.g., Google, Bing, and Yahoo!), and other related companies.

2.4 Dark Web Sites

The Dark Web represents a dark, covert side of the Internet that facilitates the sharing of illicit goods and products. Major Dark Web platforms include hacker forums, carding shops, social security number shops, paste sites, and DarkNet Marketplaces. Although these platforms have been extensively studied for proactive cyber threat intelligence activities such as malware analysis [24, 49–51, 57], enhanced vulnerability assessments [20, 25, 53], and other key activities [7, 19, 48], these platforms also provide significant privacy data and value. Hacker forums are discussion forums that allow hackers to freely share and acquire illicit goods and knowledge, including goods attained from significant data breaches. Platforms such as Raidforums post millions of breach records for free access. Carding shops facilitate the sale of stolen credit cards. Social security number (SSN) shops serve a similar purpose but for selling SSNs. Paste sites allow individuals to share large quantities of plain text anonymously. Often these pastes often contain significant, stolen, personally identifiable information, including SSNs, credit card numbers, and username and password combinations. Finally, DarkNet Marketplaces allow hackers to sell a variety of stolen goods (e.g., healthcare records, SSNs, and credit cards) [33, 35]. Collection mechanisms for Dark Web platforms require highly customized crawlers equipped with anti-crawling countermeasures.

2.5 Data Breach Services

The unfortunate and rapid growth of cyber-attacks has led to an increased quantity of breach data proliferating on the web (e.g., Dark Web sites). An emerging industry of data breach services has aimed to capitalize on these breach data by building systems that identify, collect, process, and present selected breach data for the broader public to use. For example, sites such as Have I Been Pwned and Inoitsu allow users to check if their emails have been revealed in breaches. Other sites such as BreachAlarm and Dehashed allow users to search if their username and/or passwords were released as part of a data breach. Most services provide basic search services via email or name search only; significant potential remains for executing more in-depth entity linking, matching, and resolution across multiple sites. Such analytics can help build a more comprehensive and holistic perspective of an individual's overall privacy risk profile [33, 35].

2.6 People Search Engines

The increasing quantity of publicly accessible information on the web has led to the emergence of a nascent industry similar to Data Breach Services: People Search Engines. These engines allow users to search for an individual's information. For example, Spokeo aggregates online and offline personal information (e.g., property, consumer, and court records). Search engines such as BeenVerified, MyLife, Xlek, and Whitepages allow users to conduct background checks on individuals of interest. Most of these search engines provide general data such as name, address, and email address, along with more personal information such as court records, vehicles, properties, and more. Similarly to data breach services, these search engines currently have only limited analytics powering the platforms [33, 35].

2.7 Biometrics

The rise of sensor technologies has enabled the monitoring of a person's psychological traits [32]. While commonly used in the context of precision medicine and healthcare [65, 66] these sensing technologies can be leveraged to profile selected individuals by measuring their activities directly (as opposed to self-reports) and facilitate ongoing assessments (rather than ad hoc or periodic). Fingerprints and eye movements can help ensure the security of selected physical facilities. Videos,

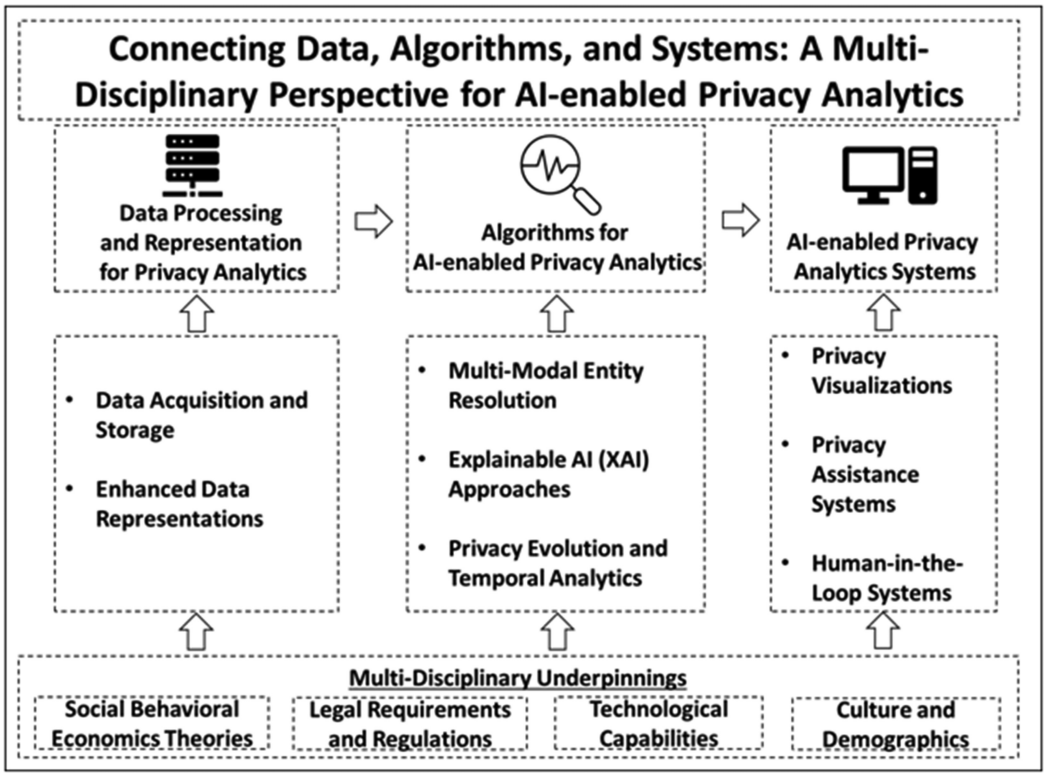

Fig. 1. Connecting data, algorithms, and systems: A multi-disciplinary framework for AI-enabled privacy analytics.

accelerometers, voice, mouse movements, and pulse rates can all provide insight into the manner in which humans behave in reponse to selected stimuli and environments. Similarly to social media providers that collect data about billions of users, Internet of Things (IoT) vendors that provide sensing technologies have tremendous potential in collecting significant biometric data pertaining to their users to enhance their product development and/or execute large-scale global surveillance or Internet measurement tasks.

3 CONNECTING DATA, ALGORITHMS, AND SYSTEMS: A MULTI-DISCIPLINARY FRAMEWORK FOR AI-ENABLED PRIVACY ANALYTICS

Each aforementioned data source can set the foundation for novel AI-enabled models for privacy analytics. However, the AI-based privacy analytics field is in its nascency. A clear roadmap for executing high impact research activities is needed to fully realize this field's potential. Past seminal works pertaining to emerging phenomena have clearly indicated that such a roadmap should account for data, theories, analytics and systems [1, 8, 13, 52]. Since no single discipline can comprehensively address all AI-enabled privacy analytics research inquiries, we present a novel multi-disciplinary framework that takes a holistic perspective of connecting data, AI-based methods, and systems. We depict the framework in Figure 1.

The proposed framework includes three major components: (1) Data Processing and Representation for Privacy Analytics that comprises Data Acquisition and Storage and Enhanced Data Representations, (2) Algorithms for AI-enabled Privacy Analytics including Multi-Modal Entity

Resolution, Explainable AI (XAI) Approaches, and Privacy Evolution and Temporal Analytics, and (3) Privacy Analytics-Driven Systems such as Privacy Visualizations, Privacy Assistance Systems, and Human-in-the-Loop Systems. Each component can draw from multi-disciplinary underpinnings including, but not limited to, theories from the social, behavioral, and economics (SBE) sciences; legal requirements and regulations; technological capabilities; and culture and demographics. We describe each major component of the framework in the following section.

3.1 Data Processing and Representation for Privacy Analytics

Each data source summarized in Table 1 holds significant promise for developing innovative privacy analytics. However, researchers face significant challenges in data acquisition, storage, and representation. We describe each challenge and associated opportunities in the following-subsections.

3.1.1 Data Acquisition and Storage. Privacy is a unique domain that necessitates careful considerations before collecting and storing data. Many data sources often contain significant details about individuals' personal information (e.g., name, birthdate, address, credit cards, and SSNs). While providing significant potential for conducting interesting, AI-based analytics, these processes should not come at the cost of safe data acquisition. The bounds, legal ramifications, country rules, and other regulations should be carefully considered and monitored when collecting relevant data. Similarly, data storage should be conducted in a sensitive manner consistent with the management requirements (e.g., HIPPA) of selected data sources. Large stores of data to support AI-enabled privacy analytics could be fruitful targets for malicious hackers to attack. Therefore, carefully designing selected security mechanisms (e.g., firewalls, backups, swipe access, and policies) is essential for maintaining the confidentiality, integrity, and availability of selected privacy data assets. Similarly, deploying privacy-preserving techniques (an emerging area in database and data mining research) and anonymization approaches are critical.

3.1.2 Enhanced Data Representations. The seven listed categories of privacy data include variables (e.g., features, attributes) that span multiple types (e.g., nominal, categorical, and continuous). The range of attribute types lends itself to interesting and clever data representations that can facilitate novel analytics. Emerging AI-based methodologies rooted in deep learning can operate directly on carefully constructed data representations (e.g., matrices, trees, graphs, tensors, and sequences) beyond the conventional flattened and manually constructed (often in an ad hoc manner) feature vector approaches prevalent in extant machine learning research [30]. Since data representations can more effectively capture the "real-world" phenomena they aim to model, researchers aiming to develop novel privacy analytics should consider various options during their analytics processes. For example, privacy policy text can be represented as graphs and sequences to capture global and/or local relationships across words, biometrics data can be grids or tensors, and so on. Representation selection can be guided based on the unique data characteristics (e.g., sparsity and population), the tasks at hand (e.g., predictive, prescriptive, and descriptive), and requirements (e.g., legal and relevant SBE theories) [55].

3.2 Algorithms for AI-enabled Privacy Analytics

The above-listed application areas, data sources, and data representations require advanced methodologies to fully uncover their potential. We note that a clear delineation of privacy analytics should be made. Privacy data can be analyzed to help a company, organization, or individual enhance their own agenda (e.g., algorithms, stalking, and investigation). Conversely, privacy data can be analyzed for the betterment of society, including raising awareness, calculating novel privacy risk scores, and supporting mitigation techniques. Given the substantial (and often negative)

focus in modern media and press on the agenda advancing analytics, we focus on how analytics can be leveraged for the "greater good" in that they can assist multiple stakeholders (e.g., legislators, regulators, and individuals). With this perspective in mind, we identify three major categories of methodologies that privacy analytics researchers and practitioners can consider when developing their work: multi-modal entity resolution, XAI, and privacy evolution and temporal changes. Each is described in further detail below.

3.2.1 Multi-Modal Entity Resolution. Prevailing data sources for privacy analytics are often disjointed and disparate. As indicated in the review of data breach services and people search engines, most analytics procedures rely on simple keyword searches. These approaches cannot identify whether records across multiple data sources are consistent, correct, and complete. As a result, identifying the overall risk profile of an individual remains a critically needed capability. Multi-modal entity resolution methods hold significant promise in resolving and matching identities across multiple data sources [33, 35]. These methods often leverage graphs as representations [58, 62] and employ multi-view, active, and/or transfer learning strategies [37, 63, 64]. The representations, architectures, and learning strategies are flexible and can accommodate numerous extensions and adaptations to help account for various real-world considerations (e.g., locations, temporal changes, etc.). Successfully resolving multiple sets of data can result in novel privacy risk scores that identify how susceptible an individual is. For example, privacy analytics can be used to estimate re-identification risk (e.g., how likely to re-identify individuals given the disclosed data) [36, 45, 60] and how multiple data sources could be linked to launch privacy attacks [6]. These resolutions can also support additional tasks such as clustering risky individuals across various demographic classifications or de-identification tasks (e.g., [29]).

3.2.2 XAI Approaches. Motivated by the black-box nature of many emerging machine learning and deep learning algorithms, XAI has gained significant traction over the past half-decade. Within the context of privacy analytics, XAI holds significant potential and utility for various stakeholders, particularly regulators and consumers. From a regulatory perspective, checking the impact of an AI-based system on consumers and verifying reliability is essential for maintaining viability and reputation. Consumers often ask questions pertaining to the impact, action, validity, and correctness of AI-based algorithms, particularly when their personal information is at stake. Providing model explainability and interpretability is essential for helping stakeholders build trust in AI-based algorithms and systems [46]. In general, model explainability can occur at the representation level (e.g., inputs into the algorithm), during algorithm processing, or post hoc. These explanations can occur at the global or local level [18]. Five major types of explanations exist [9]: (1) text, drawing upon rule-based learner and decision trees; (2) visual, using sensitivity and dependency plots; (3) local, employing example, simplification, and feature relevance approaches; (4) local, using rule-based learners, linear approximations, and counterfactuals; and (5) feature relevance, using feature importance techniques. How to leverage and employ these mechanisms to achieve explainability and interpretability for models for various stakeholders in the privacy domain requires significant research and exploration.

3.2.3 Privacy Evolution and Temporal Analytics. Privacy is an area that is undergoing significant growth and changes, especially in the past decade. As a result, questions have emerged about how privacy phenomena evolve over time, particularly after new laws and regulations (e.g., GDPR and CCPA) are released. Such information is often invaluable for guiding law, policy, and regulations. The types of data and analytics in question can help dictate the types of temporal analysis that can be executed. For example, studying how language evolves over time in legal documents can

rely on diachronic linguistics [54]. Modeling the growth of multiple Dark Web platform contents simultaneously can leverage techniques such as temporal heterogeneous information networks. Monitoring the changes in biometrics data for selected privacy applications can rely on emerging anomaly detection-based methodologies. Individual privacy behaviors can also be monitored over time to detect overall behavioral patterns. Each of these tasks, when aggregated across multiple individuals, can help determine the overall privacy of a particular demographic and/or of an overall population.

3.3 AI-enabled Privacy Analytics Systems

Privacy is of interest to most modern citizens. However, common users are unlikely to interact directly with algorithms. Therefore, systems are essential for varying types of stakeholders to engage with the collected data and developed algorithms. Data breach services and people search engines have already made significant strides in presenting privacy data and selected algorithm results (e.g., keyword searches). However, significant potential remains for developing the next-generation of privacy systems that are powered by advanced AI-based methods. We summarize three major categories of systems that can be developed for future privacy analytics research: (1) privacy visualizations, (2) privacy assistance systems, and (3) human-in-the-loop systems. We describe each in turn in the following sub-sections.

3.3.1 Privacy Visualizations. The growth of privacy data is reaching a level that is beyond the comprehension of most humans. Visualizations hold considerable promise in cutting through and succinctly summarizing large quantities of data. Developing visualizations that facilitate details on demand, explicate historical perspectives, and provide other fundamental functionalities can help summarize the key results of AI-based algorithms in a manner more intuitive than raw data. Prevailing visualizations such as trees, graphs, temporal, charts, and geo-spatial can offer great utility for various privacy tasks. These visualizations can provide details at multiple levels of granularity, including global (macro), local (meso), and individual (micro) levels [10, 56]. Relevant visualization and SBE theories and stakeholder requirements can help guide the manner in which these visualizations are developed and deployed. Ultimately, these visualizations can help support other critical analytics, systems, and decision-making processes (e.g., identifying adherence/non-adherence and fining from regulators).

3.3.2 Privacy Assistance Systems. Many citizens today are often unaware of their privacy rights and preferences. These issues are often exacerbated when considering the ever-growing technology landscape of IoT devices and systems [15]. To this end, privacy assistance systems can provide an invaluable resource for helping users manage their privacy. Such systems can incorporate selected AI-based analytics for discovering resources (e.g., privacy policies), recommending optimal configuration settings, deterring from unfavorable settings, identifying how data is being managed by companies as detailed in their privacy policies, examining how a user's privacy behaviors evolve over time, and other key privacy tasks [21, 34]. These systems can be developed for multiple data modalities and segments of the population (e.g., children, adults, and senior citizens) to maximize their reach and impact [16]. Ultimately, these systems can help multiple categories of stakeholders be nudged into particular privacy behaviors [4].

3.3.3 Human-in-the-Loop Systems. Modern information privacy has often been criticized as being a non-optional agreement between a user and a service provider. In reality, users should have a close input into their data and how analytics are used, based on users' own choices. Developing human-in-the-loop systems (i.e., augmented intelligence, human–AI interfaces) can significantly help in this regard. In general, three major categories of human-in-the-loop systems can be

developed [27, 47]: AI replaces humans (substitution), AI and humans augment each other (augmentation), or humans and AI are convened to cooperate and function symbiotically (assemblage). How privacy data and analytics can be leveraged in each setting requires significant research and exploration, particularly on how SBE theories, legal requirements and regulations, and cultural and demographic considerations can be accounted for [39].

4 MECHANISMS FOR FACILITATING AI-ENABLED PRIVACY ANALYTICS RESEARCH, EDUCATION, AND TRANSITION TO PRACTICE

Effectively and efficiently tackling key AI-enabled analytics privacy issues can often require considerable resources (e.g., technical infrastructure and human capital). The significance of privacy-related research has been met with focused attention from funding agencies at the federal, regional, and local levels to provide funding for high-impact and fundamental privacy research. These funding agencies often necessitate the development of inter-disciplinary teams cross-cutting prevailing technical disciplines such as information systems, computer science, and others along with relevant social science disciplines such as psychology, behavioral economics, cognitive sciences, and so on. The intuition behind creating such multi-disciplinary teams is to tackle grand privacy issues in a holistic and comprehensive fashion. The grant funding provided from these agencies can help develop research programs, recruit high-powered scientists, develop a strong reputation, and foster natural collaborations across government and industry that otherwise would have been difficult to attain [23, 41, 42, 44].

To help provide an overview of selected grant funding opportunities, we provide a sample of solicitations from the NSF that can help facilitate AI-enabled privacy research. NSF is often considered the gold-standard for funding transformative science and education programs. Overall, four major categories of NSF funding that can help support privacy research exist: (1) Early Career, (2) Core Research, (3) Transition to Practice, and (4) Education Oriented. We describe each major category below:

- **Early Career:** NSF is an attractive funding source for many new principal investigators (PIs) looking to launch their academic careers. For those junior faculty interested in conducting privacy analytics research, programs such as the CISE Research Initiation Initiative (CRII) and CAREER offer promising mechanisms. CRII aims to provide up to 175K of funding to junior faculty within the first three years of their new academic position to launch their research programs. CAREER offers up to $500K of funding to pre-tenured faculty to help set the foundation for a career of excellence in research and teaching in a selected area of academic inquiry. Both mechanisms are promising for faculty aiming to attain funding without competing with senior faculty in core research programs.
- **Core Research Programs:** Fundamental and applied research is essential to cultivating the next generation of privacy practice and education. NSF offers several key programs in this regard. The CORE designation under the Secure and Trustworthy Cyberspace (SaTC) program (up to $1.2M in funding) aims to serve as a vehicle to facilitate fundamental research. The Disrupting Illicit Supply Networks (D-ISN) program (up to $1M in funding) also aims to provide mechanisms for disrupting supply networks of illicit goods that can ultimately challenge the privacy of citizens. Unlike the Early Career funding opportunities listed earlier, successfully funded SaTC and D-ISN projects often require inter-disciplinary and multi-institution teams.
- **Transition to Practice:** AI-enabled privacy analytics holds significant promise in offering novel artifacts that can be used in practice. However, novel technologies can often not make it to market due to unclear pathways for transition and lack of resources. NSF offers

several mechanisms to help scholars navigate these early challenges and successfully transition technologies into practice. The Small Business Innovation Research Program and Small Business Technology Transfer Program programs (up to $1.75M in seed funding) encourage small businesses to engage in federal research and development activities with the potential of subsequent commercialization. The SaTC Transition to Practice designation (up to $1.2M in funding) can help transition research developed from the CORE designation and/or other foundational research into practice. The Convergence Accelerator program (up to $5M in funding) can also provide a mechanism to quickly convene scholars and practitioners to make a significant societal impact.

- **Education Oriented:** Knowledge derived from research and transition efforts can ultimately set the foundation for innovative pedagogical activities. NSF offers programs through the Division of Graduate Education and Education and Human Resources to support such education efforts. For example, the Scholarship-for-Service (SFS) program helps PIs train the next generation of cybersecurity and privacy professionals for eventual placement into government positions. The EDU designation under the SaTC program (up to $500K of funding) helps support innovative pedagogical efforts targeted at multiple learning levels (e.g., K-12, undergraduate, graduate, and senior citizens).

Each aforementioned funding mechanism can provide resources to generate transformative AI-enabled privacy research, education, and practical impacts. However, ensuring that the long-term health of this emerging research stream is maintained requires approaches to review and archive selected works. Conferences are an excellent approach in this regard. In general, four major categories of conferences currently exist: (1) Academic Privacy Venues, (2) Practitioner Privacy Venues, (3) Computer Science Privacy Venues, and (4) NSF Meetings. We describe each below:

- **Academic Privacy Venues** offer an excellent mechanism for scholars to share their work with the larger academic community. These venues are vital for evaluating the novelty and contribution of selected privacy research to literature. Prevailing academic venues include IEEE Security and Privacy, ACM Conference on Security and Privacy, USENIX Symposium on Usable Privacy and Security, ACM Conference on Data Application Security and Privacy, IEEE Symposium on Privacy-Aware Computing, and IEEE Intelligence and Security Informatics (ISI). Publishing in these venues can also significantly help selected scholars attain their promotion and tenure at their respective institutions.
- **Practitioner Privacy Venues** offer a mechanism for practitioners from government and the private sector to share problems they face. These venues are also very valuable for scholars to help identify research topics that are practically motivated. Selected practitioner conferences include the Conference on Applied Machine Learning for Information Security, IAPP Global Privacy Summit, DEFCON, FloCon, and BlackHat.
- **Computer Science Privacy Venues** typically offer state-of-the-art advances in foundational or applied computational related topics. Many of these venues also offer specialized workshops that tackle various aspects of privacy. Examples include Natural Legal Language Processing at ACM Knowledge Discovery from Databases, Privacy Preserving Artificial Intelligence at the Association for the Advancement of Artificial Intelligence, the Foundations on Open Source Intelligence Security Informatics at Advances on Social Network Analysis and Mining, Privacy Management in the Cyberspace at IEEE International Conference on Data Mining, and Towards Trustworthy ML at the International Conference on Learning Representations. Some conferences offer multiple workshops related to various aspects of privacy. Examples include the workshops on Trustworthy ML, SpicyFL, Building AI with Security and Privacy in Mind, and PriML at NeurIPS.

- **NSF Meetings** offer scholars an approach for identifying key areas of funding from the federal government. The selected venues that have the most relevance to analytics-driven privacy research include the SaTC PI Meeting (500+ attendees and a workshop on AI, ML, and NLP for Personalized Privacy and Security Assistants) and the SFS Job Fair (1,000+ attendees, with significant presence from many three-letter agencies).

Conferences offer a tremendous mechanism for scholars and practitioners to ground their selected research inquiries, network with potential collaborators, and acquire feedback about their progress. These activities can help develop archivable journal papers at premier venues. Prevailing venues that publish privacy-related research include *ACM Transactions on Privacy and Security, IEEE Security and Privacy Magazine, IEEE Transactions on Information Forensics and Security, IEEE Transactions on Dependable and Secure Computing*, and *Computers and Security*. Each journal is highly inter-disciplinary and has a long-standing reputation within the privacy community.

5 PAPERS IN THIS SPECIAL ISSUE

The idea for this special issue was originally conceived in conjunction with the "Analytics for Cybersecurity" (part 1 of the special issue) while at the International Conference on Information Systems in December 2018 in San Francisco, California. The guest editor team was developed in the following months, and the Call for Papers (CFP) was carefully developed and refined. Our guest editor team then promoted the CFP via various channels, including the DBWorld listserv; AISWorld listserv; the 2019 IEEE ISI Conference in Shenzhen, China; the DEFCON AI Village in Las Vegas; and personal emails in our contact networks. In total, over 60 submissions were received. Forty-four papers were sent out for review; sixteen of these were desk rejected due to lack of fit. Six were selected for the privacy-themed special issue, and nine were selected for the cybersecurity related special issue. Each paper went through at least two rounds of review. A summary of the final accepted papers in the privacy special issue is presented in Table 2. For each paper, we detail its author team, title, topic, data source(s), and methodology. We organize the papers in the table based on the last name of the first author.

Each paper made important contributions to analytics-based privacy research. In the paper entitled "Design of an Inclusive Privacy Index (INF-PIE): A Financial Privacy and Digital Financial Inclusion Perspective," Akanfe et al. employed a term frequency and topic modeling approach on privacy policies pertaining to mobile wallet and remittance applications to identify country-level relative data privacy compliance. Results of this work give an interesting perspective on country-level similarities and differences for selected privacy policy analytics. Kul et al. contributed a paper entitled "An Analysis of Complexity of Insider Attacks to Databases" that focused on conducting complexity analysis and timing experiments to identify insider attacks and user intent. Their proposed process was applied to SQL query workloads from a major national bank. In the work submitted by Ranathunga et al. entitled "Mathematical Reconciliation of Medical Privacy Policies," the authors developed a prototype system based on mathematical modeling and metagraphs to reconcile medical privacy policies. The authors demonstrated their approach in Australia's My Health Record policy to identify several critical flaws, including non-compliance issues that allow healthcare providers to access medical records by default. Sudhakar and Gavrilova co-authored an article entitled "Deep Learning for Multi-instance Biometric Privacy" that focused on leveraging support vector machines and convolutional neural networks on multi-instance iris and finger vein databases for revocable biometric systems. In the paper entitled "Optimal Recruitment in Organizations under Attribute-Based Access Control," Vaidya et al. proposed a solution based on greedy heuristics to recruit qualified employees while maintaining security and privacy. Finally, Zaeem and Barber, in their paper "The Effect of GDPR on Privacy Policies: Recent Progress and

Table 2. Summary of Papers in this Special Issue

Authors	Title	Topic	Data Sources	Methodology
Akanfe et al.	Design of an Inclusive Financial Privacy Index (INF-PIE): A Financial Privacy and Digital Financial Inclusion Perspective	Identifying country-level relative data privacy compliance	Privacy policies of mobile wallet and remittance applications	Term frequency, topic modeling
Kul et al.	An Analysis of Complexity of Insider Attacks to Databases	Insider attacks and user intent	SQL query workloads from a national bank	Complexity analysis, timing experiments
Ranathunga et al.	Mathematical Reconciliation of Medical Privacy Policies	Reconciling privacy policies	Medical privacy policies	Mathematical modeling and metagraphs
Sudhakar and Gavrilova	Deep Learning for Multi-instance Biometric Privacy	Revocable biometric systems	Multi-instance iris and finger vein databases	Support vector machines and convolutional neural networks
Vaidya et al.	Optimal Employee Recruitment in Organizations under Attribute-Based Access Control	Recruitment of qualified employees	—	Greedy heuristics
Zaeem and Barber	The Effect of GDPR on Privacy Policies: Recent Progress and Future Promise	Assigning risk levels to privacy policies	550 privacy policies	PrivacyCheck system

Future Promise," developed a PrivacyCheck system that automatically evaluates privacy policies to identify if private information is released for ten privacy factors. Based on the level of release, PrivacyCheck assigns a risk level (Green, Yellow, or Red).

6 SUMMARY

Privacy has become a critical societal concern for many citizens. AI-enabled analytics holds considerable promise in enhancing the privacy of many individuals across the globe. However, AI-enabled privacy analytics is still in its early stages. Many opportunities exist for scholars and practitioners to make significant, positive contributions in this emerging topical area. In this article, we aimed to provide a strong foundation for supporting the next generation of AI-enabled privacy analytics research. Specifically, we provided an overview of prevailing data sources to support privacy analytics, offered a multi-disciplinary and end-to-end research roadmap for tackling emerging privacy challenges, identified selected funding sources to support high-impact privacy analytics research, and summarized prevailing conference and journal venues that can be leveraged to share and archive privacy analytics research. It is our sincere hope that these contents help facilitate illuminating discussions and lead to the rapid development of high-impact privacy analytics research that positively impacts modern society.

ACKNOWLEDGMENTS

We thank the *ACM Transactions on Management Information Systems* Editor-in-Chief Dr. Daniel Zeng for guiding us through the special issue process. We thank Victoria White for her coordination of papers, reviews, and feedback. We express our gratitude to the many reviewers for providing constructive feedback on the submitted papers. We also thank Amy Lin for her assistance in

conducting selected reviews presented in this paper. We thank Dr. Heng Xu for her comments on earlier versions of this editorial.

REFERENCES

[1] Ahmed Abbasi, Suprateek Sarker, and Roger Chiang. 2016. Big data research in information systems: Toward an inclusive research agenda. *J. Assoc. Inf. Syst.* 17, 2 (February 2016), I–XXXII. DOI: https://doi.org/10.17705/1jais.00423

[2] A. Acquisti, L. Brandimarte, and G. Loewenstein. 2015. Privacy and human behavior in the age of information. *Science (80).* 347, 6221 (January 2015), 509–514. DOI: https://doi.org/10.1126/science.aaa1465

[3] A. Acquisti and J. Grossklags. 2005. Privacy and rationality in individual decision making. *IEEE Secur. Priv. Mag.* 3, 1 (January 2005), 26–33. DOI: https://doi.org/10.1109/MSP.2005.22

[4] Alessandro Acquisti, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro Giovanni Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, and Shomir Wilson. 2017. Nudges for privacy and security: Understanding and assisting users' choices online. *ACM Comput. Surv.* 50, 3 (October 2017), 1–41. DOI: https://doi.org/10.1145/3054926

[5] Idris Adjerid, Eyal Peer, and Alessandro Acquisti. 2018. Beyond the privacy paradox: Objective versus relative risk in privacy decision making. *MIS Q.* 42, 2 (February 2018), 465–488. DOI: https://doi.org/10.25300/MISQ/2018/14316

[6] Imrul Chowdhury Anindya, Harichandan Roy, Murat Kantarcioglu, and Bradley Malin. 2017. Building a dossier on the cheap: Integrating distributed personal data resources under cost constraints. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management.* 1549–1558. DOI: https://doi.org/10.1145/3132847.3132951

[7] Nolan Arnold, Mohammadreza Ebrahimi, Ning Zhang, Ben Lazarine, Mark Patton, Hsinchun Chen, and Sagar Samtani. 2019. Dark-net ecosystem cyber-threat intelligence (CTI) tool. In *2019 IEEE International Conference on Intelligence and Security Informatics (ISI'19).* DOI: https://doi.org/10.1109/ISI.2019.8823501

[8] Indranil Bardhan, Hsinchun Chen, and Elena Karahanna. 2020. Connecting systems, data, and people: A multidisciplinary research roadmap for chronic disease management. *MIS Q.* 44, 1 (2020), 185–200.

[9] Vaishak Belle and Ioannis Papantonis. 2020. Principles and practice of explainable machine learning. (September 2020). Retrieved from http://arxiv.org/abs/2009.11698.

[10] Katy Börner and David E. Polley. 2014. *Visual Insights: A Practical Guide to Making Sense of Data.* The MIT Press.

[11] Joseph R. Buckman, Jesse C. Bockstedt, and Matthew J. Hashim. 2019. Relative privacy valuations under varying disclosure characteristics. *Inf. Syst. Res.* 30, 2 (June 2019), 375–388. DOI: https://doi.org/10.1287/isre.2018.0818

[12] Zike Cao, Kai-Lung Hui, and Hong Xu. 2018. An economic analysis of peer disclosure in online social communities. *Inf. Syst. Res.* 29, 3 (September 2018), 546–566. DOI: https://doi.org/10.1287/isre.2017.0744

[13] Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. 2012. Business intelligence and analytics: From big data to big impact. *MIS Q.* 36, 4 (2012), 1165–1188. DOI: https://doi.org/10.1145/2463676.2463712

[14] Robert E. Crossler and France Bélanger. 2019. Why would I use location-protective settings on my smartphone? Motivating protective behaviors and the existence of the privacy knowledge–belief gap. *Inf. Syst. Res.* 30, 3 (September 2019), 995–1006. DOI: https://doi.org/10.1287/isre.2019.0846

[15] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. 2018. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Comput.* 17, 3 (July 2018), 35–46. DOI: https://doi.org/10.1109/MPRV.2018.03367733

[16] Anupam Das, Martin Degeling, Xiaoyou Wang, Junjue Wang, Norman Sadeh, and Mahadev Satyanarayanan. 2017. Assisting users in a world full of cameras: A privacy-aware infrastructure for computer vision applications. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* 1387–1396. DOI: https://doi.org/10.1109/CVPRW.2017.181

[17] André Deuker. 2010. Addressing the privacy paradox by expanded privacy awareness – The example of context-aware services. In *Privacy and Identity Management for Life.* 275–283. DOI: https://doi.org/10.1007/978-3-642-14282-6_23

[18] Mengnan Du, Ninghao Liu, and Xia Hu. 2019. Techniques for interpretable machine learning. *Commun. ACM* 63, 1 (July 2019), 68–77. DOI: https://doi.org/10.1145/3359786

[19] Po-Yi Du, Ning Zhang, Mohammedreza Ebrahimi, Sagar Samtani, Ben Lazarine, Nolan Arnold, Rachael Dunn, Sandeep Suntwal, Guadalupe Angeles, Robert Schweitzer, and Hsinchun Chen. 2018. Identifying, collecting, and presenting hacker community data: Forums, IRC, carding shops, and DNMs. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI).* 70–75. DOI: https://doi.org/10.1109/ISI.2018.8587327

[20] Malaka El, Emma McMahon, Sagar Samtani, Mark Patton, and Hsinchun Chen. 2017. Benchmarking vulnerability scanners: An experiment on SCADA devices and scientific instruments. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI).* 83–88. DOI: https://doi.org/10.1109/ISI.2017.8004879

[21] Pardis Emami-Naeini, Sruti Bhagavatula, Hana Habib, Martin Degeling, Lujo Bauer, Lorrie Faith Cranor, and Norman Sadeh. 2017. Privacy expectations and preferences in an IoT world. In *Proceedings of the Thirteenth USENIX Conference on Usable Privacy and Security (SOUPS'17)*. 399–412.

[22] Esther Gal-Or, Ronen Gal-Or, and Nabita Penmetsa. 2018. The role of user privacy concerns in shaping competition among platforms. *Inf. Syst. Res.* 29, 3 (September 2018), 698–722. DOI: https://doi.org/10.1287/isre.2017.0730

[23] Shirley Gregor and Alan R. Hevner. 2013. Positioning and presenting design science research for maximum impact. *MIS Q.* 37, 2 (2013), 337–355. DOI: https://doi.org/10.2753/MIS0742-1222240302

[24] John Grisham, Sagar Samtani, Mark Patton, and Hsinchun Chen. 2017. Identifying mobile malware and key threat actors in online hacker forums for proactive cyber threat intelligence. In *2017 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 13–18. DOI: https://doi.org/10.1109/ISI.2017.8004867

[25] Christopher R. Harrell, Mark Patton, Hsinchun Chen, and Sagar Samtani. 2018. Vulnerability assessment, remediation, and automated reporting: Case studies of higher education institutions. In *2018 IEEE International Conference on Intelligence and Security Informatics (ISI'18)*. DOI: https://doi.org/10.1109/ISI.2018.8587380

[26] Irina Heimbach and Oliver Hinz. 2018. The impact of sharing mechanism design on content sharing in online social networks. *Inf. Syst. Res.* 29, 3 (September 2018), 592–611. DOI: https://doi.org/10.1287/isre.2017.0738

[27] Hemant Jain, Balaji Padmanabhan, Paul A. Pavlou, and Raghu T. Santanam. 2018. Humans, algorithms, and augmented intelligence: The future of work, organizations, and society. *Inf. Syst. Res.* 29, 1 (March 2018), 250–251. DOI: https://doi.org/10.1287/isre.2018.0784

[28] Seung Hyun Kim and Juhee Kwon. 2019. How do EHRs and a meaningful use initiative affect breaches of patient information? *Inf. Syst. Res.* 30, 4 (December 2019), 1184–1202. DOI: https://doi.org/10.1287/isre.2019.0858

[29] Mehmet Kuzu, Murat Kantarcioglu, Elizabeth Ashley Durham, Csaba Toth, and Bradley Malin. 2013. A practical approach to achieve private medical record linkage in light of public resources. *J. Am. Med. Informatics Assoc.* 20, 2 (March 2013), 285–292. DOI: https://doi.org/10.1136/amiajnl-2012-000917

[30] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *Nature* 521, 7553 (May 2015), 436–444. DOI: https://doi.org/10.1038/nature14539

[31] Xiao-Bai Li and Jialun Qin. 2017. Anonymizing and sharing medical text records. *Inf. Syst. Res.* 28, 2 (2017), 332–352. DOI: https://doi.org/10.1287/isre.2016.0676

[32] Yunji Liang, Sagar Samtani, Bin Guo, and Zhiwen Yu. 2020. Behavioral biometrics for continuous authentication in the internet-of-things era: An artificial intelligence perspective. *IEEE Internet Things J.* 7, 9 (September 2020), 9128–9143. DOI: https://doi.org/10.1109/JIOT.2020.3004077

[33] Fangyu Lin, Zara Ahmad-Post, Yizhi Liu, James Lee Hu, Mohammadreza Ebrahimi, Jingyu Xin, Sagar Samtani, Weifeng Li, and Hsinchun Chen. 2020. Linking personally identifiable information from the dark web to the surface web: A deep entity resolution approach. In *IEEE International Conference on Data Mining (ICDM) Workshop on Deep Learning for Cyber Threat Intelligence (DL-CTI)*.

[34] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Zhang, Norman Sadeh, Alessandro Acquisti, and Yuvraj Agarwal. 2016. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *Proceedings of the Twelfth USENIX Conference on Usable Privacy and Security (SOUPS'16)*. 27–41.

[35] Yizhi Liu, Fang Yu Lin, Zara Ahmad-Post, Mohammadreza Ebrahimi, Ning Zhang, James Lee Hu, Jingyu Xin, Weifeng Li, and Hsinchun Chen. 2020. Identifying, collecting, and monitoring personally identifiable information: From the dark web to the surface web. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 1–6. DOI: https://doi.org/10.1109/ISI49825.2020.9280540

[36] Yongtai Liu, Chao Yan, Zhijun Yin, Zhiyu Wan, Weiyi Xia, Murat Kantarcioglu, Yevgeniy Vorobeychik, Ellen Wright Clayton, and Bradley A. Malin. 2019. Biomedical research cohort membership disclosure on social media. In *2019 Annual AMIA Symposium Proceedings*. 607–616. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/32308855.

[37] Venkata Vamsikrishna Meduri, Lucian Popa, Prithviraj Sen, and Mohamed Sarwat. 2020. A comprehensive benchmark framework for active learning methods in entity matching. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data*. 1133–1147. DOI: https://doi.org/10.1145/3318464.3380597

[38] National Science & Technology Council. 2019. *The National Artificial Intelligence Research and Development Strategic Plan: 2019 Update*. Washington, D.C. Retrieved from https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf.

[39] National Science & Technology Council. 2019. *Federal Cybersecurity Research and Development Strategic Plan*. Retrieved from https://www.nitrd.gov/pubs/Federal-Cybersecurity-RD-Strategic-Plan-2019.pdf.

[40] National Science Foundation. 2019. National Artificial Intelligence (AI) Research Institutes (2019). nsf20503 | NSF – National Science. Retrieved from https://www.nsf.gov/pubs/2020/nsf20503/nsf20503.pdf.

[41] Jay F. Nunamaker, Nathan W. Twyman, Justin Scott Giboney, and Robert O. Briggs. 2017. Creating high-value real-world impact through systematic programs of research. *MIS Q.* 41, 2 (February 2017), 335–351. DOI: https://doi.org/10.25300/MISQ/2017/41.2.01

[42] Jay F. Nunamaker, Minder Chen, and Titus D. M. Purdin. 1990. Systems development in information systems research. *J. Manag. Inf. Syst.* 7, 3 (1990), 89–106.

[43] Isabelle Oomen and Ronald Leenes. 2008. Privacy risk perceptions and privacy protection strategies. In *Policies and Research in Identity Management*. Springer US, Boston, MA, 121–138. DOI: https://doi.org/10.1007/978-0-387-77996-6_10

[44] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. 2007. A design science research methodology for information systems research. *J. Manag. Inf. Syst.* 24, 3 (2007), 45–77.

[45] Fabian Prasser, James Gaupp, Zhiyu Wan, Weiyi Xia, Yevgeniy Vorobeychik, Murat Kantarcioglu, Klaus Kuhn, and Brad Malin. 2017. An open source tool for game theoretic health data de-identification. In *2017 Annual AMIA Symposium Proceedings*. 1430–1439. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/29854212.

[46] Arun Rai. 2020. Explainable AI: From black box to glass box. *J. Acad. Mark. Sci.* 48, 1 (January 2020), 137–141. DOI: https://doi.org/10.1007/s11747-019-00710-5

[47] Arun Rai, Panos Constantinides, and Saonee Sarker. 2018. Editor's comments: Next-generation digital platforms: Toward human–AI hybrids. *MIS Q.* 43, 1 (2018), iii–ix.

[48] Sagar Samtani, Maggie Abate, Victor Benjamin, and Weifeng Li. 2020. Cybersecurity as an industry: A cyber threat intelligence perspective. In *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Springer International Publishing, Cham, 135–154. DOI: https://doi.org/10.1007/978-3-319-78440-3_8

[49] Sagar Samtani, Kory Chinn, Cathy Larson, and Hsinchun Chen. 2016. AZSecure hacker assets portal: Cyber threat intelligence and malware analysis. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. 19–24. DOI: https://doi.org/10.1109/ISI.2016.7745437

[50] Sagar Samtani, Ryan Chinn, and Hsinchun Chen. 2015. Exploring hacker assets in underground forums. In *2015 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 31–36. DOI: https://doi.org/10.1109/ISI.2015.7165935

[51] Sagar Samtani, Ryan Chinn, Hsinchun Chen, and Jay F. Nunamaker. 2017. Exploring emerging hacker assets and key hackers for proactive cyber threat intelligence. *J. Manag. Inf. Syst.* 34, 4 (2017), 1023–1053.

[52] Sagar Samtani, Murat Kantarcioglu, and Hsinchun Chen. 2020. Trailblazing the artificial intelligence for cybersecurity discipline. *ACM Trans. Manag. Inf. Syst.* 11, 4 (December 2020), 1–19. DOI: https://doi.org/10.1145/3430360

[53] Sagar Samtani, Shuo Yu, Hongyi Zhu, Mark Patton, and Hsinchun Chen. 2016. Identifying SCADA vulnerabilities using passive and active vulnerability assessment techniques. In *2016 IEEE Conference on Intelligence and Security Informatics (ISI)*. 25–30. DOI: https://doi.org/10.1109/ISI.2016.7745438

[54] Sagar Samtani, Hongyi Zhu, and Hsinchun Chen. 2020. Proactively identifying emerging hacker threats from the dark web. *ACM Trans. Priv. Secur.* 23, 4 (August 2020), 1–33. DOI: https://doi.org/10.1145/3409289

[55] Sagar Samtani, Hongyi Zhu, Balaji Padmanabhan, Yidong Chai, and Hsinchun Chen. 2020. Deep learning for information systems research. (October 2020). Retrieved from http://arxiv.org/abs/2010.05774.

[56] Ben Shneiderman, Catherine Plaisant, Maxine Cohen, Steven Jacobs, Niklas Elmqvist, and Nicholas Diakopoulos. 2016. *Designing the User Interface: Strategies for Effective Human-Computer Interaction* (6th Editio ed.). Pearson.

[57] Steven Ullman, Sagar Samtani, Ben Lazarine, Hongyi Zhu, Benjamin Ampel, Mark Patton, and Hsinchun Chen. 2020. Smart vulnerability assessment for scientific cyberinfrastructure: An unsupervised graph embedding approach. In *2020 IEEE International Conference on Intelligence and Security Informatics (ISI)*. 1–6. DOI: https://doi.org/10.1109/ISI49825.2020.9280545

[58] Runzhong Wang, Junchi Yan, and Xiaokang Yang. 2019. Learning combinatorial embedding networks for deep graph matching. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 3056–3065. DOI: https://doi.org/10.1109/ICCV.2019.00315

[59] Philipp Wunderlich, Daniel J. Veit, and Saonee Sarker. 2019. Adoption of sustainable technologies: A mixed-methods study of german households. *MIS Q.* 43, 2 (January 2019), 673–691. DOI: https://doi.org/10.25300/MISQ/2019/12112

[60] Weiyi Xia, Murat Kantarcioglu, Zhiyu Wan, Raymond Heatherly, Yevgeniy Vorobeychik, and Bradley Malin. 2015. Process-driven data privacy. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*. 1021–1030. DOI: https://doi.org/10.1145/2806416.2806580

[61] Alyson Leigh Young and Anabel Quan-Haase. 2013. Privacy protection strategies on Facebook. *Information, Commun. Soc.* 16, 4 (May 2013), 479–500. DOI: https://doi.org/10.1080/1369118X.2013.777757

[62] Dongxiang Zhang, Dongsheng Li, Long Guo, and Kian-Lee Tan. 2020. Unsupervised entity resolution with blocking and graph algorithms. *IEEE Trans. Knowl. Data Eng.* (2020), 1–1. DOI: https://doi.org/10.1109/TKDE.2020.2991063

[63] Dongxiang Zhang, Yuyang Nie, Sai Wu, Yanyan Shen, and Kian-Lee Tan. 2020. Multi-context attention for entity matching. In *Proceedings of The Web Conference 2020*. 2634–2640. DOI: https://doi.org/10.1145/3366423.3380017

[64] Chen Zhao and Yeye He. 2019. Auto-EM: End-to-end fuzzy entity-matching using pre-trained deep models and transfer learning. In *The World Wide Web Conference on - WWW'19*. 2413–2424. DOI: https://doi.org/10.1145/3308558.3313578

[65] Hongyi Zhu, Sagar Samtani, Randall Brown, and Hsinchun Chen. 2021. A deep learning approach for recognizing activity of daily living (ADL) for senior care: Exploiting interaction dependency and temporal patterns. *MIS Q.* (2021), Forthcoming. Retrieved from https://ssrn.com/abstract=3595738.

[66] Hongyi Zhu, Sagar Samtani, Hsinchun Chen, and Jay F. Nunamaker. 2020. Human identification for activities of daily living: A deep transfer learning approach. *J. Manag. Inf. Syst.* 37, 2 (April 2020), 457–483. DOI: https://doi.org/10.1080/07421222.2020.1759961