De-anonymizing, Preserving and Democratizing Data Privacy and Ownership

By

Saurabh Gupta

With the supervision of

Dr. Arun Balaji Buduru, IIIT Delhi
Dr. Ponnurangam Kumaraguru, IIIT Hyderabad

Indraprastha Institute of Information Technology Delhi

May, 2022

De-anonymizing, Preserving and Democratizing Data Privacy and Ownership

By

Saurabh Gupta

Submitted

Comprehensive report for the degree of
Doctor of Philosophy

to

Indraprastha Institute of Information Technology Delhi

May, 2022

Certificate

This is to certify that the thesis titled **"De-anonymizing, Preserving and Democratizing Data Privacy and Ownership"** being submitted by **Saurabh Gupta** to Indraprastha Institute of Information Technology Delhi, for the award of the Ph.D. degree, is an original research work carried out by him under my supervision. In my opinion, the thesis has reached the standards fulfilling the requirements of the regulations relating to the degree.

The results contained in this thesis have not been submitted in part or full to any other university or institute for the award of any degree/diploma.

May 2022

Dr. Arun Balaji Buduru
Department of Computer Science & Engineering
Indraprastha Institute of Information Technology Delhi
New Delhi, India

May 2022

Dr. Ponnurangam Kumaraguru
International Institute of Information Technology Hyderabad
Hyderabad, India

Acknowledgements

I want to thank my thesis advisors Dr. Arun Balaji Buduru, IIIT Delhi, and Dr. Ponnurangam Kumaraguru, IIIT Hyderabad, for their constant support and supervision. I am immensely grateful to them for providing me with their valuable guidance and insights through the course of the thesis. I also want to thank my colleagues at the Precog Lab, IIIT Hyderabad, who volunteered for any support or help I needed for this research project. Finally, I want to express my gratitude to my family for providing me with unfailing support and continuous encouragement through researching and writing this thesis.

May 2022

Saurabh Gupta
Computer Science and Engineering
Indraprastha Institute of Information Technology Delhi
New Delhi, India

Abstract

The fourth industrial revolution warrants a rapid change to technology, industries, and societal patterns and processes in the 21st century due to increasing interconnectivity and intelligent automation. Organizations responsible for shepherding the technology to the next level rely on data-hungry algorithms developed due to the recent advancements in machine learning and deep learning in the last decade. Often, the collected data include Personally Identifiable Information (PII) and pseudo identifiers like age, gender, zip codes, and non-PII attributes. Due to the inclusion of PII attributes, data protection and clearly defining its ownership has become paramount. Despite having several compliances in places like the Health Insurance Portability and Accountability Act (HIPAA) in the US and the National Data Health Mission (NDHM) in India for healthcare, or the more comprehensive General Data Protection Regulation (GDPR) in Europe, we witness wrongful disclosure, theft, and misuse of data by the organizations. The malpractices like selling data to third parties, using weaker anonymizations, and claiming ownership of data often lead to the loss of sensitive information that directly impacts the people whose data is collected and mishandled for the sake of providing services.

[PART I] Anonymization techniques like k-anonymity, l-diversity, t-closeness, etc., are proven to be weak and vulnerable by privacy researchers. Researchers have applied cross-linking methods to de-anonymize several publicly released datasets. For example, patients' data is deanonymized using electronic health records and other public records; the American query logs were de-anonymized to infer that 87% of the American population can be uniquely identified with just three attributes: age, gender, and zip codes. Our research presents a cross-linking attack to identify

personal identifiable information (PII), including address, family details, voter ID information with just the Twitter username, and publicly available electoral rolls. We further show how academic institutions employ weaker anonymizations to release students' information, making them vulnerable to cross-linking attacks.

[**PART II**] Several anonymized data releases failed due to cross-linking vulnerabilities it carries. We will now discuss differential privacy and how it curbs the limitations of anonymization algorithms. Differential privacy (DP) is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals. Unlike anonymization, where we reveal the actual individual samples, DP adds noise in a manner such that personal data samples are protected using randomized responses. Even if the sample gets published publicly, due to the noise added to the individual responses, no leakage of information can be claimed with absolute certainty. We curated a survey on how DP is being used with different modalities of data for privacy protections. We built a framework, imdpGAN, that can generate a synthetic dataset of private and specific images using differential privacy. The dataset can be used for healthcare research without exposing the original sensitive dataset to the public. We extend the application of private data generation to generate multi-table private samples by automatically detecting the data types, primary keys, and relationships between multiple tables. We developed a python package known as *pyadel*, which the data scientists can use to generate synthetic datasets to further their research. We also found that the idea of private data generation can be extended to creating counterfactuals that can mimic a Randomized Controlled Trial (RCT). Instead of conducting a full-scale RCT, one can use frameworks like imdpGAN, or pyadel, to generate counterfactuals that can act as controlled groups for a treatment effect study. The process will involve two steps: i) creating a robust classifier, ii) private synthetic counterfactual generation. We have created the classifier and are experimenting with the generation part.

[**PART III**] Beyond having strong privacy protections, the data generation depends on the users participating in surveys or studies. The current data collection and sharing ecosystem does not assure that the data collected by the organizations

will not be mishandled or sold to third parties. There is no consent first framework that notifies or asks the data owners (users) whether to or not to share their data. To add the notion of consent into current access control protocols and transfer of control to the individual whose data is being collected, we have developed a consent-centric platform, OWNED, that gives back the user the ownership of their data. OWNED lets its users control what all attributes are shared at a granular level, requiring the user's approval for each access, transmission, or sharing request.

Contents

Chapter 1

Introduction

We have entered the era of connected systems, and intelligent systems. We, as consumers, are producing more data every minute due to the technological advancements and we are estimated to generate approximately 450 exabytes of data each day by 2025. The data collection processes involve sharing sensitive and non-sensitive attributes. The sensitive attributes include the PII that can harm an individual if it gets into the wrong hands. The leakage carries threat to confidentiality, integrity, authentication and authorization, which are the four major pillars of cybersecurity. The individuals, if their PII is leaked, get susceptible to financial frauds [Hasham et al., 2019, Al Duhaidahawi et al., 2020, Johnson, 2011], identity theft [McCormick, 2008, Center, 2018], loss of privacy [Narayanan and Shmatikov, 2008, Sweeney, 2002], and so on.

We can confidently say that it is not news to people that protecting their data and privacy is important and is the need of the hour as their misuse can have devastating effects. Despite the awareness, several organizations still rely on age old techniques like anonymization [Sweeney, 2002, Machanavajjhala et al., 2007, Li et al., 2007], random shuffling, data obfuscation, and so on, for their data protection and privacy preservation needs. However, researchers have shown that the anonymization is weak, and can be attacked by gathering background knowledge and cross-linking it with anonymized data [Narayanan and Shmatikov, 2008, Sweeney, 2002]. Several anonymized data releases failed due to cross-linking vulnerabilities it carried. Dif-

ferential privacy [Dwork, 2006a, Dwork and Roth, 2014, Dwork, 2006b, Dwork et al., 2016], is introduced as a solution to the limitations of anonymization. Differential privacy (DP) is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals. Unlike anonymization, where we reveal the actual individual samples, DP adds noise in a manner such that individual data samples are protected using randomized responses. Even if the sample gets published publicly, due to the noise added to the individual responses, no leakage of information can be claimed with absolute certainty. Beyond having strong privacy protections, differential privacy depends on having a large curated dataset. The current data collection and sharing ecosystem does not assure that the data collected by the organizations will not be mishandled or sold to third parties [Sweeney, 2002, Narayanan and Shmatikov, 2008]. Therefore, the privacy preservation that DP can provide will not matter if data is mishandled at the time of collection itself. There is no consent first framework that notifies or asks the data owners (users) whether to or not to share their data, and

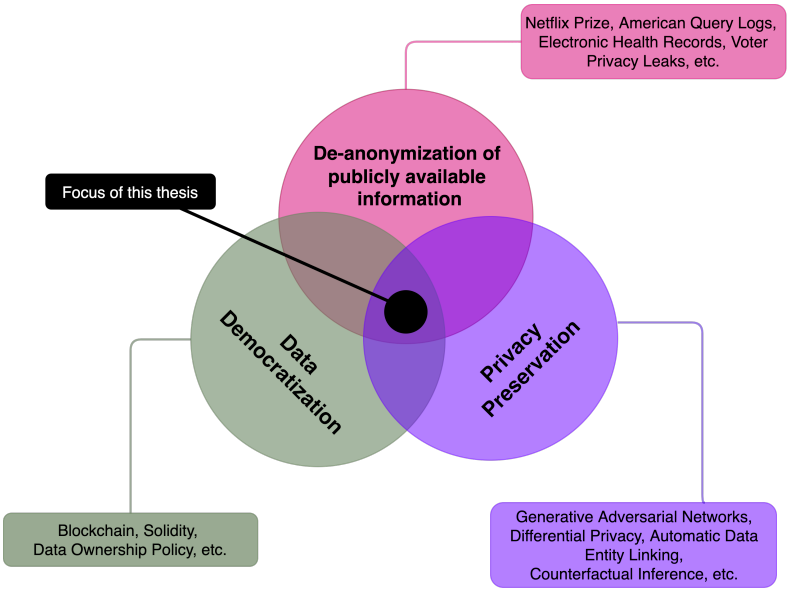

Figure 1.1: Focus of this thesis report. The emphasis is on a combination of one or more of the components to spread awareness about de-anonymization, create technique to preserve privacy, and create foundational blueprints to build a truly consent first data ownership based system.

what attributes of the data to share.

This thesis aims to address these substantial challenges by developing new tools, techniques, and paradigms that enhance individual mindfulness about privacy by presenting de-anonymization based attack scenarios, give them a way to preserve privacy at many stages using generative models in conjunction with differential privacy, and create blueprints of truly private consent first systems that give the ownership of data back to the user.

1.1 De-Anonymization attacks on data privacy

Researchers have shown cross-linking attacks that successfully re-identify individuals from an anonymized datasets that was made public with the promise that the individual data will be preserved. A few notable cross linking attacks are listed below:

- [Narayanan and Shmatikov, 2008] The authors present a new class of statistical de-anonymization attacks against high-dimensional micro-data, such as individual preferences, recommendations, transaction records and so on. They apply our de-anonymization methodology to the Netflix Prize dataset, which contains anonymous movie ratings of 500,000 subscribers of Netflix, the world's largest online movie rental service. They demonstrate that an adversary who knows only a little bit about an individual subscriber can easily identify this subscriber's record in the dataset. Using the Internet Movie Database as the source of background knowledge, the authors successfully identified the Netflix records of known users, uncovering their apparent political preferences and other potentially sensitive information.

- [Sweeney, 2013] The State of Washington sells patient-level health data for $50. This publicly available dataset has virtually all hospitalizations occurring in the State in a given year, including patient demographics, diagnoses, procedures, attending physician, hospital, a summary of charges, and how the bill was paid. It does not contain patient names or addresses (only ZIPs). News-

paper stories printed in the State for the same year that contain the word "hospitalized" often include a patient's name and residential information and explain why the person was hospitalized, such as vehicle accident or assault. News information uniquely and exactly matched medical records in the State database for 35 of the 81 cases (or 43 percent) found in 2011, thereby putting names to patient records. A news reporter verified matches by contacting patients. Employers, financial organizations and others know the same kind of information as reported in news stories making it just as easy for them to identify the medical records of employees, debtors, and others.

1.2 Privacy preservation and protection from de - anonymization using differential privacy

In an attempt to show that such cross-linking opportunities still subsist, we present an attack scenario where we start from a tweet to collect an individual's Twitter information and link that with electoral rolls to reveal their sensitive information like their voter id, address, age, gender, and so on. The cross-linking attacks reveals the caveats of using publishing data when several other data sources are available. To curb breaches due to weak anonymization, researchers have adopted differential privacy as a technique to protect individual privacy while still keeping the utlity of the datasets intact. We curated a survey on how differential privacy is being used with different modalities of data for privacy protections. We built a framework, imdpGAN, that can generate a synthetic dataset of private and specific images using differential privacy. The dataset can be used for research purposes without exposing the original sensitive dataset to the public. We extend the application of private data generation to generate multi-table private samples by automatically detecting the data types, primary keys, and relationships between multiple tables. We developed a python package known as pyadel, which the data scientists can use to generate synthetic datasets to further their research. We also found that the idea of private data generation can be extended to creating counterfactuals that can mimic a randomized controlled trial (RCT). Instead of conducting a full-scale RCT,

one can use frameworks like imdpGAN, or pyadel, to generate counterfactuals that can act as controlled groups for a treatment effect study.

1.3 Democratize data sharing by giving data ownership back to the individual

Beyond having strong privacy protections, the data generation depends on the users participating in surveys or studies. The current data collection and sharing ecosystem does not assure that the data collected by the organizations will not be mishandled or sold to third parties. To add the notion of consent into current access control protocols and transfer of control to the individual whose data is being collected, we have developed a consent-centric platform, OWNED, that gives back the user the ownership of their data. OWNED lets its users control over what all attributes are shared at a granular level, requiring the user's approval for each access, transmission, or sharing request.

Chapter 2

Background and Literature Review

To understand privacy, we first need to understand the categorization of data in the context of privacy. All kinds of heterogeneous data, be it images, videos, text, or any other format, can be categorized as PII (Personally Identifiable Information), pseudo sensitive data, and non-sensitive data. As the name suggests, PIIs are highly sensitive and must be protected from any leak. Pseudo-sensitive data is one where we can choose to protect it based on its application domain. For example, while performing anonymization on a dataset that is to be shared with an employer, the gender attribute might not be considered a PII because just knowing the gender one cannot identify the person. But if it is known that the same employer prefers male candidates over female candidates, gender can be considered as a sensitive attribute because revealing it will cause discrimination towards the candidate.

The vast amount of data about a user has created their digital footprint. The adversary can use the digital footprint to tell a lot about its user. Therefore, it is crucial to protect it using privacy-enhancing technologies. The digital footprint consists of the user's private information and public information. There are clear distinctions between what is needed to be protected and what can remain public. For example, we might easily give people our Facebook, Twitter, or LinkedIn username with little or no hesitation. Still, we will think twice before offering someone our

social security number, aadhar number, credit card details, and address. In an ideal scenario, sensitive data like social security or credit card details must be kept separate from the non-sensitive data.

2.1 Anonymization techniques and their vulnerabilities

In the preliminary work towards privacy protection, data augmentation, shuffling, and masking methods were used to obfuscate and hide sensitive attributes [Wang et al., 2004], [Martínez et al., 2012], [Archana et al., 2018]. In these methods, data is either shuffled to become random or hidden via masking two or more characters present in the data - for example, an email, abc@example.com, when masked might look like a**@example.com. Shuffling might replace the email with another email present in the dataset, xyz@example.com. These techniques are trivial, and data is still exposed, giving minimal privacy protection. In one of our ongoing works, we found a dataset where first four letters of email addresses are masked, and other information like first name, last name were given[1]. In 90% of the cases, the first four letters are found to be nothing but the first four letters of the first name of the individual. Hence, masking techniques are vulnerable to re-identification if additional information is known [Winkler, 2004].

To curb this visibility, several anonymization techniques like k-anonymity [Sweeney, 2002], l-diversity [Machanavajjhala et al., 2006], t-closeness [Li et al., 2007] were invented. k-anonymity uses suppression and generalization to divide the dataset into k groups. k-anonymity is prone to homogeneity attacks as it prevents identity disclosure but still vulnerable to attribute leakage. l-diversity uses generalization and masking within each of the k-groups created using k-anonymity to diversify sensitive attributes and overcome homogeneity attacks. l-diversity is not effective for single sensitive attributes and is difficult to achieve. t-closeness extends l-diversity by ensuring that the distribution of a sensitive attribute in any k-group is close to the distribution of a sensitive attribute in the overall distribution. While anonymization proved to be effective, researchers have shown it to be vulnerable to

[1] https://engineeringstudentsdata.com

de-anonymization attacks [Narayanan and Shmatikov, 2008], [Warner, 1965], [Homer et al., 2008], [Dinur and Nissim, 2003].

Techniques like anonymization are vulnerable and harder to apply where a complete dataset is unknown or has images and videos. Even when anonymization is successfully applied, there is a risk of de-anonymization. Therefore, with continuous efforts from researchers, differential privacy has grown to be a universal and a go-to way to release information publicly. Unlike anonymization, the idea behind differential privacy is that if the effect of making an arbitrary single substitution in the database is small enough, the query result cannot be used to infer much about any single individual, and therefore provides privacy [Dwork et al., 2016].

2.2 Synthetic data generation using differential privacy

In the works that study differential privacy in deep learning, [Abadi et al., 2016] changed the model's training algorithm to make it private by clipping and adding noise to the gradients. Authors also propose a privacy accounting technique and introduce a moments accountant that computes the privacy costs. In [Shokri and Shmatikov, 2015], authors use differential privacy with a parallel and asynchronous training procedure for a multi-party privacy-preserving neural network. It involves transmitting local parameters between server and local task, which has a high risk of information leakage. [Phan et al., 2017] models a private convolutional deep belief network by adding noise on its objective functions and an extra softmax layer. [Xie et al., 2018] leverages the moments accountant and the private training procedure from [Abadi et al., 2016] to train a differentially private generative adversarial network (DPGAN). Authors add noise to the training procedure and avoid a distributed framework to prevent any information leaks. Advantages of DPGAN's techniques over other methods made them a salient choice for privacy preservation in the proposed framework [Xie et al., 2018].

2.3 Data ownership

In the era of big data, the amount of data flowing from individual users to multiple service providers is ever increasing. The service providers, also referred to as entities, often share individual users' data with or without their consent. The shared data can be "homogeneous", i.e., similar attributes of data is being shared between entities, or it can be "heterogeneous", i.e., different attributes belonging to a particular individual are shared and then combined. For example, multiple healthcare providers sharing a patient's data with each other constitute a data sharing among homogeneous entities. Similarly, data sharing between, say, a food chain, a fitness service and a hospital form a sharing among heterogeneous entities.

Needless to say that data sharing is important and helps both the individual user and the businesses as well. For example, in a smart city, data collected about traffic patterns can be shared with different stakeholders such as transportation agencies to reduce traffic jams during rush hour [Shibata et al., 2006]. During disasters, first responders and aid workers can share data between them to better coordinate disaster relief efforts [Bharosa et al., 2010]. Therefore, data sharing among homogeneous and heterogeneous entities can help the service providers improve their quality-of-service, and help them serve individual consumers in a better way.

However, sharing of digital health data is constrained in a lot of countries and several compliances are needed to be followed while sharing such data. As we are progressing, an individual user is becoming more aware about their privacy, and therefore, the recent compliances imposed to protect digital health data define individual's whose data is being collected as **data owners**, and the healthcare providers, insurance companies, or any other service providers are termed as **data users**. In addition, these compliances like HIPAA [Atchinson and Fox, 1997] in the U.S. and NDHD [Jalan, 2019] in India focus on giving data ownership to an individual and asking for content in the complete lifecycle of their data, which starts from provenance, to storage, transmission, access, revocation, and remanence. However, the current ecosystem does not support these services and a data owner loses any control over their data as soon as they share it with a service provider.

Chapter 3

De anonymization (Motivation for the research work)

3.1 Voter Privacy Leaks

As a microblogging service, Twitter is being used by people to spread information and opinions among other users. A lot of times, people are observed reporting the ground events happening near them, making Twitter a source of getting breaking news [Arias, 2019, Sharma, 2019]. For example, when the terrorist attacks in Mumbai in 2008 were happening, Twitter users in India (especially in Mumbai) were providing an instant eyewitness account of what was happening at the ground [Beaumont, 2008]. The platform is considered so pivotal that even the Indian government recently asked Twitter to remove accounts spreading rumors about Kashmir [Rezwan, 2019]. More recently, users used Twitter for discussing the Lok Sabha elections 2019 in India to share news, opinions, facts, fake news, and making it a political playground with #LokSabhaElections2019 among the top three most tweeted hashtags in 2019 [Mathur, 2019, Gupta et al., 2020b, Manu et al., 2020].

Voter Privacy, as given in Section 39 in The Conduct of Elections Rules, 1961, is defined as the maintenance of secrecy of voting by electors within polling station and voting procedure [Dube and Jain, 1985]. Voter privacy is implemented by a method called the secret ballot. The secret ballot is a voting method in which a

voter's choices in an election or a referendum are anonymous, forestalling attempts to influence the voter by intimidation, blackmailing, and potential vote buying [Rusk, 1970]. The origin of the secret ballot in Indian elections is traced back to section 39 in the conduct of elections rules, introduced in 1961. The acts state that "Every elector to whom a ballot paper has been issued under rule 38 or any other provision of these rules shall maintain the secrecy of voting within the polling station". The voters cast their vote in isolation and exercise their right to voter privacy.

((a)) Example reveal of BJP ((b)) Example reveal of INC ((c)) Example reveal of AAP

Figure 3.1: Users reveal their votes while posting tweets on Twitter. We have taken examples of the top three most notable political parties in India, viz., BJP, INC, and AAP. The actual tweets include images of these users with their inked fingers guaranteeing that they did vote in the elections. The links to tweets are: https://twitter.com/srish1794/status/1127506793679208448/photo/1, https://twitter.com/i/web/status/1118753100481716225, https://twitter.com/i/web/status/1127540913964666880, respectively.

Twitter users lose their voter privacy by posting tweets that reveal the name of the party or the candidate they support. In this thesis chapter, we study such tweets, and refer to them as **Voter Privacy Leaks (VPL)**. A user revealing their vote warrants their eligibility to vote, and consequently, the presence of their personal information in electoral rolls. Electoral rolls are a publicly available list of all the voters curated in PDF files and categorized by their states and constituencies[1] maintained by the Election Commission of India (ECI). A single file contains the address of a part that belongs to a constituency and a complete list of voters in that part. Also, it provides several personally identifiable information like voter ID, age, gender, address, and family details along with their names. Identification of an individual using cross-linking the two data sources poses a severe threat like

[1]https://eci.gov.in/electoral-roll/link-to-pdf-e-roll/

identity theft, blackmail, reputation damage, unwanted disclosure and regret, and so on [Gupta and Kumaraguru, 2013, Lyon, 2001, Solove, 2008, Gross and Acquisti, 2005, Wang et al., 2011, Boyd and Ellison, 2007]. The individuals become prone to targeting for several political gains [Hern, 2018, Stokes, 2007] or just for threatening them based on whom they support. Considering the increasing intolerance and lynch mobs in India, such tweets successfully linked with personally identifiable information might even lead to a life and death situation [Dutta, 2018, Khanna et al., 2018].

Previous research on studying leaks on Twitter primarily explores tweets: related to everyday phenomena like vacations, drinking, and diseases [Mao et al., 2011]; exposes violation of privacy settings [Meeder et al., 2010]; build tools to detect personally identifiable information [Cappellari et al., 2017, Caliskan Islam et al., 2014, Liang et al., 2017], and so on. Unlike the general causes, breaching user privacy to target them for political gains is commonplace as several cases have unfolded in recent years [Hern, 2018, Michael Barton and McIntyre, 2018]. Therefore, in this thesis chapter, we first characterize the type of users who are vulnerable and more prone to posting a VPL to lose their voter privacy using attributes like gender, status on Twitter (verified or not), follower count, and the party they support. Second, we cross-link their Twitter details to electoral rolls to reveal several personally identifiable information. Third, we detect and prevent users from posting tweets (VPLs) that might compromise their privacy using a browser-based visual nudge. Our experiments are centered around the following research questions:

- The collected tweets can itself reveal a lot about the users who are posting the VPLs. We try to analyze user attributes that are associated with posting a VPL to ask our first research question:

  RQ1. [Characterization] *What user attributes are associated with an increased likelihood of posting a Voter Privacy Leak?*

- Users on Twitter, directly or indirectly, post VPLs. The Election Commission of India (ECI) releases electoral rolls that contain several personally identifiable information (PII) like age, gender, family details, etc. Cross-linking the two

can pose serious privacy threats, which takes us to the second research question we ask:

RQ2. [Cross-Linking] *Using a Twitter profile and a selection of tweets, can we successfully find PII of an individual in the electoral rolls?*

- When Twitter users are linked with electoral rolls, their single tweet becomes a threat to their privacy. They become vulnerable to identity theft, unwanted disclosure, discrimination, and so on [Solove, 2008]. Therefore, for awareness and mitigation, we ask the third research question:

  RQ3. [Detection and Protection] *How can we detect a voter privacy leak in real time from a tweet to inform the user and prevent it from happening in the future?*

We uncover that (1) a non-verified male Twitter user with less than 100 followers who supports the Indian National Congress (INC) and lives in Madhya Pradesh has the highest probability of losing their voter privacy by posting a VPL. We successfully cross-link (2) the Twitter data set [Gupta et al., 2020b] having 45 million tweets with a subset of electoral rolls to find 44 exact matches. We discuss the cross-linking methodology with a case study of a Twitter user. To mitigate the issue, (3) we formulate a binary classification problem with two classes: "VPL" and "Non-VPL" tweets. A random forest classifier with count vectors performs the best in all evaluation metrics. Using the trained classifier model, we develop a browser extension that nudges users whenever they are posting a tweet that might make them lose their voter privacy.

Chapter 4

Privacy Preservation

4.1 Privacy Protection for heterogeneous data releases

Recent advancements in computing have enhanced the way users and organizations interact with data. From personalized recommendations, e-commerce to industrial and scientific research, data is needed everywhere, and that too in vast abundance. There are 4.6 billion Internet users worldwide[1] (approximately 60% of the world population), each generating 1.7 megabytes of data every second, according to a report[2]. The generated data, in turn, is used with algorithms in machine learning and deep learning to create solutions for self-driving cars, recommendation engines, automated game playing, and so on.

The generation of data is diversified among several domains, therefore, making it heterogeneous. In an Internet of Things (IoT) scenario, data corresponds to values read by installed sensors across the ecosystem. At the same time, a self-driving car captures videos and images along with sensor values. Similarly, a smart meter generates time-series data of power consumption at the home it is installed. The method of parsing a video is different from that of an image. Similarly, naked sensor values are handled by entirely different tools and techniques. Due to its heterogeneous nature, coming up with a single solution to protect data privacy becomes a daunting task.

[1] https://www.statista.com/statistics/617136/digital-population-worldwide
[2] https://www.domo.com/solution/data-never-sleeps-6

Differential privacy provides a way to share information publicly (or release datasets) without compromising the privacy of individual samples present in the dataset [Xiong et al., 2014]. The former is achieved by describing the patterns experienced by the groups within a dataset. Another way differential privacy is used is when instead of revealing the full dataset, only aggregate information is released, limiting the disclosure of private information of records. For example, organizations publish statistical aggregates like percentages and mean values to ensure the confidentiality of survey responses. With its increasing popularity, differential privacy is being used in multiple domains and modified to be applied to heterogeneous data. In 2018, Facebook used differential privacy to release a dataset for researchers to study the role of social media in elections and democracy[3]. We explored the application of differential privacy for private data release with the following types of data: i) images [Xie et al., 2018], [Gupta et al., 2020a]; ii) high dimensions [Xu et al., 2021], [Xu et al., 2017]; iii) time series [Fan et al., 2013]; iv) personalized recommendations [Shen and Jin, 2014]; v) streams [Beck et al., 2017], [Wang et al., 2019]; vi) graphs [Zhang and Ni, 2019]; vii) statistical computations [Senavirathne and Torra, 2019]; viii) internet of things [Zheng et al., 2019].

4.2 ImdpGAN Architecture: for private synthetic image data release

The world of today is moving towards more personalized hardware and software, collecting sensitive information with multiple Personally Identifiable Information (PII) attributes, especially in domains like healthcare and Internet-of-Things (IoT). Often times, deep learning techniques are used to solve problems like detecting cancer patterns [Munir et al., 2019], diabetic retinopathy [Arcadu et al., 2019], and so on. But deep learning typically needs huge amount of data to achieve promising performance. However, in domains like healthcare and IoT (with a lot of PII attributes), it is impossible to get as much data as we want. Also, such models learn finer de-

[3]https://research.facebook.com/blog/2020/02/new-privacy-protected-facebook-data-for-independent-research-on-social-medias-impact-on-democracy

tails in training data and are shown to compromise privacy of individuals. One such example is successful recovery of individual samples from the training set, by using hill climbing on output probabilities [Fredrikson et al., 2015]. Therefore, enforcing privacy while using deep learning techniques to analyze such data has become an absolute necessity. In short there are two challenges: the availability of a huge amount of data and protecting the privacy of individual users.

Generative models have mitigated the data scarcity issue by successfully generating patient records, sensor data, medical records, tabular data [Choi et al., 2017, Alzantot et al., 2017, Camino et al., 2018, Guan et al., 2018, Xu and Veeramachaneni, 2018]. Using the combination of game theory and deep learning, GANs and its many other variants, have demonstrated promising performance in modeling the underlying data distribution [Goodfellow et al., 2014]. These generative models can generate high quality "fake" samples that are hard to differentiate from the real ones [Mogren, 2016, Saito and Matsumoto, 2016, Salimans et al., 2016]. Ideally, we can generate these "fake" data samples to fit our needs and conduct the desired analysis without privacy implications. Although, the generation process is random and we cannot implicitly control the variation in type or style of data we want to generate.

Privacy is being enforced on sensitive data using several anonymization techniques. Some examples include k - anonymity [Sweeney, 2002], l-diversity [Machanavajjhala et al., 2006], t-closeness [Li et al., 2007], which are effective but vulnerable to de-anonymization attacks [Narayanan and Shmatikov, 2008]. Since, these techniques do not solve the data scarcity issue, researchers are trying to introduce privacy preservation in generative models [Xie et al., 2018]. The generation of "fake" samples is not self-sufficient and is prone to disclosure of private information about the individual training samples. The adversarial training procedure with high model complexity often leads to learning a distribution that just copies the training samples. Repeated sampling from such distributions increases the chance of recovering the training samples, hence, compromising the privacy of the data. [Hitaj et al., 2017] demonstrated an inference attack that uses generated samples to recreate the training samples.

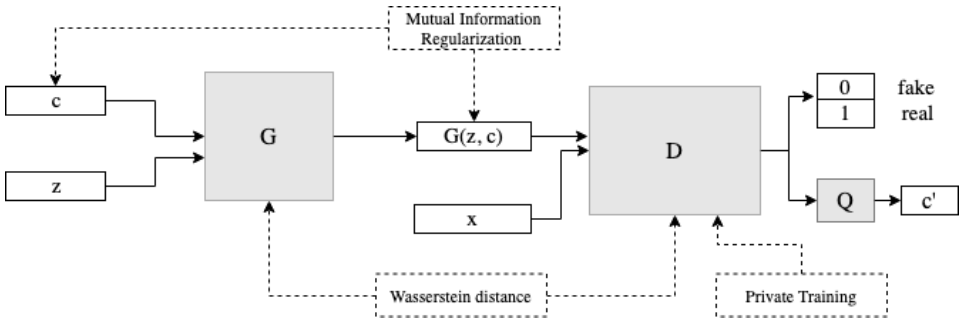

Figure 4.1: ImdpGAN Architecture: addition of the private training procedure, the mutual information regularization and the Wasserstein distance to the GAN architecture.

Our contribution. With the above considerations, we try to learn meaningful latent representations of variation, known as latent codes, to have control over the specificity of the generator output and use a private training procedure to preserve the privacy of the individual training samples. Therefore, in this thesis chapter, we present an amalgamation of techniques from Information Maximizing Generative Adversarial Network, (used to learn interpretable latent representations in an unsupervised manner) and Differentially Private Generative Adversarial Network (used to preserve privacy of the training samples). The models are built using the machine learning framework Pytorch [Paszke et al., 2017].

We propose imdpGAN, a unified framework (as shown in Figure 4.1) to:

1. *Protect privacy of training samples.* Protecting privacy in images simply means that one will not be able to recognize what is there in the image, i.e., the generator will generate blurry images as we increase privacy giving rise to a privacy versus accuracy trade-off. To demonstrate the trade-off, we train a binary classifier on digit pairs and find accuracy on corresponding test samples. Results show that as we increase privacy, the accuracy of binary classifier decreases.

2. *Control specificity of generated samples.* There are two kinds of variations in a data set: discrete and continuous. Discrete variations are represented by different classes. For example, MNIST dataset has 10 classes representing one

digit per class. Changing one class to another is a discrete variation. The dataset has digits positioned at varying angles and having different widths representing continuous variation. We learn tunable latent codes to control both types of variations.

We evaluate our proposed approach on MNIST dataset and extend the imdpGAN framework to complex CelebA [Ziwei et al., 2015] dataset. Results show that imdpGAN preserves privacy and learns meaningful latent codes, which are varied to show class and style variations while generating new images. Although, the classification accuracy decreases as we increase privacy.

As privacy concerns are rising up there are multiple use cases of our framework. For example, popular Face Recognition Systems (FRS) claim that they store only a representation of users' faces and not the actual image[4]. However, while operating they require a complete face image as input to auhenticate an user. The proposed framework, imdpGAN, can be used to create anonymized face images that are closer to the real face representations by learning meaningful latent codes while generating private faces to preserve user's privacy. Although, there will be a trade-off between the accuracy of the FRS and privacy of the face image, which can further be adjusted by tunable parameters.

4.3 Pyadel: for private multi table synthetic data release

In many cases, an organization wishes to release some data, but is restricted in the amount of data to be released due to legal, privacy and other concerns. For instance, the US Census Bureau releases only 1% of its table of records every year, along with statistics about the entire table. However, the machine learning (ML) models trained on the released sub-table are usually sub-optimal. Generating synthetic data for a single table is easy as it does not involve finding primary keys, consistently making sure that the primary key - foreign key relationship is main-

[4]Apple tweeted, "Face ID only stores a mathematical representation of your face on iPhone, not a photo.", https://twitter.com/apple/status/1215224753449066497

tained, and constraints are discovered automatically to some extent. Researchers have used generative adversarial networks (GANs) to generate tabular data in the past [Xu et al., 2019,Bourou et al., 2021,Brenninkmeijer et al., 2019]. None of them successfully maintain relationships and follow constraints when it comes to multi table synthetic data generation.

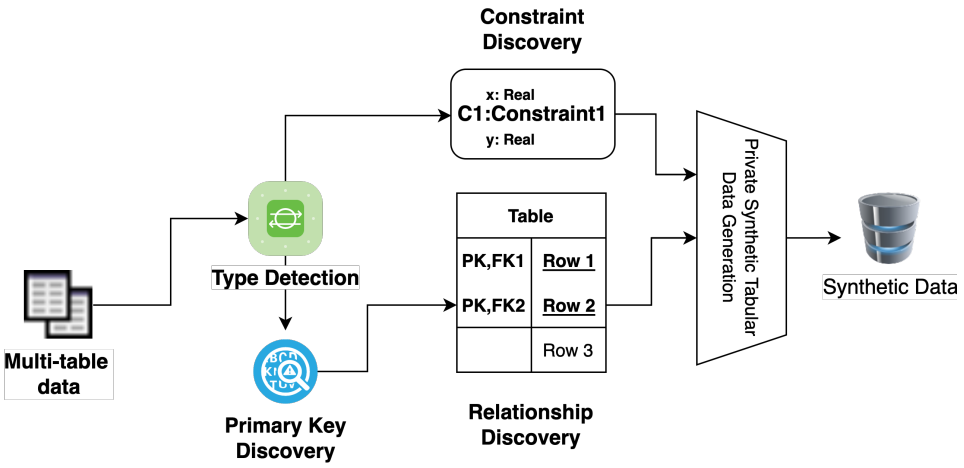

Figure 4.2: Pyadel Flow Diagram

In attempts at multi-table data generation, [Chen et al., 2019] attempts to generate a synthetic table from the released sub-table using ITS-GAN, under the constraints that the generated synthetic table (i) has similar statistics as the entire table, and (ii) preserves the functional dependencies of the released sub-table. However, the limitations are two fold: if the sub-table is not representative enough of the original data we only generate sub optimal results and the use of autoencoders to find functional dependencies can end up replicating the sub-table samples. We are trying an automatic data linkage technique to first detect the type of data present in multiple tables, then we find the primary keys, and finally the primary key-foreign key relationships. Once we figure out the schema of multiple tables, we use the synthetic data vault which includes GAN models for tabular generation and introduce differential privacy mechanisms in it. Finally, for constraint discovery we create an automatic machine learning pipeline for custom constraints discovery. The flow diagram of *pyadel* is shown in Figure 4.2. We created a python package, pyadel,

which can be used by data scientists and researchers to solve problems where data is scarce.

4.4 Counterfactual generation for treatment effect inference data release

Big data, high-performance computing, and (deep) machine learning are increasingly becoming key to precision medicine from identifying disease risks and taking preventive measures, to making diagnoses and personalizing treatment for individuals. Precision medicine, however, is not only about predicting risks and outcomes, but also about weighing interventions. Interventional clinical predictive models require the correct specification of cause and effect, and the calculation of so-called counterfactuals, that is, alternative scenarios. In biomedical research, observational studies are commonly affected by confounding and selection bias. Without robust assumptions, often requiring a priori domain knowledge, causal inference is not feasible. Data-driven prediction models are often mistakenly used to draw causal effects, but neither their parameters nor their predictions necessarily have a causal interpretation. Therefore, the premise that data-driven prediction models lead to trustable decisions/interventions for precision medicine is questionable. When pursuing intervention modelling, the bio-health informatics community needs to employ causal approaches and learn causal structures.

In the works on counterfactual inference, the authors discuss how target trials (algorithmic emulation of randomized studies), transportability (the licence to transfer causal effects from one population to another) and prediction invariance (where a true causal model is contained in the set of all prediction models whose accuracy does not vary across different settings) are linchpins to developing and testing intervention models [Prosperi et al., 2020, de Oliveira and Martens, 2021]. Treatment effect and their counterfactual are majorly part of clinical studies that involve highly sensitive healthcare data. A lot of times the healthcare compliance becomes a barrier for conducting such trials. To solve the issue, we are trying to use differential privacy and causal inference to generate counterfactual for treatment-

effect analysis. We believe that a good solution can act as a precursor to analyze clinical results that might save us from conducting full scale randomized controlled trails.

There are two components to this problem, first is to build a robust classifier and second to use generative models for counterfactuals. We have performed experiments with shallow machine learning models, ensemble models and boosting techniques. We were able to build a robust classifier using adaboost with a descision tree as a base estimator using oversampling techniques like ADASYN and SMOTE. The results of classification are shown in Table 4.1 and Table 4.2. The work for second part is under progress.

Sampling	**Classifier**	**F1-score**	**MCC**	**Accuracy**	**ROC-AUC**
Oversampling	Adaboost*	0.93±0.01	0.86±0.03	0.93±0.06	0.97±0.00
	Adaboost**	0.87±0.02	0.74±0.04	0.87±0.02	0.95±0.01
Undersampling	Adaboost*	0.83±0.03	0.68±0.06	0.83±0.03	0.92±0.02
	Adaboost**	0.84±0.03	0.68±0.06	0.84±0.03	0.92±0.02
ADASYN	Adaboost*	0.93±0.03	0.86±0.06	0.93±0.03	0.98±0.01
	Adaboost**	0.90±0.02	0.81±0.04	0.90±0.02	0.96±0.02
SMOTE	Adaboost*	**0.95±0.03**	**0.89±0.07**	**0.95±0.03**	**0.98±0.01**
	Adaboost**	0.92±0.03	0.86±0.07	0.92±0.03	0.97±0.02

Table 4.1: Classification report after applying sampling techniques like random under/over sampling, ADASYN, and SMOTE. *=XGBoost as base esimator; **=Decision Tree as base estimator; MCC=Matthew's Correlation Coefficient

Sampling	**Classifier**	**PPV**	**NPV**	**Sensitivity**	**Specificity**
Oversampling	Adaboost*	0.91±0.02	**0.95±0.01**	**0.95±0.01**	0.90±0.02
	Adaboost**	0.87±0.03	0.87±0.01	0.87±0.01	0.87±0.03
Undersampling	Adaboost*	0.85±0.03	0.83±0.06	0.82±0.08	0.85±0.04
	Adaboost**	0.85±0.01	0.83±0.06	0.82±0.09	0.86±0.02
ADASYN	Adaboost*	0.92±0.01	0.92±0.05	0.91±0.06	0.94±0.01
	Adaboost**	0.92±0.02	0.89±0.04	0.88±0.05	0.92±0.02
SMOTE	Adaboost*	**0.95±0.01**	**0.95±0.05**	0.93±0.07	**0.95±0.01**
	Adaboost**	0.93±0.02	0.92±0.06	0.92±0.07	0.93±0.02

Table 4.2: Classification report after applying sampling techniques like random under/over sampling, ADASYN, and SMOTE. *=XGBoost as base esimator; **=Decision Tree as base estimator;

Chapter 5

Democratizing data sharing and ownership (Future of Privacy)

5.1 Owned: a consent first architecture that gives ownership to the user

The current architectures of data sharing platforms either do not involve any logical agreement between entities, or even if they have a a way to enforce constraints on data sharing, the individual user, i.e., the data owner is not involved in that decision making. These architectures exist in predominantly in healthcare domain, and implemented by biomedical researchers, medical schools and healthcare providers.

We propose, OWNED, a consent-first data sharing platform to give data ownership back to the individuals. We used Ethereum blockchain network to store the smart contracts created on OWNED. The contracts are designed in a way such that the data user will request the data owner for sharing the attributes, and data moves only when the request is approved. The storage is common, therefore, at any point in future, data owner can revoke the access of attributes. We provide data owners (individuals whose data is collected) for obvious reasons and data users (service providers) all four CRUD operations as they can add additional attributes that might help in providing a better quality of service.

The framework, OWNED, has four components. We use a blockchain network

to act as a data storage, an interface that allows us to execute smart contracts, an interface for the end user to perform available operations, and an initial setup stage for identifying and classifying data attributes. The system architecture diagram is shown in Figure 5.1.

We are building the framework using smart contracts on Ethereum blockchain, and a user inferface to support the operations mentioned in Figure 5.1.

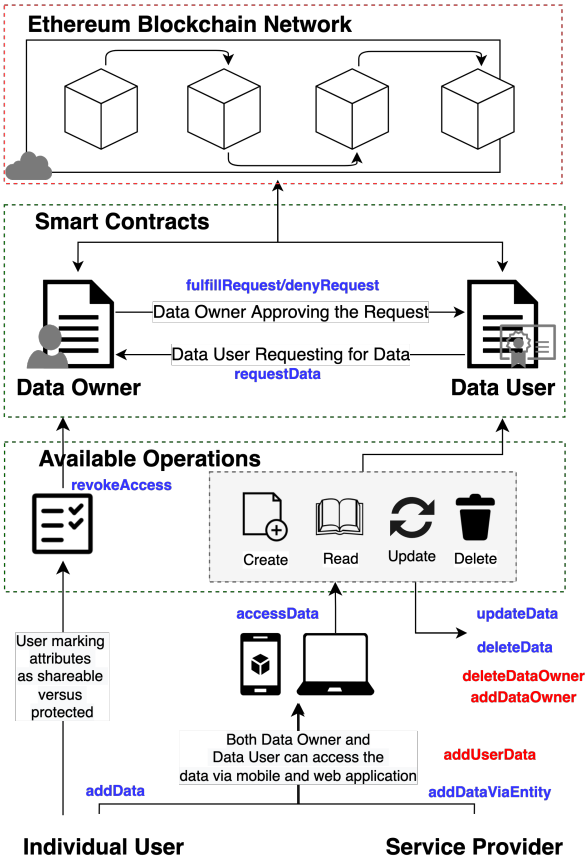

Figure 5.1: System Architecture of Proposed Framework. There are two smart contracts, one for the data owner and one for the data user. The modules for data owner are represented in blue, and the modules for data user are represented in red.

Chapter 6

Timeline

S.No.	TASK TITLE	PHASE ONE (Jun-Aug 2022)	PHASE TWO (Sept-Dec 2022)	PHASE THREE (Jan-Mar 2023)	PHASE FOUR (Apr-May 2023)
1	**Counterfactual Generation**				
1.1	Build the robust classifier				
1.2	Submit the classifier results Computers in Biology and Medicine (Impact factor: 4.5)				
1.3	Use classifier for counterfactual generation				
1.4	Results validation and experimentation				
1.5	Submit the generation results Computers in Biology and Medicine (Impact factor: 4.5)				
2	**Run Pyadel on public datasets**				
2.1	Create results on public datasets				
2.2	Write Pyadel draft				
2.3	Submit demo track paper to Cods-Comad 2023				
3	**Owned Framework**				
3.1	Feature Improvements				
3.2	Paper Improvements				
3.3	Submit results and publish				
4	**Thesis Writing and Defense**				
4.1	Writing Deanonymization part				
4.2	Writing Privacy Preservation part				
4.3	Writing Demcratization part				
4.4	Drafts to commitee				
4.5	Defense				

Figure 6.1: Timeline

Figure 6.1 shows an approximate schedule of my research and writing towards completion of my dissertation. I have broken the schedule into four phases as shown and I plan on completing my dissertation next year.

Chapter 7

Outline of the Thesis

- Chapter 1: Introduction
  - Deanonymization of anonymized public datasets
  - Privacy preservation using differential privacy as a tool
  - Democratizing data ownership
  - Impact
  - Summary including thesis statement
- Chapter 2: Background and literature review
  - Anonymization techniques and cross linking attacks on anonymization techniques
  - Differential privacy to overcome limitations of anonymization
  - Private synthetic data generation using differential privacy
  - Data democratization tools
- PART I - DE-ANONYMIZATION
  - Chapter 3: Voter Privacy Leaks
- PART II - PRIVACY PRESERVATION
  - Chapter 4: Privacy Protection for heterogeneous data releases

  - Chapter 5: ImdpGAN Architecture: for private synthetic image data release
  - Chapter 6: Pyadel: for private multi table synthetic data release
  - Chapter 7: Counterfactual generation for treatment effect inference data release
- PART III - DEMOCRATIZATION OF OWNERSHIP
  - Chapter 8: Owned: a consent first architecture that gives ownership back to the user
- Chapter 9: Conclusion and Future Work
  - Discussion of all the privacy preservation results
  - Next steps in privacy preservation and data democratization
  - Limitations of the thesis

Chapter 8

Conclusion

In this report, we explored various research attempts towards de-anonymizing multiple datasets, preserving privacy by eliminating issues that arise due to anonymization, and presenting framework that democratizes data sharing and ownership. The aim of this survey was to look at relevant literature, which could aid in studying and cross-linking public releases of datasets to prevent data frrm getting revealed and ending up in the wrong hands. Once we get an understanding of the limitations in the data release process, it is important to come up with ways that already exist in literature (or need to be invented) to preserve privacy. Finally, the current ecosystem where data brokers exist and the presence of users is continuously tracked online, we must start onsidering consent first systems that give data ownership back to the user, and has accountability on the data brokers' part.

In order to keep this survey focused, we did not cover the data breaches that occur due to hacking, unprotected sensitive information, social engineering, poor security measures or misconfigurations, but focus on scenarios where data is accidentally published or published with anonymity and deanonymized. With that assumption in mind, we explore methodologies that solve the problem of accidental publishing or de-anonymization. While finding a solution, we discover that the current systems are designed so as to give data brokers free access and opportunities to abuse users information. Therefore, we briefly discuss what the future of data privacy should look like, even though it comes at a cost of a lesser user friendly system design. Apart

from technical limitations, there exist various research gaps in existing literature, which are yet to be addressed and explored.

Bibliography

[Abadi et al., 2016] Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., and Zhang, L. (2016). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, CCS '16, pages 308–318, New York, NY, USA. ACM.

[Al Duhaidahawi et al., 2020] Al Duhaidahawi, H. M. K., Zhang, J., Abdulreda, M. S., Sebai, M., and Harjan, S. (2020). The financial technology (fintech) and cybersecurity: Evidence from iraqi banks. *International Journal of Research in Business and Social Science (2147-4478)*, 9(6):123–133.

[Alzantot et al., 2017] Alzantot, M., Chakraborty, S., and Srivastava, M. B. (2017). Sensegen: A deep learning architecture for synthetic sensor data generation. *2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 188–193.

[Arcadu et al., 2019] Arcadu, F., Benmansour, F., Maunz, A., Willis, J., Haskova, Z., and Prunotto, M. (2019). Deep learning algorithm predicts diabetic retinopathy progression in individual patients. *npj Digital Medicine*, 2(1):92.

[Archana et al., 2018] Archana, R., Hegadi, R. S., and Manjunath, T. (2018). A study on big data privacy protection models using data masking methods. *International Journal of Electrical and Computer Engineering*, 8(5):3976.

[Arias, 2019] Arias, B. (2019). How newsrooms can use twitter's latest tools to break news.

[Atchinson and Fox, 1997] Atchinson, B. K. and Fox, D. M. (1997). From the field: The politics of the health insurance portability and accountability act. *Health affairs*, 16(3):146–150.

[Beaumont, 2008] Beaumont, C. (2008). Mumbai attacks: Twitter and flickr used to break news.

[Beck et al., 2017] Beck, M., Bhatotia, P., Chen, R., Fetzer, C., Strufe, T., et al. (2017). Privapprox: privacy-preserving stream analytics. In *2017 USENIX Annual Technical Conference USENIX ATC 17*, pages 659–672.

[Bharosa et al., 2010] Bharosa, N., Lee, J., and Janssen, M. (2010). Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, 12(1):49–65.

[Bourou et al., 2021] Bourou, S., El Saer, A., Velivassaki, T.-H., Voulkidis, A., and Zahariadis, T. (2021). A review of tabular data synthesis using gans on an ids dataset. *Information*, 12(9).

[Boyd and Ellison, 2007] Boyd, D. M. and Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of computer-mediated Communication*, 13(1):210–230.

[Brenninkmeijer et al., 2019] Brenninkmeijer, B., de Vries, A., Marchiori, E., and Hille, Y. (2019). *On the generation and evaluation of tabular data using GANs*. PhD thesis, Radboud University Nijmegen, The Netherlands.

[Caliskan Islam et al., 2014] Caliskan Islam, A., Walsh, J., and Greenstadt, R. (2014). Privacy detective: Detecting private information and collective privacy behavior in a large social network. In *Proceedings of the 13th Workshop on Privacy in the Electronic Society*, pages 35–46. ACM.

[Camino et al., 2018] Camino, R., Hammerschmidt, C. A., and State, R. (2018). Generating multi-categorical samples with generative adversarial networks. *CoRR*, abs/1807.01202.

[Cappellari et al., 2017] Cappellari, P., Chun, S., and Perelman, M. (2017). A tool for automatic assessment and awareness of privacy disclosure. In *Proceedings of the 18th Annual International Conference on Digital Government Research*, pages 586–587. ACM.

[Center, 2018] Center, I. A. (2018). Global data leakage report, 2017.

[Chen et al., 2019] Chen, H., Jajodia, S., Liu, J., Park, N., Sokolov, V., and Subrahmanian, V. (2019). Faketables: Using gans to generate functional dependency preserving tables with bounded real data. In *IJCAI*, pages 2074–2080.

[Choi et al., 2017] Choi, E., Biswal, S., Malin, B., Duke, J., Stewart, W. F., and Sun, J. (2017). Generating multi-label discrete patient records using generative adversarial networks. In Doshi-Velez, F., Fackler, J., Kale, D., Ranganath, R., Wallace, B., and Wiens, J., editors, *Proceedings of the 2nd Machine Learning for Healthcare Conference*, volume 68 of *Proceedings of Machine Learning Research*, pages 286–305, Boston, Massachusetts. PMLR.

[de Oliveira and Martens, 2021] de Oliveira, R. M. B. and Martens, D. (2021). A framework and benchmarking study for counterfactual generating methods on tabular data. *Applied Sciences*, 11(16):7274.

[Dinur and Nissim, 2003] Dinur, I. and Nissim, K. (2003). Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210.

[Dube and Jain, 1985] Dube, M. and Jain, K. (1985). *Elections, law and procedures.* Vedpal Law House.

[Dutta, 2018] Dutta, P. K. (2018). 16 lynchings in 2 months. is social media the new serial killer?

[Dwork, 2006a] Dwork, C. (2006a). Differential privacy. In *Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II*, ICALP'06, pages 1–12, Berlin, Heidelberg. Springer-Verlag.

[Dwork, 2006b] Dwork, C. (2006b). Differential privacy. In *International Colloquium on Automata, Languages, and Programming*, pages 1–12. Springer.

[Dwork et al., 2016] Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2016). Calibrating noise to sensitivity in private data analysis. *Journal of Privacy and Confidentiality*, 7(3):17–51.

[Dwork and Roth, 2014] Dwork, C. and Roth, A. (2014). The algorithmic foundations of differential privacy. *Found. Trends Theor. Comput. Sci.*, 9(3–4):211–407.

[Fan et al., 2013] Fan, L., Xiong, L., and Sunderam, V. (2013). Differentially private multi-dimensional time series release for traffic monitoring. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 33–48. Springer.

[Fredrikson et al., 2015] Fredrikson, M., Jha, S., and Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1322–1333, New York, NY, USA. ACM.

[Goodfellow et al., 2014] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. D., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc.

[Gross and Acquisti, 2005] Gross, R. and Acquisti, A. (2005). Information revelation and privacy in online social networks. In *Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society*, WPES '05, page 71–80, New York, NY, USA. Association for Computing Machinery.

[Guan et al., 2018] Guan, J., Li, R., Yu, S., and Zhang, X. (2018). Generation of synthetic electronic medical record text. *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 374–380.

[Gupta et al., 2020a] Gupta, S., Buduru, A. B., and Kumaraguru, P. (2020a). imdp-gan: Generating private and specific data with generative adversarial networks. In *2020 Second IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, pages 64–72. IEEE.

[Gupta and Kumaraguru, 2013] Gupta, S. and Kumaraguru, P. (2013). Ocean: Open-source collation of egovernment data and networks-understanding privacy leaks in open government data. *arXiv preprint arXiv:1312.2784*.

[Gupta et al., 2020b] Gupta, S., Singh, A. K., Buduru, A. B., and Kumaraguru, P. (2020b). Hashtags are (not) judgemental: The untold story of lok sabha elections 2019. In *2020 IEEE Sixth International Conference on Multimedia Big Data (BigMM)*, pages 216–220. IEEE.

[Hasham et al., 2019] Hasham, S., Joshi, S., and Mikkelsen, D. (2019). Financial crime and fraud in the age of cybersecurity. *McKinsey & Company*, pages 1–11.

[Hern, 2018] Hern, A. (2018). Cambridge analytica: how did it turn clicks into votes?

[Hitaj et al., 2017] Hitaj, B., Ateniese, G., and Perez-Cruz, F. (2017). Deep models under the gan: Information leakage from collaborative deep learning. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, CCS '17, pages 603–618, New York, NY, USA. ACM.

[Homer et al., 2008] Homer, N., Szelinger, S., Redman, M., Duggan, D., Tembe, W., Muehling, J., Pearson, J. V., Stephan, D. A., Nelson, S. F., and Craig, D. W. (2008). Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays. *PLoS genetics*, 4(8):e1000167.

[Jalan, 2019] Jalan, T. (2019). Summary: National digital health blueprint proposes new body for digital health mission, health data exchanges and registries. medianama. retrieved march 7, 2020.

[Johnson, 2011] Johnson, V. R. (2011). Credit-monitoring damages in cybersecurity tort litigation. *Geo. Mason L. Rev.*, 19:113.

[Khanna et al., 2018] Khanna, P., Ghadyalpatil, A., and Das, S. (2018). Death by social media.

[Li et al., 2007] Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*, pages 106–115. IEEE.

[Li et al., 2007] Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115.

[Liang et al., 2017] Liang, H., Shen, F., and Fu, K.-w. (2017). Privacy protection and self-disclosure across societies: A study of global twitter users. *new media & society*, 19(9):1476–1497.

[Lyon, 2001] Lyon, D. (2001). *Surveillance society: Monitoring everyday life*. McGraw-Hill Education (UK).

[Machanavajjhala et al., 2006] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. (2006). L-diversity: privacy beyond k-anonymity. In *22nd International Conference on Data Engineering (ICDE'06)*, pages 24–24.

[Machanavajjhala et al., 2007] Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3–es.

[Manu et al., 2020] Manu, D., Krishnan, R., and Kumaraguru, P. (2020). Analysing how the shift in discourses on social media affected the narrative around the indian general election 2019. *Journal of Advanced Research in Social Sciences*, 3(1):21–31.

[Mao et al., 2011] Mao, H., Shuai, X., and Kapadia, A. (2011). Loose tweets: an analysis of privacy leaks on twitter. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society*, pages 1–12. ACM.

[Martínez et al., 2012] Martínez, S., Sánchez, D., Valls, A., and Batet, M. (2012). Privacy protection of textual attributes through a semantic-based masking method. *Information Fusion*, 13(4):304–314.

[Mathur, 2019] Mathur, N. (2019). Twitter celebrates 12th birthday of the hashtag.

[McCormick, 2008] McCormick, M. (2008). Data theft: a prototypical insider threat. In *Insider Attack and Cyber Security*, pages 53–68. Springer.

[Meeder et al., 2010] Meeder, B., Tam, J., Kelley, P. G., and Cranor, L. F. (2010). Rt@ iwantprivacy: Widespread violation of privacy settings in the twitter social network. In *Proceedings of the Web*, volume 2, pages 1–2.

[Michael Barton and McIntyre, 2018] Michael Barton, P. D. and McIntyre, N. (2018). Digital election: what demographics are the parties targeting?

[Mogren, 2016] Mogren, O. (2016). C-rnn-gan: Continuous recurrent neural networks with adversarial training. *CoRR*, abs/1611.09904.

[Munir et al., 2019] Munir, K., Elahi, H., Ayub, A., Frezza, F., and Rizzi, A. (2019). Cancer diagnosis using deep learning: A bibliographic review. *Cancers*, 11(9).

[Narayanan and Shmatikov, 2008] Narayanan, A. and Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy*, SP 2008, pages 111–125, Washington, DC, USA. IEEE Computer Society.

[Paszke et al., 2017] Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A. (2017). Automatic differentiation in pytorch.

[Phan et al., 2017] Phan, N. H., Wu, X., and Dou, D. (2017). Preserving differential privacy in convolutional deep belief networks. pages 1681–1704.

[Prosperi et al., 2020] Prosperi, M., Guo, Y., Sperrin, M., Koopman, J. S., Min, J. S., He, X., Rich, S., Wang, M., Buchan, I. E., and Bian, J. (2020). Causal

inference and counterfactual prediction in machine learning for actionable healthcare. *Nature Machine Intelligence*, 2(7):369–375.

[Rezwan, 2019] Rezwan (2019). Indian government asks twitter to remove accounts spreading rumours about kashmir.

[Rusk, 1970] Rusk, J. G. (1970). The effect of the australian ballot reform on split ticket voting: 1876–1908. *American Political Science Review*, 64(4):1220–1238.

[Saito and Matsumoto, 2016] Saito, M. and Matsumoto, E. (2016). Temporal generative adversarial nets. *CoRR*, abs/1611.06624.

[Salimans et al., 2016] Salimans, T., Goodfellow, I. J., Zaremba, W., Cheung, V., Radford, A., and Chen, X. (2016). Improved techniques for training gans. In *NIPS*.

[Senavirathne and Torra, 2019] Senavirathne, N. and Torra, V. (2019). Integral privacy compliant statistics computation. In *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, pages 22–38. Springer.

[Sharma, 2019] Sharma, A. (2019). Why twitter is still the best place for breaking news despite its many challenges.

[Shen and Jin, 2014] Shen, Y. and Jin, H. (2014). Privacy-preserving personalized recommendation: An instance-based approach via differential privacy. In *2014 IEEE International Conference on Data Mining*, pages 540–549. IEEE.

[Shibata et al., 2006] Shibata, N., Terauchi, T., Kitani, T., Yasumoto, K., Ito, M., and Higashino, T. (2006). A method for sharing traffic jam information using inter-vehicle communication. In *2006 3rd Annual International Conference on Mobile and Ubiquitous Systems-Workshops*, pages 1–7. IEEE.

[Shokri and Shmatikov, 2015] Shokri, R. and Shmatikov, V. (2015). Privacy-preserving deep learning. In *Proceedings of the 22Nd ACM SIGSAC Conference on Computer and Communications Security*, CCS '15, pages 1310–1321, New York, NY, USA. ACM.

[Solove, 2008] Solove, D. J. (2008). *Understanding privacy*, volume 173. Harvard university press Cambridge, MA.

[Stokes, 2007] Stokes, S. C. (2007). Political clientelism. In *The Oxford handbook of political science*. Oxford University Press.

[Sweeney, 2002] Sweeney, L. (2002). K-anonymity: A model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10(5):557–570.

[Sweeney, 2013] Sweeney, L. (2013). Matching known patients to health records in washington state data.

[Wang et al., 2019] Wang, J., Liu, C., Fu, X., Luo, X., and Li, X. (2019). A three-phase approach to differentially private crucial patterns mining over data streams. *Computers & Security*, 82:30–48.

[Wang et al., 2004] Wang, K., Yu, P. S., and Chakraborty, S. (2004). Bottom-up generalization: A data mining solution to privacy protection. In *Fourth IEEE International Conference on Data Mining (ICDM'04)*, pages 249–256. IEEE.

[Wang et al., 2011] Wang, Y., Norcie, G., Komanduri, S., Acquisti, A., Leon, P. G., and Cranor, L. F. (2011). I regretted the minute i pressed share: A qualitative study of regrets on facebook. In *Proceedings of the seventh symposium on usable privacy and security*, page 10. ACM.

[Warner, 1965] Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69.

[Winkler, 2004] Winkler, W. E. (2004). Re-identification methods for masked microdata. In *International Workshop on Privacy in Statistical Databases*, pages 216–230. Springer.

[Xie et al., 2018] Xie, L., Lin, K., Wang, S., Wang, F., and Zhou, J. (2018). Differentially private generative adversarial network. *CoRR*, abs/1802.06739.

[Xiong et al., 2014] Xiong, P., Zhu, T.-Q., and Wang, X.-F. (2014). A survey on differential privacy and applications. *Jisuanji Xuebao/Chinese Journal of Computers*, 37(1):101–122.

[Xu et al., 2017] Xu, C., Ren, J., Zhang, Y., Qin, Z., and Ren, K. (2017). Dppro: Differentially private high-dimensional data release via random projection. *IEEE Transactions on Information Forensics and Security*, 12(12):3081–3093.

[Xu et al., 2021] Xu, H., Ding, X., Jin, H., and Yu, Q. (2021). A multi-dimensional index for privacy-preserving queries in cloud computing. *Concurrency and Computation: Practice and Experience*, 33(8):e5458.

[Xu et al., 2019] Xu, L., Skoularidou, M., Cuesta-Infante, A., and Veeramachaneni, K. (2019). Modeling tabular data using conditional gan. *Advances in Neural Information Processing Systems*, 32.

[Xu and Veeramachaneni, 2018] Xu, L. and Veeramachaneni, K. (2018). Synthesizing tabular data using generative adversarial networks. *CoRR*, abs/1811.11264.

[Zhang and Ni, 2019] Zhang, S. and Ni, W. (2019). Graph embedding matrix sharing with differential privacy. *IEEE Access*, 7:89390–89399.

[Zheng et al., 2019] Zheng, Z., Wang, T., Wen, J., Mumtaz, S., Bashir, A. K., and Chauhdary, S. H. (2019). Differentially private high-dimensional data publication in internet of things. *IEEE Internet of Things Journal*, 7(4):2640–2650.

[Ziwei et al., 2015] Ziwei, L., Ping, L., Xiaogang, W., and Tang, X. (2015). Deep learning face attributes in the wild. In *Proceedings of International Conference on Computer Vision (ICCV)*, page 3730–3738, Washington, DC, USA.