Is Privacy a Myth for Facebook Users?

Vishwas T. Patil[a] and R. K. Shyamasundar

Information Security Research and Development Center, Department of Computer Science and Engineering,
Indian Institute of Technology Bombay, Mumbai 400076, Maharashtra, India

Keywords: Online Social Networks, Personally Identifiable Information, Privacy, Right-to-be-forgotten.

Abstract: The management of personal information has become an insurmountable problem. The reasons are multi-fold and intertwined: technological, legal, regulatory, commercial, and behavioural. The proliferation of online social networks like Facebook has made the problem even more acute because of its business model where users' personally identifiable information is monetised via advertisements. One of the distinctive features of privacy policies is that users' data can be shared with their consent for specified purposes; but users do not have effective control over that data once it is shared with third-parties. There is a race to collect user data as it helps in building unique insights about the users. These insights help in matching the users to advertisements with high success. As advertisers seek a maximum return on investments and the data platforms thrive to achieve this expectation. With the current sophistication levels of data platforms in collecting and processing user data, we highlight why it appears futile to achieve privacy despite correct privacy settings enabled. The business model of monetizing of user data and a slow evolution (or absence in some jurisdictions) of legal frameworks to control proliferation of user data has lead to a power asymmetry in the data ecosystem between the motivated data processors and hapless end-users; thus making the users anxious about their participation in the ecosystem. Erosion of user trust has economic consequences. And a lack of continuous flow of data (volume, variety, velocity, and veracity) into the ecosystem will starve the emergence of data-driven innovations with profound societal impact. We elaborate approaches that could help restore the sense of privacy.

1 INTRODUCTION

We have entered into a digital era where every aspect our lives is interacting with a digital service. The data and metadata of these interactions reflect upon our personality. Personality based models help advertisers to predict consumer response to an advertisement/information. Therefore, users' data and metadata has garnered immense monetary potential. The transactional data is governed by privacy policies and service agreements; whereas, the metadata, which is the data collected/observed by ISPs, DNSs, payment providers, et al., is usually ungoverned. The observers of a transaction may not be privy to the data within the transaction but they witness a portion of the digital service, which allows them to make inferences. The transactional and observational data that we produce, share, and consume has given rise to a data economy, which has its own dynamics. The commodity of this new economy is of a very special type: it does not deplete, produces more of it upon processing, is cheaper to store/transport, and at times it is difficult to ascertain its provenance – thus making the problem of personal information management a technological, behavioural, legal, economic, and regulatory challenge. Furthermore, with the advent of AI, one of the biggest concerns users have in current era is how their data is being analyzed and put to use. The limited explanation users get about what inferences and predictions are made about their online activities is making them anxious (Wachter and Mittelstadt, 2019). With the privacy settings they need to configure on each service they use, keeping track of their own data-trails is burdensome, time-consuming, complicated and sometimes impossible. Many a times users are unaware of where their data is being collected, stored, shared, and processed; thus a large amount of user data remaining ungoverned by the privacy settings they employ. Privacy should not be studied in the purview of the actual parties that are transacting with each other under an agreed upon policy and terms of data governance; but should be studied in a tripartite setup where the third party is an observer of the data and metadata of transactions between actual parties bound by a legal con-

[a] https://orcid.org/0000-0001-7714-2291

Patil, V. and Shyamasundar, R.
Is Privacy a Myth for Facebook Users?.
DOI: 10.5220/0008018805100516
In *Proceedings of the 16th International Joint Conference on e-Business and Telecommunications (ICETE 2019)*, pages 510-516
ISBN: 978-989-758-378-0

tract. This notion of an observer is especially relevant in perforated data platforms of OSNs like Facebook – an App, Friends, Advertisers are among the observers of user actions on and off the platform.

OSNs are the data platforms where users voluntarily share their information in exchange of a unique online (social) experience. All the user actions on the platform are recorded by the platform with user's consent and partially by the secondary observers (like ISPs, DNS, PKIs, trackers) to whom the user might not have given explicit consent. As the business model of Facebook revolves around monetizing user profiles through advertisements, it has built an extensive data collection apparatus called *social graph* through a symbiotic app ecosystem (authorized/consented co-observer) that partners with the underlying platform in collecting and contextually labeling user actions on or off the platform. For example, an app of type health will allow the platform to record user actions in the context of health, and similarly for other categories of apps a contextual labelling is done. Veracity of data is a critical condition in the success of AI/ML models used for monetizing of data and almost all data platforms are not only recording the transactional, observational data but also other dimensions of data in which the data is generated – contextual information. What data the platform collects and how it does that is communicated to the user before obtaining user's consent for data collection. But, how the data is labelled, interpreted, acted upon is not known to users and non-remediable, as of today. The data platforms use approximation algorithms to predict about users' future inclinations/actions/interests with impunity. The use of user data for that user's online experience is done for platform's financial gain and the profiling algorithms are always tuned for better matching rates of an advertising campaign rather than accuracy of the prediction made by the algorithm. A wrong guess by the algorithm used does not cost the platform much, but it may have serious ramifications for the user.

Foreseeing such scenarios and to assuage user apprehensions about privacy, laws like GDPR (European Union, 2018) were introduced to curb data collection/processing without user consent and helping EU users to re-mediate their privacy. However, the law covers personally identifiable information (PII) in its traditional sense and we argue in this paper that in presence of a powerful observer even non-PII information is potentially privacy revealing; we will use a term *infon* to address both of these types of data. In order to provide impetus to the tenets of the GDPR new tools/interfaces need to be devised that: i) help users (laymen) to understand collection, propagation, usage of their infons; ii) guide users to interfaces where they exercise their legal rights related to privacy and subsequent invocation at locations where user infons have propagated; iii) platforms should extend & open up their APIs to privacy regulators where they can verify the compliance of rights invoked by a user; iv) platforms should incorporate a by-design feature to retain user data for a period mandated by local jurisdiction for law enforcement purpose. We shall delve in these for an effective PII management.

2 BACKGROUND

The desire to predict future is an innate trait of humans. They do so to mitigate potential risks emanating from future events. Forecasting weather, traffic, crude/commodity prices, disease spread, agricultural output, et al. are some of the examples we routinely come across. Approaches to build these models may vary but their utility is unquestionably useful. Their accuracy over a period of time gives credence to their predictions and gets accepted as new knowledge to rely upon. Statistical, heuristic models developed for one domain find their use in other domains. During the WWI and WWII, models were devised to match job roles (e.g., who should be posted to back offices or to the trenches) for new recruits, where a questionnaire provided to the recruits at the time of enlisting helped the decision makers to profile them into distinct psychometric categories that are suitable for certain job profiles. Such psychometric models were later used in conducting surveys, poll outcomes, impact of advertisement campaigns et al.

What changed in the past two decades is that the cost to conduct surveys fell due to the Internet and their scope increased beyond from geographic constraints. Statistical/heuristic data models were there even before the Internet era but the ability to reach beyond geographic boundaries and cheaper processing started data-driven decision making. The insights on the users found its usage in matching those users (buyers) with products through online advertisement (Matz et al., 2017; Youyou et al., 2015). The advertisement model was so successful due to its return-on-investments, a race to reach more and more users began in order to profile them against products and services from prospective sellers – some of the Fortune top 10 companies have this method as their business model (Esteve, 2017). In the absence of legal regulations on data collection, exchange, processing, and usage, we have reached to a stage where the type of one's mobile and the area where she resides determines loan approval process. Though the data economy model has potential to serve all of its constituents fairly, the balance is tilted towards the platform owners in the absence of enforceable, verifiable regulation. Efforts need to be done at each constituent of the data economy: the technology platforms, policy makers, regulators, and the users. Understanding the data ecosystem is the first step.

3 NOTION OF AN OBSERVER

Every online service is composed of entities that play a role in delivery of that service. The entities vary from hardware devices like computers/mobiles from which end users access services, software components like browsers/apps doing data representation, ISPs providing connectivity, a DNS helping in end point discovery, a PKI authenticating end points, API services helping in payments, et al. The actual service provider with whom the end user makes a service agreement for data collection, storage, and its usage is called *primary observer*. All the other *interchangeable* entities that enable a service are called *secondary observers*, who may have separate service agreements with the end points. As a thought experiment, we recommend the reader to envision some of the online services they use and the terms of services they have entered into with their secondary observers. Accessing a service from three different locations like home, office, cafe will reveal those locations to primary and secondary observers, who in turn may monetize their respective observations (Chaabane et al., 2012). Assume the service agreement with primary observer does not cover location information. Assume the type of the bank the user uses is known for providing its services to customers with certain financial strata. Imagine the logs at the service provider end receiving user device information like iPhone X vs Android 4.1.

The current legal frameworks provide protection to users' PII collection, storage, sharing, and processing. Whereas, the large amount of infons that get generated and observed during the delivery of any online service are difficult to govern as they fall outside the ambit of PII's legal definition. Apart from the online services, even the offline activities of users (e.g., the purchases at retail shops) are recorded and traded at data exchanges legally – without the end user being a party to the trade. From the users' perspective, in order to start addressing the management of their infons it is imperative to classify the infons so their treatment becomes easier. In presence of observers, Figure 1 coarsely classifies (Gurevich et al., 2016) a user's infons into:

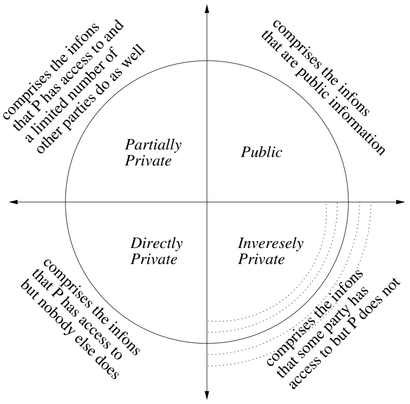
comprises the infons that P has access to and a limited number of other parties do as well
comprises the infons that are public information
Partially Private
Public
Directly Private
Inveresely Private
comprises the infons that P has access to but nobody else does
comprises the infons that some party has access to but P does not

Figure 1: Classification of PII (infon) of user P.

- *Public:* name, email, phone, vehicle number;
- *Partially Private:* salary, installed apps, call logs;
- *Directly Private:* passwords, sexual orientation;
- *Inversely Private:* CLSI logs, WiFi-beacon logs;
  - *independently inferable:* biometrics, DNA;
  - *context sensitive:* location, server logs;
  - *remediable:* habits, behavioral logs;

Depending on the position of the secondary observer in a service composition, user infons make their way into different classes. A primary observer may have access to all the classes except *Directly Private* infons. By definition, only the user should have control over the *Directly Private* infons. However, with the advent of AI/ML, it is possible to determine sexual orientation of a user with a very high probability– just from a photograph. With a consent from user for an improved service experience the primary observer starts making inferences (Kosinski et al., 2013; Kristensen et al., 2017) about user's future likes and dislikes using proprietary functions. The accuracy of predictions determines user engagement and advertisement revenue. In other words, the more the primary observer observes users, the better it is for user insights[1].

4 OSN: ACCESS CONTROL OF PII & ITS LIMITATIONS

OSNs like Facebook are at the forefront of user engagement through social interaction services like

[1] Google has been making its foray into several free services that keep its users as close to its platform as possible: Chrome browser, Android OS, 8.8.8.8 DNS, Google Trust Services, Public WiFi – thus reducing the exposure of its user infons to secondary observers and becoming an omniscient observer itself.

Messenger, WhatsApp, Instagram – all having their own terms of services with end users but funneling the user infons to the same data apparatus where reams of user profiles are perfected to provide a unique user experience. It has a symbiotic data ecosystem where users are provided with an online social experience through the core services offered by the platform along with its collaborators: apps, website buttons, event APIs. The platform acts as a primary observer for core services and the collaborators act as secondary observers when users are online. In the context of Facebook apps, as per GDPR, the apps are data controllers and the platform is the data processor – keeping the primary onus of violations with the controllers. However, through the analytic service that apps/websites use for audience measurement, a stream of user behavioral infons make their way to the platform, which the powerful observer (Facebook) links to the individuals. Facebook introduced the concept of Local_ID such its collaborators do not link their respective user actions.

The privacy settings of Facebook and those of its collaborators (apps) are disjoint and are set independently. Thus, a personal attribute that a user does not want to share with Facebook but shares with the app makes its way to Facebook's platform, which is not liable for its protection as a data processor. Another peculiar characteristic of social interactions is that the users share infons with other users, which can be observed by other users and apps (secondary observers) based on the access control specified on that interaction. The access control on Facebook's platform is specified using labels like Friends, Friend-of-Friends, etc., which may resolve to different set of users at different times – keeping no trail of who has accessed the information in the past. The users and their information is represented in terms of a social graph where connectivity between the nodes is the primary criteria for accessing the information and the secondary check is done by the platform against the access control policy specified at the node by its owner. By deleting a friend on the platform will not allow the deleted friend to visit a content protected by label Friends. Facebook also tracks its users off-the-platform through its Pixel trackers, which website owners and app developers integrate with their content for audience measurement (analytic) and advertisement composition. User infons via the analytical data makes their way to the platform, where it is acting as a processor but has equal observation capability as the primary observer (i.e., the website/app, the data controllers).

The inversely private infons from users' on-platform and off-platform interactions are used to generate actionable intent markers (Beaudin et al., 2019) so that advertisers can identify prospective buyer/consumers. Thus, the inversely private information about users is made accessible to advertisers upon payment and service terms that do not allow an advertiser to compose a *too narrow* campaign. But the advertisers can submit several well-crafted campaign request and later perform intersection on the audiences returned by those requests. Advertisers too are provided with analytical services which generate inversely private infons about the users who interacted with the advertisements. This convoluted flow of infons helps the platform in continuously improving its knowledge-base about users, apps, advertisers. Figure 2 depicts the hierarchy of access control to user infons: among users, between apps and their users – across the "User Space", and indirectly between the Advertisers and apps – across the "Paid Space". Facebook being the owner of the platform and in service terms with all the entities on the platform acts as a "Superuser" and can traverse across the social graph without any access restriction. In other words, though the platform protects user PII in letter but in spirit it makes the labelled infons of the users to advertisers for a fee. Figure 3 depicts the administrative scope of data governance as per GDPR, which deals only with the PII and the context (Barth et al., 2006) for which user agreed at the time of enrollment and excludes linked inferences and observed contexts that were used to reinforce user profiles.

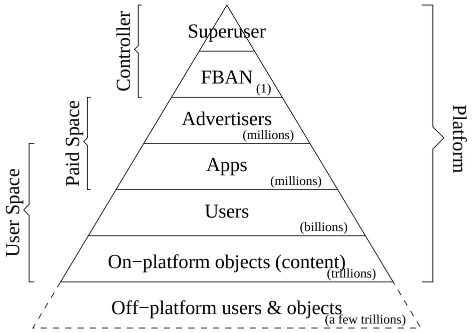

Figure 2: Access Hierarchy on Facebook platform.

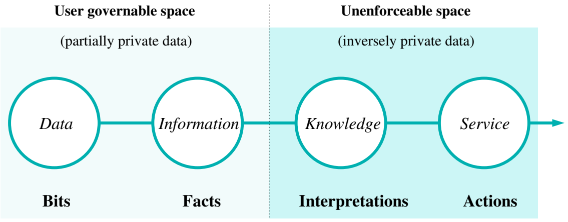

Figure 3: Scope of regulatory governance on infons.

5 TOWARDS AN EFFECTIVE PII MANAGEMENT ON OSNS

The right-to-be-forgotten tenet of the GDPR is the most challenging tenet in the current pretext of data economy where user infons are pervasively making their way beyond the ambit of stipulated governance contours as shown in Figure 3. The objective of this tenet is to revoke user's consent to collect, store, process, share her infons and the service provider should erase all PII about the user thus far collected. Ideally, the user should get an online experience as if the user had not provided her consent to the service provider. For example, if a user had ordered flowers for her marriage anniversary and provided spouse details to be printed on the card, the app and the platform should not target these two users as couple once the user has invoked her right-to-be-forgotten on the app that she had used to order the flowers. But, between the order and the invocation of her right-to-be-forgotten, for the sake of unique user experience the app and the platform start treating the two users as a couple and another unrelated apparel app that is looking for prospective customers will be presented with these two users as prospective customers. If one of the users interact with the advertisement from apparel app, that users future infon will be treated in the context of a married person. This trail of causal actions and inferences are difficult to undo in spirit through legal interpretations of GDPR. In order to enforce the right-to-be-forgotten tenet *in letter and spirit* we propose the following architecture that is practical and probably acceptable to all the entities that have a role in user's privacy preservation. We identify 4 such entities as important pillars of a vibrant digital economy:

1. users: act as a primary source of infons and may limit their participation in the digital economy in the face of continuing apprehension about their privacy and lack of remedial measures
2. data platforms: data aggregators, controllers, processors, and compete with each other for making better use of user infons at their disposal, including legal participation from data brokers
3. policy makers: elected representatives of the users who may face backlash in the absence of better regulations that contain and govern the platforms
4. regulators, law enforcement: regulators audit platforms against the regulations enacted by policy-makers. Law enforcement may require data-trails from platforms for national security or forensics.

These pillars need to work in tandem for effective PII management (FTC, 2012; McCallister et al., 2010). We make the following conjectures.

For users: A misnomer is created (Athey et al., 2017) by saying users are in full control of their own data, whereas inversely private probabilistic dossiers are created. Users simply do not have legible interfaces to express their privacy choices in lay terms.

- Users may voluntarily and accurately label themselves (Leon et al., 2013) into a limited categories the platform is interested in and user is agreeable to – a white-list. Thus, the user understands what type of personalizing improvements to expect.
- A *privacy-by-design* implementation of "do-not-track" should be standardised and made available to paid users who do not want to be tracked across apps, websites, and devices. The ad supported websites will present advertisements to such users based on the content of that website alone.
- An infon management interface to the user should provide *causal graphs* representing the transition of infons into different categories due to the online actions the user has taken from the last time the user visited the interface.
- The infon management interface should also provide a list of external collaborators to whom user dossiers are made available. This will be useful while invoking her right-to-be-forgotten from an app or the platform, informing user where else she needs to invoke her rights consequently.

For data platforms: there is a need to steer away from user data monetizing model to a transparent, equitable model where users can understand how their infons are reflecting back on them and could take remedial steps if something is undesirable.

- the platform (Facebook) should start treating the user dossiers as a mutually shared resource (Lessig, 1999) on which users can specify category labels like health, finance, sports. These categories can be access controlled using extensible labels similar to Friends or Friend-of-Friends – like ANA (Association of National Advertisers).
- the platform should introduce a special label "Frozen". The utility of such a label would be to address two contradictory requirements that arise when a user invokes her right-to-be-forgotten whereas the law enforcement mandates the platform to retain all records for forensics purpose. Thus, this label will be tagged to appropriate nodes of the user who is invoking her right-to-be-forgotten. The friends or public/apps connected to that user can no more query to "Frozen" nodes, whereas the law enforcement app can still traverse through such nodes for a stipulated time.

- the platform should mandate its apps to maintain a “Merkle root” of its users’ USERIDs who have invoked their right-to-be-forgotten on the app. A blockchain like chain of snapshots of Merkle roots for an app will keep a trail of actions the app has taken so far. This will help users to verify and track the status of their requests efficiently.

For policy makers: the policy makers will have to keep themselves abreast with the technological advancements and figure out ways to introduce laws that help govern the data economy while providing sufficient flexibility to allow innovation.

- persuade the data platforms to design verifiable mechanisms to help users manage their infons.
- persuade the platforms to provide tools/interfaces for regulators in order to help end users verify whether the app or platform has complied with the legal request made by its users.

For regulators: they have to ensure/audit compliance of laws in their jurisdiction. In this fast evolving ecosystem a feedback loop from regulators to policy makers is necessary for timely legislative evolution.

- regulators may build sandbox environments with different privacy policy configurations with dummy users and evaluate data controllers/processors claims about those dummy users’ remediation status with the platform.
- regulators may also design and get approved their apps on the data platform and ask privacy aware users to interact with these apps to analyse percolation of their infons in the data ecosystem.

In this paper, we have argued about the notion of infons, which is a super-set of PII data and associated contextual data that can be reflected on a user’s privacy by an observer. We have explored few scenarios to highlight how non-PII data (infons) from a user are potentially privacy undermining in view of two types of observers: primary (trusted) and secondary (plausibly invisible). Data platforms of OSNs like Facebook are the most complex and convoluted data collection, process, and usage systems – a mechanism that serves this type of architecture can be easily used elsewhere. Clearly, there is lot that is required to be done before we can start building effective tools and frameworks for privacy management; i) standards for data classification and labeling, and ii) quasi-uniform legal framework for infon treatment/management. In this section, though we have listed out plausible steps for effective management of user privacy, we would like to highlight a peculiar trait of social communications, that is, the *ownership and control of metadata* between independent entities. Consider an app like Truecaller (a reverse lookup provider for phone numbers) is installed by two of the Facebook friends. Truecaller is integrated with Facebook’s social graph to obtain various information about its user: name, photo, online/offline status, et al. Each incoming/outgoing call generates analytic traffic using Facebook app_event hooks. Each user’s call logs (infons) make their way to Facebook. Now, assume that one of the users invokes her right-to-be-forgotten on Truecaller; subsequent calls of that user to other Truecaller users will inadvertently get her infons (metadata and analytic) to Facebook. Who among the caller and the callee is the owner of that phone call log made after right-to-be-forgotten was invoked? Should one user’s restriction on such legally undefined data type be enforced on the counter-party, who is using Truecaller? If yes, it is too costly to enforce, if no, there is a leakage of infons of the caller.

This brings us to a very useful insight: effective privacy management may not be achieved by treating it as a data leakage/containment problem. ***We should start addressing privacy as a usage control problem instead of data leakage problem.*** Analogous to the approach taken in handling tax evasion cases, where a revenue officer seeks answers from a suspected tax evader about the sources of her income. Borrowing this analogy will be effective to force the data monetizing platforms to source infons from entities who have obtained informed consent from its users. The platform will have to prove to the user and regulator about the source of infons the platform has used to predict suitability of an advertisement to a user. In other words, when an advertisement/experience is presented to a user, it should carry a proof with it explaining the user how a decision is made to match a particular advertisement/experience to her. This will also nudge the data platforms to innovate in the direction of explainable AI instead of AI models that are developed to maximise their profits. With the current prediction models, it does not matter to the platform owners if the model is wrong in its predictions or approximations about a user, because it does not cost anything to the platform – but the user has no redressal mechanism at hand when the prediction/approximation goes wrong.

6 CONCLUSIONS

The notion of an observer is an important notion because for any given service the data and metadata about users is being observed not only by the entities that compose that particular service but also by the secondary observers about whom the user may not have complete knowledge. The user consents for its data to the primary observer and by adjusting the privacy settings derives a false sense of privacy control. Whereas, the secondary observers that participate in user transactions either by users' choices or by service providers' service composition choices are ignored for their potential to collect, infer, and monetize user data. Despite the users being provided with legal rights to protect themselves from online tracking, profiling by services, the users do not have a comprehensive view of their personal infons scattered across the data ecosystem and thus fail to exercise their rights. We discussed the works that need to be done by the four pillars of digital ecosystem in order to bring back the user trust in online services for a greater good of the ecosystem. We emphasize that the notion of an observer is a helpful notion for users, system designers in order to make informed decisions about privacy settings, online behaviour and privacy-preserving system designs. This will enable users to understand and remediate their perceived privacy violations by the environment in which they operate. Regulators may seek an explainable proof of platform's decision making in building an audience for targeted advertisement. This will inhibit data processors from relying on data sources to which the targeted user had not provided consent. The legal rights like right-to-be-forgotten or right-to-consent cannot be effectively exercised if the users do not have ability to identify and locate their inversely private infons.

ACKNOWLEDGEMENTS

The work was carried out as part of research at IS-RDC, supported by 15DEITY004, Ministry of Electronics and Information Technology, Govt of India.

REFERENCES

Athey, S., Catalini, C., and E. Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk. *SSRN Electronic Journal*.

Barth, A., Datta, A., Mitchell, J. C., and Nissenbaum, H. (2006). Privacy and contextual integrity: Framework and applications. In *IEEE S&P'06*, pages 184–198.

Beaudin, L., Downey, S., Hartsoe, A., Renaud, C., and Voorhees, J. (2019). Breaking the marketing mold with machine learning. In *MIT Tech Review Insights*.

Chaabane, A., Kaafar, M. A., and Boreli, R. (2012). Big friend is watching you: Analyzing online social networks tracking capabilities. In *Proc. of ACM Workshop on Online Social Networks*, pages 7–12. ACM.

Esteve, A. (2017). The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law*, 7(1):36–47.

European Union (2018). 2018 reform of EU data protection rules. `online`.

FTC (2012). Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers. `online`.

Gurevich, Y., Hudis, E., and Wing, J. M. (2016). Inverse Privacy. *Communications of ACM*, 59(7):38–42.

Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805.

Kristensen, J., lbrechtsen, T., Dahl-Nielsen, E., Jensen, M., Skovrind, M., and Bornakke, T. (2017). Parsimonious data: How a single facebook like predicts voting behavior in multiparty systems. *PLOS ONE*, 12(9):1–12.

Leon, P. G., Ur, B., Wang, Y., Sleeper, M., Balebako, R., Shay, R., Bauer, L., Christodorescu, M., and Cranor, L. F. (2013). What Matters to Users?: Factors That Affect Users' Willingness to Share Information with Online Advertisers. In *SOUPS*, pages 7:1–7:12. ACM.

Lessig, L. (1999). *Code and Other Laws of Cyberspace*. Basic Books, Inc., New York, NY, USA.

Matz, S. C., Kosinski, M., Nave, G., and Stillwell, D. J. (2017). Psychological targeting as an effective approach to digital mass persuasion. *Proc. of the National Academy of Sciences*, 114(48):12714–12719.

McCallister, E., Grance, T., and Scarfone, K. A. (2010). SP 800-122. Guide to Protecting the Confidentiality of Personally Identifiable Information (PII). Technical report, National Institute of Standards & Technology.

Nilizadeh, S., Kapadia, A., and Ahn, Y.-Y. (2014). Community-enhanced de-anonymization of online social networks. In *Proceedings of the 2014 ACM CCS*, pages 537–548.

Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review, Vol. 57, p. 1701, 2010.*

Patil, V. T. and Shyamasundar, R. K. (2017). Privacy as a Currency: Un-regulated? In *Proceedings of the 14th SECRYPT*, pages 586–595. SciTePress.

Patil, V. T. and Shyamasundar, R. K. (2018). Efficacy of GDPR's Right-to-be-Forgotten on Facebook. In *Information Systems Security*, volume 11281, pages 364–385. LNCS, Springer International Publishing.

Wachter, S. and Mittelstadt, B. (2019). A Right to Reasonable Inferences: Re-Thinking Data Protection Law in the Age of Big Data and AI. In *Columbia Business Law Review, 2019(1)*. SSRN.

Xu, Y., Frahm, J.-M., and Monrose, F. (2014). Watching the Watchers: Automatically Inferring TV Content From Outdoor Light Effusions. In *Proceedings of the 2014 ACM CCS*, pages 418–428. ACM.

Youyou, W., Kosinski, M., and Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. *Proceedings of the National Academy of Sciences*, 112(4):1036–1040.