Privacy for All: Ensuring Fair and Equitable Privacy Protections

Michael D. Ekstrand, Rezvan Joshaghani , Hoda Mehrpouyan

FAT* 2018

Introduction

- recent research on ensuring sociotechnical systems are fair and nondiscriminatory to the privacy protections those systems may provide
- privacy scheme protects all persons uniformly, irrespective of membership in protected classes or particular risk in the face of privacy failure
  - so also privacy regimes may disproportionately fail to protect vulnerable members of their target population
- We propose a research agenda that will illuminate this issue, along with related issues in the intersection of fairness and privacy, and present case studies that show how the outcomes of this research may change existing thinking and research on privacy and fairness.

Motivation

- ethical, legal, and social effects of technology do not exist in isolation, but often interact in complex ways
  - We argue here for expanding the lens to include both concepts (fairness and privacy) together.
  - We seek to understand how fairness and privacy interact and complement or compete with each other
- Contemporary analyses of fairness do not have as long a history, though they are grounded in more than fifty years of legal work on fairness and nondiscrimination, with precursors reaching further back in scholarly discourse

Questions

1. Are technical or non-technical privacy protection schemes fair, under contemporary definitions of fairness?

2. When and how do privacy protection technologies or policies improve or impede the fairness of the systems they affect?

3. When and how do technologies or policies aimed at improving fairness enhance or reduce the privacy protections of the people involved?

- We expect the answers to these questions to vary based on domain, technology, and the specific definitions of privacy and fairness under consideration

Privacy-fairness Intersection

- This fairness different from algorithmic fairness in information systems
  - Privacy is often linked to the ethical language of fairness, particularly in the U.S. regulatory context; since 1973, fair information practices have been the guiding paradigm for managing privacy and considering regulations around data protection
- Extend privacy so far as we can guarantee it to all subjects of an information system.
  - We expand this issue into a broad agenda at the intersection of privacy and fairness that considers the entire sociotechnical system in which a technical, legal, or social privacy mechanism is deployed and situating it in the current language of algorithmic fairness.

Privacy Definitions

- privacy is a right and define it as ”the right to be let alone.”, This stance identifies privacy with seclusion. Under the seclusion definition, perfect privacy is achieved through complete solitude, e.g. living alone on a deserted island (at least prior to the invention of spy satellites) ☺
- Another definition of privacy regards it as being free from intrusion or interference; we can call this the non-intrusion view of privacy
  - ”right of the individual . . . to be free from unwarranted government intrusion”
- The limitation theory of privacy defines privacy as individuals keeping information to themselves. In this theory privacy is defined as limited and contextually bounded information access.
- The control theory promotes privacy by enabling users to exert control over private information. In control theory, a person has privacy if they have control over their information
  - ”the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others”.
- Contextual privacy: in each information transition context there are different variables that define privacy and the norms in that specific context determine the privacy violation.

Algorithmic Fairness

- Algorithmic fairness
  - as embodied in the literature surrounding Fairness, Accountability, and Transparency, discrimination-aware data mining and related research threads
  - on fairness as nondiscrimination: a person's experience with an information system should not irrelevantly depend on their personal characteristics, especially their membership in groups subject to historical discrimination.
- There are at least two meaningful dimensions on which we can organize approaches to ensuring and evaluating fairness:
  - for whom fairness being considered
  - how fairness is operationalized so it can be measured or assured.
  - Easy for credit applications and hard for recommendation systems

Types of Fairness/Discrimination

- Individual fairness says that similar individuals should receive similar treatment.
- Group fairness says that different groups of users should receive similar statistical treatment, by experiencing similar classification accuracy or error.
  - group and individual fairness often cannot be achieved simultaneously and that different group fairness measures are not simultaneously achievable in many realistic settings
  - Achieved in pre-processing, learning, after learning, etc.
- **Disparate <u>impact</u>** and **<u>mistreatment</u>**
  - **DI:** This doctrine says that a practice is discriminatory if it has disproportionate adverse effects on protected groups without a compelling business need.
  - **DM:** this model examines decision errors, particularly errors that harm the protected class.
- Disparate impact and disparate mistreatment used for most of the specific concerns regarding the fairness of privacy systems.

Differential Privacy

- Differential privacy provides a strong guarantee of privacy by incorporating random noise calibrated to nullify the impact of the presence or absence of any one on the final result.
  - If a data access mechanism is **(Ɛ,δ)**-differentially private, then it generally bounds the distinguishability of two databases, one containing an individual's record and the other not, by Ɛ, with a δ (usually cryptographically small) probability of total privacy failure.
  - Dwork et al. (2014) refer to Ɛ as the "knowledge gain ratio from one dataset over the other." Hence, the higher the value of Ɛ, the weaker the privacy guarantee
  - Usually **0.01 or 0.1**, Apple **1 or 2**
  - Dwork et al. (2012) later adapted the mathematical machinery of differential privacy to provide certain fairness properties, considering (individual) fairness to be a generalization of differential privacy.

Privacy vs Fairness Tradeoffs

- Consumers often need to trade information about themselves for goods and services.
  - For example, online personalization can improve the relevance of product recommendations and advertising, thereby reducing the number of irrelevant ads a customer sees, but requires data on customer behavior and preferences. Attackers can exploit that.
- Limitations of theory of privacy, these users are trading privacy for personalization
  - **Under a control theory**, they may be exercising their right to privacy by choosing to participate in the exchange, but only to the extent that they have sufficient notice and knowledge to make an informed decision.
  - **Under contextual integrity**, the user may consent to the advertiser having their information but not to its use in targeting advertising.
  - There remains much work to be done in characterizing under what circumstances and definitions privacy and fairness are simultaneously achievable, and when they compete such that a joint approach must solve a multicriteria optimization problem and trade off privacy or fairness for the other.

Fair Privacy

- Q1: Does the system provide comparable privacy protections to different groups of subjects?
- Q2: Are privacy attacks more effective against members of protected classes?
- Q3: Does the system require disparate effort from its subjects in order to enjoy privacy protection?
- Q4: Is the fairness of privacy guarantees robust to shifts in threat model
- Q5: What properties of a problem setting or privacy mechanism make fair privacy easier or harder to achieve?
- Q6: Are there identifiable properties of a privacy mechanism and problem setting that form necessary or sufficient conditions for fair privacy?
- Q7: What other properties may need to be sacrificed to achieve fair privacy?

Fair Privacy (Cont.)

- Q8: Does a privacy-protection scheme impede the ability to ensure or audit the fairness of decision-making processes or information systems?
- Q9: Can privacy protection technologies or policies be used or adapted to enhance the fairness of a system?
- Q10: Does a fairness auditing or enhancement scheme diminish the privacy of its subjects?
- Can fairness-enhancing technologies be used to provide privacy guarantees?
  - it is crucial to carefully define the kinds of privacy and fairness under consideration.
  - We need to map out what kinds of fairness and privacy may intrinsically support each other or contradict each other, much like the existing impossibility results for fairness definitions.
  - **Definition:** Fair Privacy Protection. Proposed that a privacy system be deemed to provide fair protection if the probability of failure and expected risk are statistically independent of the subject's membership in a protected class.

Case Studies

- Differential privacy
  - Differential privacy, on its face, provides fair privacy, as all users' privacy loss is bounded by Ɛ and δ.
  - However, there may remain subtle ways in which differential privacy may fail to provide fair privacy. If omitting a protected class of users from the database admits a lower bound on the privacy loss of a differentially private mechanism with equivalent accuracy, then the system may be unfair (Q1).
- Deanonymizing risks
  - auxiliary data that relates to identifying attributes of the database. This extra data can be from another database, a certain behavioral pattern of people, or some background information on the target person. Usually, a combination of these methods is used to identify a person in the dataset.
  - Netflix vs IMDB data
  - Worryingly, in some cases, deanonymization attacks are easier to carry out against members of a particular group
  - NYC taxi drivers
  - Under Q2, it appears that deanonymization attacks may be disparately successful against minority groups in some cases, and when and how this occurs should be studied carefully.

Case Studies (Cont.)

- In some fields such as precision medicine, too much anonymization can compromise data quality, and the anonymized dataset is useless for the study
  - it is also likely difficult to auditing fairness in such a setting (Q8).
  - if the anonymization particularly decreases the effectiveness of treatment for minority groups, it could cause a Q3 fairness problem by imposing higher costs (reduced medical effectiveness) for insisting on privacy in the data set
- Recommendation Systems
  - privacy attacks that allow an attacker with some information about a person's transaction history to observe the public outputs of a live recommender system and infer other transactions made by the target individual.
  - This is similar in spirit to a deanonymization attack against a recommender system data set but is feasible on a live system only observing the recommender's output.
  - Recommender systems can also give away users' identity information.
  - we can ask whether members of protected classes are at greater risk of a recommender system disclosing their identity or other information than less vulnerable users (Q1)

Case Studies (Cont.)

- Genetic data
  - The interaction of genetic privacy, or the privacy of other health data, with fairness is subtle. Members of historically-vulnerable groups may be at greater risk of genetic privacy breach (Q1)
  - prevent some forms of discrimination in health care, employment, and other domains (Q9)

Conclusion

- For privacy protection mechanisms to advance a just and equitable society, it is necessary that they

  (1) provide their protections equitably to all their subjects and

  (2) that they integrate positively with other important concerns such as fairness and non-discrimination in the information systems deployed in their sociotechnical setting.

- Math definiens for differential privacy are good ☺
  - But there remains much to be done, particularly in understanding the implications of privacy and fairness on each other in practical settings
  - They require meaningful, user-interpretable privacy and fairness properties, and on understanding the ways in which implementation details, human factors, and legal concerns may hinder one or the other of privacy and fairness when they are both pursued.

References

- Ekstrand, M.D., Joshaghani, R. and Mehrpouyan, H., 2018, January. Privacy for all: Ensuring fair and equitable privacy protections. In *Conference on Fairness, Accountability and Transparency* (pp. 35-47).

Thank you ☺