ALEKSANDRA KOROLOVA RESEARCH STATEMENT

Email: korolova@princeton.edu, Web: www.korolova.com

Privacy, algorithmic fairness, accountability and transparency are currently at the center of key debates across academia, industry and policy. My research sits at the intersection of these topics and aims to leverage algorithmic thinking in order to provide new solution spaces that allow for a better balance between individual interests, societal goals, and technical innovation.

Novel and daring uses of individual personal data by companies and governments, combined with the overall failure of many of these entities to protect this data, have positioned tech innovation and individual privacy as contradictory goals. **The first goal of my research is to develop algorithmic and systems advances that can enable data-driven innovations while preserving individual privacy**, defined in the paradigm of differential privacy (DP). Towards this goal, my research has significantly advanced the feasibility of adoption of differential privacy-preserving technologies. I have developed DP algorithms for key Internet data-mining applications that work within existing system and engineering constraints [1, 2, 4, 16], including the first algorithm for sharing a subset of user search data that is now core to Google's open source library. My algorithmic and modeling work on RAPPOR [5] has pioneered the use of DP in industry, and brought attention to the local model of DP (in which a user's data undergoes a privacy-preserving transformation before it leaves their device) as a desirable and feasible goal. It has paved the way for other industry DP deployments, and motivated significant interest in developing algorithms for this model of privacy in the academic literature. My binary code analysis of Apple's deployment of DP [6] has led to their revelation and public commitment to improved privacy loss values, and set industry precedent for making such choices public. My collaboration with Mozilla introduced the hybrid model of DP [7], in which the data of individuals contributing with local DP is augmented with a small number of contributions via a trusted curator. This model is aligned with current industry practices and user expectations; and we demonstrate that it can make DP analyses feasible for entities whose user base is not in the billions [10, 11]. Most recently, I am exploring the power of learning-augmented algorithms for improving privacy-utility trade-offs [17]. In parallel, my research has identified novel privacy vulnerabilities in targeted advertising systems [3, 8], motivating the need for deploying DP in practice and advancing the privacy protections deployed by Facebook and Google.

The private data of individuals collected by companies may negatively impact individual lives and societies not only through the data's explicit revelation to others but also by its implicit use. Specifically, individual's data, often taken without direct permission, is used towards decisions with significant implications while the algorithms behind these decisions are not known. Moreover, these algorithms are often developed to optimize the companies' interests which may be at odds with individual and societal interests and established law. **The second goal of my research is to understand how opaque algorithmic decision-making systems may be affecting individuals and society, and to develop techniques for mitigating their negative consequences.** My research developed new black-box audit methodologies for isolating the role of ad delivery algorithms from other confounding factors such as economic forces and differences in behavior across demographic groups, addressing a question open since 2013. Our application of these methodologies demonstrated that Facebook's ad delivery algorithms lead to discriminatory outcomes in housing and employment advertising even when advertisers are targeting inclusively [9, 14] and to filter bubbles in political ad delivery [13]. Our findings led to a settlement between the Department of Justice and Meta (formerly Facebook) in 2022, requiring Meta to modify its ad delivery system to address algorithmic discrimination and to the inclusion of ad delivery algorithms in proposed regulation of political advertising by the European Commission in 2023. Most recently, I have proposed a new notion of fairness for algorithmic decision-making systems in contexts where individuals have diverse preferences over outcomes [12]. The notion non-trivially relaxes classic notions of individual fairness and envy-freeness and meaningfully expands the space of solutions for multi-sided optimization settings such as ad delivery. I have also quantified and proposed algorithms for achieving better trade-offs between diversity, fairness and utility in such settings [15, 18].

My immediate research plan lies at the intersection of the above two goals. The desire to protect privacy is often used as a reason for limited transparency and accountability about the data and models powering today's machine learning systems. I plan to develop new paradigms for achieving meaningful auditability while protecting individuals' privacy and companies' interests. I have initiated a new approach for studying social media algorithms that gives auditors query-level privileged DP access to the relevance scores computed by the algorithms [19]. In addition to proving that privacy imposes only a small cost on auditability in this paradigm, the work provides a feasible technical

implementation for the dual privacy and transparency goals of the European Union's Digital Services Act and proposed U.S. Platform Accountability and Transparency Act.

REFERENCES

[19] **Having your Privacy Cake and Eating it Too: Platform-supported Auditing of Social Media Algorithms for Public Interest**, B. Imana, A. Korolova, J. Heidemann. 26th ACM Conference On Computer-Supported Cooperative Work And Social Computing (CSCW), 2023.

[18] **Fairness in Matching Under Uncertainty**. S. Devic, D. Kempe, V. Sharan, A. Korolova. 40th International Conference on Machine Learning (ICML), 2023.

[17] **Pushing the Boundaries of Private, Large-Scale Query Answering**, B. Avent, A. Korolova. In The Fourth AAAI Workshop on Privacy-Preserving Artificial Intelligence (PPAI), 2023.

[16] **Differentially-Private "Draw and Discard" Machine Learning**, V. Pihur, A. Korolova, F. Liu, S. Sankuratripati, M. Yung, D. Huang, R. Zeng. In 6th International Symposium on Cyber Security, Cryptology and Machine Learning, 2022.

[15] **Robust Allocations with Diversity Constraints**, Z. Shen, L. Gelauff, A. Goel, A. Korolova, K. Munagala, Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS), 2021.

[14] **Auditing for Discrimination in Algorithms Delivering Job Ads**, B. Imana, A. Korolova, J. Heidemann. Proceedings of the 30th Web Conference (WWW), 2021.

[13] **Ad Delivery Algorithms: The Hidden Arbiters of Political Messaging**, M. Ali, P. Sapiezynski, A. Korolova, A. Mislove, A. Rieke. In 14th ACM International Conference on Web Search and Data Mining (WSDM), 2021.

[12] **Preference-Informed Fairness**, M. P. Kim, A. Korolova, G. Rothblum, G. Yona. 11th Conference on Innovations in Theoretical Computer Science (ITCS), 2020. 3rd ACM Conference on Fairness, Accountability, and Transparency (FAT*), 2020.

[11] **The power of synergy in differential privacy: Combining a small curator with local randomizers**, A. Beimel, A. Korolova, K. Nissim, O. Sheffet, U. Stemmer. 1st Conference on Information-Theoretic Cryptography (ITC), 2020.

[10] **The Power of The Hybrid Model for Mean Estimation**, B. Avent, Y. Dubey, A. Korolova. Proceedings of the 20th Privacy Enhancing Technologies Symposium (PETS), 2020.

[9] **Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes**, M. Ali, P. Sapiezynski, M. Bogen, A. Korolova, A. Mislove, A. Rieke. Proceedings of the 22nd ACM Conference on Computer Supported Cooperative Work (CSCW), 2019.

[8] **Facebook's Advertising Platform: New Attack Vectors and the Need for Interventions**, I. Faizullabhoy, A. Korolova. Workshop on Technology and Consumer Protection (ConPro) @S&P 2018.

[7] **BLENDER: Enabling Local Search with a Hybrid Differential Privacy Model**, B. Avent, A. Korolova, T. Hovden, D. Zeber, B. Livshits. Journal of Privacy and Confidentiality, Vol. 9 (2), 2019. USENIX Security, 2017.

[6] **Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12**, J. Tang, A. Korolova, X. Bai, X. Wang, X. Wang. Poster at 3rd Workshop on the Theory and Practice of Differential Privacy @ CCS, 2017.

[5] **RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response**. Ú. Erlingsson, V. Pihur, A. Korolova. In proc. of the 21st ACM Conference on Computer and Communications Security (CCS), 2014.

[4] **Privacy via the Johnson-Lindenstrauss Transform**. K. Kenthapadi, A. Korolova, I. Mironov, N. Mishra. Journal of Privacy and Confidentiality (JPC), Vol. 5: Iss. 1, 2013.

[3] **Privacy Violations Using Microtargeted Ads: A Case Study**. A. Korolova. Journal of Privacy and Confidentiality (JPC), Vol. 3, Iss. 1, 2011.

[2] **Personalized Social Recommendations - Accurate or Private?** A. Machanavajjhala, A. Korolova, A. Das Sarma. In proc. of the 37th International Conference on Very Large Databases (VLDB), 2011.

[1] **Releasing Search Queries and Clicks Privately.** A. Korolova, K. Kenthapadi, N. Mishra, and A. Ntoulas. 18th International World Wide Web Conference (WWW), 2009.