Internship proposition: Build a platform for increasing the transparency of social media advertising.

Keywords: audit, privacy, online targeted advertising, Twitter, Facebook, browser extension, data collection and analysis, machine learning, inference, statistics
Lab: Laboratoire d'Informatique de Grenoble (LIG), Grenoble, France (head: Eric Gaussier)
Team in the lab: SLIDE (head: Sihem Amer-Yahia)
Advisor: Oana Goga (CNRS & Univ. Grenoble Alpes, LIG) oana.goga@mpi-sws.org https://people.mpi-sws.org/~ogoga/

Project Description:
In recent years, targeted advertising has been source of a growing number of privacy complaints from Internet users [1]. At the heart of the problem lies the opacity of the targeted advertising mechanisms: users do not understand *what data advertisers have about them* and *how this data is being used* to select the ads they are being shown. This lack of transparency has begun to catch the attention of policy makers and government regulators, which are increasingly introducing laws requiring transparency [2].

The project consists in building a tool that provides *explanations* for why a user has been targeted with a particular ad that does not need the collaboration of the advertising platform. The key idea is to crowdsource the transparency task to users. The project consists in building a collaborative tool where users donate in a privacy-preserving manner data about the ads they receive. The platform needs to aggregate data from multiple users (using machine learning) to infer (statistically) why a user received a particular ad. Intuitively, the tool will group together all users that received the same ad, and look at the most common demographics and interests of users in the group. The key challenge is to identify the limits of what we can statistically infer from such a platform.

Throughout the project the student will be able to familiarize himself with the online targeted advertising ecosystems, learn to collect data from online services and apply machine learning and statistics concepts on real-world data.

Requirements:
Strong coding skills and strong background in statistics and data mining**.**

References:
[1] J. R. Mayer and J. C. Mitchell. Third-party web tracking: Policy and technology. In *IEEE S&P*, 2012.
[2] The "loi pour une Republique Numerique" adopted in France in October 2016 and the European "General Data Protection Regulation adopted" in April 2016.
[3] J. M. Carrascosa, J. Mikians, R. Cuevas, V. Erramilli, and N. Laoutaris. I always fell like somebody's watching me. measuring online behavioral advertising. In *CoNEXT'15*.
[4] A. Datta, M. C. Tschantz, and A. Datta. Automated experiments on ad privacy settings: A tale of opacity, choice, and discrimination. In *PETS'15*.
[5] M. Lécuyer, G. Ducoffe, F. Lan, A. Papancea, T. Petsios, R. Spahn, A. Chaintreau, and R. Geambasu. Xray: Enhancing the web's transparency with differential correlation. In *USENIX Security'14*.
[6] M. Lecuyer, R. Spahn, Y. Spiliopolous, A. Chaintreau, R. Geambasu, and D. Hsu. Sunlight: Fine-grained targeting detection at scale with statistical confidence. In *CCS'15*.
[7] B. Liu, A. Sheth, U. Weinsberg, J. Chandrashekar, and R. Govindan. Adreveal: Improving transparency into online targeted advertising. In *HotNets'13*.
[8] J. Parra-Arnau, J. P. Achara, and C. Castelluccia. Myadchoices: Bringing transparency and control to online advertising, 2016. arXiv:1602.02046.
[9] E. Zheleva and L. Getoor. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In *WWW '09*.