Research Statement

Ron Hirschprung Ph.D.

Department of Software and Information System Engineering

Faculty of Engineering Sciences

Ben-Gurion University

hrony@post.bgu.ac.il

**1. Synopsis**

The digital era introduces significant privacy issues (risks and fairness) which are mainly an outcome of the machines computational power. The user is unable to manage the privacy by himself. My research interests are at the adoption of machines by implementing AI-based algorithms to mitigate and to control those issues. I plan to establish a research group that will develop methodologies and technologies to carry out: a) some transformations on published datasets in order to minimize privacy risk by increasing anonymization while maximizing the relevancy of the dataset to its designated purpose; b) development of a proxy server that isolate data analyzer the data itself; c) providing an automated mechanism to tune the trade-off between utility and privacy cost (after it has been optimized by the transformations), so that a user can still elicit his preferences to the technologically complicated environment. These advanced methodologies should address both the requirements of legislators / regulators and the demand for ‘trust’ which is a viability feature for many IT systems.

**2. The Trade-Off between Utility and Privacy-Loss in the Digital** **Era**

An inherent trade-off between the utility that is provided by IS and the cost of privacy is a growing problem and in the current digital era empowered to a level that may even threatened the process of further adoption of those systems. The increase of data collection means, the feasibility (mainly economically) of mass storage, the availability of computational power (e.g. discovering some “hidden” facts about an individual by implementing machine learning) – yield a growing awareness and thus concerns regarding privacy issues. The phenomena can be demonstrated in a variety of domains. One of them may be A*utomated Decision Making*, by implementing AI-based algorithms. In this case, an agent that acts on behalf of the user, or as a service of another entity requires a significant amount of information about the user in order to carry out its tasks. The agent’s function yields decisions from a given information, however, when many decisions are given the function might be reversed yielding the source information. Another important application domain is *Medical Informatics*  where currently a clinical data of relatively large groups of patients is analyzed and can yield significant findings. Also key attributes are omitted from the dataset (anonymization), the individual can sometimes be identified by the quasi identifiers or even by the sensitive data. Legislators are aware of the privacy violation risk, which perceived as human rights, and make research difficult. In this case the trade-off is stretched radically since on one side of the scale lies a lifesaving factor and on the other side a disclosure of maybe the most sensitive information that an individual hold.

A common approach to handle this type of trade-offs, is to provide the user with a mechanism that enable its regulating, e.g. by configuring the system. This task however, as simply as it may seem to be, holds some difficulties which actually prevent its implementation. First, the complexity of those sophisticated systems, not to mention its indirect privacy violation effect is beyond the literacy of the layman user. Second, users incline to cognitive laziness thus avoiding such tasks. And finally, user behavior is characteristic by risk aversion and not by maximizing expectancy. The direct outcome of those insights is that **a human cannot handle privacy issues that are introduced by machines**, and machines must be harnessed to successfully carry out this task. In My researches I’m intend to develop algorithms based mainly on mathematical models that will provide a solution to this trade-off in two layers: a) by *Mitigating the trade-off*. b) by *Controlling the trade-off*. The above mentioned methodologies are applicable at PbD (Privacy by Design) approach, as required by regulators (e.g. GDPR).

**3. Mitigating the Privacy Loss**

A published dataset, even when anonymized can still be a source for privacy disclosure. However, given the (legitimate) purpose of the dataset publication, it has shown that by applying some **transformations to the data, e.g. data perturbation, the purpose can still be served but privacy risks are reduced**. My research mission in this section is to find and define those transformations as AI machine learning agents that can be applied on-line (as required in contemporary information systems). I distinguish two levels of protections: a) against inference attacks that relies on aggregated data published from the dataset; b) against inference attack that relies also on auxiliary information that cannot be controlled by the defender. The transformation can be implemented in a proxy server for example, and an authorized administrator should have the ability to set boundaries to the privacy disclosure risks while under those constrains the published dataset is optimized to provide maximal purpose achieving efficiency.

In an advanced phase of this research, we seek for a novel methodology to apply Machin Learning on Hidden Data (I called it for now: ML-HD). The concept is to create a development methodology of researches based only on conveying logic to a proxy without access to the raw data. This methodology which belongs to the circle of PETs solutions (Privacy-Enhancing Technologies) have great advantage over existing methods such as Obfuscation. This way, machines can for example provide a personal data-mining to an individual without a significant risk of data disclosure.

These models can empirically be tested by sampling real data and applying both to the original dataset and the sanitized dataset: a) inference attacks to measure the amount of privacy disclosure reduction; and b) processing the data for providing the purpose to measure the amount of efficiency loss.

**4. Controlling Privacy**

This layer of privacy protection should be implemented as a second phase after the trade-off was optimized, and Its purpose is to enable the user to tune the trade-off according to personal preferences. In My previous research, I developed an algorithm to reduce the configuration space (that may control this trade-off) and thus providing the user with a more efficient choice architecture to elicit preferences. The algorithm was empirically tested on Facebook real data (n= 266 users ; 21,950 posts), proofed to provide a significantly better choice architecture than current Facebook’s defaults. By adopting a different approach, we developed a methodology to quantify the value of privacy in terms of intrinsic valued (e.g. Dollars). By doing so, it is possible to accommodate also average utilities and social fairness in the objective function.

The ability to quantify privacy loss open opportunities for a wide range of implementations that can automatically configure digital systems on behalf of a user. I am interested in developing methodologies to establish Intelligent-Agents (IA) that will carry out this mission. The IA should have the ability to respond to the dynamic changes both of the environment and of the user’s preferences. Intuitively It seems that the IA design is domain oriented, However I seek to generalize the problem as a step towards creating a universal data disclosure tuning IA.