**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 the result of machines’ computational power, and users are no longer able to manage their online privacy effectively. My research interests focus on using machines to implement AI-based algorithms to mitigate and control these challenges. I plan to establish a research group that will develop methodologies and technologies to: a) carry out some transformations on published datasets in order to minimize privacy risks by increasing anonymization, while maximizing the relevancy of the dataset to its designated purpose; b) develop a proxy server that isolates data and enables analysis of the data itself; c) provide an automated mechanism to balance the trade-offs between utility and privacy cost (after it has been optimized by the transformations), so that a user can still elicit his or her preferences to the technologically complex 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 provided by Information Systems (IS) and the cost of privacy is a growing problem, and in the current digital era may even threaten the process of further adoption of those systems. The increase of data collection technologies, the feasibility (mainly economic) of mass storage, and the availability of computational power (e.g. discovering “hidden” facts about an individual by implementing machine learning) yield a growing awareness, and also raise concerns regarding privacy issues. These phenomena can be demonstrated in a variety of domains. One of them may be *Automated Decision Making*, by implementing AI-based algorithms. In this case, an agent acting 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 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 current clinical data of relatively large groups of patients is analyzed and can yield significant findings. Even though key attributes are omitted from the dataset (anonymization), the individual can sometimes still be identified by quasi-identifiers or by other sensitive data. Legislators are aware of the risks associated with violation of privacy, regarded as a human right, and this constraint presents challenges for research. In this case the trade-off is extended radically since on one side of the scale lies a lifesaving factor and on the other side a disclosure of what may be the most sensitive information an individual possesses.**

**A common approach to handling such trade-offs is to provide the user with a mechanism that enables its regulation, e.g. by configuring the system. However, this task, as simple as it may seem, entails some difficulties which actually prevent its implementation. First, the complexity of these sophisticated systems, in addition to the issue of indirect privacy violation, is beyond the literacy of the typical lay user. Second, users may be prone to cognitive laziness, thus avoiding such tasks. Finally, user behavior is characterized by risk aversion and not by maximizing expectancy. The direct outcome of these insights is that a human being cannot handle privacy issues that are introduced by machines, and machines must be harnessed to successfully carry out this task. In my research, I intend to develop algorithms based mainly on mathematical models that will provide a solution to these trade-offs at two levels: a) by *mitigating the trade-off*; and b) by *controlling the trade-off*. The above-mentioned methodologies employ the PbD (Privacy by Design) approach, as required by regulators (e.g. GDPR).**

**3. Mitigating Privacy Loss**

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

**At an advanced phase of this research project, we intend to develop a** **novel methodology for applying Machine Learning on Hidden Data (ML-HD). The concept is to create a research development methodology based only on conveying logic to a proxy without access to the raw data. This methodology, which belongs to the category of PETs (Privacy-Enhancing Technologies), has great advantages over existing methods such as Obfuscation. This way, machines can, for example, provide personal data-mining to an individual without significant risk of data disclosure.**

**These models can be tested empirically by sampling real data and applying: a) inference attacks to measure the degree of privacy disclosure reduction; and b) processing the data for providing the purpose, in order to measure the amount of efficiency loss.**

**4. Controlling Privacy**

**This layer of privacy protection should be implemented as a second phase once the trade-off is optimized. 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 (which can control this trade-off) and thus provide the user with a more efficient architecture to elicit preferences. The algorithm was empirically tested on Facebook real data (n= 266 users; 21,950 posts), and provided a significantly better 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 value (e.g. Dollars). By doing so, it is also possible to accommodate average utilities and social fairness in the objective function.**

**The ability to quantify privacy loss provides 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.**