PERSONAL INFORMATION FACTOR: A tool to tell *if* and *how* to share your datasets

CYBER SECURITY COOPERATIVE RESEARCH CENTRE

PROJECT SNAPSHOT

The Personal Information Factor (PIF) Tool measures the risk associated with releasing a dataset, proposing recommendations for sharing data that pivot around privacy considerations. When risks are high, the AI-enabled tool analyses attack vectors and automatically transforms the data, making it suitable for publication. The tool is geared for sharing data across sectors like healthcare, transport, financial services and smart cities, at scale.

WHO'S INVOLVED

This project is a collaboration between the Cyber Security CRC, CSIRO's Data61, the Australian Computer Society (ACS) and the NSW and WA Governments.

WHAT'S THE ISSUE?

Data sharing offers huge potential for innovation in service delivery and economic efficiency, but personal privacy concerns remain a hindrance to effective data sharing. Simple data de-identification has proven ineffective in protecting data privacy, as shown by high profile data re-identification attacks. To increase the confidence of data custodians when sharing data, users should be aware of the re-identification risks and more effective ways to mitigate these.

WHAT WERE THE OBJECTIVES?

Analyse privacy risks involved with sharing data;

Suggest recommendations for protecting sensitive data;

Raising awareness of the risks associated with data release to allow for better informed decision making

TECHNOLOGY DESCRIPTION

PIF uses Information Theory to compute privacy risk in a dataset. The tool suggests the associated risks and proposes recommendations for sharing data, for example suppression of certain attributes. The analysis results are also displayed as visuals, which makes interpretation easier. Based on the associated risks, the tool uses a provable privacy technique to perturb data. Unlike traditional tools which choose design parameters in an ad hoc fashion, the AI-based framework considers various attack vectors, the risk appetite of the user, and required level of accuracy, to select the design parameters.

APPLICATION

Healthcare: PIF has already been used by NSW Government to publish Covid data.

Transport: PIF can access the risk associated with publishing Opal Card data, hence protecting the privacy of individuals.

Domestic violence: NSW Government is using the PIF tool to share domestic violence data.

Smart Cities: PIF can be used to share energy usage of individuals without disclosing behavioural patterns.