Ethics, Privacy, Security

What is ‘PII’

“any information relating to an [...] natural person [...] who can be identified, directly or indirectly, in particular by reference [...] to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity.”

-Data Protection Directive

PII and Privacy Protection Technologies

Examples: *k*-anonymity, *l*-diversity

“these methods aim to make joins with external datasets harder by anonymizing the identifying attributes.”

The methods modifies quasi-identifiers to satisfy various syntactic properties to prevent

Problem: Simply not enough to do the job😅

Re-identification without PII

It turns out there’s a wide spectrum of human characteristics that enable re-identification as long as they satisfy the following key properties:

1. They are reasonably stable across time and contexts
2. The corresponding data attributes are sufficiently numerous and fine-grained that no two people are similar, except with a small probability.

Re-identification algorithms take advantage of such properties to re-identify individuals using other attributes.

Differential Privacy

Differential privacy (DP) is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals in the dataset.