ACS Report on Privacy Preserving Data Sharing Frameworks

Five Safes data analytics framework is developed for safe and non-intrusive data sharing. This framework focusses on setting up the environment for overcoming issues faced in trusted data sharing. The framework offers a quantifiable way of ensuring that the shared data is deidentified and the environment of data sharing is conducive to maintaining anonymity. It addresses the practical aspect of data sharing where identification whether the data being shared contains any personal information is necessary. The framework introduces the concept of safe thresholds which can be modified as per the project need while maintaining a control on the privacy of data.

Modern day AI algorithms are prone to violating data privacy and inhering bias present in training data. These threats imply that the algorithms should be passed through rigorous checks such as those employed by the Five Safes frameworks. This framework extends the concepts that are applicable to a human entity and apply them to an artificially intelligent algorithm with additional scrutiny:

Safe Algorithms: This principle suggests monitoring the training of the AI algorithm to identify the behaviour of model in various scenarios. This also includes monitoring any biases which might develop in training an AI model.

Safe Projects: This principle evaluates the end goal of the AI algorithm and considers the ethical and practical implications of how the algorithm operate in real world.

Safe Setting: This principle suggests ideal environment under which the algorithms should interact with the training data. This includes limiting access to additional details related to a data point which is not relevant to the scope of the algorithm.

Safe Data: This principle limits the algorithm ability to interact with other similar or dissimilar data which might lead to introducing unforeseen biases in the algorithm.

Safe Outputs: This principle is concerned with identifying the risks when sensitive user data is published.

References:

[1] https://www.acs.org.au/insightsandpublications/reports-publications/data-sharing-frameworks.html