



Data Science Ethics – Checklist

Version 1.0 (Date: 2019/08/17)

This checklist is being drawn by the professional service team of Smart Data Institute Limited (SDI). This is a reference material for all data science practitioners.

Informed Consent

- Is there any informed consent to the related stakeholders during collection of any personal information?
- Is it transparent to the stakeholders for the purpose of data collection?

Data Ownership

- Do we have a mechanism for an individual to request their personal information be removed?

Privacy & Anonymity

- Have we considered ways to minimize exposure of personally identifiable information (PII)?
- Have we ensured that data with PII not being used or displayed unless necessary for the analysis?
- Have you applied reasonable protection (such as encryption/ masking) on top of personal data like medical records, personal identity?
- Is there a plan to delete the historical data no longer needed?
- Is it possible to re-identify the underlying person by the data or information disclosed? (Anonymous/ masked data)

Data Validity

- Have we examined the data for possible sources of bias and taken steps to mitigate these biases?
- Is the analysis process being documented and available to audit?
- Are the people using our models aware of its shortcomings?
- Do we ensure all visualizations, summary statistics, and reports designed to honestly represent the underlying data?
- Have we considered possible bias for the data collected for analysis?
- Is there a way to suspend or roll back the model in production if necessary?

Address: Unit 1104A, 11/F Kai Tak Commercial Building, No 317-319 Des Voeux Road Central, Sheung Wan, Hong Kong
Tel.: +852 3110 6816
Fax: +852 3110 6813
Email: info@smartdatainstitute.com

Algorithmic Fairness

- Shall we take any result just favorable for our own organization benefits rather than social interests? Do we have a plan for response if users are harmed by the results?
- What are the consequences of not acting on false negatives and acting on false positives?
- Have we taken steps to monitor unintended uses and abuse of the statistical model?
- Can we explain and describe the decision using the model and result drawn?
- Do we maintain the model overtime to ensure its fairness?

SDI – Professional Service Team
service@smartdatainstitute.com

