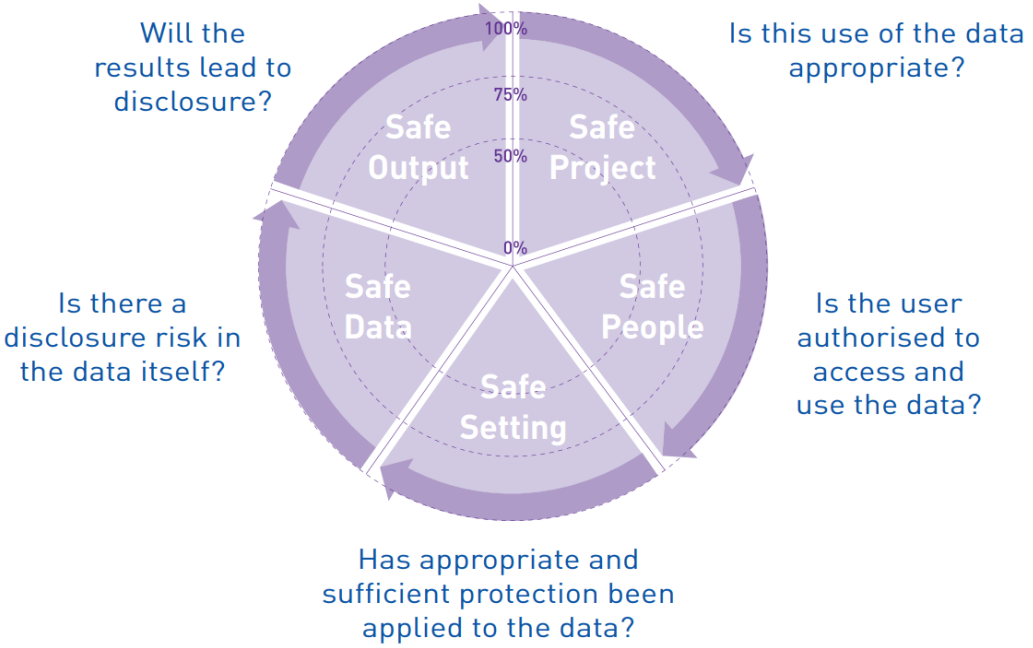
SIT719 Security and Privacy Issues in Analytics

Pass Task 10.1: ACS Report on Privacy Preserving Data Sharing Frameworks

‘Five Safes’ data analytics framework:



The Five Safes is a risk assessment framework for data access: safe people, safe projects, safe settings, safe data and safe outputs.

- Safe projects: Is the use of the data appropriate?
  Use of the data is legal, ethical and the project is expected to deliver public benefit.
- Safe people: Can the users be trusted to use it in an appropriate manner?
  Researchers have the knowledge, skills and incentives to act in accordance with required standards of behaviour.
- Safe data: Is there a disclosure risk in the data itself?
  Data has been treated appropriately to minimise the potential for identification of individuals or organisations.
- Safe settings: Does the access facility prevent unauthorised use?
  There are practical controls on the way the data is accessed – both from a technology perspective and considering the physical environment.
- Safe output: Are the statistical results non-disclosive?
  A final check can be required to minimise risk when releasing the findings of the project.

The Table 1 below illustrates how a Five Safes framework risk assessment supports the application of controls for data access. The table illustrates the four most common modes by which the AIHW shares and releases data and their associated controls.

	Open access **Website data files, tables and publications.**	Delivered access **Providing data directly to particular users.**	Secure remote access **Providing access to data through a secure remote connection.**	Secure on-site access **Providing access to data within the security of the AIHW data lab.**
Safe projects Is the use of the data appropriate?	**No control** Anyone can use the data for their own purposes.	**Moderate control** Users sign a declaration regarding the purpose for which they will use the data.	**Considerable control** Users can only use the data for the stated purpose; their access and use is controlled and monitored.	**High control** Project proposals are subject to a comprehensive evaluation by the AIHW.
Safe people Can the users be trusted to use it in an appropriate manner?	**No controls** Anyone can access the data.	**Very high control** Users sign legally binding undertakings.	**Considerable control** Authorised users sign legally binding undertakings.	**High control** Available to authorised expert users who agree to attend the Data Lab and sign legally binding undertakings.
Safe data Is there a disclosure risk in the data itself?	**Very High control** Data are highly aggregated and treated to protect privacy and confidentiality.	**High control** Data are treated by the AIHW to minimise the likelihood of identifying individuals.	**Considerable control** Treatments are applied to protect privacy and confidentiality while supporting the aims of the project.	**Moderate control** Treatments are applied to protect privacy and confidentiality while maximising the utility of the data.
Safe settings Does the access facility prevent unauthorised use?	**No controls** There no controls.	**Moderate control** Users are required to store the data securely and use it in their own physical and IT environment in accordance with a signed agreement.	**Considerable control** Access control is password based, physical security is specified in an agreement, data cannot be removed, and use of the data can be monitored and audited.	**Very high Control** The AIHW Data Lab is within the AIHW premises and subject to physical security, IT security, as well as monitoring and auditing capabilities. Data cannot be taken from the Data Lab.
Safe output Are the statistical results non-disclosive?	**No controls** There are no controls.	**Moderate control** The outputs are controlled by the user, but are governed by agreements with the AIHW.	**High control** Outputs can be audited by the AIHW and users are required to comply with the AIHW confidentialisation policy and practices.	**Very high control** Outputs meet project objectives, AIHW confidentialisation policy and practices, and are assessed by the AIHW before being released.

- Table 1. Five Safes framework controls applied under different modes of data sharing and release

AI algorithms implications on the framework:

The Safe People dimension may be replaced with AI algorithms that process data supplied for analytical purposes (such as clustering or classification) or for the purpose of delivering smart services (such as smart lighting or smart message routing).

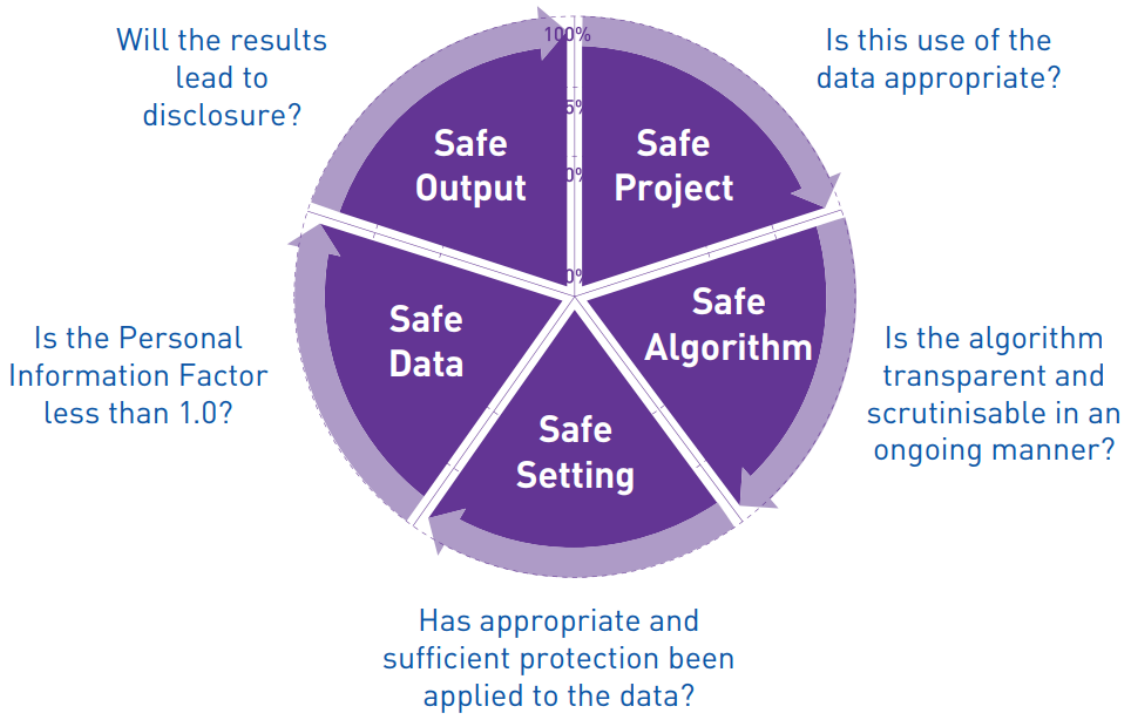


Figure 28. Five Safes Framework for algorithms

The Five Safes Framework is a system model and is intended to be considered in the context of all the elements. The answer to whether a researcher (or algorithm) is permitted to access a dataset assumes that all other necessary conditions are in place. If secure facilities do not exist, this does not seem like an appropriate way to use the data.

- SAFE ALGORITHMS: An artificially intelligent algorithm, the behaviours and associated access conditions can be enforced under many circumstances more easily than for a person but will need supervision if adapting over time. Any biases that develop also need to be monitored.
- SAFE PROJECTS: The safeness of the project that an AI algorithm undertakes should be known before application of the algorithm to the data. The challenge, however, is in discovery as the project progresses or if the project is a continuous operation rather than a discrete event.
- SAFE SETTING: When the researcher is an AI algorithm, the operating environment can be locked, disconnecting the algorithm from other sources of input. This does not, however, allow for any biases in the algorithm itself being evaluated or the implications of these being understood.

- SAFE DATA: When the observer is an AI algorithm, the context which the algorithm brings to the data can be limited through limiting access to other datasets, strictly limiting the personal Information Factor to be less than 1.
- SAFE OUTPUTS: There is a distinct difference to be further examined as to a single discrete output from an AI algorithm and something that feeds an operational loop (such as a steering algorithm or cruise control algorithm).

The underpinning concepts of the Five Safes Framework are significantly stretched when ‘person’ or ‘researcher’ is extended to an artificially intelligent algorithm. However, the basic considerations of the risk framework remain, including the Safe People and Safe Projects dimensions.