Appendix 2: Addressing re-identification of data in AI

To support responsible data handling and privacy preservation in our use of AI, we will adopt the following to assess the potential of re-identifying individuals from data which we use in AI. This scale informs decisions around anonymisation, access controls, data sharing, and AI usage.

Level	Risk Level	Description
1	Very Low	Data is fully anonymised using irreversible methods and contains and/or contains no direct or indirect identifiers. Risk of re-identifications remains negligible even when paired with external data sources. This applies equally when data is shared with or processed by AI services.
2	Low	Data has been pseudonymised or aggregated, with weak indirect identifiers. Re-identification is possible only with substantial external data and expertise. Use in Ai services may present low residual risk.
3	Moderate	Data includes combinations of indirect identifiers (quasi-identifiers) that could allow re-identification, especially when processed by AI models that retain memory or context. Care is needed when using external platforms.
4	High	Data contains sensitive information or detail identifiers, making re-identification likely without strong controls. Use in existing Ai services (especially black-box or public tools) could expose individuals to privacy risks.
5	Very High	Data includes clear personal identifiers or data where re-identification is trivial. Even using secure AI tools may lead to inadvertent exposure or misuse.

As we operationalise this risk scale and embed it into our responsible practices, we commit to the following actions across relevant projects and processes:

Data classification before use

- All datasets must be assessed using the re-identification risk scale before being:
    - Uploaded into or processed by AI tools and systems
    - Use to train or fine-tune internal AI models
    - Share with external collaborators or vendors
- Risk classification should be documented and included in the projects' data usage register or project details.

High-risk data used in tools

- Prohibit High-Risk data in unprotected external AI tools.
    - Level 4 and Level 5 data must not be entered into publicly available or black-box AI services (e.g., chatbots, generative models, only analytics tools) without:
        - A legal basis (e.g., explicit consent)
        - Strong contractual safeguards (e.g., which limit training)
        - Technical risk mitigation measures
    - Usage of such data in external systems must be documented and require written approval from the relevant line of responsibility as outlined in this policy or equivalent authority.

Mandatory 'safe space' for Level 3 – 5 data

- For Level 3 and above, data must be processed within a designated 'safe space'. This means a secure, access-controlled environment with the following characteristics:
    - No outbound API or cloud access without approval
    - Logging of all data access and user actions
    - Clear data retention and deletion policies
- Data should never be exported, copied, or reused outside of its intended purpose without reassessment.

Risk review before sharing or model training

- Before training on any dataset, a re-identification risk review must be conducted to:
    - Identify whether de-identification is sufficient
    - Determine if differential privacy, data minimisation, or synthetic data alternatives are required
- For external providers, documentation of these controls may be required with respective processing agreement.