

# Anonymization Procedure and Risk Assessment for Confidence Without Clarity

## 1. Overview:

This document outlines the procedures used to anonymize data collected in the study to protect participant confidentiality while maintaining research integrity.

## 2. Steps Taken:

### - Personal Identifiers:

- Removed direct identifiers (e.g., names, emails, IP addresses).
- Replaced participant identities with unique IDs (e.g., Participant\_001).

### - Demographic Aggregation:

- Grouped ages into 5-year intervals.
- Used general descriptors for geography (e.g., "Midwestern US").

### - Variable Aggregation:

- Combined education levels into categories (e.g., "post-secondary education").
- Calculated average accuracy scores for subgroups instead of individuals.

## 3. Data Masking:

- Ensured no rare combinations of demographics could re-identify participants.
- Masked unique participant traits through randomization.

## 4. Validation:

- Conducted a re-identification risk assessment using simulated attempts.
- Reviewed by an independent data privacy expert.

## 5. Documentation:

- A README file accompanies the dataset to explain anonymization steps.