



Safe Data Processing Outcome

Identified Privacy Risks

- **ML-Based Linkage:** AI models matched **88%** of ground truth records.
 - **High t-Closeness:** Education–Income relationship exceeds privacy thresholds.
 - **Quasi-Identifier Risk:** Small city populations are prone to re-identification.
 - **Poor I-Diversity:** The Health Condition field lacks sufficient diversity.
-

Applied Mitigation Measures

- **Differential Privacy Noise** added to quasi-identifiers.
 - **Generated Synthetic Data** for low-population regions.
 - **Generalized Education Levels** into fewer categories.
 - **Swapped Disease Values** between records for diversity.
-

Utility & Compliance

- **Mean Indicator Variation:** 0.38% — minimal analytical bias introduced.
- **Legal Alignment:** Meets **privacy-by-design** principles and SafeData AI framework standards.