STATATHON 2025

TITLE PAGE

- Problem Statement ID: **01**
- Problem Statement Title: **Data Security and Compliance**
- PS Category- Software/Hardware: **Software**
- Team ID: **3861**
- Team Name (Registered on portal): **Noise Injectors**

Noise Injectors

QuasiShield

STATATHON 2025

Proposed Solution (Describe your Idea/Solution/Prototype)

Detailed explanation of the proposed solution

An AI system that **secures anonymized data** from re-ID attacks and **auto-fixes risks** for privacy compliance.

- Enabled **API integration** so other systems can analyze and secure datasets remotely.
- Simulated multiple **re-identification attacks** to evaluate dataset safety.
- Calculated **risk levels** using recognized privacy metrics.
- Applied protection via **masking, generalization, noise, and synthetic data**.
- Validated protections by **re-running attacks** to confirm risks are removed.
- Generated detailed compliance **reports with risks, fixes, and outcomes**.

How it addresses the problem?

- Stops **linkage attacks** by breaking identifier connections.
- Prevents **attribute leaks** using diversity and noise.
- **Blocks composition risks** with synthetic and swapped data.

Innovation and uniqueness of the solution

- Self-learning system **adapts** to **new attack methods**.
- AI chooses the **best privacy fix** per risk.
- **Real-time protection** for streaming and large datasets.

CLICK HERE TO VIEW SOLUTION DEMO

TECHNICAL APPROACH

TECHNOLOGY STACK

Core Development & Programming:

Python, JavaScript – Core application development.

Data Privacy & Processing:

Pandas, NumPy, OpenDP, anonypy – Data handling and privacy protection.

Attack Simulation & ML:

scikit-learn, XGBoost, recordlinkage – Risk analysis and predictive modeling.

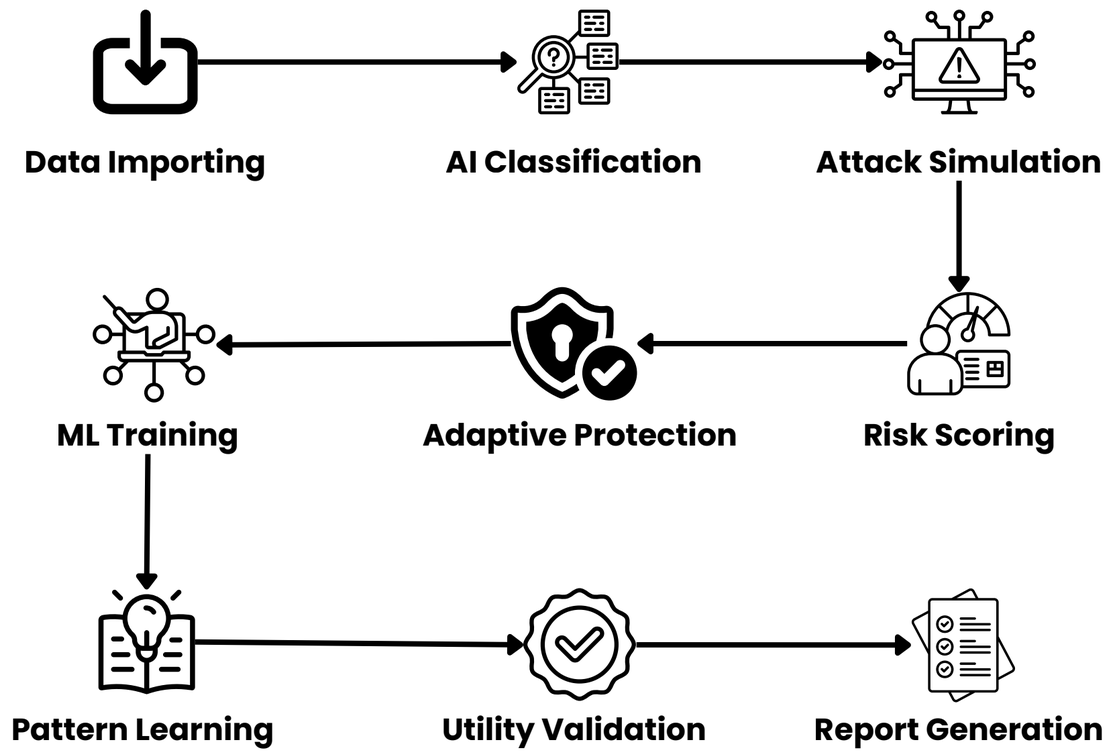
Frontend & Visualization:

React.js, Tailwind CSS, Plotly – Web interface and data visualization.

Reporting & Compliance:

Jinja2 & WeasyPrint – Automated privacy compliance reporting.

FLOW CHART



Noise Injectors

FEASIBILITY AND VIABILITY

STATATHON 2025

Analysis of the feasibility of the idea

- **Record linkage** can replicate **deterministic** and **probabilistic** matching effectively.
- **OpenDP** framework enables applying differential privacy with tunable privacy budgets.
- Synthetic data generators can **hide sensitive values** while keeping patterns intact.
- Graph analysis tools detect **network-based linkage** risks in datasets.
- Risk scoring methods like **k, l, t, δ** are already well-established.

Potential challenges and risks

- ML-based linkage attacks may **bypass** basic **anonymization** methods.
- Attribute disclosure risk when sensitive values **show clear patterns**.
- Composition attacks from **merging multiple anonymized datasets**.

Strategies for overcoming these challenges

- Use **adaptive anonymization** that strengthens after each attack simulation.
- Apply **l-diversity** and **t-closeness** to break sensitive value patterns.
- Add **synthetic data** and **swapping** to disrupt dataset merging links.

CLICK HERE TO VIEW SOLUTION DEMO

Noise Injectors

IMPACT AND BENEFITS

STATATHON 2025

Potential impact

- **Detects hidden identifiers** using AI to stop indirect re-identification before any data release.
- **Protects sensitive information** from advanced linkage and reconstruction attacks.
- Reduces re-identification risk while keeping **essential patterns** for **safe use**.
- Strengthens data compliance with evolving **privacy and protection** laws.
- Supports **UN SDG Goal 16** by promoting strong institutions through safe and transparent data sharing.
- **Blocks composition attacks** that combine multiple datasets for identity leaks.

Benefits of the Solution

- Improves **public trust** by ensuring shared data **cannot be misused** for harmful identity exposure.
- **Reduces legal costs** through automated compliance with strict privacy and protection regulations.
- Supports **secure research** by providing safe datasets without risking individual privacy breaches.
- **Prevents data misuse** in industries handling sensitive information like healthcare and finance.
- **Encourages data sharing** that can help solve social problems without revealing personal details.
- Supports **UN SDG 9** by promoting innovative privacy technologies.

CLICK HERE TO VIEW SOLUTION DEMO

RESEARCH AND REFERENCES

- https://www.nber.org/system/files/working_papers/w32905/w32905.pdf?utm_source=PANTHEON_STRIPPED
- https://www.cs.purdue.edu/homes/ninghui/papers/t_closeness_icde07.pdf
- https://www.nowpublishers.com/article/DownloadEBook/DBS-008
- https://arxiv.org/pdf/2211.10459

CLICK HERE TO VIEW SOLUTION DEMO