📜 Campus Entity Resolution & Security Monitoring System (Prototype)

Saptang Labs Product Development Challenge 2025

Team: INFINITYv69

1. Introduction and Problem Statement

Modern campus environments generate vast and fragmented data sources such as card-swipe logs, Wi-Fi connections, library checkouts, and CCTV events.
Security and administrative teams struggle to unify this information, making it difficult to track individuals or assets across multiple systems.

This project presents a **Campus Entity Resolution & Security Monitoring System** — a unified, privacy-aware platform that links fragmented identifiers (card IDs, device hashes, emails) into a **single entity identity**.

The system aims to:

- Consolidate multiple data sources into one view
- Build chronological activity timelines
- Detect and predict anomalous behavior
- Trigger proactive alerts for inactivity (e.g., no activity for > 12 hours)

This prototype demonstrates all **Round 1 objectives**:
Entity Resolution, Cross-Source Fusion, Timeline Generation, Predictive Monitoring, and Security Alerting.

2. System Architecture (Simplified Prototype)

The prototype follows a **three-layer modular architecture**, designed for clarity and future scalability.

2.1 Data Sources Layer (Simulation)

Simulated datasets emulate real campus logs:

- **Profiles:** master list of students and staff (entity_id, name, email, card_id, device_hash, dept)
- **Swipe Logs:** card-access events (card_id, location, timestamp)

- **Wi-Fi Logs:** device association events (device_hash, access_point, timestamp)
- **Free-Text Notes:** simulated help-desk or internal notes

These synthetic datasets ensure privacy while representing realistic data diversity.

2.2 Fusion Layer (Entity Resolution)

Core processing layer performing:

- **Deterministic Matching:** exact joins on card_id, device_hash, email
- **Probabilistic Matching:** fuzzy name similarity using SequenceMatcher ratio
- **Confidence Scoring:** weighted scoring (card match > device > email > name)

The layer outputs a **Fused Activities Dataset** — the unified record of all known entity events.

2.3 Application Layer (Streamlit Dashboard)

A responsive **Streamlit dashboard** provides:

- Entity search and selection
- Chronological activity timeline
- Confidence visualization for linked data
- Predictive monitoring of next likely location
- 12-hour inactivity alerts

3. Entity Resolution and Fusion Algorithm

3.1 Deterministic Entity Resolution

Entity identities are resolved through direct foreign-key matching.

Dataset	Matching Key	Master Key	Action
swipes.csv	card_id	profiles.card_id	Merge → entity_id
wifi.csv	device_hash	profiles.device_hash	Merge → entity_id
notes.csv	name, email	profiles.name, profiles.email	Merge → entity_id

3.2 Probabilistic Matching

If deterministic links fail, fuzzy name similarity (ratio ≥ 0.6) provides partial matches.

All logs are standardized into a common schema:
(entity_id, timestamp, source, type, description, match_confidence)
and concatenated into one comprehensive **Activities DataFrame**.

3.4 Confidence Scoring

Weighted heuristic:

- Card ID → 5
- Device hash → 4
- Email → 3
- Name similarity → 2

Final confidence = weighted sum / total weight — producing an explainable linkage score.

4. Predictive Monitoring and Security Alerting

4.1 Predictive Monitoring

A simple pattern-based predictor analyzes previous movement sequences to infer the **next probable location** of an entity.
Example: if typical path = *Cafeteria → Library → Dorm*, and last = *Cafeteria*, predicted next = *Library*.
Output includes:

- Text prediction ("Likely heading to Library")
- Confidence (0 – 1) based on frequency of recurrence

4.2 Proactive Alert – "Missing for 12 Hours"

Each entity's last log time is compared with system time; a gap > 12 hours triggers an alert.

Alert includes:

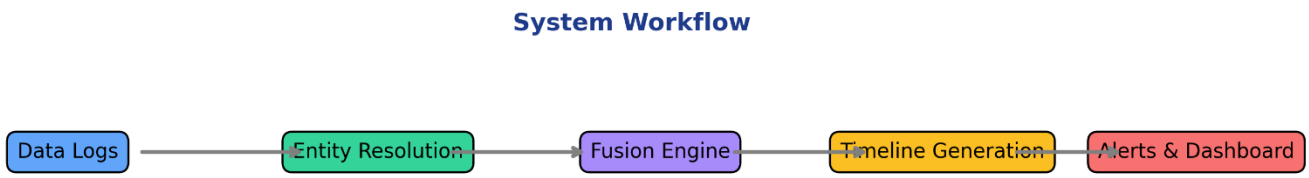
- Duration of inactivity
- Last activity timestamp
- Entity details (name, department)

4.3 Explainability & Provenance

Every timeline event lists:

- **Source:** Swipe, Wi-Fi, Note
- **Type:** CardAccess or AP_Connection
- **Timestamp:** Exact time

When an alert triggers, the **last known event** is highlighted as evidence.



5. Privacy Safeguards and Robustness

5.1 Privacy Safeguards

- **Entity Abstraction:** analysis on anonymized entity_id values
- **Controlled Access:** dashboard queries limited to authorized users
- **Data Minimization:** only relevant, non-PII fields displayed

5.2 Robustness & Failure Mode

- Duplicate logs removed via drop_duplicates()
- Missing foreign keys handled gracefully
- Future Enhancements:
    - Fuzzy string matching for misspelled names/emails
    - Confidence-weighted probabilistic matching
    - Real-time data stream integration (Kafka, Firestore)

6. Key Features Summary

Feature	Description	Impact
🔗 **Entity Resolution**	Links card, Wi-Fi, and note data into unified IDs	Eliminates fragmentation
🧠 **Predictive Monitoring**	Learns and predicts user movement patterns	Supports proactive security
⏰ **12-Hour Alert System**	Detects abnormal inactivity	Enables timely response
🧩 **Explainability Layer**	Lists evidence for each prediction	Builds trust & transparency
🔒 **Privacy Controls**	Uses anonymized entity IDs	Protects PII & compliance

7. Privacy Safeguards and Robustness

7.1 Privacy Safeguards

- Operates only on anonymized entity_ids
- Restricted dashboard access simulating SOC environment
- Data minimization — shows only necessary fields

7.2 Robustness

- Handles duplicates with drop_duplicates()
- Gracefully skips missing identifiers
- Future enhancements:
    - Fuzzy matching for text fields
    - Real-time ingestion with Kafka or Firebase

Keywords: *Entity Resolution, Multi-Modal Fusion, Predictive Monitoring, Campus Security, Streamlit Prototype*

8. Future Scope and Scalability

The prototype can be scaled into a **full-fledged Smart Campus Monitoring System** with:

- 🔍 **CCTV & IoT Integration:** video and sensor fusion
- 🌐 **Real-Time Event Streaming:** using Kafka or AWS Kinesis
- 🧮 **Machine Learning Models:** anomaly prediction and location clustering
- 📊 **Dashboard Enhancements:** role-based access, live maps, and risk scoring

These extensions would make it production-ready for enterprise-level deployment.

9. Conclusion

This Streamlit prototype demonstrates the **core goals** of a Campus Entity Resolution & Security Monitoring System:

- Unified cross-source identity linking
- Chronological timeline generation
- Predictive monitoring and inactivity alerts
- Privacy and explainability at the core

It provides a scalable foundation for future multi-modal security analytics integrating video feeds, real-time alerting, and adaptive ML models.