TED UNIVERSITY

CMPE 492 / SENG 492 Senior Project

General AI Safety Systems Test Plan Report

Spring 2025

Team Members

Mustafa PINARCI 18853734706 Computer Engineering

Ege İZMİR 12584814676 Computer Engineering

Egemen Doruk SERDAR 71155167474 Software Engineering

Mustafa Boğaç MORKOYUN 44764509874 Software Engineering

Supervisor: Gökçe Nur YILMAZ

Jury Members

Eren ULU

Tansel DÖKEROĞLU

Tolga Kurtuluş ÇAPIN

General AI Safety Systems: A Multidisciplinary

1. Executive Summary

The General AI Safety System is a comprehensive platform designed to enhance safety, efficiency, and accountability in school transportation through a fusion of facial recognition, route optimization, real-time notifications, and data management. This report presents a multidisciplinary view of the system, not only assessing its technical performance but also analyzing its ethical, legal, human-centered, and policy-related implications. By bridging engineering practices with real-world societal needs, this work aims to contribute meaningfully to the development of responsible and impactful AI systems.

2. Technical Foundation

The system comprises modular components responsible for attendance tracking via facial recognition, route recalculations based on external APIs, and communication through messaging services.

Testing Objectives:

- **Unit Testing:** Conducted using JUnit and Pytest for verifying methods such as `verifyStudent(image_path)`.
- **Integration Testing:** Focused on data flow between modules like FaceRecognition and AttendanceLogger.
- **System Testing:** Simulated real-life boarding scenarios.
- **Performance Testing:** Benchmarked key operations such as face recognition (<2s) and route recalculation (<5s).
- **UAT & Beta Testing:** Engaged end-users (drivers, school staff) to assess usability.

Tools & Environment:

- Dockerized backend, GPU-enabled face recognition, AES-256 encrypted PostgreSQL database.
- Simulated APIs (Google Maps, OpenWeather) and messaging via Twilio, SendGrid, and Firebase.

3. Ethical and Legal Considerations

In building a safety-critical AI system, ethical vigilance is paramount.

Data Privacy & GDPR:

- The system complies with GDPR by ensuring encryption of sensitive data and respecting user consent.

Fairness & Bias:

- Facial recognition systems inherently risk demographic bias. Mitigation involves continuous evaluation of training datasets and fairness metrics.

Transparency:

- The ACM Code of Ethics emphasizes transparency and accountability. Users should be informed about how their data is used, and have the right to access and correct it.

Security Standards:

- Alignment with ISO/IEC 27001 ensures that the system's security architecture adheres to globally accepted standards.

4. Human Factors and Usability

Technology must serve people with empathy.

Psychological Safety:

- Visible surveillance may cause discomfort or behavioral shifts in students. Transparency and clear communication are essential.

User Experience (UX):

- UAT highlighted the need for intuitive interfaces, especially for non-technical users such as school administrators and drivers.

Feedback Loops:

- Structured feedback during testing allowed iterative improvements, aligning system functionality with user expectations.

Accessibility:

- Future iterations should include multi-language support and accessibility features for differently-abled users.

5. Public Safety and Policy Implications

The system's deployment extends beyond the technical domain into public policy.

School Bus Safety:

- By automating attendance and enabling live tracking, the system directly contributes to safer, more accountable school transportation.

Policy Integration:

- There is potential for collaboration with ministries of education and transport to standardize such safety systems nationally.

Community Trust:

- Transparent data practices and public engagement will be crucial in fostering trust and adoption at scale.

6. Risk Assessment and Mitigation Strategies

Technical Risks:

- *Face recognition failure:* Mitigated via image preprocessing and fallback alert systems.
- *API downtime:* Handled with local caching.

Human Risks:

- *Inadequate training:* Addressed through user manuals and training sessions.

Legal Risks:

- *Data misuse:* Prevented with strict access control and regular audits.

Business Risks:

- *Rejection due to poor UX:* Minimized by prioritizing UAT and stakeholder reviews.

7. Future Work & Recommendations

Technical Enhancements:

- Real-time model updates, edge device support, and mobile-first design.

Ethical Upgrades:

- Integrate bias auditing tools and model explainability dashboards.

Educational Outreach:

- Create digital literacy workshops for users to understand the system's capabilities and boundaries.

Scalability & Commercialization:

- Partner with local governments or NGOs for pilot deployments in rural or underserved areas.

---

8. References

[1] ACM Code of Ethics and Professional Conduct

[2] General Data Protection Regulation (GDPR), European Union

[3] National Highway Traffic Safety Administration, U.S. Department of Transportation

[4] ISO/IEC 27001 - Information Security Management

[5] YOLOv3: Real-Time Object Detection, J. Redmon

[6] Dlib Face Recognition Documentation

[7] OWASP Web Security Testing Guide

[8] Object-Oriented Software Engineering, B. Bruegge and A. H. Dutoit

[9] Dlib HOG Face Detection Overview