# Investigating Contextual Variations in Aviation Social Intention Recognition: A Scenario Design Framework for Evaluating Intelligent Systems

**Student:** Omer Arslan (Arslan@student.tudelft.nl)
**Responsible Professor:** Hayley Hung, Vitaliy Popov, Arthur Mercier
**EEMCS, Delft University of Technology, Netherlands**

TU Delft

## 1. Introduction and Background

- **Problem:** Identical observable human behavior can generate multiple valid intention interpretations depending on the environmental context and observer's role
- **Problem:** Current intelligent systems assume single ground-truth intentions, collapsing legitimate ambiguity [1]
- **3Cs framework:** we interpret the world by noticing **cues**, identifying their **characteristics**, and organizing them into **classes** [2].
- **Script theory:** human behaviour falls into patterns because they function the way a written script does, by providing a program for action [3]
- **Goal:** advise researchers who want to make a model by giving them a framework for the right way of designing case studies

## 2. Research Question

**How can contextual variations in aviation be systematically characterized to guide AI researchers in designing data collection scenarios that capture multiple valid interpretations of social behavior?**

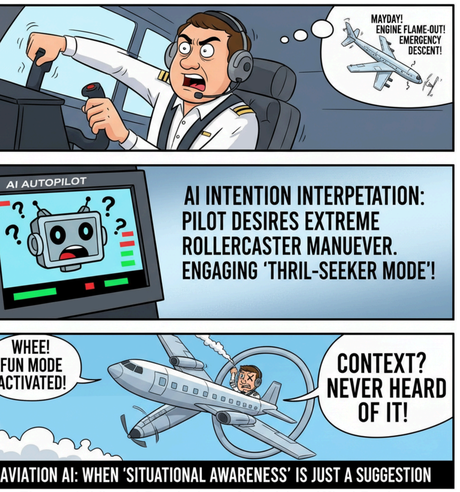


Figure 1: Visual Depiction of the Identiffied Problem

**SQ1:** contextual dimensions identified in literature influencing interpretation difference

**SQ2:** how external/internal scripts used to analyze divergent intention narratives

**SQ3:** how conflicts between internal and external scripts be systematically represented in aviation scenarios to reveal interpretive ambiguity

**SQ4:** limitations of using contextual framework and script-based scenario

## 3. Methodology

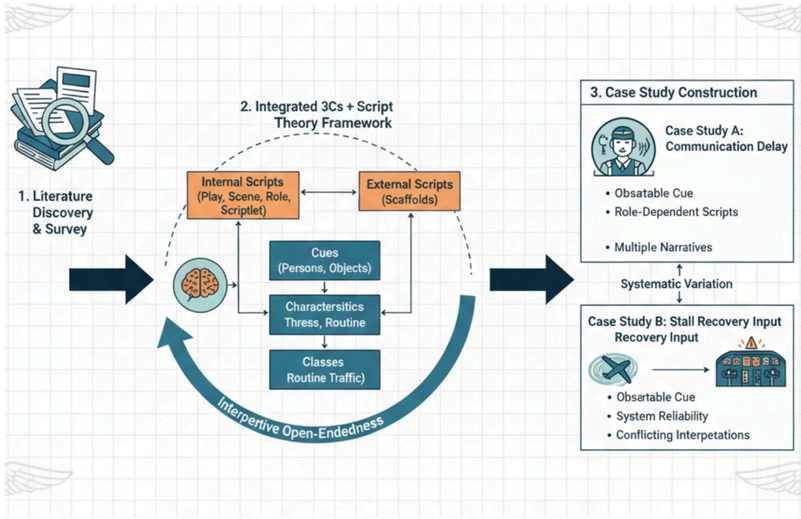


Figure 1: Visual Demonstration of the Steps Taken to Develop Case Studies

## 4. Results

| Observer | Internal Scripts | External Scripts | Characteristics | Classes |
|---|---|---|---|---|
| Experienced ATC Routine Context | Prioritizes traffic flow, predictability, and immediate compliance for system efficiency. | Institutional procedures requiring standardized phraseology and timely pilot responses. | **Non-compliance.** A minor, but notable, deviation from the norm. | Routine Traffic Violation |
| Experienced ATC Emergency Context | Prioritizes threat detection and airspace safety; non-compliance is a potential indicator of a critical event. | Emergency protocols that elevate the significance of any deviation from expected behavior. | **Potential Incapacitation.** A critical indicator of a possible loss of crew situational awareness. | Undeclared Emergency |
| Experienced Pilot High-Stress Context | Prioritizes the "Aviate, Navigate, Communicate" hierarchy; aircraft control is paramount. | Emergency checklists and company policies that mandate stabilizing the aircraft before external communication. | **Appropriate Task Prioritization.** A deliberate and disciplined action. | Effective Workload Management |
| Novice Pilot High-Stress Context | Tends toward rigid procedural adherence; may struggle to prioritize under high cognitive load. | Training-based procedures that may not have been fully internalized for dynamic, high-stress situations. | **Hesitation/Indecision.** A sign of being overwhelmed or tasksaturated. | Potential Task Saturation |

Table 1: Multi-Dimensional Script-Based Interpretations of the 8-Second Delay

| Observer | Internal Scripts | External Scripts | Characteristics | Classes |
|---|---|---|---|---|
| Pilot / Instructor (Assuming Reliable Sensors) | Built on the ingrained "push, roll, power" mantra for stall recovery; prioritizes immediate procedural execution. | Standard Operating Procedures and flight manuals that explicitly mandate a nose-down input to recover from a stall. | **Gross Pilot Error.** A catastrophic failure to apply fundamental training. | Loss of Control by Pilot Error |
| Pilot / Instructor (Suspecting Faulty Sensors) | Experience-based knowledge that high-altitude sensor errors can produce misleading warnings and data. | Specialized training on system limitations and unreliable airspeed checklists that override basic stall recovery mantras. | **Misguided Procedural Adherence.** A rational attempt to fly the aircraft based on faulty instrument readings | System Induced Pilot Error |
| Maintenance Engineer (Assuming Reliable Sensors) | System-first diagnostic mindset; assumes pilot competence and searches for external causes or anomalous human factors. | Maintenance logs and system performance data showing no history of sensor faults for this specific aircraft. | **Anomalous Pilot Action. A puzzling behavior that contradicts expected competency.** | Unexplained Pilot Deviation |
| Maintenance Engineer (Suspecting Faulty Sensors) | Diagnostic mindset that prioritizes potential hardware/software failures as root causes for pilot actions. | Diagnostic fault trees and airworthiness directives that provide knowledge of potential sensor failures. | **Logical Response to Faulty Data. A rational action given the high probability of erroneous sensor data.** | Suspected Sensor Failure |

Table 2: Multi-Dimensional Script-Based Interpretations of the Nose-Up Input

## 5. Conclusion and Future Work

- Intention recognition in aviation is inherently open-ended: identical behaviors support multiple valid interpretations across roles and contexts.
- Holding behavior constant while varying context reveals unrealized intentions and cross-role conflicts.
- Future work should do empirical validation with pilots, ATCs, and maintenance professionals in real and simulated scenarios is needed. It should also handle cross-cultural analysis of professional scripts and interpretation norms.

**References**
[1] Anonymous, "Evaluating model explanations without ground truth," arXiv:2505.10399, 2025, preprint. [Online]. Available: https://arxiv.org/abs/2505.10399 [2] J. F. Rauthmann and R. A. Sherman, "Conceptualizing and measuring the psychological situation," in Measuring and Modeling Persons and Situations, D. Wood, S. J. Read, P. D. Harms, and A. Slaughter, Eds. Academic Press, 2021, pp. 427–463, doi: 10.1016/B978-0-12-819200-9.00009-0. [3] R. C. Schank and R. P. Abelson, Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures. Hillsdale, NJ, USA: Lawrence Erlbaum Associates, 1977.