



BUCKINGHAMSHIRE NEW UNIVERSITY

EST. 1891

Downloaded from: <https://bnu.repository.guildhe.ac.uk/>

This document is made available with permission. Copyright remains with the author(s).

Use of items from Buckinghamshire New University's institutional repository must comply with the usage guidelines.

This work is made available under a **Creative Commons Attribution-NonCommercial 4.0 International Licence (CC BY-NC 4.0)**: <https://creativecommons.org/licenses/by-nc/4.0/>

Please note that you must also do the following;

- cite the authors, title and full bibliographic details of the item
- include a hyperlink/URL to the repository record in the citation
- keep all files required for use of the item together

You may not

- sell any part of an item
- refer to any part of an item without citation
- use or adapt the work in a way that misrepresents the author or impugns their reputation
- remove or alter the copyright statement on an item

If you need further guidance contact the Library and Learning Services
at Library@bnu.ac.uk

Human Machine Interaction in Automated Flight Operations through Accident-based Evidence on Post-disruption Recovery

Naveed Kapadia

College of Leadership and Innovation
Buckinghamshire New University
High Wycombe, HP11 2JZ, UK
naveed.kapadia@bnu.ac.uk

Qichun Zhang

College of Creative Arts, Technology & Engineering
Buckinghamshire New University
High Wycombe, HP11 2JZ, UK
kit.zhang@bnu.ac.uk

Abstract—Accident investigation remains one of aviation's principal mechanisms for safety learning, yet its value depends not only on timely and credible Final Reports, but also on how those reports are interpreted beyond simple attributions of pilot error. This paper examines human machine interaction in automated flight operations through a selective comparison of official accident and serious incident reports, supported by a focused literature analysis. It argues that the critical safety question in highly automated operations may arise when normal system stability ends, and the flight crew must rapidly rebuild understanding, control, and coordination under uncertainty. Across the literature and case material, the most operationally significant difficulty appears to lie not only in disruption itself but also in the period that follows, when crews must interpret degraded or conflicting cues, revise their understanding, and coordinate recovery within a compressed decision window. This paper shows that flight crew preparedness for post-disruption recovery appears to be less explicitly developed in the literature than disruption itself and uses accident-based evidence to demonstrate why this issue merits closer attention in safety learning and training design.

Keywords— Human machine interaction, automated flight operations, accident investigation, post-disruption recovery, aviation safety, crew-resilience.

I. INTRODUCTION

Accident investigation remains one of aviation's principal mechanisms for system learning. Under Annex 13 to the Convention on International Civil Aviation, the State conducting the investigation should make the *Final Report* publicly available as soon as possible, if possible, within twelve months. If that timeline is not met, an interim statement should be issued on each anniversary of the occurrence [1]. The *Final Report* is intended to include factual information, analysis, conclusions, and safety recommendations, which make it central to the production of credible safety knowledge [1].

Recent industry concern suggests that this learning chain is not always functioning as intended. A joint position paper by Airports Council International (ACI) Europe, the European Cockpit Association (ECA), the Flight Safety Foundation (FSF), the International Air Transport Association (IATA), the International Coordinating Council of Aerospace Industries Associations, the International Federation of Air Line Pilots' Associations (IFALPA), and the International Federation of Air

Traffic Controllers' Associations reported that, as of 30 June 2024, only 140 of 268 accidents recorded in IATA data between 2018 and 2023 had completed investigations. The same paper also reported that Flight Safety Foundation data showed only 160 of 448 accidents and serious incidents between 2018 and 2022 had been completed [2]. It argues that Final Reports must be complete, accessible, and timely, and that interim statements are not an adequate substitute for a comprehensive report [2].

This matters because accident and incident investigations sit within the wider safety management architecture. The ICAO Safety Management Manual states that accident and incident databases, together with Final Reports, support the analysis of safety deficiencies and the development of preventive action. It also notes that mandatory occurrence reporting tends to capture more technical information than human performance aspects [3]. In automated flight operations, where breakdowns may involve system transparency, cue interpretation, monitoring, control action, and crew coordination, official reports therefore provide an important, though still incomplete, basis for understanding how performance degrades and how recovery unfolds after disruption [3].

The literature on unexpected events and automation surprise reinforces the significance of this problem. Landman, Groen, van Paassen, Bronkhorst, and Mulder argue that the situations requiring human intervention in highly automated operations are often unforeseen and complex, may arise after long periods of low workload, and can be shaped by reduced process transparency and automation surprise [4]. De Boer and Dekker further suggest that automation surprise is better understood through a distributed and sense-making oriented perspective than through narrow attribution to individual failure alone [5]. Current IATA guidance on Competency Based Training and Assessment (CBTA) defines resilience as the ability of a flight crew member to recognise, absorb, and adapt to disruptions, and links this to observable behaviours such as monitoring aircraft state, validating information, responding to reduced situation awareness, and making timely decisions [6].

This paper examines a purposive set of official accident and serious incident reports to explore whether recurring patterns

in those reports point to a potentially under-examined area concerning flight crew preparedness for post-disruption recovery in highly automated flight operations. This paper uses accident based evidence to show why this area merits further investigation in safety learning and training design.

II. LITERATURE AND THEORETICAL FRAMING

A. Automation and unexpected events

The literature relevant to this paper suggests that the challenge in highly automated flight operations is not simply the occurrence of disruptions, but the cognitive and operational demands that arise when routine system behaviour is no longer reliable, transparent, or readily understood. Landman, Groen, van Paassen, Bronkhorst, and Mulder argue that increased automation has improved aviation safety while also creating new challenges, because situations requiring human intervention are often unforeseen and complex, may arise after long periods of automated flight, and demand rapid judgment and decision making [4]. They also note that automation may reduce process transparency, contribute to automation surprise, weaken active monitoring through complacency, and erode manual flying skills through extensive use [4].

This starting point can be strengthened by empirical work on automation conflict. Dehais, Peysakhovich, Scannella, Fongue, and Gateau describe automation surprise as a condition in which pilots detect a conflict with the automation but do not understand it and show that this can trigger excessive and inefficient visual search [7]. This is useful for the present paper because it supports the view that breakdowns in automated flight operations are not only about system malfunction or pilot error in isolation, but about the difficulty of interpreting automation behaviour quickly enough to recover effectively. A more cautious supporting perspective comes from Volz, Yang, Dudley, Lynch, Dropps, and Dorneich, whose study of information automation found that the automation group experienced the greatest performance degradation and the highest workload when returning to manual work, while the manual group experienced the least degradation and workload [8]. Although their study was not conducted in airline operations, it lends support to the broader concern that frequent reliance on automation may reduce the regular exercise of the cognitive work required when the pilot must suddenly reassume a more active problem-solving role.

B. Sense-making and re-framing

A useful way to understand post-disruption performance is through sense-making and re-framing rather than through a narrow focus on isolated errors. Landman, Groen, van Paassen, Bronkhorst, and Mulder explain that surprise arises when feedback no longer aligns with the crew's active frame, necessitating a rapid shift in interpretation before effective action can follow [4]. Here, the value of this perspective lies in its implication that disruption is not the endpoint of analysis.

The more important question is whether the flight crew can quickly revise their understanding to support recovery once the original frame no longer matches the situation.

This interpretation is also supported by empirical and cognitive research beyond the initial conceptual framing. Enriquez-Geppert, Smit, de Boer, Daneshnia, Lafont, and Dehais extend this by showing that flying performance depends, in part, on Executive Functions (EFs), such as working memory updating, set shifting, response inhibition, and conflict monitoring [9]. Collectively, these studies suggest that post-disruption recovery depends not only on noticing that something is wrong but also on interpreting conflicting cues, discounting inappropriate responses, and restructuring understanding in a rapidly developing situation.

C. Cognitive demands and performance adaptation

The cognitive demands of this problem can be developed further through the review by Enriquez-Geppert, Smit, de Boer, Daneshnia, Lafont, and Dehais. Their systematic narrative review argues that piloting is a complex task in a constantly changing environment and examines the role of Executive Functions (EFs), particularly working memory updating, set shifting, response inhibition, and conflict monitoring, in flight performance [9]. The review notes that working memory updating appears important for communication management and decisions requiring mental flexibility, while other Executive Functions remain comparatively understudied.

This is useful in the context of this paper because it provides a more explicit cognitive account of what recovery after disruption may require. The same review explains that under high stress, cognitive flexibility may degrade, and response inhibition becomes critical when a routine or automatic response must be inhibited in favour of a controlled one. It also explains that conflict monitoring is important when pilots must detect and resolve discrepancies between competing information sources under pressure [9]. These observations map closely onto the kinds of difficulties described in accident reports, such as misleading cues, delayed intervention, or uncertainty about the correct next action. They also help explain why post-disruption recovery may warrant closer analytical attention than a narrow focus on the initial disruption alone.

D. Systems thinking and performance assessment

A broader systems perspective strengthens this position. Systems-Theoretic Process Analysis (STPA) is described in ICAO guidance as a proactive, holistic hazard analysis and risk management tool derived from the Systems-Theoretic Accident Model and Processes (STAMP). Unlike methods that focus mainly on task level risks associated with human operators, it seeks to identify hazards and risks across an entire system and to address interactions between system components rather than only component failures. The same guidance states that safety is an emergent property of work

systems and not merely a property of isolated components [10]. This framing is useful for this study because the selected accident reports are unlikely to be adequately explained by a single failed action. They involve combinations of automation behaviour, degraded or misleading cues, monitoring demands, role allocation, procedural response, and recovery under time pressure.

Current training literature also highlights the need for more structured methods to observe and interpret pilot performance in complex operational settings. IATA guidance on Competency Based Training and Assessment (CBTA) links assessment to Observable Behaviours (OB) by asking how many behaviours were demonstrated when required, how often they were demonstrated, and what the outcome of Threat and Error Management (TEM) was for the competency being assessed [6]. This is important because it provides a structured vocabulary for linking training observation to broader questions of preparedness and performance robustness. Sun, Zhou, Zhang, Liu, Lu, Huang, and Song provide an example of how this logic can be operationalised through a competency-based assessment of pilots' manual flight performance during instrument flight training [11]. Their study developed observable behaviour indices and quantified how many and how often Observable Behaviours (OB) were demonstrated, with the resulting model showing strong agreement with expert evaluation. At the same time, their work is centred on manual flight performance in training rather than on post-disruption recovery in highly automated airline operations. Viewed together, these sources suggest that structured observation of pilot performance is feasible, but that the specific question of flight crew preparedness for post-disruption recovery in highly automated environments remains a distinct area deserving closer study.

This literature framing therefore supports the paper's central position. The issue is not only that disruption occurs in automated flight operations, but also that recovery depends on how the flight crew interpret changing cues, revise their understanding, coordinate action, and adapt under pressure within a wider socio-technical system. The next section, therefore, turns to a purposive set of official accident and serious incident reports to examine whether recurring patterns in those reports reinforce this interpretation.

III. METHOD AND CASE SELECTION LOGIC

A. Analytical approach

This paper adopts a qualitative and purposive analytical approach. It is not a systematic review or a statistical meta-analysis. It combines a brief literature review with a focused reading of selected official accident and serious-incident reports to examine whether recurring patterns can be identified regarding human-machine interaction and post-disruption recovery. This approach is consistent with the paper's aim, which is exploratory and interpretive rather than predictive.

Official investigation reports were selected because, under Annex 13, the Final Report is intended to include factual information, analysis, conclusions, and safety recommendations [1]. These reports therefore provide a structured basis for comparing how different investigation authorities describe the interaction between flight crew, aircraft systems, procedures, and operational context after disruption has occurred. At the same time, this paper recognises that '*Final Reports*' are written for safety learning rather than for testing human factors theory directly, and that the depth and style of analysis may vary between authorities [2], [3].

A systems perspective also informs the analytical stance taken here. Systems-Theoretic Process Analysis (STPA) treats safety as an emergent property of socio-technical systems and places emphasis on interactions, constraints, control, and feedback rather than on single component failures alone [7]. Although this paper does not conduct a formal Systems-Theoretic Process Analysis, that perspective is helpful in interpreting official reports, as the events considered here are unlikely to be adequately explained by isolated pilot action alone.

B. Case selection criteria

The case set was selected purposively rather than representatively. The aim was not to capture the full population of automation related events, but to assemble a small set of official reports that would allow analytical comparison across different phases of flight and different forms of human machine interaction. The selected reports were therefore chosen according to four criteria.

First, each case had to be documented in an official accident or serious incident report issued by a recognised investigation authority. Second, each case had to involve an event in which interaction between flight crew and aircraft systems, cockpit controls, operational cues, or procedural logic was central to understanding the occurrence. Third, the set was intended to include variation in the type of disruption, so that the paper did not become limited to one narrow category of event such as unstable approach or mode confusion alone. Fourth, each case needed to provide sufficient analytical detail to support the comparison of post-disruption interpretation, action, coordination, and recovery.

The case set was selected purposively to support analytical comparison rather than statistical representation. While this introduces the possibility of selection bias, the selection criteria were defined in advance, applied consistently across all cases, and designed to include variation in event type, flight phase, and interaction mechanism rather than a single accident category.

C. Coding framework

The purpose of the coding was therefore not to confirm a predetermined conclusion, but to test whether similar

recovery related themes could be identified across cases with different operational characteristics. To support comparison across cases, each report was read using a common thematic coding framework. The framework was designed to remain simple enough for a short conference paper while still capturing the main dimensions relevant to the argument. The coding focused on whether the report showed evidence of misleading or degraded system cues, monitoring or interpretation breakdown, control or action error, coordination or transfer difficulty, procedural misapplication, recovery in a compressed decision window, and explicit training or safety learning implications.

This coding structure was informed by the literature reviewed in Section II. The emphasis on cues, interpretation, and re-framing reflects the discussion of unexpected events and automation surprise [4], [7], [9]. The attention to interaction, rather than isolated failure, is consistent with the systems and socio-technical perspective described in ICAO guidance on Systems-Theoretic Process Analysis [10]. The decision to keep the coding categories concise also reflects the practical need to communicate the case comparison clearly to a mixed conference audience.

The resulting case table should therefore be read as a descriptive analytical aid rather than as a statistical statement about accident prevalence. Its purpose is to show whether similar interaction and recovery related themes recur across a small set of official reports, and whether those patterns help justify further research into flight crew preparedness for post-disruption recovery in automated flight operations.

IV. OFFICIAL REPORTS AND CROSS CASE COMPARISON

Table 1. presents a simplified cross case comparison. A tick indicates that the issue was explicit, central, or strongly evidenced in the official report analysis.

Case	Year	Misleading cues	Monitoring and interpretation breakdown	Control action error	Coordination or transfer issue	Procedural issue	Recovery under time pressure	Training relevance
Atlas Air 3591 [12]	2019	✓	✓	✓	✓		✓	✓
Qantas Flight 72 [13]	2008	✓	✓				✓	✓
Boeing 777 G-VIIT [14]	2024		✓	✓			✓	✓
Wizz Air A6-WZG [15]	2023	✓	✓	✓	✓	✓	✓	✓
A340 Clariden [16]	2018	✓	✓	✓	✓	✓	✓	✓

Table 1. Recurring themes across selected official reports

This table should be read as a descriptive analytical aid rather than as a statement about prevalence across aviation. Its purpose is to show that, despite variation in event type and flight phase, similar recovery related themes recur across the selected official reports.

A. Cross case overview

The selected official reports span different phases of flight and different forms of disruption, but they show a recurring cluster of issues relevant to Human Machine Interaction (HMI) and post-disruption recovery. Across the cases, the most prominent themes are misleading or degraded cues, monitoring and interpretation breakdown, inappropriate or delayed control action, coordination or transfer difficulty, procedural misapplication, and recovery under time pressure. The point of comparison is therefore not that the reports describe identical events, but that they reveal similar pressures on flight crew interpretation and adaptation once the situation no longer conforms to expectation.

B. Brief case illustrations

Atlas Air 3591 (2019) [12] provides a strong example of a time critical descent event in which unexpected automation state change, degraded interpretation of cues, and ineffective recovery become central to the accident sequence. In the context of this paper, its value lies in showing how human machine interaction problems can emerge when the crew’s understanding of system behaviour is overtaken by events.

Qantas Flight 72 (2008) [13] contributes a different mechanism. Rather than a discrete action slip or approach decision trap, it illustrates how erroneous system data and cascading cues can place flight crew in a high workload recovery situation in the cruise stage of the flight. It therefore broadens the case set beyond approach and landing events and strengthens the argument that post-disruption recovery in automated operations is not confined to one operational phase.

The Boeing 777 G-VIIT (2024) report [14] is especially useful because it shows that human machine interaction problems are not limited to advanced automation anomalies. During takeoff, the copilot began retarding the thrust levers at V_1 instead of removing his hand from them, then momentarily re-advanced them before committing to the Rejected Takeoff (RTO). The report characterises this as an action slip in a highly learned motor sequence and notes the operator’s emphasis on methodical control selection, mental rehearsal, and recurrent training on focus and mis-selections [14]. This case therefore shows that recovery can depend not only on understanding a complex system state, but also on recognising and correcting an initially incorrect control action under severe time pressure.

The Wizz Air A6-WZG (2023) report [15] adds a particularly rich recovery case. The investigation found a low energy go-around initiated after touchdown in gusty conditions, with inadvertent reverse selection, missed monitoring cues, unclear communication, role confusion, and a mistaken perception of excess energy despite low airspeed. It also records that tail strike protection callouts were inhibited during the go-around phase once Takeoff/Go-Around (TOGA) mode was engaged, and it recommends simulator training on low energy rejected

landing close to the ground, stronger emphasis on the pilot monitoring role, and continued attention to Crew Resource Management (CRM) issues during checks [15]. This makes it one of the clearest examples in the paper of how cue interpretation, coordination, and recovery preparation intersect during a dynamically evolving disruption.

The A340 Clariden (2018) report [16] is useful because it shows how weather, automation logic, procedural response, and crew cooperation can combine at high altitude. The investigation found that a rapidly changing wind caused an over-speed condition, after which the Pilot in Command disconnected the autopilot, later had two Air Data Inertial Reference Unit (ADIRU) switched off during incorrect application of the Operations Engineering Bulletin (OEB) procedure, and thereby lost auto-thrust with a temporary reversion to Alternate Law. The report further states that the crew's cooperation and implementation of procedures during over-speed and stall recovery were erroneous. In analytical terms, this case is important because it links misleading cues, misunderstood system logic, procedural misapplication, and degraded recovery in one sequence.

Across the selected cases, the reports do not merely show isolated pilot errors. They point to a recurring pattern in which disruption is followed by difficulties with cue interpretation, reframing, control selection, role clarity, or procedural execution. The combined significance of the findings lies less in the individuality of each event than in the way they suggest that post-disruption recovery in automated flight operations warrants more direct attention as a research and training focus.

V. DISCUSSIONS

A. Recurring patterns across the case set

The selected official reports suggest that the problem in automated flight operations is not limited to the initial disruption itself. Across the cases, difficulties emerged in the period immediately after the disruption, when flight crew had to interpret changing cues, reassess what the aircraft or automation was doing, select an appropriate response, and coordinate action under time constraints. Although the mechanisms differed across the reports, the cross-case pattern was broadly similar. Recovery became more difficult when cues were misleading or degraded, when the crew's understanding of the situation lagged the aircraft state, when control inputs did not align with the emerging problem, or when role clarity and coordination weakened at a critical moment.

This pattern is consistent with the literature reviewed earlier. Landman, Groen, van Paassen, Bronkhorst, and Mulder argue that unexpected events challenge the active mental frame and require rapid reframing before effective action can follow [4]. Dehais, Peysakhovich, Scannella, Fongue, and Gateau show that when pilots detect a conflict with the automation but do not understand it, attentional resources may be drawn into inefficient visual search [7]. Enriquez-Geppert, Smit, de Boer,

Daneshnia, Lafont, and Dehais similarly indicate that adaptation in dynamic flight conditions depends on Executive Functions such as updating, inhibition, shifting, and conflict monitoring [9]. In tandem, the official reports support the interpretation that the core problem is not only disruption but also the crew's capacity to quickly regain an accurate and workable understanding to support recovery.

B. From disruption to recovery

One of the most important observations from the case set is that disruption and recovery should not be treated as analytically identical. The accident sequence may begin with an automation anomaly, an action slip, a degraded energy state, or a procedural error, but the decisive safety question often becomes what happens next. In Atlas Air 3591 (2019) and Qantas Flight 72 (2008), the reports point to situations in which the relationship between aircraft behaviour and crew interpretation became difficult to manage in a fast moving environment [12], [13]. In Boeing 777 G-VIIT (2024), the event began with an action slip in a highly learned motor sequence, yet the safety significance of the occurrence depended on how the crew recognised and responded to that slip during the Rejected Takeoff (RTO) [14]. In Wizz Air A6-WZG (2023) and the A340 Clariden (2018) events, the reports show how monitoring, communication, procedure use, and control logic became tightly coupled during a deteriorating situation, making recovery itself the central analytical issue [15], [16].

This matters because it shifts the focus from asking only why the event began to asking how well the crew were positioned to recover once the original frame no longer matched the situation. That question is particularly relevant in highly automated flight operations, where long periods of routine performance may be interrupted by short phases of high consequence uncertainty. It also helps explain why a narrow reading of these events as isolated pilot error is insufficient. A systems perspective suggests that recovery difficulty may arise from the interaction among automation behaviour, interface logic, procedural expectations, monitoring burden, and coordination demands, rather than from a single failed act.

C. Research and training

The selected reports indicate that flight crew preparedness for post-disruption recovery may deserve more explicit attention than it currently receives in parts of the literature and training discourse. Sun, Zhou, Zhang, Liu, Lu, Huang, and Song further show that structured observation of pilot behaviour can be formalised in a way that supports evaluation and targeted training response [11].

Even so, the cases reviewed here suggest that post-disruption recovery is not captured fully by the initial recognition of a problem alone. It may also involve reframing, suppression of inappropriate responses, interpretation of conflicting cues, rapid re-prioritisation, and coordination under pressure. This suggests that future work may benefit from

looking more directly at how such recovery related behaviours are observed, interpreted, and supported in automated flight environments. In that respect, the contribution of this paper is not to offer a validated model, but to use official reports and existing literature to argue that the period after disruption occurs may be analytically and operationally significant.

VI. FUTURE DIRECTION AND OPERATIONAL IMPLICATIONS

Across the selected reports, the central difficulty often lies not in the onset of disruption alone, but in the period that follows, when flight crew must interpret changing cues, revise an increasingly unstable understanding of the situation, and coordinate action under time. This pattern suggests that post-disruption recovery merits more explicit attention as an operational and research problem in highly automated flight operations.

The literature reviewed in this paper already provides relevant concepts through automation surprise, re-framing, Executive Functions (EFs), and socio-technical interaction. The contribution of the case material is to sharpen their operational significance. The reports indicate that recovery may become vulnerable where cue interpretation degrades, procedural application becomes unstable, control inputs no longer align with the developing situation, or crew coordination weakens at critical moments. This suggests that future work may usefully focus on the transition from disruption to recovery, rather than concentrating primarily on disruption onset.

The findings also point to a practical opportunity in training and assessment. IATA on Competency Based Training and Assessment (CBTA) already defines resilience in relation to recognising, absorbing, and adapting to disruption, and links assessment to Observable Behaviours and Threat and Error Management outcomes [6]. This provides a basis for examining whether existing training environments are already generating observable indicators of post-disruption recovery, even if these are not yet being treated analytically in those terms.

VII. CONCLUSION

This paper examined human machine interaction in automated flight operations through selected official accident and serious incident reports, supported by a focused literature analysis. The evidence suggests that the most operationally significant challenge may lie not only in disruption itself, but in the period that follows, when flight crew must interpret degraded or conflicting cues, revise their understanding, and coordinate recovery within a compressed decision window. The events considered here are therefore better understood not as purely technical failures or isolated instances of individual underperformance, but as interaction problems shaped by automation behaviour, operational context, procedural logic, and crew coordination. The paper argues that while automation may support routine stability, the breakdown of that stability can expose a demanding form of human

vulnerability that is not always made explicit in accident interpretation or training discourse. It therefore positions flight crew preparedness for post-disruption recovery as an area warranting closer examination, particularly where accident analysis, training observation, and performance interpretation can be brought into closer alignment.

REFERENCES

- [1] *Annex 13: Aircraft Accident and Incident Investigation*, 13th ed., International Civil Aviation Organization (ICAO), 2024.
- [2] Airports Council International Europe, European Cockpit Association, Flight Safety Foundation, International Air Transport Association, International Coordinating Council of Aerospace Industries Associations, International Federation of Air Line Pilots Associations, and International Federation of Air Traffic Controllers Associations, "Joint position paper on publication of final reports," Position Paper, 2025.
- [3] International Civil Aviation Organization (ICAO), *Safety Management Manual*, 4th ed. Montreal, Canada: International Civil Aviation Organization, 2018, doc 9859, AN/474.
- [4] A. Landman, E. L. Groen, M. Van Paassen, A. W. Bronkhorst, and M. Mulder, "Dealing with unexpected events on the flight deck: A conceptual model of startle and surprise," *Human factors*, vol. 59, no. 8, pp. 1161–1172, 2017.
- [5] R. De Boer and S. Dekker, "Models of automation surprise: results of a field survey in aviation," *Safety*, vol. 3, no. 3, p. 20, 2017.
- [6] "Competency assessment and evaluation guidelines for pilots, instructors, and evaluators," International Air Transport Association (IATA), Montreal, Canada, Guidelines/Technical Report, 2025.
- [7] F. Dehais, V. Peysakhovich, S. Scannella, J. Fongue, and T. Gateau, "Automation surprise" in aviation: real-time solutions," in *Proceedings of the 33rd annual ACM conference on Human Factors in Computing Systems*, 2015, pp. 2525–2534.
- [8] K. Volz, E. Yang, R. Dudley, E. Lynch, M. Dropps, and M. C. Dorneich, "An evaluation of cognitive skill degradation in information automation," in *proceedings of the human factors and ergonomics society annual meeting*, vol. 60, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2016, pp. 191–195.
- [9] S. Enriquez-Geppert, D. Smit, M. C. de Boer, N. Daneshnia, A. Lafont, and F. Dehais, "A systematic narrative review of the involvement of executive functions in flying performance of pilots," *Frontiers in Neuroergonomics*, vol. 5, p. 1462304, 2024.
- [10] *Safety Risk Management Methodologies (SRM): STPA – System Theoretic Process Analysis*, International Civil Aviation Organization (ICAO), 2024.
- [11] H. Sun, X. Zhou, P. Zhang, X. Liu, Y. Lu, H. Huang, and W. Song, "Competency-based assessment of pilots' manual flight performance during instrument flight training," *Cognition, Technology & Work*, vol. 25, no. 4, pp. 345–356, 2023.
- [12] National Transportation Safety Board (NTSB), "Rapid descent and crash into water: Atlas air inc., flight 3591, boeing 767-375bcf, n1217a, near trinity, texas, february 23, 2019," National Transportation Safety Board, Aviation Accident Report NTSB/AAR-20/02, 2020.
- [13] Australian Transport Safety Bureau, "In-flight upset, 154 km west of Learmonth, western australia, 7 october 2008, vh-qpa, airbus a330-303," Australian Transport Safety Bureau, Aviation Accident Report AO-2008070, 2011.
- [14] Air Accidents Investigation Branch, "AAIB bulletin: 5/2025, boeing 777236, g-viit," Air Accidents Investigation Branch, Aviation Accident Report AAIB-30154, 2025.
- [15] Air Accident Investigation Sector, "Serious incident final report, aifn/0006/2023: Tail strike, wizz air abu dhabi, airbus a321-271nx, a6wzg, abu dhabi international airport, united arab emirates, 29 april 2023," Air Accident Investigation Sector, Aviation Accident Report AIFN/0006/2023, 2024.
- [16] Bundesstelle für Flugunfalluntersuchung, "Investigation report," Bundesstelle für Flugunfalluntersuchung, Aviation Accident Report BFU18-1626-FX, 2023.