*Proceedings of the 35th European Safety and Reliability & the 33rd Society for Risk Analysis Europe Conference*
*Edited by* Eirik Bjorheim Abrahamsen, Terje Aven, Frederic Bouder, Roger Flage, Marja Ylönen
 *Published by* Research Publishing, Singapore.
doi: 10.3850/978-981-94-3281-3_ESREL-SRA-E2025-P7204-cd

# Air Accident Analysis with AI: Reassessing Flight BA 5390 Using the Accimap and STAMP/CAST Methodologies

João Guilherme Fowler

*PhD Student, Aeronautics Institute of Technology, Brazil. E-mail: joaofowlergmail.com*

Marina Ronconi de Oliveira

*Master Student, Aeronautics Institute of Technology, Brazil. E-mail: marina.oliveira.101443@ga.ita.br*

Moacyr Machado Cardoso Júnior

*PhD, Professor, Aeronautics Institute of Technology, Brazil. E-mail: moacyr@ita.br*

This article explores the potential of artificial intelligence (AI) in aiding the investigation of aviation accidents, particularly through the application of the Accimap and STAMP/CAST methodologies. These frameworks are recognized for their ability to map complex causal relationships in aviation incidents, providing insights into the structural, mechanical, organizational, and human factors involved. The study focuses on the BA 5390 accident, utilizing AI to generate textual and schematic representations of the events and contributing factors. The AI was guided through structured prompts to perform tasks such as hazard identification, control structure modeling, and the generation of safety recommendations. While AI demonstrated strengths in text-based analysis and logical structuring, challenges emerged in generating accurate graphical representations, especially in the Accimap methodology. Human intervention was required to validate and refine both the textual outputs and the diagrams, ensuring factual accuracy and addressing gaps in the AI's understanding of complex system interactions. The findings suggest that AI can be a valuable tool in accident investigations, offering efficiency in organizing and processing large datasets, but it requires human oversight to mitigate potential inaccuracies and to deepen the analysis of human factors. The study concludes that AI is most effective when used in conjunction with expert judgment, particularly in scenarios where human factors and decision-making processes are critical to understanding the full scope of an accident.

*Keywords*: Accimap, STAMP/CAST, investigation, artificial intelligence, accidents, aviation.

## 1. Introduction

Aircraft accident investigation provides a unique opportunity to understand the multitude of factors contributing to such events and to enhance safety systems not only in structural and mechanical terms but also in organizational and human factors domains.

To support investigations, national specialized agencies typically adopt methodologies that reveal both direct and indirect correlations of events, enabling a narrative from an impartial perspective. The primary objective of these methodologies is to promote safety rather than assign blame.

The Accimap (Rasmussen, 1997) and STAMP/CAST (Leveson, 2004) frameworks are two well-established methods that, in addition to their textual contributions, offer analysts the capability to illustrate events through schematic diagrams. These visual tools facilitate pattern recognition and the synthesis of all existing information into coherent representations, highlighting the inherent complexity of such occurrences (Viitanen and Reiman, 2023).

However, constructing maps and diagrams using these methods can be a tedious task, prompting the question of whether generative language artificial intelligence (AI) can serve as a valuable ally in this context. If so, the reliability of the information provided by AI becomes a critical consideration.

This article seeks to evaluate the potential of AI to structure the investigation of the BA 5390 accident, which occurred in 1990, through the application of the Accimap and STAMP/CAST methodologies. Furthermore, it aims to identify the main strengths and limitations encountered in this process.

# 2. Literature review

## 2.1. *Artificial intelligence at civil aviation*

The application of artificial intelligence (AI) is extensively studied across various industrial sectors, ranging from systemic machine analysis for maintenance prediction (Bharadiya 2023) to supporting medical diagnostics (Sufyan et al. 2023) and conducting technological risk assessments (You et al. 2021).

In civil aviation, interest in the use of artificial intelligence is prominently highlighted in existing literature. It can be applied in air traffic control, flight safety, aircraft maintenance, facilitating and operating autonomous aircraft, and has drawn notable interest from regulatory agencies regarding the training and application of artificial intelligence within the realm of civil aviation (Kabashkin et al. 2023).

This view is supported by Tselentis et al. (2023), who report studies in which artificial intelligence can assist flight management, for instance, by integrating information to avoid turbulence, aiding in landings under strong wind conditions. Deniz et al. (2024) further identify AI as a potential solution for traffic management in the context of urban airworthiness.

Conversely, Youseftorkaman et al. (2023), while acknowledging several contributions of artificial intelligence, especially in repetitive tasks, also emphasize the risks and the need for improvements in reliability and safety, particularly in traffic management and activities that could be vulnerable to malicious actions.

In accident investigation, the potential of AI was already recognized in 1999 when Mussone et al. (1999) employed artificial neural networks to analyze and understand urban traffic accidents.

Currently, open-source AI demonstrates potential for event analysis applications, as exemplified by the work of Ziakkas and Pechlivanis (2023), who analyzed aviation accidents using HFACS, Accimap, Bowtie, and STAMP methodologies, achieving promising results. They concluded that AI can be useful in accident analysis and investigation, though ChatGPT performs better with linear models, such as BowTie. Ray et al. (2023), however, caution that event analyses conducted by generative language AI require careful consideration, emphasizing the critical role of human systems in validating generated information.

These findings establish a precedent for continuous testing and experimentation involving the application of different investigative methodologies, supported by open-source artificial intelligence, to assess the accuracy of such analyses.

## 2.2. *Accimap and STAMP/CAST*

Accident investigation models play a pivotal role in elucidating highly complex events with impartiality and minimal bias. Effective methods should be capable of representing interactions and multiple causalities, while also highlighting opportunities for improvement in a self-explanatory manner (Stanton and Young 2003).

The Accimap framework (Rasmussen 1997) initially maps key actors and their hierarchical levels within a system. It then illustrates the interactions among various events related to these actors, allowing for the visualization of influences, constraints, and imposed limitations, as well as system disturbances and feedback responses. These feedback mechanisms, moving both bottom-up and top-down, aim to restore system equilibrium (Qureshi 2008).

In contrast, the Systems-Theoretic Accident Model and Processes (STAMP), developed by Leveson (2004), is grounded in the Systems Theoretic Process Analysis (STPA) and facilitates the identification of hazards and risks, along with the investigation of accidents via Causal Analysis based on Systems Theory (CAST). The theoretical foundation of STAMP asserts that undesirable events arise when system disturbances, component failures, or inaccurate interactions are not effectively countered by existing control mechanisms. Such issues may be linked to organizational processes, operational structures, or adherence to legal and regulatory requirements.

The STAMP investigation process begins with the identification of actors and hierarchical levels, echoing the methodology of Accimap as proposed by Rasmussen (1997). These actors can include organizations, individuals, legal frameworks, governmental policies, or any elements that influence system operations and potentially affect organizational culture (Hulme et al. 2021). The analysis then delves into the control and feedback structures present in interactions among system components. This enables the identification of causal failures, their subsequent consequences, and a deeper understanding of how and why these control failures occur.

## 3. Method

The initial phase involved analyzing the event using the STAMP/CAST methodology proposed by Leveson (2004). To guide this process, the CAST Handbook (Leveson, 2019) was employed, providing the foundational structure for the analysis. Subsequently, open-source artificial intelligence (ChatGPT 3.5) was utilized to carry out and define the following steps: the basic composition of information, the analysis of loss components, the identification of control structure failures, and the generation of recommendations.

To achieve this, the AI was prompted to act as an expert in aviation accidents, explicitly following the methodology outlined in Nancy Leveson's STAMP/CAST framework (2004) and tasked with producing a comprehensive and detailed analysis of the BA 5390 accident. While the AI provided some preliminary insights, these were primarily employed to contextualize the system.

The detailed analysis progressed through the input of specific prompts, with the results being entirely revised or partially supplemented to ensure adherence to factual accuracy. The prompts and their corresponding outputs, along with modifications made, are summarized in Table 1.

Table 1. Prompts and provided information used to model STAMP/CAST in GPT Chat

| CAST step used as prompt at Chat GPT 3.5 | Description of any complementation |
|---|---|
| Define the system involved and the boundary of the analysis; | Partially complemented |
| Describe the loss and hazardous state that led to it | Partially complemented |
| From the hazard, identify the system-level safety constraints required to prevent the hazard (the system safety requirements and constraints). | Entirely made by AI |
| Describe what happened (the events) without conclusions nor blame. Generate questions that need to be answered to explain why the events occurred | Partially complemented |
| Analyze the physical loss in terms of the physical equipment and controls, the requirements on the physical design to prevent the hazard involved, the physical controls (emergency and safety equipment) included in the design to prevent this type of accident, failures and unsafe interactions leading to the hazard, missing or inadequate physical controls that might have prevented the accident, and any contextual factors that influenced the events | Partially complemented |
| Model the existing safety control structure for this type of hazard | The entire structure was created by the authors |
| Examine the components of the control structure to determine why they were not effective in preventing the loss: Starting at the bottom of the control structure, show the role each component played in the accident and | Partially complemented |

| CAST step used as prompt at Chat GPT 3.5 | Description of any complementation |
|---|---|
| the explanation for their behavior (why they did what they did and why they thought it was the right thing to do at the time). | |
| Identify flaws in the control structure as a whole (general systemic factors) that contributed to the loss. The systemic factors span the individual system control structure components. | Partially complemented |
| Create recommendations for changes to the control structure to prevent a similar loss in the future. If appropriate, design a continuous improvement program for this hazard as part of your overall risk management program. | Partially complemented |

For the elaboration of the Accimap, in the same interaction window, the following steps were requested:

1. Define the actors of the accident at their hierarchical levels considering Rasmussen (1997);
2. Make correlations between facts linked to each of the actors that contributed to the occurrence of the accident, providing feedback from top to bottom and from bottom to top;
3. Organize the previous information into a diagram to illustrate the Accimap of this event.

All existing information was validated and supplemented through consultation of the official occurrence report issued by the Air Accidents Investigation Branch (AAIB, 1992) and the audiovisual work on the incident Mayday: Air Disaster (2021). This approach ensured the detailed reconstruction of the event and the presentation of relevant specifics necessary for the investigation.

# 4. Results and dicussion

## 4.1. *STAMP/CAST*

In the context of presenting the responses generated by the AI, the definition of the Systems Levels Constraints was considered acceptable without any human intervention. The AI's return to the command "*From the hazard, identify the system-level safety constraints required to prevent the hazard (the system safety requirements and constraints*" resulted in the following answers:

SSC1: Clear and detailed specification of components in maintenance manuals;
SSC2: Inclusion of detailed checklists in maintenance procedures;
SSC3: Regular and updated training for the maintenance team;
SSC4: Periodic refresher training;
SSC5: Second-person checks for critical components;
SSC6: Use of detailed inspection checklists;
SSC7: Regular and detailed audits by supervisors;
SSC8: Robust feedback systems for communicating deviations;
SSC9: Maintenance and inspection procedures without time pressures;
SSC10: Design with safety redundancies;
SSC11: Installation of critical components in controlled environments;
SSC12: Detailed pre-flight checks by the flight crew;
SSC13: Immediate communication of pre-flight anomalies to maintenance.

Regarding the control structure, there was partial human intervention, which resulted in the creation of Figure 1.

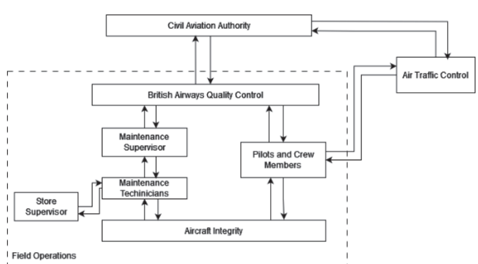


*Fig. 1. Control structure considering STAMP/CAST analysis*

The analysis carried out using the STAMP/CAST method was enhanced with the help of artificial intelligence. It is important to highlight that the inputs to the system were made freely, without specific coding to optimize the AI's learning, and followed the steps described in Nancy Leveson's Cast Handbook (2019).

The results of the analysis conducted by the generative language AI were satisfactory, considering that the investigative process already had consolidated clarity through access to the AAIB report (1992). In other words, there were no uncertainties regarding the causes of the event, nor was AI used to infer situations where the investigative process was still under analysis or pending.

Qualitatively, it was observed that the AI demonstrated good potential for use, understanding the taxonomic characteristics and logical process of the methodology, corroborating the findings of Ziakkas and Pechlivanis (2023). However, some points of attention were identified. During the analysis via ChatGPT, there was a limitation in the scope, which was more restricted than desired. In the first attempt, the AI limited its response to maintenance technicians and related procedures. When the interaction was expanded to include the participation of the airline's quality control and the civil aviation authority, the AI provided a control structure more aligned with the one generated by the authors of this article.

Nevertheless, the AI did not consider the role of air traffic control as relevant and excluded this actor from the analysis, although it positioned the other actors in a hierarchically coherent manner. In the third approach, the AI generated a control structure in which air traffic control was subordinated to the crew, resulting in a hierarchical error. Although it is possible to generate control structures closer to the idealized one, it was observed that the aspects of control and interaction between systems still require evaluation and human supervision for validation.

A relevant example in the investigation was the divergence regarding the part number presented in the model generated by the AI compared to the actual information in the AAIB official report.

The contribution of STAMP/CAST, using Rasmussen's (1997) abstraction concepts, allowed for an acceptable mapping of the actors. However, the approach by Mello et al. (2007) suggests a broader historical context that could have contributed to the event. They highlight the period in England, under the influence of strong economic liberalism and changes in civil aviation regulatory guidelines, coinciding with the privatization of airlines and the relaxation of oversight. This aspect could have been included in the official report, but it is understandable that the AAIB investigation (1992) focused on hierarchical levels below the central government.

Regarding control failures and recommendations, the analysis generated by Leveson's methodology (2019) was similar to the official report, emphasizing the inadequacy of internal and external oversight systems, as well as poor management of fatigue and work overload.

Considering that aviation accidents often involve significant human factors (Youseftorkaman et al., 2023), the combination of methodologies for performance and cognition evaluation represents a valuable strategy for a more comprehensive understanding of events.

## 4.2. *Accimap*

The results of the investigation using the Accimap provided feasible textual descriptions of the event. It is significant to discuss that previous inputs had already been made for the analysis through STAMP/CAST, which may have helped reduce system hallucinations. It cannot be ruled out that different results could have been obtained if originating from other commands or in different orders than those used in this work.

Notably, the responses to questions 1 and 2 did not require corrective actions, presenting credible results. The systemic correlations in top-down and bottom-up flow are presented below:

- *"Top-Down (Descending):*

*Economic policies, weak regulations, and organizational decisions converged to create an environment tolerating or neglecting inadequate maintenance practices. This reveals a "disconnect" between the safety expectations set at higher levels and the operational reality at lower levels.*

- *Bottom-Up (Ascending):*

*The incident highlighted systemic and structural vulnerabilities requiring organizational, regulatory, and policy changes to mitigate future risks. Operational failures fed back up the hierarchical chain, exposing the limitations of efficiency-driven policies at the expense of safety."*

However, there was considerable difficulty in the graphical aspect of demonstrating the event, especially in the positioning of the actors and connections that refer to the top-down and bottom-up feedbacks.

Figure 2 presents the analysis performed by the AI.

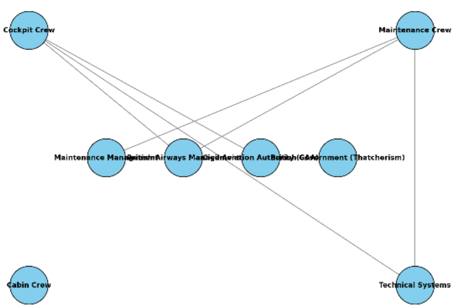
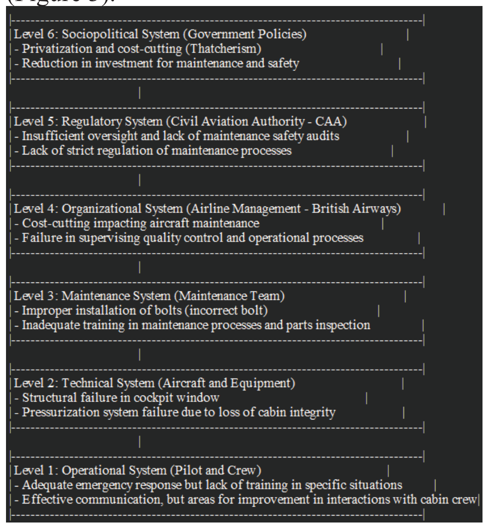



Fig. 2. First Accimap analysis made by GPT

Notably, the diagram presented was insufficient. After some feedback, the AI made structural adjustments to the diagram, but these maintained the superficiality of the analysis (Figure 3).



Level 6: Sociopolitical System (Government Policies)
- Privatization and cost-cutting (Thatcherism)
- Reduction in investment for maintenance and safety

Level 5: Regulatory System (Civil Aviation Authority - CAA)
- Insufficient oversight and lack of maintenance safety audits
- Lack of strict regulation of maintenance processes

Level 4: Organizational System (Airline Management - British Airways)
- Cost-cutting impacting aircraft maintenance
- Failure in supervising quality control and operational processes

Level 3: Maintenance System (Maintenance Team)
- Improper installation of bolts (incorrect bolt)
- Inadequate training in maintenance processes and parts inspection

Level 2: Technical System (Aircraft and Equipment)
- Structural failure in cockpit window
- Pressurization system failure due to loss of cabin integrity

Level 1: Operational System (Pilot and Crew)
- Adequate emergency response but lack of training in specific situations
- Effective communication, but areas for improvement in interactions with cabin crew

Fig. 3. Second attempt to ilustrate Accimap's diagram

The results obtained corroborate with the analysis of Ziakkas and Pechilivanis (2023), who considered Chat GPT insufficient for more complex analyses, such as in the Accimap model.

### 4.3. *AI evaluation*

These findings do not invalidate the use of AI but highlight the need for human supervision in the information-gathering process and emphasize the importance of transparency regarding the system's limitations. This also indicates that appropriate questioning is essential for the analysis (Padovan et al., 2023).

Another important aspect is the need to understand the emotional factors and the influence of human factors in complex sociotechnical systems and their correlation with disastrous events. In this study, there was no evaluation of the AI's capacity to map emergent factors, remaining limited to an objective analysis based on facts already documented. In the case of this accident, with a consolidated investigation and a large data set, the replication of information related to human factors (such as maintenance team fatigue, lighting issues in the maintenance area, work shifts, and task overload) corresponded to what was reported officially. However, given the necessary detail required to map human factors and the biases that human evaluators can introduce into the analysis (Ray et al., 2023), tools specifically designed for this analysis and conducted by human experts remain more effective for examining cognitive aspects and performance.

## 5. Conclusion

The analysis of this work led to the conclusion that the use of artificial intelligence can be useful in the investigative analysis process, especially in taxonomic structuring, logical processes, and textual construction. The graphical aspects were insufficient for both methodologies tested, but the development of the STAMP/CAST control structure required fewer subsequent interventions than the Accimap diagram.

In general, human supervision is crucial for validating the technical information extracted, especially regarding numerical values and quantifications, which may result from the software's hallucinations.

It is reiterated that the accuracy of the analysis provided by artificial intelligence also varies according to the publicity of the data. The event of flight BA 5390 already has various pieces of information on the internet that facilitate the verification of the facts.

Investigative analyses through systemic methods provide a holistic mapping of events, seeking to answer what happened, why it happened, and how the event occurred. However, to understand the reasons behind the decisions made by human components in the system, a deeper evaluation can be achieved by complementing the investigation with specific tools designed for this purpose.

### Acknowledgement

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

### References

Air Accidents Investigation Branch (1992). Aircraft Accident Report 01/92: Report on the accident to BAC One-Eleven, G-BJRT over Didcot, Oxfordshire on 10 June 1990. Royal Aerospace Establishment.

Bharadiya, J. P. (2023). Artificial Intelligence in Transportation Systems: A Critical Review. American Journal of Computing and Engineering 6, no. 1, 35–45.

Deniz, S., Wu, Y., Shi, Y., and Wang, Z. (2024). A Reinforcement Learning Approach to Vehicle Coordination for Structured Advanced Air Mobility. Green Energy and Intelligent Transportation 3, no. 2, 100157.

Hulme, A., Stanton, N. A., Walker, G. H., Waterson, P., and Salmon, P. M. (2021). Are Accident Analysis Methods Fit for Purpose? Testing the Criterion-Referenced Concurrent Validity of AcciMap, STAMP-CAST, and AcciNet. Safety Science 144.

Kabashkin, I., Misnevs, B., and Zervina, O. (2023). Artificial Intelligence in Aviation: New Professionals for New Technologies. Applied Sciences 13.

Leveson, N. (2004). A New Accident Model for Engineering Safer Systems. Safety Science 42, 237–270.

Leveson, N. G. (2019). CAST Handbook: How to Learn More from Incidents and Accidents.

Mayday: Air Disaster. (2021, September). Blow Out. YouTube. Available at: https://www.youtube.com/watch?v=7xfaDr0nhoQ. Accessed June 20, 2024.

Mello e Souza, M., and Nunes, E. (2007). A Aviação Civil nos Estados Unidos: Um Estudo sobre o Papel do Estado na Regulação do Setor Aéreo: Documento de Trabalho nº 71. Observatório Universitário.

Mussone, L., Ferrari, A., and Oneta, M. (1999). An Analysis of Urban Collisions Using an Artificial Intelligence Model. Accident Analysis & Prevention 31, no. 6, 705–718.

Padovan, P. H., Martins, C. M., and Reed, C. (2023). Black is the New Orange: How to Determine AI Liability. Artificial Intelligence and Law 31, no. 1, 133–167.

Qureshi, Z. H. (2008). A Review of Accident Modelling Approaches for Complex Critical Sociotechnical Systems. In Proceedings of the 12th Australian Workshop on Safety Related Programmable Systems (SCS'07), Adelaide, vol. 86, 47–59.

Rasmussen, J. (1997). Risk Management in a Dynamic Society: A Modelling Problem. Safety Science 7, 183–213.

Ray, A. T., Bhat, A. P., White, R. T., Nguyen, V. M., and Fischer, O. J. P. (2023). Examining the

Potential of Generative Language Models for Aviation Safety Analysis: Case Study and Insights Using the Aviation Safety Reporting System (ASRS). Aerospace 10, no. 770.

Stanton, N. A., and Young, M. S. (2003). Giving Ergonomics Away? The Application of Ergonomics Methods by Novices. Applied Ergonomics 34, no. 5, 479–490.

Sufyan, M., Shokat, Z., and Ashfaq, U. A. (2023). Artificial Intelligence in Cancer Diagnosis and Therapy: Current Status and Future Perspective. Computers in Biology and Medicine 165.

Tselentis, D. I., Papadimitriou, E., and van Gelder, P. (2023). The Usefulness of Artificial Intelligence for Safety Assessment of Different Transport Modes. Accident Analysis and Prevention 186.

Viitanen, K., and Reiman, T. (2023). Network Visualisation in Supply Chain Quality and Safety Assurance of a Nuclear Power Plant Construction Project. In Visualising Safety: An Exploration of Drawings, Pictures, Images, Videos and Movies, edited by J.-C. Le Coze and T. Reiman, 83–92. Springer.

You, M., Li, S., Li, D., and Xu, S. (2021). Applications of Artificial Intelligence for Coal Mine Gas Risk Assessment. Safety Science 143.

Youseftorkaman, M., Kangaranifarahani, A., and Dana, D. (2023). Applications of Artificial Intelligence in the Aviation Industry and Air Accidents. Mathematical Statistician and Engineering Application 72, no. 2, 178–192.

Ziakkas, D., and Pechlivanis, K. (2023). Artificial Intelligence Applications in Aviation Accident Classification: A Preliminary Exploratory Study. Decision Analytics Journal 9.