

Addressing the challenges for air safety Investigators:
**Employing Artificial Intelligence to Mitigate Professional Bias in Air Safety
Investigation**

ISASI Rudolph Kapustin Memorial Scholarship 2023

Niall Paul Miranda

Cranfield University
MSc in Safety and Human Factors in Aviation
School of Aerospace, Transport and Manufacturing

Supervisor: Dr Wen-Chin Li

Employing Artificial Intelligence to Mitigate Professional Bias in Air Safety Investigation

Introduction

Perfecting air safety requires perfecting the outcome from investigations. However, striving for perfection undermines the human factors that affect the human investigators tasked with analysing and making decisions in air accident and incident investigations. Challenges in air safety investigations include interpretation of evidence, and deduction of causal factors and recommendations. Interpretation and deduction in various domains of decision making are at times derived by heuristics. They simplify the tasks of evaluating and estimating information but also engender Cognitive Biases (Tversky & Kahneman, 1974). Decisions influenced by cognitive biases often defy logic and rationality and yet are not attributed to motivation or penalization. Despite the merits of experience and training accorded to air safety investigators to provide fair and impartial investigations, they too are obliviously susceptible to Cognitive Biases. At times, limitations such as, thinking capacity or availability of evidence and time, investigators may use the approach of bounded rationality to dispense satisfactory decisions as opposed to following a systematic process to yield best decisions (Gobet et al., 1997). Such scheme of deriving imperfect decisions is a characteristic of indecisiveness which ultimately leads to informational tunnel vision where investigators would seek to prove a chosen set of hypotheses rather than disproving a broader range of hypotheses, thereby falling prey to confirmation bias (Rassin et al., 2008). Apart from confirmation bias, investigators are unconsciously confronted with a string of professional biases that can impact the accuracy of investigations. Notably, investigative conclusions can be significantly clouded by hindsight bias where investigators tend to believe that a consequence was anticipated before its occurrence (Dekker, 2012). Hindsight bias characteristically renders substitution tests ineffective. This prompts investigators to employ regressive reasoning to the plausible causes from the conclusions rather than to disprove plausible hypotheses.

Bias in accident investigation processes

Although the purpose of air safety reports is not to underscore blame or liability but rather to impel air safety in the industry, professional bias attributes in investigators have, on few occasions, led to turmoil whilst inferring the causal factors attributed to air accidents and incidents in their reports. The accident report of Asiana Airlines Flight 214 by the National Transportation Safety Board (NTSB) contains dissenting statements from four members, drawing attention to their perspective of recommendations and causal factors (NTSB, 2014). Whilst the decision to publish all four statements on the report is commendable, the indecisiveness to strike a cordial agreement on recommendations and causal factors signifies attentional tunnelling as a derivative of cognitive bias. Similarly, the fatal crash of Arrow Air 1285R drew split opinions between the nine investigators working for the same safety board. Whilst the findings and conclusions of five members made it to the Canadian Aviation Safety Board's (CASB) primary report (CASB, 1988), the dissenting findings and conclusions of the other four members appeared in a separate document (Bobbitt et al., 1988). The scepticism of the public from the indecisive reports ultimately led to the collapse of the CASB (Watson, 2008). The crash of Singapore Airlines Flight 006 also presented contesting findings. The investigators of the Taiwan Aviation Safety Council (ASC) emphasized on pilot error as the primary cause (ASC, 2002) whilst the investigators of the Singapore Ministry of Transport laid prominence on deficiencies in the airport's infrastructure (Ministry of Transport, 2002). Such instances of bias in aviation safety reports not only undermines the industry's confidence on the authenticity of contending reports but also negatively impacts the reputation of investigative agencies.

Adopting an Artificial Intelligence (AI) solution:

The objective of this solution is to digitize the decision making process in an investigation by employing an AI agent to develop a host of tangible hypotheses using a heterogenous and dynamic database for a given scenario. The only human intervention afterward would be for the investigators to disprove the hypotheses thereby negating the effects of bias in making a hypothesis in the preliminary stage. Currently, no digital investigation, backed by an AI solution, exists to develop hypotheses in air safety investigation domain, however, digital investigation is established in the field of forensics to provide digital analysis of evidence and to identify robust and impartial solutions to deal with criminal or cyber incidents (Casey, 2014). Perhaps the greatest benefit of digital investigation lies in its inability to conduct biased analyses and produce biased decisions. In an ideal world, to negate the effect of human factors, the sole usage of AI and digitization of investigation would be favoured. But in an air safety investigation, it is difficult to understate the contribution of human investigators to analyse and function in a manner that is unique to the logic processes governing AI. Therefore, the scope of digital investigation in air accidents is best limited to developing automated hypotheses as an auxiliary tool rather than substituting human investigators.

Envisioned Architecture of The AI Agent

This solution takes the approach of deep learning (Bisong, 2019) to get the AI agent to classify evidence directly from the source, such as, data from a flight data recorder, sound from a cockpit voice recorder, or images from a crash site, to formulate a paradigm for applying computational logic on ontologies with the database. The database comprises of two domains- one domain consisting of data from past accident and incident reports as well as a dynamic data from flight data monitoring reports whilst the other domain consists of safety assessment and hazard analysis obtained from the original equipment manufacturers and independent safety assessment service providers. The AI engine would search for suitable inferences from the database corresponding to the evidence obtained to produce a knowledge representation (Davis et al., 1993) of a heterogenous digital asset, thereby presenting a host of hypotheses to the air safety investigators. Fig. 1 shows the envisioned architecture of the AI agent.

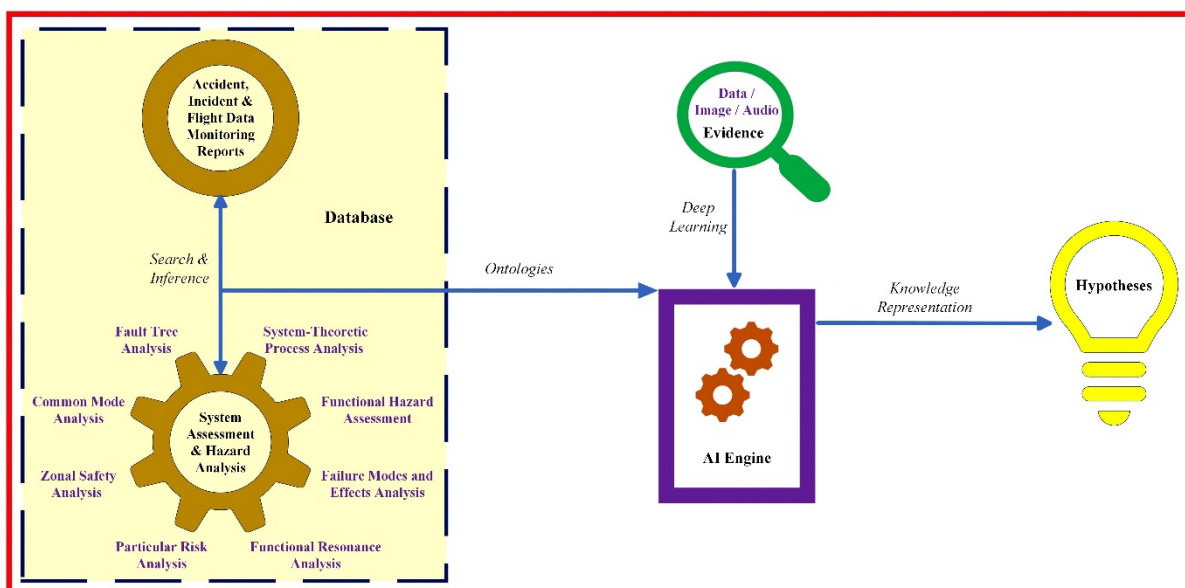


Figure 1. Envisioned architecture of an AI agent developing hypotheses related to causal factors of an accident to facilitate air safety investigators

Advantages of employing an AI agent to develop hypotheses

1. Redundant to cognitive bias

The inability of an AI agent to be influenced by human traits makes it redundant to cognitive bias. The logic governing the agent can only yield an outcome of its design which is limited to analysing and inferring data to generate a host of hypotheses.

2. Accuracy

The reliability of the data in reports improves not just due to the absence of cognitive biases but also the ability of the AI agent to accurately analyse information from an enormous database that far outweighs an investigator's ability to analyse or recollect similar data due to human psychophysiological and memory limitations.

3. Efficient utilization of resources

Time and funds are two crucial resources for an air safety investigation. The ability of a single AI agent, to accurately perform analysis of evidence and generate hypotheses, can save hours of manpower during the initial phases of the investigation. This facilitates more resources for the advance stages of the investigation which can ultimately aid in enhancing the legitimacy of the reports.

4. Encourages open reporting

The database forms the memory of the AI agent, and it is essential to keep it updated dynamically to maintain precision of hypotheses generation. This can only be accomplished by wilful sharing of data by various agencies and operators around the globe. Successful implementation can encourage open reporting and inculcate frequent monitoring.

5. Convenience

Digitization of investigation enables a single AI based agent to perform tasks in multiple investigative agencies. This eliminates the inconvenience to obtain the availability of air safety investigators to explore the evidence or the availability of subject matter experts to analyse the obtained data.

Conclusion and Future Application

Biases among air safety investigators is not limited to just cognitive biases. Conflict of interest can also yield bias in investigators and as argued by Moore et al., (2004), they can occur autonomously, without conscious awareness, just as in the case of cognitive biases. Therefore, it is essential to combat such biases with technology rather than placing solitary reliance on the efforts of behavioural scientists. Apart from negating biases, understanding human investigators and the human factors associated with them is critical to establishing an accurate research platform for conducting investigations.

The challenge for air safety investigators addressed by the AI agent does not limit its functions to providing assistance to the investigators, keeping 'human in the loop', during the initial phase of developing hypotheses. With an established reliability and within a suitable architecture, the collaboration between AI and the investigators can progress to more autonomy, with the investigators at design and oversight phases, thereby keeping 'human in control' of the AI investigation (EASA, 2020). This can benefit the advance phases of investigation such as generating unbiased conclusions and recommendations. Currently, however, the characteristic distinctiveness and intellectual capability of the human investigator cannot be undervalued, particularly, to deal with challenges in AI agent's reliance and compliance. Despite the reliability and accuracy of the AI agent as previously established, inaccuracy and bias in the database can negatively influence the authenticity of the AI agent's inference. Therefore, it is crucial for human investigators to rework contending investigation reports or disregard them from the database for successful implementation of the AI agent.

References

- ASC. (2002). Aircraft Accident Report, Crashed on a Partially Closed Runway during Takeoff, Singapore Airlines Flight 006, Boeing 747-400, 9V-SPK, CKS Airport, Taoyuan, Taiwan. Retrieved March 23, 2023. <https://www.tsb.gov.tw/media/4252/sq006-executive-summary.pdf>
- Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. <https://doi.org/10.1007/978-1-4842-4470-8>
- Bobbitt, N., Filotas, L., Mussallem, D., Stevenson, R. (1998). Arrow Air Inc., Douglas DC-8-63, N950JW, Gander International Airport, Newfoundland. *Dissenting Opinion*. Retrieved April 10, 2023. https://web.archive.org/web/20050404060705/http://www.avsaf.org/reports/CANADA/1985.12.12_ArrowAirInc_MF1285R%28DissentingOpinion%29.pdf
- CASB. (1988). Aviation Occurrence Report, Arrow Air Inc., Douglas DC-8-63, N950JW, Gander International Airport, Newfoundland. Retrieved April 10, 2023. <http://flightopsresearch.org/data/files/arrow1285.pdf>
- Casey, E. (2014). *Digital evidence and computer crime: Forensic science, computers, and the internet* (3rd ed.). Elsevier Science.
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What Is a Knowledge Representation?. *AI Magazine*, 14(1), 17-33. Retrieved March 23, 2023. <https://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>
- Dekker, S. (2012). *Just Culture: Balancing Safety and Accountability* (2nd ed.). CRC Press. <https://doi.org/10.4324/9781315251271>
- EASA. (2020). EASA Artificial Intelligence Roadmap, A human-centric approach to AI in aviation. Version 1.0. Retrieved April 12, 2023. <https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-10-published>
- Gobet, F., Richman, H., Staszewski, J., & Simon, H. A. (1997). Goals, Representations, and Strategies in a Concept Attainment Task: the EPAM Model. In *Psychology of Learning and Motivation* (pp. 265–290). Elsevier. [https://doi.org/10.1016/s0079-7421\(08\)60504-6](https://doi.org/10.1016/s0079-7421(08)60504-6)
- Ministry of Transport. (2002, April 26). Singapore MOT's Comments To The Final Report Of The Investigation Into The SQ006 Accident. National Archives of Singapore. Retrieved April 10, 2023. <https://www.nas.gov.sg/archivesonline/data/pdfdoc/2002042607.htm>
- Moore, D.A., Loewenstein, G. (2004). Self-Interest, Automaticity, and the Psychology of Conflict of Interest. *Social Justice Research* 17, 189–202. <https://doi.org/10.1023/B:SORE.0000027409.88372.b4>
- NTSB. (2014). Aircraft Accident Report, Descent Below Visual Glidepath and Impact With Seawall Asiana Airlines Flight 214 Boeing 777-200ER, HL7742, San Francisco, California. Retrieved April 10, 2023. <https://www.nts.gov/investigations/accidentreports/reports/aar1401.pdf>
- Rassin, E., Muris, P., Booster, E., & Kolsloot, I. (2008). Indecisiveness and informational tunnel vision. *Personality and Individual Differences*, 45(1), 96–102. <https://doi.org/https://doi.org/10.1016/j.paid.2008.03.006>
- Tversky, A., & Kahneman, D. (1974). Judgment under Uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131. <http://www.jstor.org/stable/1738360>

Watson, B. (2008, August). The Transportation Safety Board Taking centre stage to advance aviation safety. Transportation Safety Board of Canada. Retrieved April 10, 2023.
https://web.archive.org/web/20100925140925/http://www.tsb.gc.ca/eng/medias-media/articles/aviation/2008/ht_20080701.asp