Deliverable 8.4

Safety Culture Safeguards for Aviation Intelligent Assistant Systems

Author: Barry Kirwan (ECTL-DBL)

Abstract:

This report develops two provisional sets of safeguards to prevent the erosion of safety culture in future Human-AI Teaming (HAT) systems and their parent organisations. This is achieved firstly by considering the likely impacts – positive and negative – of Intelligent Agents (IAs) in aviation systems, based on a review of a validated safety culture questionnaire used with air traffic controllers and pilots. This speculative analysis by three aviation safety culture experts identified six safety culture concerns and six safety culture 'affordances' that could enhance safety culture.

The second approach involved developing and testing a safety culture monitoring system, via a reduced question set curated from the established and validated safety culture questionnaire. This shorter, AI-focused questionnaire was tested via participants from three HAIKU use cases (UC1, 2 and 4), with licensed pilots and controllers who had used the respective Use Case IAs in realistic real-time simulation exercises.

Based on these two studies, two sets of higher level safeguards are proposed, one for the HAT development process, the other for aviation organisations who aim to implement HAT in their operational systems. These safeguards aim to maintain and even augment the level of safety culture in future HAT environments, leading to high overall safety and resilience in future AI-assisted flight operations.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Information Table

Deliverable Number	8.4
Deliverable Title	Safety Culture Safeguards for Aviation Intelligent Assistant Systems
Version	1.0
Status	Final
Responsible Partner	ECTL-DBL
Contributors	DBL
Contractual Date of Delivery	August 31st, 2025
Actual Date of Delivery	August 6th, 2025
Dissemination Level	Public

Document History

Version	Date	Status	Author	Description
0.1	Dec 30th, 2024	First Draft	Barry Kirwan (ECTL)	First Draft for Review
0.2	March 2025	Second Draft	Barry Kirwan (ECTL)	Second draft, addressing DBL comments
0.3	April 2025	Third draft	Barry Kirwan (ECTL)	Addition of updated UC4 responses
0.4	June 2025	Fourth Draft	Barry Kirwan (ECTL)	Review by Vanessa Arrigoni with comments and integrations (DBL)
0.5	July 31st, 2025	Fifth Version	Barry Kirwan (ECTL)	Final version for quality check
1.0	August 6th, 2025	Final version	Vaness Arrigoni (DBL)	Final quality check and submission

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Table of Contents

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

List of Acronyms

Acronym	Definition
AI	Artificial Intelligence
ANSP	Air Navigation Service Provider
ATCO	Air Traffic Controller
ATM	Air Traffic Management
CRM	Crew Resource Management
EASA	European Union Aviation Safety Agency
HAT	Human-AI Teaming (also known as Human Autonomy Teaming)
IA	Intelligent Assistant (also known as Digital Assistant)
IFATCA	International Federation of Air Traffic Controllers' Associations
LLMs	Large Language Models
ML	Machine Learning
NLP	Natural Language Processing
OpXAI	Operational Explainability
SCQ	Safety Culture Questionnaire
SMS	Safety Management System
SOPs	Standard Operating Procedures
TEM	Threat & Error Management
TRL	Technology Readiness Level
TRM	Team Resource Management
UAM	Urban Air Mobility
UC	Use Case

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Executive Summary

Artificial Intelligence (AI) is already impacting aviation via Machine Learning (ML) applications in air traffic control (EC, 2022), but much more is envisaged, including AI supporting future (>2030) single pilot operations. At the moment, ML can be considered 'just more automation' (Kaliardos, 2023), and as such does not detract from the human pilot's or air traffic controllers (ATCO's) duty and responsibility for the safety of both crew and passengers. Today, flight crew (including cabin crew) and ATCOs keep flights safe, especially when things go wrong. Flight crew are literally at the front of the plane, and controllers know they have hundreds of lives in their hands as they direct dozens of aircraft every hour through sometimes congested airspace and busy airports.

In the European air traffic network, safety culture is measured regularly in different countries using a standardised 50-item scientifically validated questionnaire developed and coordinated by EUROCONTROL, based on eight safety culture dimensions. The questionnaire, with minor modifications, has also been applied to airlines and several airports, and is generally seen as valid for aviation. This questionnaire, together with a newly developed AI-specific version, is the focus of this report.

In the first step, the full questionnaire was applied by three aviation safety culture experts to the Intelligent Assistant (IA) concept to see which facets of safety culture might be affected by future Human-AI Teaming (HAT) scenarios. The results identified six high-level concerns, but also six instances where the Intelligent Assistant could potentially reinforce or improve safety culture, providing new 'safety affordances'.

In the second step, following this preliminary speculative analysis, the questionnaire was refined to deliver seventeen more focused questions aimed specifically on Human-AI Teaming. This reduced-set questionnaire was then tested with participants from three HAIKU use cases where pilots and controllers had the opportunity to interact with prototype IAs: UC1 (Startle Response – 12 pilots), UC2 (Flight Re-Routing – 9 pilots) and UC4 (Tower Control – 8 ATCOs). The results gave insight into controller and pilot reactions to, and perceptions of, the impact of AI on their safety culture, following their exposure to prototype Intelligent Assistants in realistic simulations. The results confirmed the viability of the questionnaire and show significant agreement between ATCOs and pilots on both desirable and undesirable attributes of future Human-AI Teaming (HAT) system characteristics and arrangements. These results have led to the identification of provisional safeguards for the design and development of future HAT systems, including the human-AI interaction.

The third step involved returning to the potential benefits and threats of AI in aviation raised in the speculative questionnaire review, and the derivation of a set of organisational guard-rails to help shore up and even advance safety culture in operational aviation organisations. This guidance set is stratified into four layers of 'risk owners', corresponding to the operational, safety management, middle management and corporate leadership structures of typical operational aviation organisations.

These two sets of safeguards (design-focused and organisation-focused) are necessarily provisional, as there is a general lack of experience in IA application in aviation, and so no benefit of 'hindsight'. However, they can already be adopted/adapted and followed by system developers and product design teams, as well as aviation organisations seeking to integrate AI-based systems into their operations, until better ones developed as more experience accumulates. In addition, the safety culture questionnaire plus additional interviews/focus groups represents a resources-light and agile way to 'take the pulse' of safety culture and AI impacts, serving as a check to see if there is any drift or decrement in safety culture as a function of material steps towards AI integration into aviation operations.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

1 Safety Culture in Future Aviation Human-AI Teaming Scenarios

In aviation today, safety in terms of accident rates is seen as better than the other three transport modes (rail, sea and road). This is in part attributed to the level of safety culture in aviation, hard won due to various spates of accidents during the early years (1950 onwards) of commercial aviation, and following the introduction of 'fly-by-wire' cockpits in the early 1980s. Since commercial aviation accidents in flight are almost always fatal, there is both strong public awareness of air crashes (they make headlines) and appropriately strong regulation on safety across the aviation industry spectrum. Hence, safety culture (which can be summarised as '*the way we do safety around here*') has generally been strong in aviation, whether in aircraft manufacturing, air traffic control, airport operations or in the airlines themselves.

Safety culture is a property of the people working in an industrial system, and while this is largely true (without people there is no safety culture, or indeed culture), in practice it is an emergent property of the system, including the technology. Artificial Intelligence (AI) is certainly going to be a feature of the future technology used by aviation personnel. Currently, it is common in culture discussions to talk of *artefacts* (i.e. tools) that people use, with current aviation examples being flight strips or track data blocks (air traffic), various cockpit displays, alarms and electronic flight bags (airlines), weekly safety reports and safety messaging via airport community 'Apps' (airports), and drone control interfaces (drones and urban mobility), etc.

AI in the form of Machine Learning (ML) can provide information that can be considered as an artefact, or feeding into an artefact, and in most practical respects can be simply considered as additional automation (Kalliardos, 2022). But future (yet to be realised) Intelligent Assistants (IAs), anticipated to be operational in the 2030-2040 timeframe, can and will go further. Current AI support doesn't 'answer back', whereas there will be dialogue with future IAs, elevating the role of AI from information support to team player and in some cases, team supervisor. This means the future IA must not only make sense of their task, but it must also to a degree make sense of human culture, so it knows how to behave with humans, to be effective.

A useful definitions of safety culture is the following (ACSNI, 1993):

> "*The safety culture of an organization is the product of individual and group values, attitudes perceptions, competencies, and patterns of behaviour that determine the commitment to and the status and proficiency of organization's health and safety management.*"

Safety culture has been shown to be a key predictor of safety performance in several industries (including nuclear, chemical, oil & gas, and rail). It refers to the attitudes of personnel to the company's approach to safety, their perceptions about the magnitude of the risks they face, and their beliefs about how to control those risks. Importantly, it affects what they say and do (and don't do) about safety during their daily work. Safety culture concerns everyone. The most obvious critical candidates are the front-line people such as air traffic controllers, pilots and ground handling staff. But everyone contributes to the state of safety in an organisation, including CEOs, senior and middle managers, people in support functions, cabin crew, flight ops, drivers on the apron, fuellers, ramp operators and baggage handlers, caterers and cleaners, security and police, etc. At an airport, almost all operations are interconnected, involving multiple partners from different organisations, so it is even more important that all these companies and their staff have a positive safety mindset. The bottom line is that people make safety, so no one should be excluded.

The origins of Safety Culture are usually traced back to the Chernobyl nuclear power plant accident in 1986. Just as the Three Mile Island nuclear power plant accident in 1979 demonstrated unequivocally the importance of Human Factors in the design of human-machine interfaces in high systems, Chernobyl showed that the prevailing operational culture could catastrophically defeat established safety procedures and processes. It appeared that if, as Management legend Peter Drucker famously said, "culture eats strategy for breakfast," then unfortunately it could eat safety for lunch, too. A spate of high-profile accidents from different domains (see Figure 1), including space, oil and gas, and rail – all seen as 'safety culture accidents' – only served to emphasise the enduring importance of this newly identified organisational trait. Certain high profile public enquiries into key accidents such as the Piper Alpha disaster (Cullen, 1990) and Clapham Junction rail crash (Hidden 1989), and key safety thought leaders at the time (e.g. Turner and Pidgeon, 1997; Reason, 1997), and a number of accidents at least partly attributed to safety culture ever since, have ensured that safety culture has endured as a critical and non-negotiable attribute for any high risk organisation.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

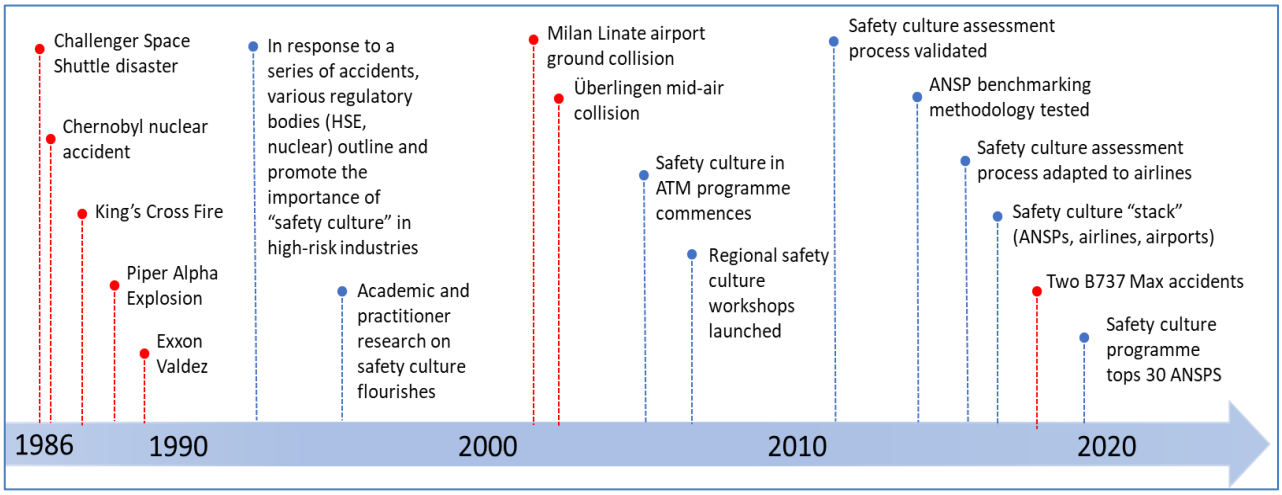


Figure 1: Safety Culture Timeline for European Aviation (including key US accidents)

Figure 1 highlights the fact that safety culture wasn't initially seen as being of too much concern for aviation. This was despite the Kegworth air crash in 1989 (AAIB, 1990), which had certain safety culture aspects, and growing concern over Civil Aviation Authorities (CAAs) having dual roles of protecting the flying public but also protecting the industry (most CAAs have now become principally regulators for safety). The thought at the time was that the strong training and design in the cockpit and air traffic operations rooms, as well as Safety Management Systems (SMS) and Standard Operating Procedures (SOPs) were sufficient. In Europe, this notion was shattered with the mid-air collision over Lake Constance in Uberlingen in 2002, following shortly after the Milan Linate runway collision a year earlier. As Chernobyl did for nuclear, these accidents triggered a rethinking that SMS and SOPs were not enough, and that safety culture was crucial. Since the two accidents were primarily related to air traffic control (ATC), the safety culture assurance 'legwork' was carried out in that sector of the industry, with almost all European Air Navigation Service Providers (ANSPs) engaging in one of several independent safety culture evaluations of their organisation. This led later to safety culture spreading to some airlines and airports, though not nearly as systematically as the ongoing EUROCONTROL-led European ANSP programme on safety culture.

Management and Designers – 'Upstream' Drivers of Safety Culture?

Just as the focus in aviation safety occurrence investigation has generally shifted its focus from 'human error' to 'systemic' failure in recent years, the concept of safety culture must also consider the critical importance of management action regarding safety, based on management's collective values, beliefs and decisions, as well as the safe behaviour and attitudes of operational personnel, engineering and other support staff. As the wide-ranging investigation into the two Boeing 737 Max accidents has shown (Zwiefel and Vyal, 2021; Dias et al, 2020), even with the best engineering and a strong track record in safety performance, a compromised organisational safety culture can precipitate disaster. And whilst HAIKU is generally focused on Human-AI Teaming at the 'sharp end', i.e. operational contexts, the safety culture of the designers and developers of AI systems – the manufacturers – is likely to be an important consideration. This will be returned to in more depth in section 4.

Measuring Safety Culture

Since the Uberlingen mid-air collision in 2002 (Nunes and Laursen, 2004), there has been a focus on safety culture in the air traffic industry. EUROCONTROL, originally in partnership with the University of Aberdeen and more recently the London School of Economics (LSE), developed a validated survey approach (Reader et al, 2015) that has since been applied in over thirty countries in Europe, with a number of Air Navigation Service Providers (ANSPs) carrying out surveys every 3-5 years. These ANSPs have found the process useful in understanding their complete risk picture, including where their safety culture strengths and weaknesses lie and how to improve them (Kirwan et al, 2021). Since 2016, this successful approach has been extended to deal with other sectors of aviation including airlines and airports (e.g. Kirwan et al, 2019). The EUROCONTROL Safety Culture questionnaire is based around eight safety culture 'dimensions':

1. Management Commitment to Safety
2. Collaboration & Involvement
3. Just Culture & Reporting
4. Communication & Learning

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

5. Colleague Commitment to Safety
6. Risk Handling
7. Staff and Equipment
8. Procedures & Training

Each of these dimensions is supported via a set of carefully worded questions, tailored to different segments of the aviation industry (ANSP, airline, airport, and airframe manufacturer), e.g. '*My colleagues are committed to safety*,' and '*If I saw unsafe behaviour by one of my colleagues, I would talk to them about it*.' The answers to these questions build a picture of the safety culture of an organisation, including an understanding of how it may differ in an organisation's various sub-cultures.

Safety Culture & the Intelligent Assistant

Safety culture surveys in Europe have generally shown a high degree of safety culture in both controllers (Kirwan and Shorrock, 2015) and pilots (Reader et al, 2016). But if AI is acting as a second pilot in the cockpit, and if AI begins carrying out more operational functions in the aerodrome control tower and air traffic control centres, how might this affect safety culture? Could professionals become more complacent about safety, trusting AI-based systems to do the 'heavy lifting'? Could operational staff find themselves too 'out of the loop' to take over when things go wrong and are beyond the AI's capabilities? Or could AI in some ways enhance safety culture, by both broadening and sharpening situation awareness, or by positively influencing crew resource management, even speaking up for safety when required? Safety culture does not usually change quickly, but it is unwise to adopt a 'wait and see' attitude with regard to the potential impacts of AI on safety culture in future AI-assisted aviation scenarios. What is needed is an approach to test the early indications of how AI might influence safety culture, along with a set of provisional safeguards or 'guard rails' to ensure that positive safety culture is 'designed in' and then maintained and monitored during the aviation AI-support development and integration processes.

Given that there is as yet no experience of the impacts of AI-based intelligent agents (IAs) on safety culture in aviation (or in any other industry, for that matter), it makes sense to leverage our existing knowledge on aviation safety culture, based on the validated EUROCONTROL questionnaire survey approach. To this end, each of the 50 questions from the EUROCONTROL questionnaire was considered in the context of a future Intelligent Assistant, e.g. commitment to safety could reduce if the IA was judged to be handling safety, and this could affect both human operators at the 'sharp end' as well as managers running an organisation (see Appendix A). This set was Initially reduced to a set of eighteen potential concerns and benefits, or affordances (see Appendix B). Each was further analysed by three aviation safety culture experts (see also Kirwan, 2024), and amalgamated into higher level constructs, leading to six safety culture concerns or threats (which could reduce safety culture), and six 'safety affordances' (which could enhance existing safety culture). These are defined below.

Six Safety Culture concerns

i. People make safety, don't they?

In aviation today, it is often said that 'people make safety', whether in the cockpit, the air traffic Ops room or tower, or on the ground at the airport. But if Intelligent Assistants become (in)valuable safety assets, might this change? Currently in aviation safety culture (and other domains), the notion that safety is everyone's responsibility has replaced the decades-ago refrain that 'safety is only the job of the safety manager'. Could the reliance on human commitment to safety be replaced by strong safe AI performance characteristics? Is this a trade we want to make (because there may be no coming back from it)?

ii. Who's flying the plane?

At what point would passengers and businesses be happy to have an AI fly a plane or sky-taxi? The comparison with self-driving cars or automated train systems is not the best comparison, due to the vastly superior number of people at risk in air crashes, and the fact that there is no safe 'default' mode in an aircraft once in flight (it cannot simply stop or apply the brakes). One major 'AI-induced' air crash might delay AI-controlled flight for years. Probably AI's controlling role in other sectors outside aviation will weigh on such a decision (e.g. in self-driving cars, trucks, and trains), as well as taking a cautious stepwise approach towards full automation, e.g. AI plus single pilot during the cruise phase of flight (while one pilot is resting); single pilot plus AI for cargo operations (no additional pilot); AI plus single pilot for short-haul commercial flights; AI-piloted sky-taxis; etc.

iii. Who's to blame?

It is unlikely that any AI manufacturer will guarantee that its Intelligent Assistants will never make a mistake. It would probably be unreasonable to do so, as any AI relies on data which may be incomplete or even skewed, and so unable to anticipate every eventuality that could come to pass. Nevertheless, humans working with Intelligent Assistants may fear a double-bind: if they fail to take safety critical advice from an AI that turns out to be correct

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

and there is an accident, or if they take such advice and it turns out to be flawed (and there is an accident), how will they be judged? Judicial bodies have a mission that 'justice must be served,' and relatives of victims inevitably want someone to blame, someone to sue, someone to pay. This dilemma is likely to apply to many domains, and will be at the forefront of professional unions' concerns about the introduction of AIs into human teams with safety-critical roles.

iv. Digital Cop

The AI should be able to advise, but also point out if an unsafe course of action is being taken by its human 'colleagues'. Depending how this is done, and how frequently it happens, resentment could occur between the human and Intelligent Assistant, as the human may feel (s)he is being 'policed'. A related aspect is that the DA could be used to check on the human's competence. This could be advantageous if it is done confidentially to the individual human user, but could also be disadvantageous if it reports to management on the human's performance, in which case it could be seen as a 'snitch.'

v. Turn me up, turn me down

Intelligent Assistants may not be solely concerned with safety; there may be other 'drivers' impacting their decision-making, such as environmental considerations, (e.g. carbon footprint, noise near airports, etc.), but also performance considerations related to efficiency (on-time-performance) and even competitiveness against other airlines. Today, people in aviation already face these sometimes-competing priorities, and make the necessary trade-offs. However, when safety comes to the fore, they will put such factors to one side and 'put safety first', in order to minimise the risk of an accident or its consequences. Humans can make this judgement call, based on experience and their instincts and values. It is not clear whether an AI could do the same (we do not really know how humans do it). The problem is that the human receiving AI advice may not be able to unpack the optimising balancing act that underpins the AI's advice. It also means that the AI's own weighting of competing demands could be set by its 'owners', and with respect to the priority of safety, turned up or down.

vi. Fewer People, More AI

If AI is automation by another name, then it has to be recognised that automation inevitably leads to fewer human workers, e.g. single pilot operations after 2030, and perhaps fewer air traffic controllers as well as less ground handling staff as more automation and robotics supports ground activities. In the post-COVID world wherein several aviation sectors are finding it hard to attract staff, whether airport ground staff or pilots, this may seem less of a problem than before. And yet aviation has always attracted people who are passionate about their jobs, and this has no doubt supported a strong safety culture. If the job becomes less challenging, with fewer people, how will this affect overall motivation and commitment to safe work?

Six Safety Culture Affordances

i. Don't panic…

One of the prime intended uses is to help pilots in what are known as flight upset conditions, in which something goes drastically wrong, and the flight crew have very little time to diagnose the event and recover the aircraft. Accidents such as AF 447 and the two B737 Max accidents fall into this category. Whilst it can take precious minutes for flight crew to work out what is happening and how to correct it, a Intelligent Assistant plugged in to all the sensors, with a vast database of conditions and recovery actions, could arguably make such a diagnosis in seconds. This use of AI (Intelligent Assistant in the cockpit) is therefore a priority in aviation safety research.

ii. The Living Black Box

All commercial aircraft carry two black box recorders, recording all control inputs and voice recordings of flight crew prior to an accident. Investigators nevertheless are not always clear from this information as to why pilots chose a particular course of action over another, and accident investigations can take many months piecing together what really happened. A Intelligent Assistant, particularly one that is interacting with a human crew, could in theory record in far richer detail the reasoning and decision-making during an event (whether recovered or not) than is possible today.

iii. The Dispassionate Oracle

Today there are various aviation risk models and 'top 5' or 'top 10' hitlists for safety area improvement. These are always a mixture of operational data (from accidents and incidents or near misses) and subjective judgements from experienced safety managers or directors. Such data are always 'lagging', since they relate to the past few months or even years and depend on detailed and painstaking analysis that can itself take months or years. Furthermore, such safety management tools can be relatively 'blunt' as they may miss weak signals, or common underlying factors that could make a larger safety difference if they were improved. An AI that sifted through all the data, including live operational data, and 'knew' all the risk models and could do the 'safety calcs' (including

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

use of Bayesian statistics to avoid certain human judgemental biases), could in theory determine a better way forward for safety enhancement and accident prevention. It could help aviation organisations see around the corner and take action 'upstream' before incidents turned into accidents.

iv. Mr Know It All

A Intelligent Assistant can be a vast source of knowledge for aviation flight crew, controllers, and ground staff, on procedures and hazards, and if networked into live operations and multiple teams, can give live, up to date advice related e.g. to weather, problems other teams might be encountering etc. Such a DA could also store daily NOTAMs (Notices to Airmen) and other briefings, and transmit them to the teams, and ensure that they don't miss any key points during subsequent operations. Such a DA could be seen as a valuable repository by human teams.

v. Looking after the little guy

Teamwork can have adverse effects, such that someone's advice is over-ruled, even though later it might turn out that such advice should have been listened to. A number of aviation accidents which manifested this problem led to the development of Crew Resource Management (CRM) training (and its Team Resource Management equivalent in air traffic). The DA could in effect be integrated into CRM practices, ensuring that all voices are heard and listened to, before a decision occurs (depending on time constraints) in the scenario. The DA could help flatten 'authority gradients' in the cockpit, and ensure that the best (safest) decision is taken.

vi. Reducing the work as done vs. work as imagined divide

In most safety culture surveys, some people complain that what they need to do in practice to get the job done often disagrees with what it states in the formal job procedures. Such procedures, they will say, have been written by people sitting in offices who have never done the job, or haven't done it for a long time. This leads to a gap between 'work as done' and 'work as imagined'. An IA could dispassionately observe how the work is done in real operational contexts, and determine if it is less or more safe compared to the official procedures (including whether they are practicable in real working conditions). This information could be used to determine where existing procedures should be maintained, and where they need updating.

This analysis of the EUROCONTROL questionnaire has identified the facets of safety culture likely to be affected – positively and negatively – by the introduction of IAs into aviation operations. These insights give us useful foresight to inform safety culture safeguards (see section 4 onwards). However, the analysis remains 'speculative', as it was carried out early in the HAIKU project timeline (see Figure 2), before any IAs had been developed for the various HAIKU use cases. To gain more specific and contextual insights, from pilots and controllers themselves rather than safety culture experts, a different approach is required. What is needed is to see how pilots and air traffic controllers (ATCOs) think and feel about safety culture after being exposed to realistic Human-AI Teaming (HAT) simulations. This approach is indicated in the upper part of Figure 2 ('Ask the People') and became possible once three of the HAIKU use cases (UC1, UC2 and UC4) reached sufficient maturity to run realistic simulations with licensed controllers and pilots. This required the development of a new questionnaire approach that could be answered by HAIKU user participants from these simulations. This process is described in the next section.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

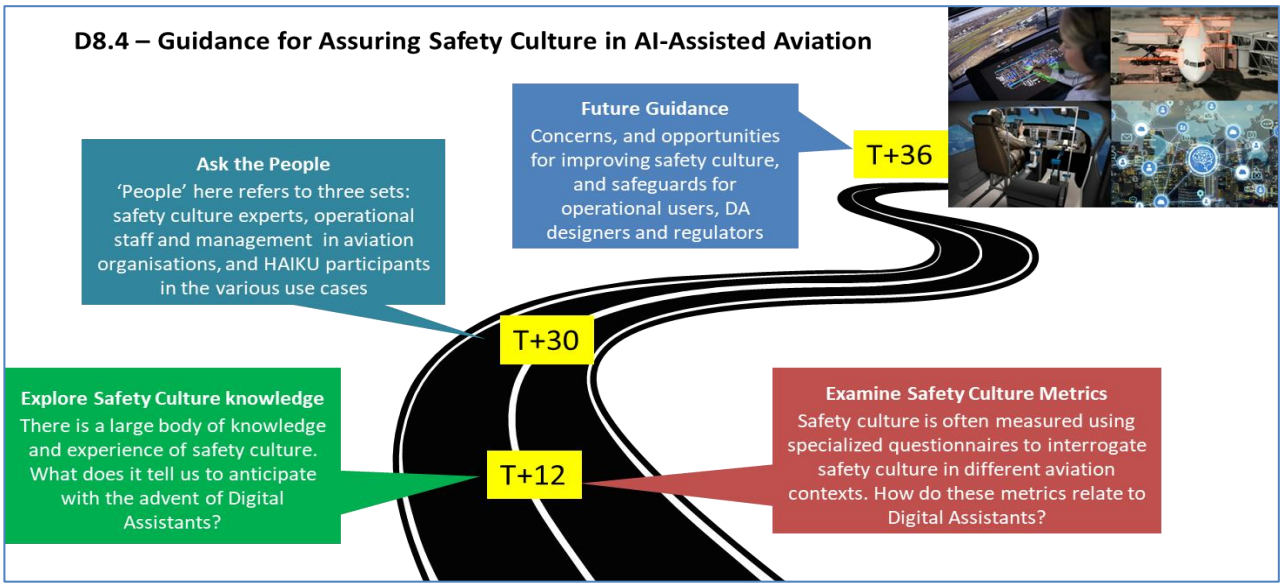


Figure 2: Roadmap for considering the impacts of Intelligent Assistants on safety culture in aviation

During the period between the first speculative analysis of the questionnaire (Months 0-6) and the application of a new questionnaire to Use Case simulation participants (Months 24-32), the scientific literature was reviewed and used to inform future safety culture guidance and guardrails (see Kirwan 2024, 2025), as also indicated in Figure 2.

The remainder of this deliverable therefore charts the development of the new safety culture questionnaire (Section 2), the application of the questionnaire to the two use cases (Section 3) along with the insights gained from controllers and pilots, and the resulting guidance on Safeguards for Safety Culture (Section 4).

2 A Safety Culture Questionnaire for Human-AI Teaming Systems

In order to relate pilot and controller experience of IA usage with potential impacts on their safety culture, a specialised safety culture tool is desirable. This tool was developed by considering which of the questions from the EUROCONTROL questionnaire could be used to gauge future safety culture with HAT systems. A shorter set was desirable, one that was adapted to HAT and could be applied after simulations such as those foreseen in HAIKU, but one that still represented the core safety culture dimensions in the larger validated question set. This analysis (including review by two additional aviation safety culture survey experts) resulted in the distillation of a set of 17 questions, shown in Table 1.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Table 1 - Safety Culture Questions for Human-AI Teaming

No.	Original Questionnaire Item	Dimension	IA Impact	New Question Wording	Comment on Human-AI Teaming safety culture relevance
1.	B01 My colleagues are committed to safety.	Colleague commitment to safety	The IA would effectively be a digital colleague. The IA's commitment to safety would likely be judged according to the IA's performance. Human-supervised training, using domain experts with the IA would help engender trust. The concern is that humans might 'delegate' some of their responsibility to the IA. A key issue here is to what extent the IA sticks rigidly to 'golden rules' such as aircraft separation minima (5NM lateral separation and 1000 feet vertical separation) or is slightly flexible about them as controllers may (albeit rarely) need to be. The designer needs to decide whether to 'hard code' some of these rules or allow a little leeway (within limits); this determines whether the IA behaves like 'one of the guys' or never, ever breaks rules.	**I feel safer working with the AI switched on.**	This is an over-arching statement for Human-AI Teaming. If the answer is 'no', it should be a deal-breaker for operating with the AI in a safety critical work environment.
2.	B04 Everyone I work with in this organization feels that safety is their personal responsibility.	Colleague commitment to safety	Since an IA cannot effectively take responsibility, someone else may be held accountable for an IA's 'actions.' If a supervisor fails to see an IA's 'mistake', who will be blamed? HAIKU use cases may shed light on this, if there can be scenarios where the IA gives 'poor' or incorrect advice. If an IA is fully autonomous, this may affect the human team's collective sense of responsibility, since in effect they can no longer be held responsible.	**The AI system is accountable for its own mistakes.**	The 'AI system' here can mean the developer or those managing the AI day-to-day, all the way up to the Chief Technical Officer (CTO). What it means is that the frontline operator is **not** responsible for the AI's mistakes.
3.	B07 I have confidence in the people that I	Colleague commitment to safety	As for B01, this will be judged according to performance. Simulator training with IAs should	**Relying on the AI erodes my safety skills.**	This is a little different to the original B07 intent but needs to be covered as skills loss and complacency need to

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

	interact with in my normal working situation.		help pilots and others 'calibrate' their confidence in the IA. This may overlap significantly with B01.		be captured. 'Safety awareness' was also considered instead of 'safety skills', but people may be less aware of what they are not aware of. B07 is also captured in the new B01.
4.	B02 Voicing concerns about safety is encouraged.	Just culture and reporting	The IA could 'speak up' if a key safety concern is not being discussed or has been missed. This could be integrated into Crew Resource Management (CRM) and Threat and Error Management (TEM) practices, and CRM's corollary, Team Resources Management (TRM) in air traffic management. However, then the IA may be considered a 'snitch', a tool of management to check up on staff. This could also be a two-way street, so that the crew could report on the IA's performance.	**Working with the AI encourages speaking up for safety.**	The impact of the AI on speaking up is a double-edged sword – it could increase or decrease its likelihood, so needs to be captured. Possibly some will see it as having neutral/no impact.
5.	B14 If I see an unsafe behaviour by a colleague I would talk to them about it.	Just culture and reporting	[See also B02] The IA can 'query' behaviour or decisions that may be unsafe. Rather than 'policing' the human team, the IA could possibly bring the risk to the human's attention more sensitively, as a query. However, this would have to be very advanced AI. More likely, the human notices aberrant behaviour by the IA, and should report it so that it can be analysed offline and fixed. Otherwise, the IA will develop 'bad habits'.	**If I see the AI is making an error, I will report it.**	This is a core requirement of the EU AI Act (European Parliament, 2023), and necessary for maintaining a safe AI, especially if it is a continuously learning AI. Note that formal reporting of errors is different to 'talking to my manager' as in B16 below.
6.	B16 I would speak to my manager if I had safety concerns about the way that we work.	Just culture and reporting	If the IA is behaving systematically in conflict with how human crews expect it to behave, possibly due to bias in its programming, the human crews need to raise this with management rather than let it continue, which would ultimately affect their own behaviour and values / standards.	**If I have concerns about the AI I will speak to my manager.**	This is larger than noting an 'error' by the AI, and can refer to issues about AI bias for productivity rather than safety, for example, or even sticking too rigidly to the rules when it is not the safest course of action.
7.	C06 I am satisfied with the level of confidentiality of the reporting and	Just culture and reporting	As for B16, the use of IA recordings as information or even evidence during investigations needs to be considered. Just Culture policies will need to adapt/evolve to the use of IAs in operational contexts. Again, otherwise the IA could be	**I would have concerns over the confidentiality of incident reporting when working with the IA**	If confidentiality is not guaranteed, reporting will decrease. Trust can easily be lost. Organisations should develop a **code of practice** for investigating incidents involving human-AI teaming.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

	investigation process.		considered a 'snitch', telling management about the errors its human counterparts have made.		
8.	C09 A staff member prosecuted for an incident involving a genuine error or mistake would be supported by the management of this organization.	Just culture and reporting	This largely concerns management attitudes to staff and provision of support. However, the term 'genuine error or mistake' needs to encompass the human choice between following IA advice which turns out to be wrong and ignoring such advice which turns out to be right, since in either case there was no human intention to cause harm. This can be enshrined in Just Culture policies, but judiciaries (and the travelling public) may take an alternative viewpoint. In the event of a fatal accident, black-and-white judgements sharpened by hindsight may be made which do not sufficiently reflect the complexity of IA's and Human-AI Teams' operating characteristics and the local rationality at the time.	**A staff member involved in an incident when working with AI would be supported by the management of their organization.**	This is core to Just Culture. If the statement is true, it does not fully protect the individual, but at least shows management support.
9.	B03 We have sufficient staff to do our work safely.	Staff and equipment	Despite many assurances that AI will not replace humans, many see strong commercial imperatives for doing exactly that (e.g. a shortage of commercial pilots and impending shortage of air traffic controllers, post-COVID low return-to-work rate at airports, etc.).	**The IA could be used to replace staff who have valuable safety experience.**	This is in fact key to any automation that reduces staff, and hence the overall 'safety capital' in the operational arm of the organization.
10.	C02 We have the equipment needed to do our work safely.	Staff and equipment	The perceived safety value of IAs will depend on how useful the IA is for safety and will be a major question for the HAIKU use cases. One 'wrong call' could have a big impact on trust.	**The AI plays a critical role in safety.**	This counterpoints the previous statement (note that agreement/disagreement with both statements is viable).
11.	C18 Operational staff are sufficiently involved in system changes.	Communication	There is a risk that if the IA is a very good information collector, people at the sharp end might be gradually excluded in updates to system changes, as the systems developers will consult data from the IA instead. However, in discussions during HAIKU, what became clear was that operational staff should know about updates and changes to the AI system's performance parameters (because these affect its goal setting and alignment with its human counterpart's goals).	**Staff are involved in changes to the AI's performance parameters.**	This will be important for trustworthiness, even if it may not always be practicable to have direct involvement.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

12.	B13 My involvement in safety activities is sufficient.	Collaboration	As for C15 and C18. Given that the IA concept in HAIKU allows a certain degree of 'Agency' of the IA (it can have and set goals, and have executive control), there is the central question of 'who's in charge?' at the end of the day.	**My safety judgment always takes precedence over that of the IA.**	This relates to the EU Act on AI's notion of human oversight in safety critical systems, and EASA's notion of Collaboration and negotiation between human and AI (especially with Category 2B systems [EASA, 2024]).
13.	B25r There are people who I do not want to work with because of their negative attitude to safety.	Collaboration	There could conceivably be a clash between an IA and a team member who, for example, was taking significant risks or continually overriding / ignoring safety advice, or an IA that was giving poor advice. If the IA is a continual learning system, its behaviour may evolve over time, and diverge from optimum or that of its human counterparts , even if it starts off safe when first implemented.	**I don't want to work with the IA because of the way it handles safety.**	This relates to B01, but is useful as a check since it is a very direct question. In general safety culture surveys this question is somewhat polemic, in that it often triggers useful conversations about safety if subsequent workshops are held.
14.	B18 Changes to the organization, systems and procedures are properly assessed for safety risk.	Risk Handling	The IA could have an underpinning model of how things work and how safety is maintained, so any changes will need to be incorporated into that model, which may identify safety issues that may have been overlooked or played down. This is like current use of AIs for continuous validation and verification of operating systems, looking for bugs or omissions. Conversely, the IA may give advice that makes little sense to the human team, or the organization, and be unable to explain its rationale. Humans may find it difficult to adhere to such advice.	**I always know why the AI behaves as it does.**	Automation sometimes 'surprises' people. Unpredictability in safety critical situations is not good for human end users, as it increases workload and stress (unless the IA surprises in a good way).
15.	C07r We often have to deviate from procedures.	Risk Handling	The IA will observe (and perhaps be party to) procedural deviation and can record associated reasons and frequencies (highlighting common 'workarounds'). Such data could identify procedures that are no longer fit for purpose, or else inform retraining requirements if the procedures are in fact still fit for purpose.	**The AI always follows procedures.**	This is a key question, particularly for a learning AI that may adapt and go beyond procedures, including safety ones.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

16.	C14r I often have to take risks that make me feel uncomfortable about safety.	Risk Handling	The IA will likely be unaware of any discomfort on the human's part (unless emotion detection is employed), but the human can probably use the IA's advice to err on the side of caution. Conversely, a risk-taker or someone who puts productivity first, may consult an IA until it gets around the rules (human ingenuity can be used for the wrong reasons).	**The IA puts safety first.**	Another possibility here is 'the AI always errs on the side of caution' but this is slightly archaic language. The new statement also interrelates strongly with B05 and B25r. It basically asks if the AI ever takes risks.
17.	C11 Adequate training is provided when new systems and procedures are introduced.	Procedures and training	As for C08.	**I have received adequate training to work with the IA.**	A straightforward interpretation.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The full safety culture questionnaire (SCQ) was represented on the digital secure *SurveyMonkey* platform so it could be accessed independently and confidentially with a QR code or on an iPad, using a 5-point Likert scale for responding to each statement, i.e. strongly disagree, disagree, neutral, agree and strongly agree.

The 17 questions address the eight safety culture dimensions of the original EUROCONTROL questionnaire as indicated below.

- Colleague commitment to safety (3)
- Just Culture & Reporting (5)
- Staff & Equipment (2)
- Management Commitment to Safety (0)
- Communication (1)
- Collaboration (2)
- Risk Handling (3)
- Procedures & Training (1)

There are no direct *Management Commitment* questions, because these were felt to be too hypothetical at a time when almost no aviation organisations have implemented HAT into safety-critical operations (therefore ATCO/Pilot responses would be pure speculation). However, questions JC3-JC5 indirectly relate to their perception of how management would support them in a future HAT environment. This is also one reason the *Just Culture* dimension has most questions, aside from the fact that some major staff representative organisations (e.g. IFATCA, the International Federation of Air Traffic Controller Associations) have raised questions about Just Culture in future HAT environments, specifically who might be blamed after an incident or accident involving a Human-AI partnership. *Communication* has only one question (CO1), as most HAT prototypes do not yet use oral (natural language processing) communication. Procedures & Training only has one question (PT1), related to training, as procedures are very notional at this stage (though RH2 relates to procedure following and CC3 relates to training). Colleague Commitment has three questions (CC1-3), as in this case 'colleague' refers to the IA. Overall, the requisite dimensions are represented in the new shortened questionnaire.

3 Application of the Questionnaire to Controllers and Pilots

The questionnaire was first administered to eight controllers who had participated in UC4 simulations, then some months later to twelve airline pilots who had participated in UC1 simulations, and several months afterwards to nine UC2 pilots (see Figure 3). The three concepts range from TRL 4-5 and EASA Cat 1B-2A, with UC1 being the most safety-critical as it provides advisory support during an in-flight emergency.

Figure 3: Visualisations from 2 HAT simulations: UC1 (left), UC4 (centre-right) and UC2 (right)

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

3.1 ATCO Survey Results

Figures 4 summarises the responses received from the eight ATCOs, and Table 2 summarises the approximate priority of the issues according to their responses. Overall, Figure 4 highlights a generally positive (green) perception of the AI tool, though with some clear concerns (pink/red) on certain safety culture statements.

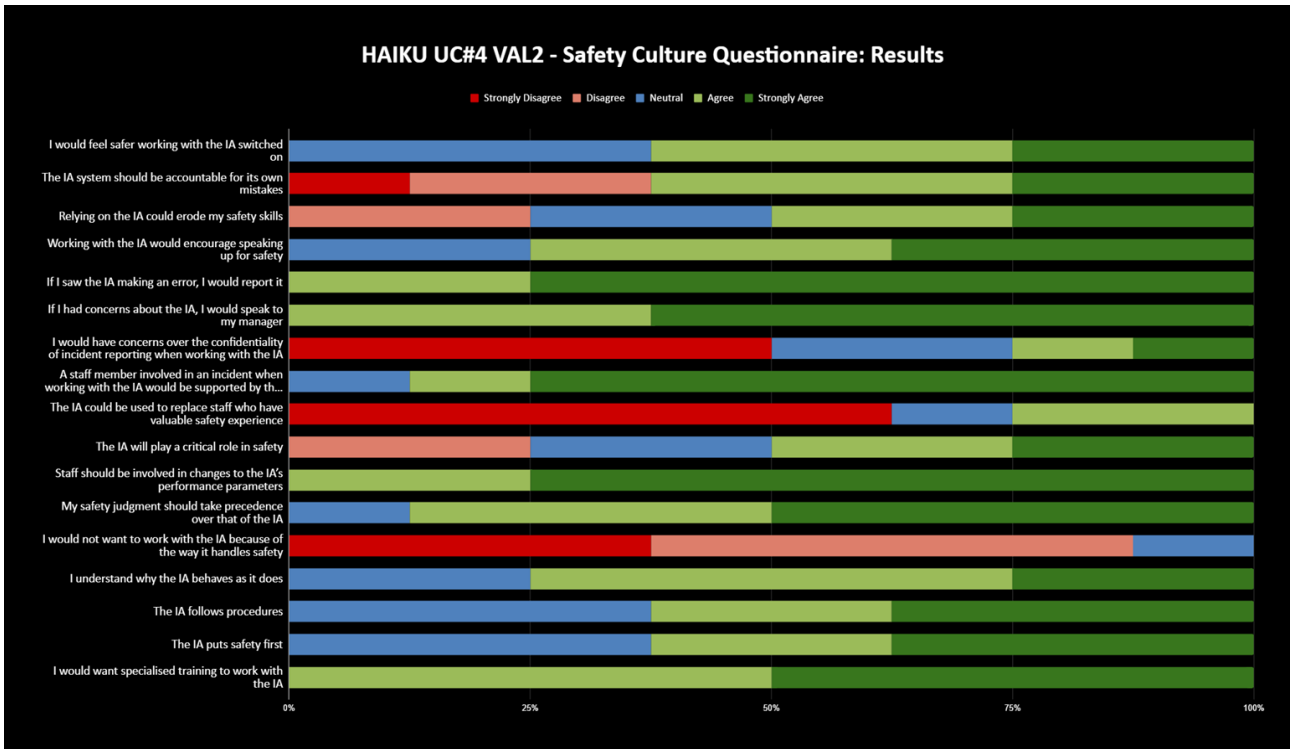


Figure 4: Responses to Questions by ATCOs (green = strongly agree (dark) / agree (light), blue = neutral, pink = disagree, red = strongly disagree.

Table 2 - ATCO Responses Overview

Strongly Desirable "Must-Haves"	Intermediate	Undesirable "Deal-breakers"
If I saw the IA making an error, I would report it. If I had concerns about the IA, I would speak to my manager. I would want specialized training to work with the IA. The IA system should be accountable for its own mistakes. Staff should be involved in changes to the IA 's safety performance parameters. A staff member involved in an incident when working with the IA would be supported by management.	The IA follows procedures. I would feel safer working with the IA switched on. The IA puts safety first. Working with the IA would encourage speaking up for safety. I understand why the IA behaves as it does.	I would not want to work with the IA because of the way it handles safety. The IA could be used to replace staff who have valuable safety experience. I would have concerns over the confidentiality of incident reporting when working with the IA. Relying on the IA could erode my safety skills. The IA will play a critical role in safety.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

My safety judgment should take precedence over that of the IA.		

What comes across clearly is the willingness to report if there are problems with the IA (which incidentally is also a requirement coming from the EU Act on AI), as well as wanting to know if the IA's parameters are changing, having specialised training on the IA and how it works, ensuring the IA (its developers/managers) have accountability, and having human safety judgement take precedence over the IA's decision-making/advice. Many of these 'must-haves' are related to Just Culture as well as learning and ensuring the IA 'gets it right', but if it doesn't, the human ATCO can step in and take charge, which itself requires understanding and training on how the AI works and how it can go wrong.

On AI accountability, a post-questionnaire comment from an ATCO asked "*How can we blame the IA that gets it wrong when we designed it?*". However, whilst ATCOs may be involved in an IA's training and have input into its interaction design, they cannot and should not be held responsible for its inner workings, which will be opaque to anyone save data scientists, and not even them if the AI is complex.

This latter training aspect is also reflected in the undesirability of safety skills erosion and non-understanding of how the IA works, suggesting the need for clear operational explainability (OpXAI). In verbal comments after the questionnaire, ATCOs suggested IA systems should not be introduced during ATCO training or certification phases, as it is crucial for controllers to first develop and consolidate the relevant skills. The IA could then be considered at later stages, once the ATCO has demonstrated an adequate level of expertise. Determining this adequate level would itself require developing and applying specific behavioural markers.

Other candidate 'deal-breakers' relate to the loss of human safety expertise, and a concern over confidentiality issues related to incident reporting. The latter is due to the IA's theoretical ability to record all interactions it has with ATCOs. Such an evidence trail could potentially be used against crew members in an incident investigation. The AI system should be designed with current incident reporting practices in mind. Specifically, therefore, the AI should provide neutral information to avoid creating fear of punishment, thereby supporting a continued culture of reporting. This is essential to preserve the original purpose of reporting systems: enabling continuous improvement rather than assigning blame to individuals.

With respect to the potential erosion of safety skills, some of the ATCOs (who also filled in the questionnaire) voiced the opinion that they would not want the IA 'on' all the time, in part to ensure that the ATCOs retained their core skills, which also meant they could take over if required (e.g. due to IA failure, aberrant behaviour or unavailability).

With respect to the 'erosion of safety skills' and 'the IA playing a critical role in safety', which were viewed negatively, some additional post-questionnaire comments illuminated what was behind the negative ratings, e.g.; "*If we put ISA in low traffic airports, skills will be eroded. If we put it in high traffic ones, it will be a big help*". Similarly, on the safety critical role of the IA: "*We should try to avoid giving critical roles to AI*" and "*The critical part is the human itself*".

Lastly, there were several noteworthy comments made by controllers, post-questionnaire, on the general desirability and utility of AI in AM operations:

> "*Let people understand how the system works to avoid fear and negative thinking. People fear what they don't understand.*"
>
> "*Introduce it in the simulator first to allow people familiarise with it before using it into operations. Avoid just giving documentations thinking that this is enough for people to understand and trust the system.*"
>
> "*Always leave the last word on safety culture and just culture to human - AI does not understand emotions.*"
>
> "*Design AI that helps humans see what they struggle to see.*"

Such comments should be useful to middle managers responsible for overseeing the introduction of AI into ATM systems, and if heeded, should help the chances of successful technology integration and acceptance into the controller's workspace.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

3.2 Pilot Survey Results

The safety culture questionnaire (SCQ) survey results for the 22 pilots who took part in VAL2 simulations at ENAC in Toulouse in October/November 2024 (UC1 - n. 12) and at TAVS in Bordeaux in January 2025 (UC2 - n. 10) are respectively shown in Figures 5 and 6.

Figure 5: UC1 Pilot (n=12) Responses to SC Questions

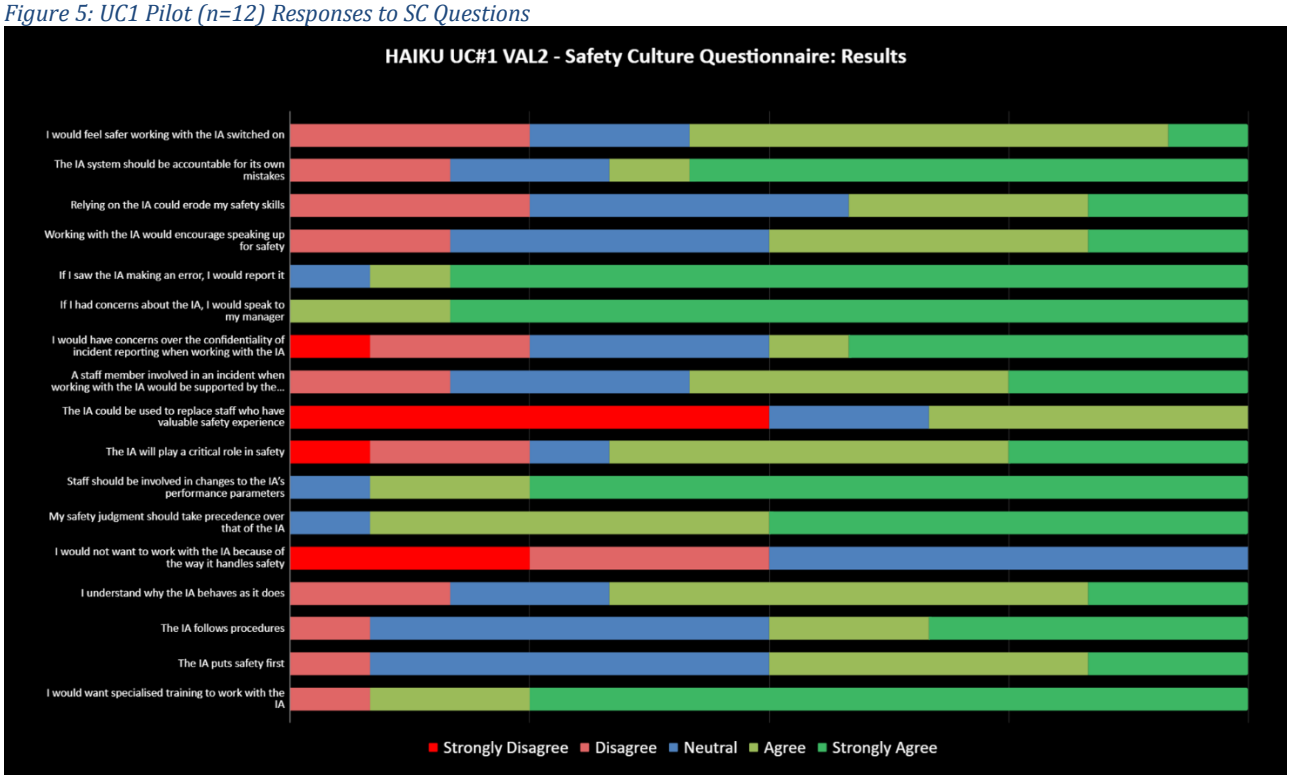
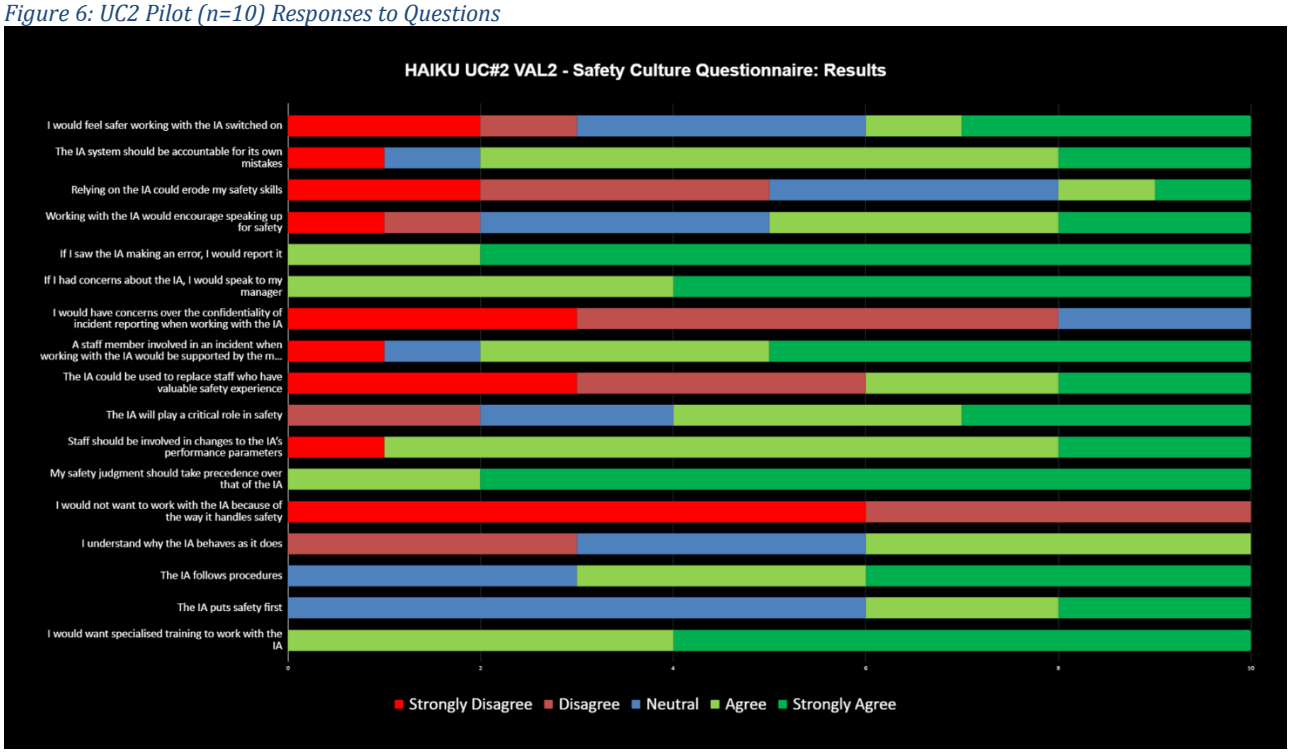


Figure 6: UC2 Pilot (n=10) Responses to Questions



© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

The prioritization of issues across both sets of pilots is shown in Table 3. Interestingly, although the pros and cons generally align, UC1 pilots exhibited more 'neutral' ratings, suggesting they were less sure about the items than the pilots from UC2. Also, the key issues are similar for ATCOs and pilots.

Table 3 – Pilot Responses Categorised

Strongly Desirable	Intermediate	Undesirable
If I had concerns about the IA, I would speak to my manager. If I saw the IA making an error, I would report it. Staff should be involved in changes to the IA's parameters. My safety judgement should take preference over that of the IA. I would want specialised training to work with the IA. I would ~~not~~ want to work with the AI because of the way it handles safety	The IA puts safety first. The IA follows procedures. I understand why the IA behaves as it does. Working with the IA would encourage speaking up for safety. The IA system should be accountable for its mistakes. I would feel safer working with the IA switched on.	The IA could be used to replace staff who have valuable safety experience. I would have concerns over the confidentiality of incident reporting when working with the IA. The IA will play a critical role in safety. Relying on the IA could erode my safety skills.

As noted, there is significant similarity between the pilots and ATCOs responses, for example Just Culture, the desire for specialised training and staff, and being involved in changes to the IA's parameters are strongly desirable, and that pilots feel their safety judgement should take precedence over that of the IA. Furthermore, a willingness to "work with the AI because of the way it handles safety" has been reported, clearly indicating that, overall, the potential of AI is well received and can be accepted. However, pilots emphasized that this acceptance largely depends on two key factors. First, how the system is introduced: it should be fully integrated into the aircraft as a core component, rather than simply added on as a tool to replace human tasks. Second, the role the system plays: it should act as a support tool, enhancing pilot performance without overriding their decisions or intentions.

The undesirable items for pilots also share strong overlap with the ATCOs, for example erosion of safety skills, concern about losing people with valuable safety expertise. On this point, pilots emphasized the crucial role of regulators in setting clear boundaries to ensure that AI is used as a support mean rather than a replacement. Furthermore, they brought up three fundamental questions: "How can we train AI on safety culture?", "How can we teach AI to detect, prevent, and recover from its own errors?", and "How do I perform safely when AI becomes another "crew member" to oversee and cross-check actions with?". These are core human competencies, further reinforced in the two-pilots configurations. Addressing these questions should be a pre-requisite before starting considering AI as a potential replacement for personnel with valuable safety expertise.

Similarly to ATCOs, pilots also rated items related to the confidentiality of incident reporting as undesirable, particularly in UC1. This is likely because UC1, by definition, involves more critical and error-prone scenarios, and includes the monitoring of physiological data, which raises additional concerns.

The sentiment that the IA should not play a critical role in safety is also echoed for similar reasons as for the ATCOs, that safety is reserved for pilots. One notable difference is that pilots are more negative about 'feeling safer with the IA switched on' compared to ATCOs. Overall, many pilots found this question difficult to answer, stating "Don't ask me if I would feel safer with AI - It is impossible to say right now." They also reflected on the point that the "Openness to adopting AI onboard may depend on the aircraft type I fly" highlighting that pilots flying highly automated aircraft may be more open to AI integration than others. In addition, pilots identified several key preconditions for the successful implementation of AI. These include: the need for adequate and timely training (starting now), the expectation that AI should serve as a cross-check for pilot reasoning and provide consistent support during periods of high workload, fatigue, or other altered mental states, and the requirement that pilots must retain the ability to switch off the AI whenever they deem it necessary.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

In addition to the confidential questionnaire, the 22 pilots from UC1 and UC2 were also asked to answer a few 'open' questions about both the validity of the questionnaire and the more general perception of the potential impact of AI on safety culture.

All questions regarding the validity of the questionnaire received positive feedback, indicating that the questionnaire is well-structured and on the right track.

As for the overall impact of AI on Safety Culture, pilots expressed a generally positive outlook, with many stating that "AI could have a positive impact on Safety Culture, but only if properly regulated." This highlights that appropriate regulation is seen as a key factor in ensuring AI enhances rather than undermines safety culture. In addition to regulation, pilots identified other critical conditions for a successful integration of AI: system reliability, a gradual and step-by-step implementation, structured training on AI usage, and a stronger focus on safety culture education starting from flight school. Additionally, many pilots highlighted specific design principles and practical expectations as essential preconditions for a positive contribution of AI to safety culture, and more generally to aviation. Among the most frequently mentioned were: "Start with AI handling the tasks I find boring," "Involve me in its design and introduce it gradually", "Make AI systems intuitive and easy to use," and "Design systems based on the assumption that pilots don't always perform perfectly."

Overall, it can be concluded that pilots are open to the integration of AI in the cockpit, provided that "what truly matters is that AI gives me a genuine sense of control, does not override my intentions or decisions, and can be switched off whenever necessary."

3.3 Discussion of Questionnaire Results

Overall, the results suggest a cautious optimism for AI and future HAT environments. It was recognised by some of the participants that some of these questions are early, as we do not yet have sufficient experience of HATs and how they might be integrated into operational systems. However, there are some clear guidelines and even 'red lines' already appearing, suggesting the following preliminary safeguards:

1. HAT development needs to be a human-centred process, focusing on the end user.
2. The human should stay in control for now[1], especially in the decision-taking loop[2].
3. Pilots/ATCOs need oversight of the IA's key parameters, especially if they can change dynamically.
4. There needs to be effective Operational Explainability with HAT systems, tuned to the context and the human's situation awareness.
5. Training on the IA – how it works and how to master its usage – will be key and should start from now to ease understanding and acceptance.
6. Strategies for skill retention need to be developed and implemented, including consideration of operating without the IA 'on' all the time.
7. The quality of Human-IA interaction will be key – IA interactions must be relevant, helpful, timely and contextual.
8. There should not be loss of critical human safety expertise.
9. Just Culture assurances will be key to IA technology acceptance.
10. The development and introduction of AI into real operational systems should be gradual rather than sudden, with extensive testing and evaluation, allowing sufficient 'headspace' to learn how best to implement HAT.

These are further represented in Figure 7 below, in terms of six core safeguard areas for HAT development.

[1] This may change in the future, e.g. For 'auto-land' due to pilot incapacitation.
[2] Note this is for EASA Categories 1A to 2B; 3A and 3B are by definition higher AI autonomy systems.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

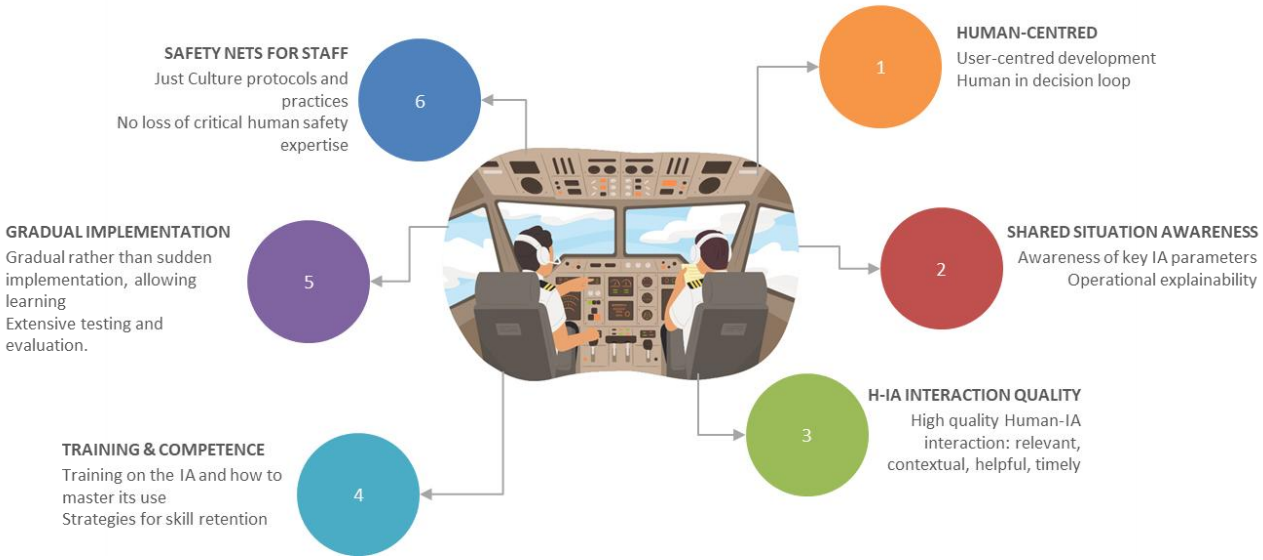


Figure 7: Six Core Development Safeguards Resulting from SCQ Questionnaire Application

One interesting remark by one of the interviewee was that the IA needed to be *as good as a good co-pilot*, or even *as good as having an instructor sitting next to you*. Yet currently IAs are very far from such aspirations, observing a vast difference between the situation awareness of the pilot and that of an IA. Currently rooted in Machine Learning, IAs are experts at certain aspects and dummies in others (hence the term 'Narrow' AI), and it is difficult to see how they can have the breadth of expertise of a good First Officer, let alone an instructor or trainer. A possible future way forward could involve a digital twin representing an amalgamation of many experienced pilots or trainers, coupled with a Natural Language Processing (NLP) interface that had the fluidity of current leading Large Language Models (LLMs), but without the attendant hallucinations. Another option is a 'domain-limited LLM', i.e. one trained on the aviation domain, which would embody more context and be less prone to hallucinations. It seems more likely, therefore, that in the mid-term IAs develop at the EASA 1B or 2A level. It may be some time before a genuine IA capable of performing satisfactorily ('holding its own') at EASA level 2B appears, as it may require a new hybrid AI architecture.

On the subject of the methodology, namely the Safety Culture Questionnaire, this should be considered as a provisional but viable method, and is probably best coupled with interviews or focus groups with several pilots or ATCOs. This would help ensure that all relevant issues are captured, and that their opinions are not 'anchored by the questionnaire items.

4 Safety Culture Guardrails for Aviation Organisations

The earlier analysis in this report raised a number of potential threats to future safety culture, and a more or less equivalent number of safety culture 'affordances' wherein safety culture could be enhanced. In this sense the overall impact of AI and HAT on safety culture will depend on how it is researched, designed, developed, deployed and managed in actual operational environments. The issues identified, whether positive or negative, give rise to provisional safeguards to prevent safety being diminished due to the introduction of advanced AI systems into aviation. However, for safeguards to be effective, those who can enact them – the organisational stakeholders – also need to be identified and enlisted.

The various potential impacts are diversely spread across different human 'levels' in organizations; some relating to front-line staff, some to middle management, and some to senior or executive levels. Safety culture always works best in aviation when those at the top – CEOs, VPs and Executive Boards, firmly believe in and support safety as a priority. There needs to be continued safety stewardship by senior executives, to maintain the human as the principal safety agent, which can then be translated by middle management throughout the organization into satisfactory actionable outcomes. A critical risk owner will be the safety department, typically the hub of safety learning for an organization, as well as the key working interface with external regulators. Next are the front-line

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

and support staff, who are the lifeblood of safety culture in any operational organization. A four-layer model is shown in Figure 8, with the identified safeguards inserted at corresponding 'risk owner' levels.

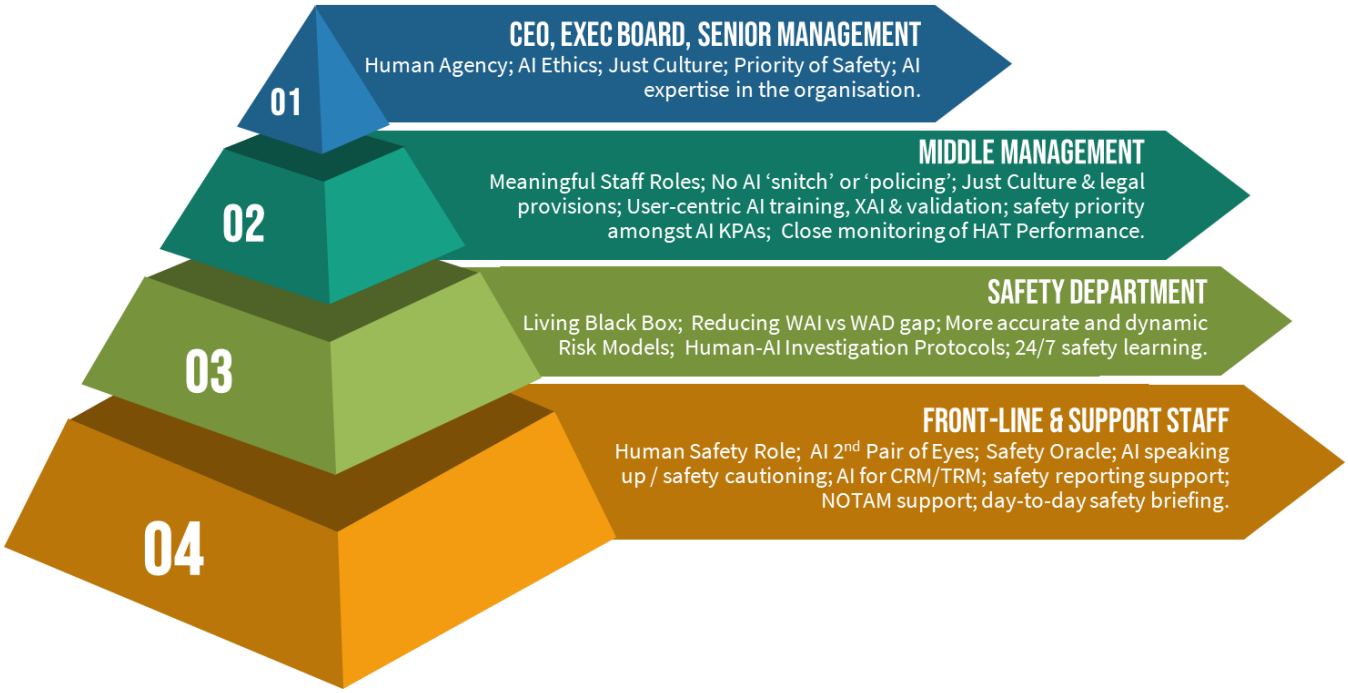


Figure 8 – Four Layer Model of Stakeholder Safety Culture Engagement

At the bottom layer are the principal safeguards related to **front-line and support staff**. First is the need to maintain *human agency for safety*, i.e. a valid safety role. Second is the fact that the IA can act as a *second pair of eyes* (similar to the notion of the IA being like a good co-pilot) whether aiding in an emergency or noting a safety issue or deviation or risky course of action by the human operator. This leads to a third useful aspect of an IA, that it can be a ready-to-hand *safety oracle* (again, similar to the notion of having a trainer sitting next to you) that the human team can consult at any point when considering the best course of action and the safety risks it might entail. The IA could also be programmed to *'speak up'* for safety if warranted, and this can be embedded into human *Crew Resource Management (CRM)* and *Team Resource Management (TRM* – the ATM equivalent of CRM) practices and training. The IA could be a useful aid for *safety reporting*, able to quickly capture events, their precursors, signals and actions, to which the human could then add a narrative (their version of events). *NOTAM*s (Notices to Airmen) could be automatically uploaded into the IA, which could remind human crews if they have forgotten or overlooked any relevant aspects during operations. Similarly, the IA could be useful as a *day-to-day briefing* tool, letting the oncoming shift know of anything unusual, or changes to procedures, or the status of ongoing maintenance, etc. that occurred or started on previous shifts. Taken together, these eight safeguards could keep safety at the human's fingertips, eyes or ears, whilst guarding against simple omissions and reckless acts.

At the next organisational layer is the **safety department**. A first safeguard is the notion of the IA serving as a *living black box*, such that after an event the IA could reproduce the detailed flow of events, signals, interactions, decisions made and even the thinking underpinning those decisions, prior to and during the incident. This could paint a much more detailed picture than investigators currently have. Such an 'annotated timeline' could also be very useful in safety learning and training. In parallel, investigators will undoubtedly need to develop new *Human-AI investigation protocols* to deal with Human-AI Teaming events, particularly when relating to the double-bind type of scenarios raised earlier. These protocols should include non-pejorative (i.e. non-judging or blaming) language and be informed by Just Culture principles adopted at higher levels in the organization and enshrined in European law.

The IA could also compare ways of working (what is actually done) against procedures and rules; not to police, but as a way of defining *the gap between real operational practices and the official rules and procedures*. If the gaps are unsafe, then this can lead to more training, but in many cases, it is likely that the official rules are either inefficient or unworkable in real operational conditions, or else need updating as operations and technology has moved on. Similarly, *risk models* in aviation are often seen as not reflecting operational reality or being at too high a level of abstraction. The IA could record interactions in both safety-related events as well as 'when things go

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

right'. Such information could be fed into risk models to render them more operationally relevant, giving them a more detailed level of description. If this can be achieved, then such models can become useful to operations departments, and not just seen as being only for safety departments and regulators. The goal here would be that day-to-day operations are feeding *dynamic risk models* so that live safety performance can be seen, including when things may be drifting towards danger, or when new hazards are emerging. This could pave the way for true 24/7 safety monitoring and real-time learning.

At the next level up in the pyramid model is **Middle Management**, who have the challenge of exercising senior management aspirations within real world operational and resource constraints. Part of their mission regarding safety culture and AI is to ensure that *staff have meaningful jobs*, that *IAs do not act as 'snitches' on staff or police them*, and that *Just Culture ideals* can be translated into effective and trusted principles and practices enacted at lower levels, in agreement with social partners (unions etc.). A key role will be the overseeing of introducing autonomous AI systems into the organization, ensuring that they are *user centric.* This is especially the case where 'explainability' of the AI's advice or decisions (OpXAI) to the human crews, as well as *human-supervised-learning*, *user validation* and *Human-AI Team training* prior to operational deployment. Since 'user-centeredness' is likely to be a common safeguarding theme, the discipline of Human Factors is likely to be a critical determinant of success in these activities.

If the IA, as is likely, has more key performance areas (*KPAs*) than safety alone (e.g. productivity, green-ness, etc.), then middle management must ensure that safety retains priority when the IA is making *trade-offs*. Last, middle management must closely *monitor the IA's performance* as it is deployed, as it will evolve both in its dealings with humans and other IAs.

The top level in the pyramid is senior management, including **CEOs, VPs, Directors and Executive Boards**. Here is where there needs to be an authentic message that safety is the priority and that '*people create safety*', albeit backed up and supported by AI. There should also ideally be a *code of ethics* related to the use of AI in the organization, as well as a *Just Culture policy and framework* which deals with AI accountability in the case of an accident, developed with social partners. It is also at this level that decisions need to be made on having *internal AI expertise* in the organization, so that organizational leaders can maintain a basic understanding and realistic expectations of their AI 'assets' and be prepared to face and answer the media when things go wrong.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Conclusions

This deliverable has reviewed the potential impact of AI on safety culture and identified potential threats and benefits ('affordances') that could arise via the introduction of AI into the cockpit and air traffic tower and Ops Room. Adding to this theoretical or speculative review, it has explored controller and pilot reactions to, and perceptions of, the impact of AI on their safety culture, following their exposure to prototype IA in three realistic simulations. This has led to the identification of provisional safeguards for the development of future HAT systems (Figure 7). The latter part of this report returned to the potential benefits and threats of AI in aviation and has developed a set of organisational guard-rails to help shore up and even advance safety culture in operational aviation organisations (Figure 8).

These two sets of safeguards are provisional. However, they can be adopted and followed by system developers and product teams, as well as aviation organisations seeking to integrate AI-based systems into their operations, until better ones based on more experience are developed. In the meantime, the safety culture questionnaire plus additional interviews/focus groups represents a resources-light and agile way to take the pulse of safety culture and AI impacts, serving as a check to see if there is any drift or decrement in safety culture as a function of material steps towards AI integration into aviation operations.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

References

AAIB (1990) Report on the accident to Boeing 737-400 G-OBME near Kegworth, Leicestershire on 8th January, 1989. Air Accidents Investigation Branch Report 4/90. London: HMSO.

Advisory Committee on the Safety of Nuclear Installations (ACSNI) Study Group. Third Report: Organizing for Safety. Sheffield: H.M. Stationery Office; 1993

Cullen, D. (1990) The public enquiry into the Piper Alpha Disaster. London: HMSO.

Dias, M., Teles, A, and Lopes, R. (2020) Could Boeing 737 Max crashes be avoided? Factors that undermined project safety. Global Scientific Journals: Volume 8, Issue 4, April, Online: ISSN 2320-9186.

EASA (2023) EASA Concept Paper: first usable guidance for level 1 & 2 machine learning applications. February. https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-20-published

European Commission (2022) CORDIS Results Pack on AI in air traffic management: A thematic collection of innovative EU-funded research results. October 2022. https://www.sesarju.eu/node/4254

European Parliament (2023) EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

Hidden, A. (1989) Investigation into the Clapham Junction Railway Accident. London: HMSO.

Kaliardos, W. (2023) Enough Fluff: Returning to Meaningful Perspectives on Automation. FAA, US Department of Transportation, Washington DC. https://rosap.ntl.bts.gov/view/dot/64829

Kirwan, B. and Shorrock, S.T. (2015) A view from elsewhere: safety culture in European air traffic management. In Waterson, P. (Ed.) Patient Safety Culture. Aldershot, UK: Ashgate. pp. 349-370.

Kirwan, B., Reader, T.W., Parand, A., Kennedy, R., Bieder, C., Stroeve, S., and Balk, A. (2019) Learning curve: interpreting the results of four years of safety culture surveys. Aerosafety World, Flight Safety Foundation, January.

Kirwan, B. Shorrock, S.T. and Reader, T. (2021) The future of safety culture in European ATM – a White Paper. EUROCONTROL. https://skybrary.aero/bookshelf/future-safety-culture-european-air-traffic-management-white-paper

Kirwan, B. (2024) The Impact of Artificial Intelligence on Future Aviation Safety Culture. Future Transportation, 4, 349-379. http://dx.doi.org/10.3390/futuretransp4020018

Kirwan, B. (2025) Human Factors Requirements for Aviation Human-AI Teaming. *Future Transp*. 2025, 5(2), 42; DOI: https://10.3390/futuretransp5020042

Nunes, A. & Laursen, T. (2004) Identifying the factors that led to the Uberlingen mid-air collision: implications for overall system safety. Proceedings of the 48th Annual Chapter Meeting of the Human Factors and Ergonomics Society, September 20 - 24, 2004, New Orleans, LA, USA.

Reader, T. W., Noort, M. C., Kirwan, B., & Shorrock, S. (2015). Safety sans frontiers: An international safety culture model. Risk Analysis, 35, 770-789.

Reader, T., Parand, A. and Kirwan, B. (2016) European pilot's perceptions of safety culture in European aviation. Future Sky Safety Report D5.4, November. DOI: 10.13140/RG.2.2.14285.51686

Reason, J.T. (1997) Managing the risks of organisational accidents. Aldershot: Ashgate.

Turner, R. and Pidgeon, N. (1997) Man-made disasters (2nd edition) Oxford: Butterworth-Heineman.

Zweifel, T.D. and Vyal, V. (2021) Crash: BOEING and the power of culture. Journal of Intercultural Management and Ethics Issue No. 4, 13-26.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Appendix A: Detailed Analysis of Questionnaire Items According to Human-AI Team Integration

Table 4: Assessment of the Impact of Intelligent Assistants on Safety Culture Survey Items

Questionnaire Item	Dimension	IA Impact	Team / Org / Society	H/M/L
B01 My colleagues are committed to safety.	Colleague commitment to safety	The IA would effectively be a digital colleague. The IA's commitment to safety would likely be judged according to the IA's performance. Human-Supervised Training, using domain experts with the IA would help engender trust. The concern is that humans might 'delegate' some of their responsibility to the IA. A key issue here is to what extent the IA sticks rigidly to 'golden rules' such as aircraft separation minima (5NM and 1000 feet) or is slightly flexible about them as controllers may sometimes (need to) be. The designer needs to decide whether to 'hard code' some of these rules or allow them to be more plastic (within limits); this determines whether the IA behaves like 'one of the guys' or never, ever breaks rules.	Team	High
B04 Everyone I work with in this organization feels that safety is their personal responsibility.	Colleague commitment to safety	Since an IA cannot effectively take responsibility, someone else may be held accountable for a IA's 'actions'. If a supervisor fails to see an IA's 'mistake', who will be blamed? HAIKU use cases may shed light on this, if there can be scenarios where the IA gives 'poor' advice. If an IA is fully autonomous, this may affect the human team's collective sense of responsibility, since in effect they can no longer be held responsible.	Team / Org	High
B07 I have confidence in the people that I interact with in my normal working situation.	Colleague commitment to safety	As for B01, this will be judged according to performance. Simulator training with IAs should help pilots and others 'calibrate' their confidence in the IA. May overlap significantly with B01.	Team	High
B02 Voicing concerns about safety is encouraged.	Just culture and reporting	The IA could 'speak up' if a key safety concern is not being discussed or has been missed. This could be integrated into Crew and Team Resource	Team / Org	High

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

		Management and Threat and Error Management practices. However, then the IA may be considered a 'snitch', a tool of management to check up on staff. This could also be a two-way street, so that the crew could report on THE IA's performance.		
B08 People who report safety related occurrences are treated in a just and fair manner.	Just culture and reporting	The IA could monitor and record in real time and would be akin to a 'living' Black Box recorder. This could affect how humans behave and speak around the IA.	Org	High
B12 We get timely feedback on the safety issues we raise.	Just culture and reporting	The IA could significantly increase reporting rates, depending on how its reporting threshold is set, and also record and track how often an issue is raised.	Org / Team	Medium
B14 If I see an unsafe behaviour by a colleague I would talk to them about it.	Just culture and reporting	[See also B02] The IA can 'query' behaviour or decisions that may be unsafe. Rather than 'policing', the IA could possibly do it more sensitively.	Team	High
B16 I would speak to my manager if I had safety concerns about the way that we work.	Just culture and reporting	The IA should not become a 'snitch' for management, as this would be a deal-breaker for honest teamworking.	Team / Org	Low
C01 Incidents or occurrences that could affect safety are properly investigated.	Just culture and reporting	As for B08, the IA's record of events could shed light on the human colleagues' states of mind and decision-making.	Org / Team	High
C06 I am satisfied with the level of confidentiality of the reporting and investigation process.	Just culture and reporting	As for B16, the use of IA recordings as information or even evidence during investigations needs to be considered. Just Culture policies will need to adapt/evolve to the use of IAs in operational contexts.	Org / Team	High
C09 A staff member prosecuted for an incident involving a genuine error or mistake would be supported by the management of this organisation.	Just culture and reporting	This largely concerns management attitudes to staff and provision of support. However, the term 'genuine error or mistake' needs to encompass the human choice between following IA advice which turns out to be wrong, and ignoring such advice which turns out to be right, since in either case there was no human intention to cause harm. This can be enshrined in Just Culture policies, but judiciaries (and the travelling public) may take an alternative viewpoint. In the event of a fatal accident, black-and-white judgements sharpened by hindsight may be made which do not reflect the complexity of IA's and Human-AI Teams'	Team / Org / Society	High

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

		operating characteristics, nor the over-riding benefits to the industry.		
C13 Incident or occurrence reporting leads to safety improvement in this organisation.	Just culture and reporting	This is partly administrative and depends on financial costs of safety recommendations. Nevertheless, the IA may be seen as adding dispassionate evidence and more balanced assessment of severity, and how close an event actually came to being an accident (e.g. via Bayesian statistical analysis). It will be interesting to see if the credence given to the IA by management is higher than that given to its human counterparts.	Org	High
C17 A staff member who regularly took unacceptable risks would be disciplined or corrected in this organisation.	Just culture and reporting	As for C09, an IA may be aware of an individual who takes more risks than others. However, there is a secondary aspect, linking to B07, that a IA may be 'trained' by humans, and may be biased by their own level of risk tolerance and safety-productivity trade-offs. If an IA is seen as offering solutions judged as too risky, or 'too safe', nullifying operational efficiency, it will need 're-training' or re-coding in some way.	Team / Org	High
B03 We have sufficient staff to do our work safely.	Staff and equipment	Despite many assurances that AI will not replace humans, many see strong commercial imperatives for doing exactly that (e.g. an endemic shortage of commercial pilots, post-COVID low return to work rate at airports, etc.).	Team / Org / Society	High
B23 We have appropriate support from safety specialists.	Staff and equipment	The IA could serve as a 'safety encyclopaedia' for its team, with all safety rules, incidents and risk models stored in its knowledge base.	Team	Medium
C02 We have the equipment needed to do our work safely.	Staff and equipment	The perceived safety value of IAs will depend on how useful the IA is for safety, and will be a major question for the HAIKU use cases. One 'wrong call' could have a big impact on trust.	Team	High
B05 My manager is committed to safety.	Management commitment to safety	The advent of IAs needs to be discussed with senior management, to understand if it affects their perception of what is keeping their organisation safe. They may come to see the IA as a more manageable asset than people, one that can be turned up or down with respect to safety.	Org	High
B06 Staff have a high degree of trust in management with regard to safety.	Management commitment to safety	Conversely, operational managers may simply be reluctant to allow the introduction of IAs into the system, due to both safety and operational concerns.	Org	Medium
B10 My manager takes action on the safety issues we raise.	Management commitment to safety	See C13 above.	Org	Low

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

B19 Safety is taken seriously in this organization.	Management commitment to safety	Depends on how much the IA is designed to focus on safety. The human team will watch the IA's 'behaviour' closely and interpret for themselves whether the IA is there for safety or for other purposes. These could include profitability, but also a focus on environment issues. Ensuring competing priorities do not conflict may be challenging.	Org	Medium
B22 My manager would always support me if I had a concern about safety.	Management commitment to safety	See B16, C09, C17. If the IA incorporates a dynamically updated risk model, concerns about safety could be rapidly assessed and addressed according to their risk importance.	Team / Org	Low
B28 Senior management takes appropriate action on the safety issues that we raise.	Management commitment to safety	See B12. A further aspect is whether the management supports getting the IA 'fixed' if its human teammates think it is not behaving safely.	Org / Team	Low
B09 People in this organization share safety related information.	Communication	The IA could become a source of safety information sharing, but this would still depend on the organisation in terms of how the information would be shared and with whom. The IA could however share important day-to-day operational observations e.g. by flight crew, who can pass on their insights to the next crew flying the same route, for example, or by ground crew at an airport (some airports already use a 'Community App' for sharing such information).	Org / Team	Medium
B11 Information about safety related changes within this organisation is clearly communicated to staff.	Communication	The IA could again be an outlet for information sharing, e.g. notices could be uploaded instantly and the IA could 'brief' colleagues or inject new details as they become relevant during operations. The IA could also upload daily NOTAMs (Notices to Airmen) and safety briefings for controllers, and could distil the key safety points, or remind the team if they forget something from procedures / NOTAMs / briefings notes	Team / Org	Medium
B17 There is good communication up and down this organisation about safety.	Communication	An IA could reduce the reporting burden of operational staff if there could be an IA function to transmit details of concerns and safety observations directly to safety departments. An IA 'network' could be useful for safety departments to quickly assess safety issues, and prepare messages to be cascaded down by senior/middle management.	Org / Team	Medium
B21 We learn lessons from safety-related incident or	Communication	The IA could provide useful and objective input for safety investigations, including inferences on causal and contributory factors. Use of Bayesian inference and	Org	High

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

occurrence investigations.		other similar statistical approaches could avoid some typical human statistical biases, to help ensure the right lessons are learned and are considered proportionately to their level of risk. Alternatively, if information is biased or counterfactual evidence is not considered, the way the IA judges risk may be incorrect, leading to a lack of trust by operational people. It could also leave managers focusing on the wrong issues.		
B24 I have good access to information regarding safety incidents or occurrences within the organisation.	Communication	IAs or other AI-informed safety intelligence units could store a good deal of information on incidents and accidents, with live updates, possibly structured around risk models, and capturing more contextual factors than are currently reported (this is the aim of HAIKU Use Case 5). Information can then be disseminated via an App or via the IA itself to various crews / staff.	Team / Org	High
B26 I know what the future plans are for the development of the services we provide.	Communication	No impact identified.	Org / Team	Low
C03 I read reports of incidents or occurrences that are relevant to our work.	Communication	The IA could be used to store incidents, but this would not require anything so sophisticated as an IA. However, if the IA is used to provide concurrent (in situ) training, it could bring up past incidents related to the current operating conditions.	Team	Low
C12 We are sufficiently involved in safety risk assessments.	Communication	Working with an IA might give the team a better appreciation of underlying risk assessments.	Org / Team	Low
C15 We are sufficiently involved in changes to procedures.	Communication	The IA could build up evidence of procedures that regularly require workarounds or are no longer fit for purpose. The IA could highlight gaps between 'work as designed', and 'work as done'.	Org / Team	Medium
C16 We openly discuss incidents or occurrences in an attempt to learn from them.	Communication	[See C03] Unless this becomes an added function of the IA, it has low relevance.	Org / Team	Low
C18 Operational staff are sufficiently involved in system changes.	Communication	There is a risk that if the IA is a very good information collector, people at the sharp end might be gradually excluded in updates to system changes, as the systems developers will consult data from the IA instead.	Org / Team	Medium
B13 My involvement in	Collaboration	As for C15 and C18.	Org / Team	Low

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

safety activities is sufficient.				
B15r People who raise safety issues are seen as troublemakers.	Collaboration	It needs to be seen whether an IA could itself be perceived as a trouble-maker if it continually questions its human team-mates' decisions and actions.	Team	Medium
B20 My team works well with the other teams within the organization.	Collaboration	The way different teams 'do' safety in the same job may vary (both inside companies, and between companies). The IA might need to be tailored to each team, or able to vary/nuance its responses accordingly.	Team	Medium
B25r There are people who I do not want to work with because of their negative attitude to safety.	Collaboration	There could conceivably be a clash between an IA and a team member who, for example, was taking significant risks or continually overriding / ignoring safety advice, or an IA that was giving poor advice.	Team	High
B27 Other people in this organization understand how my job contributes to safety.	Collaboration	The implementation of an IA in a particular work area (e.g. a cockpit; an air traffic Ops room; an airport/airline operational control centre) itself suggests safety criticality of human tasks in those areas. If an IA becomes an assimilator of all safety relevant information and activities, it may become clearer how different roles contribute to safety. This may become evident in UC5.	Team	Medium
C05 Good communication exists between Operations and Engineering/ Maintenance to ensure safety.	Collaboration	If Eng/Maintenance 'own' the IA, i.e. are responsible for its maintenance and upgrades, then there will need to be good communication between these departments and Ops/Safety. A secondary aspect is that IAs used in Ops could transmit information to other departments concerning engineering and maintenance needs observed during operations.	Org	Medium
C10 Maintenance always consults Operations about plans to maintain operational equipment	Collaboration	It needs to be determined who can upgrade an IA's system and performance characteristics. E.g. if a manual adjustment is made to the IA to better account for an operational circumstance that has caused safety issues, who makes this adjustment and who needs to be informed?	Org	Medium
B18 Changes to the organisation, systems and procedures are properly assessed for safety risk.	Risk Handling	The IA could have a model of how things work and how safety is maintained, so any changes will need to be incorporated into that model, which may identify safety issues that may have been overlooked or played down. This is similar to current use of AIs for continuous validation and verification of operating systems, looking for bugs or omissions.	Org	High

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

		Conversely, the IA may give advice that does not make sense to the human team or the organisation, yet be unable to explain its rationale. Managers may find it difficult to adhere to such advice.		
C07r We often have to deviate from procedures.	Risk Handling	The IA will observe (and perhaps be party to) procedural deviation, and can record associated reasons as well as frequencies (highlighting common 'workarounds'). Such data could be used to identify procedures that are no longer fit for purpose, or retraining requirements if the procedures are in fact still fit for purpose.	Org	High
C14r I often have to take risks that make me feel uncomfortable about safety.	Risk Handling	The IA will likely be unaware of any discomfort on the human's part, but the human can probably utilise the IA's advice to err on the side of caution. Conversely, a risk-taker or someone who puts productivity first, may consult an IA until it finds a way to get around the rules (human ingenuity can be used for the wrong reasons).	Team	High
C04 The procedures describe the way in which I actually do my job.	Procedures and training	People know how to fill in the gaps when procedures don't really fit the situation, and it is not clear how an IA will do this. [This was in part why the earlier Expert Systems movement failed to deliver, leading to the infamous 'AI winter']. Also, the IA could conceivably record *work as done* and contrast it to *work as imagined* (the procedures). This would, over time, create an evidence base on procedural adequacy (see also C07r).	Org / Team	High
C08 I receive sufficient safety-related refresher training.	Procedures and training	The IA could take note of human fluency with the procedures and how much support it has to give, thus gaining a picture of whether more refresher training might be beneficial.	Org / Team	Medium
C11 Adequate training is provided when new systems and procedures are introduced.	Procedures and training	As for C08.	Org / Team	Medium
C19 The procedures associated with my work are appropriate.	Procedures and training	When humans find themselves outside the procedures, e.g. in a flight upset situation in the cockpit, an IA could rapidly examine all sensor information and supply a course of action for the flight crew.	Org / Team	High
C20 I have sufficient training to understand the procedures associated with my work.	Procedures and training	As for C08 and C11.	Org / Team	Medium

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

Appendix B Preliminary Safety Culture Impacts Derived from Analysis of Safety Culture Questionnaire

Table 5: Preliminary Safety Culture Concerns and Affordances Identified

Safety Concerns	Safety Affordances
Humans may become less concerned with safety if the IA is seen as handling safety aspects. This is an extension of the 'complacency' issue with automation, and may be expected to increase as the IA's autonomy increases.	The IA could 'speak up' if it assesses a course of action as unsafe.
Humans may perceive a double-bind: if they follow 'bad' IA advice or fail to follow 'good' advice, and there are adverse consequences, they might find themselves being prosecuted. This will lead to lack of trust in the IA.	The IA could be integrated into Crew Resource Management practices, helping decision-making and post-event review in the cockpit or Ops Room.
If the IA reports on human error or human risk-taking or other 'non-nominal behaviour' it could be considered a 'snitch' for management, and may not be trusted.	The IA could serve as a living black box recorder, recording more of decision-making strategies than is the case today.
If IA recordings are used by incident and accident investigators, Just Culture policies will need to address such usage both for ethical reasons and to the satisfaction of the human teams involved. Fatal accidents in which an IA was a part of the team are likely to raise new challenges for legal institutions.	If the IA is able to collect and analyse day-to-day safety occurrence information it may be seen as adding objective (dispassionate) evidence and a more balanced assessment of severity, as well as an unbiased evaluation of how close an event came to being an accident (e.g. via Bayesian analysis).
An IA that is human-trained may adopt its human trainers' level of risk tolerance, which may not always be optimal for safety.	The IA could significantly increase reporting rates, depending on how its reporting threshold is set, and could also record and track how often an issue is raised.
The introduction of Intelligent Assistants may inexorably lead to less human staff. Although there are various ways to 'sugar-coat' this, e.g. current shortfalls in staffing across the aviation workforce, it may lead to resentment against IAs. This factor will likely be influenced by how society gets on with advanced AI and IAs .	The IA could serve as a safety encyclopaedia, able to give instant information on safety rules, risk assessments, hazards, etc.
If the IA queries humans too often it may be perceived as policing them, or as a troublemaker.	The IA can upload all NOTAMs and briefings etc. to be able to keep the human team updated, or to advise them if they have missed something.
If the IA makes unsafe suggestions, trust will be eroded rapidly.	If the IA makes one really good 'save', its perceived utility and trustworthiness will increase.
The IA may have multiple priorities (e.g. safety, environment, efficiency/profit). This may lead to advice that humans find conflicted or confusing.	The IA could share important day-to-day operational observations, e.g. by flight crew, controllers, or ground crew, who can pass on their insights to the incoming crew.
Management may come to see the IA as a more manageable safety asset than people, one where they can either 'turn up' or 'tone down' the accent on safety.	The IA could reduce the reporting 'burden' of operational staff by transmitting details of human concerns and safety observations directly to safety departments. An IA 'network' would allow safety departments to quickly assess safety issues and

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332

	prepare messages to be cascaded down by senior/middle management.
Operational managers may simply be reluctant to allow the introduction of IAs into the system, due to both safety and operational concerns.	The IA could provide objective input for safety investigations, including inferences on causal and contributory factors. Use of Bayesian inference and other similar statistical approaches could help avoid typical human statistical biases, thereby ensuring the right lessons are learned and are considered proportionately to their level of risk.
If information is biased or counterfactual evidence is not considered, the way the IA judges risk may be incorrect, leading to a lack of trust by operational people. It could also have managers focusing on the wrong issues.	IAs could store information on incidents and associated (correlated) contextual factors, with live updates structured around risk models, and disseminate warnings of potential hazards on the day via an App or via the IA itself communicating with crews / staff.
There is a risk that if the IA is a very good information collector, people at the sharp end may be gradually excluded in updates to system changes, as the systems developers will consult data from the IA instead.	The IA might serve as a bridge between the way operational people and safety analysts think about risks, via considering more contextual factors not normally encoded in risk assessments.
There could conceivably be a clash between an IA and a team member who, for example, was taking significant risks or continually over-riding / ignoring safety advice, or an IA that was giving bad advice.	The IA could build up evidence of procedures that regularly require workarounds or are no longer fit for purpose. The IA could highlight gaps between 'work as designed', and 'work as done'.
IAs may need regular maintenance and fine-tuning, which may affect the perceived 'stability' of the IA by Ops people, resulting in loss of trust or 'rapport'.	IAs used in Ops could transmit information to other departments concerning engineering and maintenance needs observed during operations.
The IA may give advice that does not make sense to the human team or the organisation yet be unable to explain its rationale. Managers may find it difficult to adhere to such advice.	The IA could have a model of how things work and how safety is maintained, so that any changes will need to be incorporated into the model, which may identify safety issues that have been overlooked or played down. This is similar to current use of AIs for *continuous validation* and verification of operating systems, looking for bugs or omissions.
A human risk-taker or someone who puts productivity first, may consult ('game') an IA until it finds a way to get around the rules.	The human can utilise the IA's safety advice to err on the side of caution, if she or he feels pressured to cut safety corners either due to self, peer or management pressure.
People know how to fill in the gaps when procedures don't really fit the situation, and it is not clear how an IA will do this. Their advice might not be so helpful unless they are human-supervisory-trained across an exhaustive range of scenarios.	When humans find themselves outside the procedures, e.g. in a flight upset situation in the cockpit, an IA could rapidly examine all sensor information and supply a course of action for the flight crew.

© Copyright 2025 HAIKU Project. All rights reserved

This project has received funding by the European Union's Horizon Europe research and innovation programme HORIZON-CL5-2021-D6-01-13 under Grant Agreement no 101075332