

Article

Human–AI Teaming in the Cockpit: From Theory to Practice—Views from the Flightdeck

Esther Kox *, Jeroen van Rooij , Rolf Zon and Barry Kirwan 

Department Aerospace Operations: Safety & Human Performance, Netherlands Aerospace Centre (NLR), Anthony Fokkerweg 2, 1059 CM Amsterdam, The Netherlands; jeroen.van.rooij@nlr.nl (J.v.R.); rolf.zon@nlr.nl (R.Z.); barrykirwan@bkhf.onmicrosoft.com (B.K.)

* Correspondence: esther.kox@nlr.nl

Abstract

Crew Resource Management (CRM) in aviation is grounded in principles derived from human–human teamwork, such as the Big Five model of Teamwork (B5T). With the advent of AI-based systems, which could one day take on increasingly complex roles in the cockpit, it remains unclear whether these principles will remain applicable or risk becoming outdated. This study examines how human teamwork concepts translate to use-case-based human–AI collaboration. Five contemporary aviation human–AI teaming use cases are examined by two groups (four Human Factors experts and four professional pilots who are also CRM experts) using B5T and CRM as frameworks. Although the two expert groups are small and are reviewing use cases rather than observing actual pilot behaviour, these early findings suggest that teamwork (B5T) elements and CRM core competencies can be useful for identifying potential positive and negative impacts of AI-based systems on human teamworking, and hence operational risk. Candidate augmentations to B5T are identified to increase its adaptation to human–AI teaming system assessment, and preliminary new CRM training avenues are proposed to enable pilots to better accommodate future AI assistance in the cockpit.

Keywords: human–AI teaming; artificial intelligence; aviation; Crew Resource Management (CRM); teamwork; big five teamwork model

1. Introduction

1.1. Background

In commercial aviation flight crews are often composed of individuals with little or no prior experience of working together. Given the large pool of pilots qualified on a given aircraft type, recurrent ‘same crew’ pairings may occur only once or a few times over a career. Nevertheless, crews are expected to function as a highly coordinated unit from the moment they enter the cockpit. In this high-stakes environment, effective teamwork is essential for safety, operational efficiency, and flight success.

Decades of research in psychology have focused on understanding and supporting human–human teamwork, resulting in models such as Salas et al.’s Big Five Teamwork (B5T) model [1] and Crew Resource Management [2]. Central to both are mechanisms such as shared mental models, closed-loop communication, and mutual trust, which enable coordination, alignment, and effective team performance.

These same mechanisms are frequently referenced in human–AI teaming (HAT) literature [3–8], despite the fact that their application to the collaboration between humans



Academic Editors: Douglas D. Boyd and Elizabeth L. Blickensderfer

Received: 31 July 2026

Revised: 14 September 2026

Accepted: 24 September 2026

Published: 29 September 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

and AI agents is not straightforward. Unlike human teammates, an AI-based system (AS) does not possess shared cognition, intent, or social reciprocity, which complicates the direct transfer of established teamwork principles. While many criteria for effective teamwork, such as clear communication, shared situational awareness, effective workload management, and appropriate use of authority remain relevant in human–AI interaction, achieving these outcomes becomes fundamentally more complex when one of the actors is non-human.

In terms of closed-loop communication (i.e., ensuring that information is received, understood, and appropriately acted upon), AI does not “understand” information in the human cognitive sense, although AI may receive information and appropriately act upon it. Likewise, AI does not form mental models as humans do, instead often relying on stochastic analysis of extensive ‘trained’ datasets. AI ‘computes’ rather than thinks, and at any one time can be said to have a *representation* of what is happening, rather than a mental model in the human sense, and hence often requires ‘explainability’ to help human crews discern if they and the AI are aligned or ‘on the same page’. Similarly, for AI reciprocally ‘understanding’ human intent, AI infers and predicts intent from human behavioural cues and patterns, analogous to behaviourist learning, in which agents respond to learned associations. Finally, interpersonal mutual trust, rooted in intent, empathy and reciprocity, does not map cleanly onto interactions with non-sentient systems such as AI.

Similarly, the concept of human–AI teaming (HAT) is widely used in the literature, yet its validity remains contested. Teaming is traditionally understood as a fundamentally social phenomenon, characterised by interdependence, shared intent, and affective and motivational dynamics [9]. From this perspective, it can be argued that many current forms of human–AI interaction do not fully meet the criteria of genuine teaming, and may be more accurately described as advanced forms of automation [10].

At the same time, simply rejecting the notion of HAT does not resolve the issue. In practice, humans increasingly interact and coordinate with AI systems in ways that resemble collaboration with other agents, particularly in complex domains such as aviation. Moreover, existing operational frameworks, such as CRM, are grounded in human–human teamwork theory. The introduction of AI systems raises the question of whether these principles remain applicable, require adaptation, or risk becoming outdated in human–AI work contexts.

Early work by researchers at NASA explored the relationship between CRM and HAT, arguing that established CRM principles could inform the design of effective HAT concepts, while also recognising that CRM itself may need to evolve to accommodate increasingly capable AI systems [11]. More recently, Salmon et al. [12] examined the applicability of Salas’ B5T model in HATs and highlighted the lack of empirical research investigating whether its five teamwork behaviours and three coordinating mechanisms can meaningfully be observed in real-world human–AI interaction. Together, these studies point toward a common challenge: although existing teamwork theory continues to shape the way HAT is conceptualised, its practical applicability in contemporary operational settings remains insufficiently understood.

1.2. Study Aims

As developing HAT use cases increase in maturity, offering new AI-based systems (ASs) concepts to assist pilots in the cockpit, the challenge from both Human Factors and flight crew perspectives is to ensure that teamwork is not being degraded, and is preferably being enhanced, or at least safeguarded. Approaches are therefore required to examine teamwork in such use cases at the design and development stage. Consequently, this study aims to evaluate whether B5T and CRM approaches are relevant to

HAT AS assessment, either as they exist today or with minor refinement, and whether they yield useful insights on the impacts—both positive and negative—of AI on crew performance. Specifically, in this study we use B5T and CRM principles as two ‘lenses’ to assess five contemporary aviation use cases involving AI-assisted operations. Our aim is twofold:

- To assess which aspects of the B5T theory remain applicable when applied to the HAT use cases, and to identify areas where the B5T model may need updating to better accommodate such future operational AI-based systems (ASs).
- To determine the impact of HAT use cases on CRM and determine how CRM might need to adjust in relation to these use cases, to at least maintain the same level and quality of teamwork in the cockpit and ideally improve it.

The first ‘lens’ is more theoretical and can lead to the identification of key research areas as well as practical insights to help assure the design and safety of future teamworking in HATs. The second is more practical, and ultimately of more interest to airlines and flight crew and can result in procedural and training recommendations.

1.3. Organisation of the Paper

Section 2 describes in more detail the B5T model and CRM, and what they have in common. Section 3 outlines the five use cases, adapted from two recent European aviation HAT research projects. Section 4 describes the qualitative methodology adopted, and Section 5 details the results of the two analyses. Section 6 discusses the results and their implications for B5T theory and CRM practices, as well as study limitations and further research. Conclusions are summarised in Section 7.

2. Description of Two Teamwork Lenses (B5T and CRM)

2.1. Theoretical Framework-The Big Five Model (B5T)

Research on interpersonal teamwork has produced a wide range of models aimed at explaining the factors underlying effective collaboration. One of the most prominent within the Human Factors and Ergonomics domain is the B5T, developed by Salas and colleagues [1,13]. The framework was originally developed through a synthesis of a large number of existing teamwork models. The model conceptualises teamwork as a set of interrelated behavioural processes that contribute to team effectiveness. It distinguishes between five key behavioural dimensions that support team effectiveness: leadership, adaptability, team orientation, back-up behaviour, and mutual performance monitoring. These behaviours are enabled through three underlying coordinating mechanisms: shared mental models, mutual trust, and closed-loop communication, which facilitate alignment and coordination between team members (Table 1).

Table 1. Overview of components and definitions of Salas et al.’s B5T on human–human teamwork [1].

Type	Component with Definition
Teamwork behaviour	Team Leadership: the ability to direct, coordinate, and motivate team members while establishing goals, expectations, and task priorities.
	Mutual Performance Monitoring: the ability to develop common understandings of the team environment and apply appropriate task strategies to accurately monitor teammate performance.
	Back-up Behaviour: the ability to anticipate other team members’ needs through accurate knowledge about their responsibilities.

Table 1. *Cont.*

Type	Component with Definition
	Adaptability: the ability to adjust strategies based on information gathered from the environment through the use of back-up behaviour and reallocation of intrateam resources. Altering a course of action or team repertoire in response to changing conditions. Team Orientation: the propensity to take other's behaviour into account during group interaction and the belief in the importance of team goals over individual members' goals.
Coordinating mechanisms	Shared Mental Models: an organising knowledge structure of the relationships among the task the team is engaged in and how the team members will interact. Closed-loop Communication: the exchange of information between a sender and a receiver irrespective of the medium. Mutual Trust: the shared belief that team members will perform their roles and protect the interests of their teammates.

Since its introduction, the model has been widely cited across a broad range of domains. Yet recent work has highlighted two key gaps: limited empirical evidence linking B5T components to team effectiveness, and a lack of studies applying the framework to human–AI teaming contexts [12]. Accordingly, further research is needed to strengthen the empirical basis of the model and to examine its validity in HAT settings.

2.2. Practical Framework—Crew Resource Management (CRM)

The origin of CRM (originally Cockpit Resource Management, updated later to Crew Resource Management) is usually traced to a National Aeronautics and Space Administration (NASA) workshop in 1979 tackling the human error causes of aviation accidents, seen principally as interpersonal communications, decision-making, and leadership [2]. The term CRM was adopted to encompass training crews, and concerns the reduction of “pilot error” by making better use of the human resources on the flight deck. Since those early origins, CRM has become one of the major Human Factors ‘success stories’, credited with significantly reducing accidents, with CRM a mandated requirement for flight and cabin crew in the US (Federal Aviation Administration, FAA) and Europe (the European Union Aviation Safety Agency, EASA).

Since its inception CRM has passed through six ‘generations’ [2,14–17]:

1. First generation: an emphasis on individual psychology, including behavioural corrections, leadership style, authority gradients, and assertiveness.
2. Second generation: a transition from the individual to cockpit group dynamics, including briefings, situation awareness, stress and decision-making.
3. Third generation: a broader range of training topics, with an emphasis on crew dynamics both in and out of the cockpit, including automation management, and an extension to dispatch personnel, maintenance personnel, and cabin crew.
4. Fourth generation: the integration and proceduralisation of CRM into technical training, with FAA and EASA enabling airlines to customise their training programmes.
5. Fifth generation: acknowledgement that human error is inevitable and provides information to improve safety standards (e.g., avoid, trap, and mitigate errors).
6. Sixth generation: extension of Threat and Error Management (TEM) to include environmental and system threats, and a focus on anticipatory behaviour.

A ‘seventh generation’ of CRM has been suggested [18] concerning bringing CRM to encompass the entire organisation (e.g., airline) via ‘corporate CRM’. This is in order to make maximum use of resources in the organisation (including for example, senior management), and not only in the cockpit and related operational elements. However, this is not yet widely embraced, and therefore the model is considered to have six generations.

Nowadays CRM is incorporated in widely used aviation training methodologies such as Competency-Based Training and Assessment (CBTA), and Evidence-Based Training (EBT) [19]. With the use of well-described competencies, a standard is set for pilot performance, covering both technical competencies and CRM competencies. These are used throughout the entire training curriculum of pilots and, within EBT, and for assessment. Not only do individual pilots receive feedback on these competencies, but the whole training design is based, evaluated and adjusted through feedback on the competencies gathered in training, line-flights (safety data) or by identifying trends outside of the airline [20]. Within CBTA and EBT the competencies are called ‘pilot core competencies’ and are defined by a description and corresponding ‘observable behaviours’ (OBs), which serve as behavioural markers [21]. Below, the pilot core competencies are given in Table 2.

Table 2. Overview of the EASA pilot core competencies and descriptions (EASA, 2025 [21]) *.

Type	Pilot Core Competency with Description
CRM competency	Leadership and Teamwork: influences others to contribute to a shared purpose. Collaborates to accomplish the goals of the team.
	Situation awareness and management of information: perceives, comprehends, and manages information and anticipates its effect on the operation.
	Communication: communicates through appropriate means in the operational environment, in both normal and non-normal situations.
	Problem solving and decision-making: identifies precursors, mitigates problems, and makes decisions. Workload management: maintains available workload capacity by prioritising and distributing tasks using appropriate resources.
	Application of procedures and compliance with regulations: identifies and applies appropriate procedures in accordance with published operating instructions and applicable regulations.
Technical Competency	Aeroplane Flight Path Management, manual control: controls the flight path through manual control.
	Aeroplane Flight Path Management, automation: controls the flight path through automation.
General competency	Application of Knowledge: demonstrates knowledge and understanding of relevant information, operating instructions, aircraft systems and the operating environment.

* Application of procedures and compliance with regulations could be in both CRM and Non-CRM domains.

The two models clearly have overlaps, and a potential mapping of the two is given in Figure 1, developed by one of the CRM pilot-trainers who participated in the study.

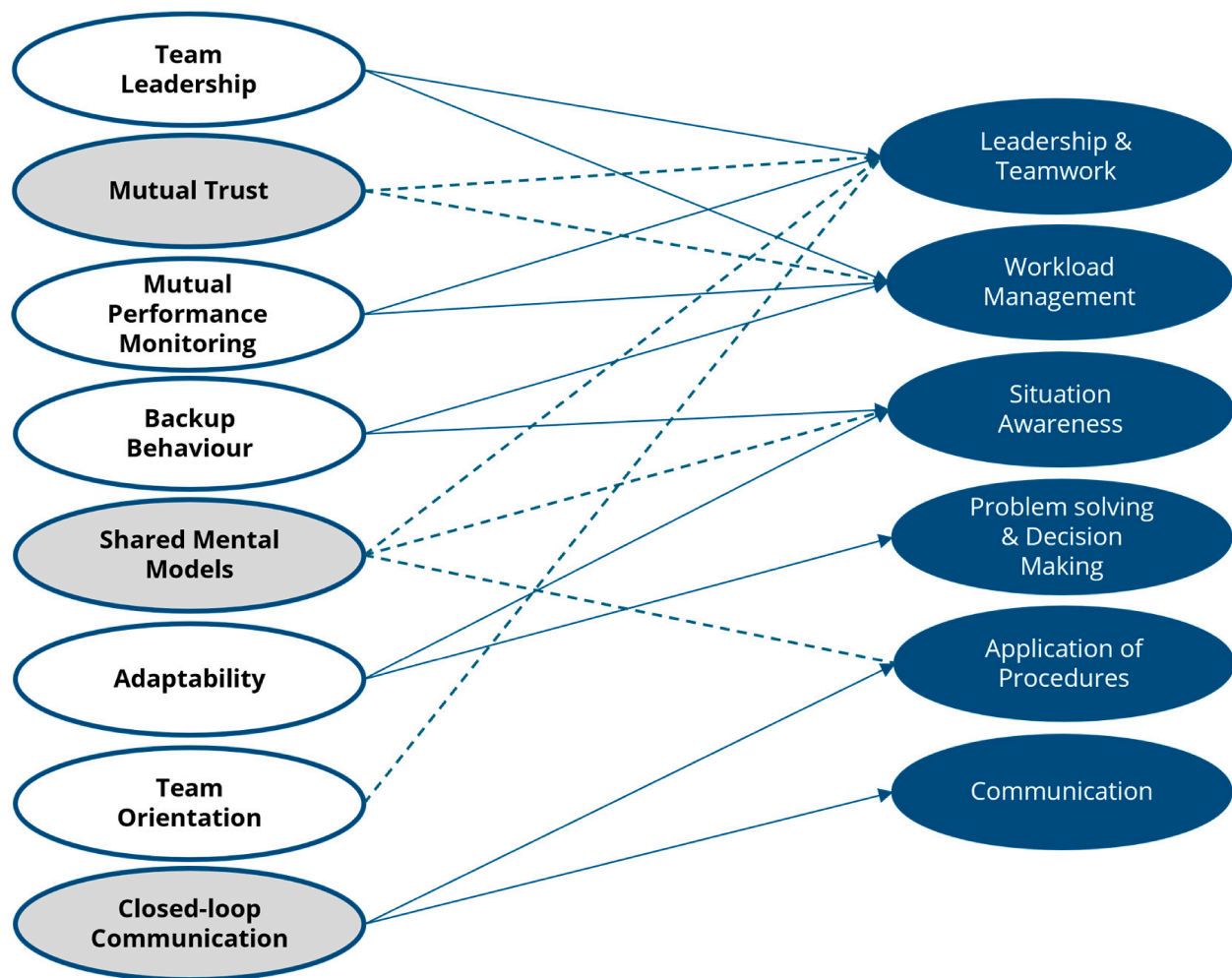


Figure 1. Relationships between B5T and CRM pilot core competencies. Direct relationships have a solid arrow, and indirect relationships a dashed arrow. On the left, the team behaviours of B5T are shown in balloons with a white background, and the coordinating mechanisms in balloons with a grey background. To the right are the CRM pilot core competencies.

3. HAT Use Cases

Having described the two teamwork ‘lenses’ that exist today, the next step is to contextualise HAT via use cases where AI prototypes have been developed and tested with licenced operational pilots, air traffic controllers and airport personnel. Since AI-based systems are not yet in operation, it is necessary to draw from leading-edge research projects. Two such EU Horizon Europe-funded projects were selected: HAIKU (2022–2025) [22] and JARVIS (2023–2026) [23]. Both of these projects developed use cases to Technology Readiness Levels (TRLs) 4–6, with supervised learning AI-based systems, and were tested with controllers and pilots in dynamic simulations using realistic scenarios in formal validation experiments.

In HAIKU, JARVIS and elsewhere, various terms have been used to denote the AI element: Digital Assistant, Intelligent Assistant, AI-based system, etc. However, cautions exist against applying anthropomorphic terminology [24,25], and hence the more neutral term ‘AI-based system’ is employed (abbreviated to ‘AS’) during the examination of use cases and the ensuing results and discussion.

Before describing the use cases in more detail, the European regulatory perspective on AI-based systems requires elucidation, as any new system (AI-based or otherwise) to be deployed in European airspace must comply with regulatory requirements from EASA in order to be certified for deployment and use.

3.1. Regulatory Perspective on AI in Operational Aviation Settings

3.1.1. EASA's Level of AI Autonomy Categorisation

Before describing and analysing the use cases it is relevant to categorise them in relation to the EASA human–AI teaming classification system [26]. EASA envisages six categories of future human–AI partnerships:

- 1A—Enhanced information via Machine Learning support.
- 1B—Cognitive assistant (equivalent to advanced automation support).
- 2A—Cooperative agent, able to complete tasks as demanded by the operator.
- 2B—Collaborative agent—an autonomous agent that works with human colleagues, but which can take initiative and execute tasks, as well as being capable of negotiating with its human counterparts.
- 3A—AI executive agent—the AI is basically managing the operation, but there is human oversight, and the human can intervene (sometimes called management by exception).
- 3B—the AI is running everything, and the human cannot intervene.

The five use cases inspired by HAIKU and JARVIS and described below concern EASA Levels 1A to 2B, so are representative of the full range of AI currently envisaged for aviation systems (Level 3 guidance is not yet addressed by EASA. It may be noted that EASA has issued new draft guidance (Issue 3) which is currently out for public consultation, but as it is a draft and may change, the authors have used the current published guidance issued in 2024.

The FAA in its AI Roadmap [25] notably prefers to talk of advanced automation rather than human–AI teaming, and eschews any personification or anthropomorphism related to AI. The trend in European aviation HAT research, despite use of the term ‘Teaming’, also avoids personification, anthropomorphism and emotional mimicry [27]. Rather, as stated earlier, the word teaming is maintained because a high level of AI autonomy implies collaboration and interdependence with human agents, and so the traits of teamwork (and therefore teaming) may apply.

3.1.2. When AI Is No Longer ‘Just More Automation’

As already noted, it has been argued that AI innovation, for all its benefits, is essentially advanced automation supporting the human operator [10]. The critical threshold in AI autonomy where this may no longer hold appears to be between EASA Level 2A and 2B [26], since 2B in particular is considered different from what we have today in civil aviation cockpits. An EASA Level 2A AS supports the end user (pilot) in achieving their goals (it can execute tasks); it is directed by the pilot and does not necessarily imply a shared situation awareness. Level 2B, in comparison, is a collaborative process in which the pilot and the AI-based system work together to achieve a predefined shared goal and solve a problem through a co-constructive approach. Collaboration implies the capability to share situation awareness and to readjust strategies and task allocation in real time. Communication is paramount to share valuable information needed to achieve the goal. This is highlighted in the EASA figure below (Figure 2).

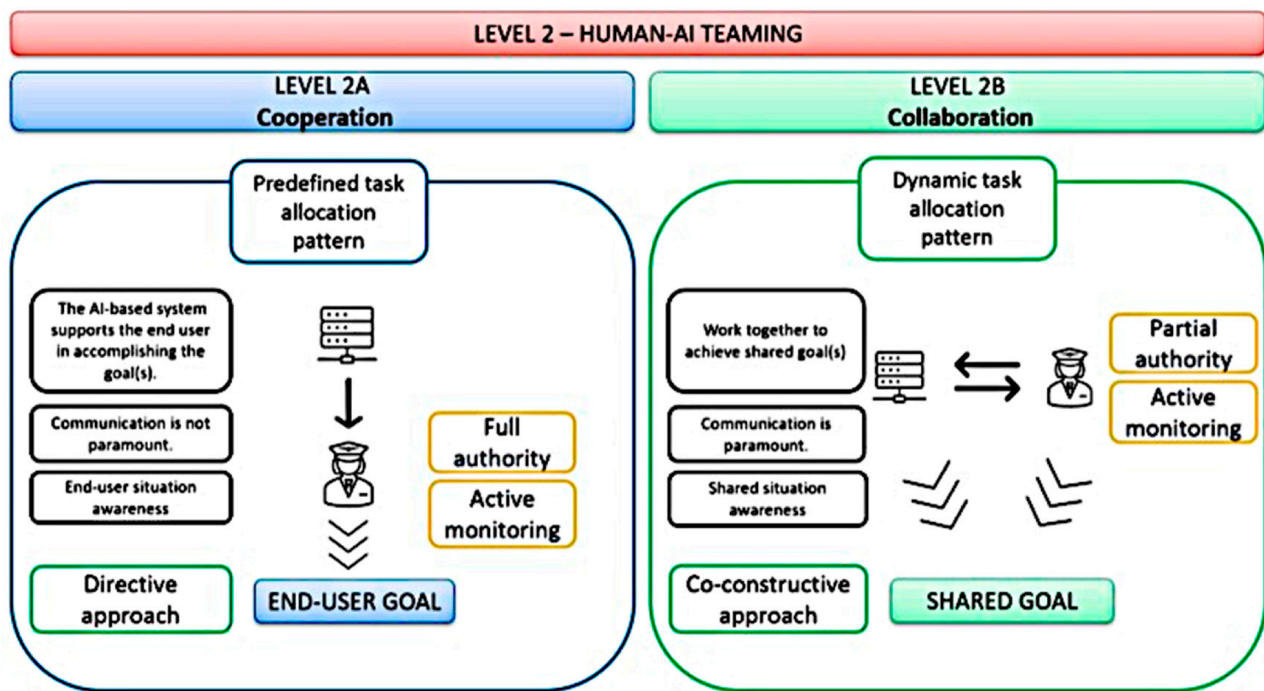


Figure 2. The difference between human–AI cooperation (level 2A) and human–AI collaboration (level 2B). Figure from [26].

3.1.3. EASA's Existing Guidance on CRM for AI-Based Systems

EASA, in their guidance on AI-based systems, adopt the acronym HAIRM—see Objective HF-27 in the EASA Guidance-Human–AI Resource Management [26]. Under this heading they outline potential future CRM issues:

- Failure to respect task allocation pattern(s) between human and AI.
- Teamwork breakdown, e.g., due to trust or communication issues.
- Incorrect or incomplete cross-check process (especially for single-pilot operations).
- Human–AI communication issues.
- Misalignment in situation awareness between flight crew and AI-based system.
- Decision-making authority issues (only for EASA Cat 2B).

Whilst these are useful 'headline' issues related to CRM (indeed they appear in the EASA guidance to be 'placeholders' pending further research), they are not detailed enough to specify the CRM accommodations required. Therefore, the next section sets out five HAT use cases so that future CRM requirements can be elaborated.

3.2. Use Cases

Four of the use-cases utilised for analysis are based on the original HAIKU and JARVIS use cases, with certain amendments as described in the individual use case (UC) sections below. One (UC4–Safety Advisor) was not an actual use case in either of these two projects but is currently under development by one of the HAIKU partners. The five use cases as described below are judged to give sufficient 'bandwidth' for the expert review of B5T and CRM aspects (see Table 3), and for a preliminary determination of the utility and operational relevance of the two perspectives for contemporary HAT system designs.

Table 3. Overview of five use cases *.

Use Case	Function	Shared Goal	Interaction Mode	Action Execution	Autonomy of End User (Pilot)	EASA AI Level of Autonomy
UC1 Voice to Text	Translate oral ATC instructions into text, so that the pilot can read the messages and the messages [1A] can also be forwarded into the aircraft systems, e.g., to automatically update radio frequency after receipt of ATC text [2A].	Error-free understanding of ATC instructions. Upon receipt of ATC message to change radio settings, AI will execute, saving pilot workload.	[Visual] CPDLC text display area in cockpit.	Pilot via normal controls.	Partial	1A/2A
UC2 RWY/TWY Predictor	Predict the departure or arrival RWY and time (with confidence levels), and the TWY route to take in order to travel from gate to RWY or vice versa.	Begin approach, landing and taxi preparations earlier when approaching a less familiar airport with multiple runways where approach RWYs change frequently. TWY guidance to support situation awareness if flight crew unfamiliar with the airport TWY layout.	[Visual] Multi-Functional Display (MFD).	Pilot via normal controls.	Full	1B
UC3 Flight Re-Route Advisor	Identify three candidate airports when a major diversion is required in-flight. Candidates selected must optimise pre-defined criteria.	Determine the best alternate airport for aircraft, crew, passengers and airline, pilots negotiating with AI. Once selected, AI changes course.	[Visual] Alternate airports shown on a moving map display. Criteria and optimisations on MFDs. Keyboard Interaction.	Pilots elect alternate and AS enters it into the FMS and changes course.	Partial	2A-2B
UC4 Safety Advisor	Support pilots when things go wrong in-flight and the causes, consequences and required responses are not clear, helping them understand the problem and gain rapid situation awareness of what is happening, along with the provision of guidance on how to rectify the fault or mitigate its consequences so the flight can safely continue.	Determine (collaboratively) the causes, consequences and solutions for an unclear system disturbance (electrical, mechanical, environmental). This will require pilot–AI interaction via ‘dialogue’, including the ability for the pilots to pose counterfactual questions.	[Visual] Text and diagrams on both the Text area and MFD. Textual dialogue (pilots asking the Safety Advisor questions and receiving answers in text / diagrams).	Pilot via normal controls. Text entry via cockpit keyboard. At a later development point, pilots can ask oral questions (via AI NLP).	Partial	1B-2B
UC5 Startle Response Support	Physiological detection of startle response, support in overcoming the cognitive impairment, and directed situation awareness so that the flight can be brought under stable control.	Recover from sudden and unexpected flight upset/loss of control event.	[Visual] Pulsing green light to coordinate breathing. [Visual/Audio] Highlighting of key flight instruments using red boxes and audio callouts.	Pilot via normal controls.	Full	1B

* N.B: in all 5 use cases the pilot(s) can override or disable the AI at any point.

For each use case, we describe the current operational procedure, the proposed AI assistance concept, its intended benefits, and the AI capabilities required for successful implementation. Figure 3 illustrates where visual information for the flight crew could hypothetically appear in the cockpit (this figure is based on a dual pilot configuration with Pilot Flying (PF) in the left seat and Pilot Monitoring (PM) in the right-hand seat).



Figure 3. Highlighted B787 cockpit display elements for the 5 use cases. Highlighted Areas: A and B = Text area (e.g., for CPDLC (Controller-Pilot Data-Linked Communication)), C = Multifunction Display (MFD), D = Primary Flight Display. Human–AI Interaction Areas by Use Case: UC1: Voice to Text = A + B; UC2: RWY and Taxiway = A + B, C for TWY map; UC3: Major Re-Routing = A, B, C for route map and XAI; UC4: Safety Advisor = A, B, C; UC5: Startle Response Support = C + D.

3.2.1. Use Case 1: Voice to Text

Current System: For the pilot the current way of working comprises listening to the message that the ATCO speaks over the radio, considering their request, deciding to accept or not, and to take action accordingly. Many aircraft also can receive certain standardised and routine messages via CPDLC, effectively text messages displayed in the cockpit.

AI-Assistance Concept: The cockpit-AS comprises a feature for ATC instructions and clearance management, translating ATC speech into text, so that the pilot can read the messages and the messages can also be forwarded into the aircraft systems. In the latter case, for example, the system can autonomously set the standby radio frequency, while the pilot remains in control and can override the change. This automation uses Machine-Based Learning (Natural Language Processing [NLP]) and is considered to be primarily EASA Level 1A (speech recognition component), with an element of 2A (changing radio settings). This AI concept of use has not been modified for the analysis (i.e., it is as it was in the JARVIS project).

The text message will appear in different locations depending on the type of aircraft. Figure 3 above shows where it would appear for a Boeing 787, a comparatively advanced cockpit, similar in level of automation to an Airbus A350. Usefully in the B787, the message can appear in two places, for the Pilot Flying (PF) and the Pilot Monitoring (PM). On some other aircraft the message appears on instrumentation in the console between the two flight crew members.

Intended Benefits: There are a number of potential advantages to this kind of AS:

- It can help prevent and recover from miscommunications, hear-back/readback errors, and callsign confusion (where two aircraft have similar ATC callsigns, e.g., BAW 232 and BAW 332).
- It can lead to less need for clarifications or requests to repeat the ATC instruction, thereby making air traffic control more efficient and reducing ATC workload.
- It can help overcome difficulties in comprehension caused by noise, static, accents, distortion, etc.

AI high-level requirements: The speech-to-text system must be near-perfect and seamless to use (e.g., not requiring separate log-in procedures by the crew for each new ATC sector they enter).

3.2.2. Use Case 2: Runway (RWY) and Taxiway (TWY) Predictor

Current system: Currently pilots must take into account multiple possible departure and arrival routes when planning their flights. When they contact ATC (both for departure as well as arrival) they will receive instructions to go to their departure RWY, or when they approach, what their arrival RWY and gate will be. Taxiing on an airport surface that a pilot is not familiar with requires significant attention, and during this taxiing phase other processes may be taking place in the cockpit (e.g., reviewing the take-off checklist). Depending how familiar pilots are with the airport they might have limited knowledge about what to prepare for, whether for the take-off or the arrival.

AI Assistance Concept: The AS predicts the departure or arrival RWY and the TWY route to take to travel from gate to the RWY or vice versa. The AS also offers a confidence level for RWY prediction for multi-RWY airports and can explain how it came to its prediction. Information is merged with the navigation display so that flight crew can integrate the information from the AS into their normal way of working. In relation to the EASA categorisation of level of autonomy, this is EASA Level 1B. This JARVIS use case originally mainly concerned RWY selection—it has been extended here to include TWY guidance, as this is a useful function for pilots, and can be based on similar AI architecture.

Intended Benefits: Such a prediction enables the flight crew to start preparations earlier, e.g., during cruise prior to contacting approach. This could prove useful when approaching an airport that pilots are less familiar with, or an airport with multiple runways where approach RWYs change frequently. TWY guidance (both from landing and from pushback) will support situation awareness if the flight crew are unfamiliar with the airport TWY layout system and can reduce cockpit-ATC instruction and clarification requests, and could ultimately help prevent RWY incursions and selection of incorrect TWYs which can lead to TWY blockages or even TWY collisions.

AI High-Level Requirements: The AS needs a great deal of information from different sources. It needs the flight plan and information about the airport: factual, or fixed, information such as location, RWYs TWYs; variable information such as NOTAMS (Notices to Airmen); and a database of which RWYs and TWYs are normally used at the airport under a range of circumstances.

3.2.3. Use Case 3: Flight Re-Route Advisor

Current System: It is standard practice for commercial flights to have one or two 'alternate' airport destinations in case the original destination airport becomes unavailable. These alternates are usually in the same geographical area as the destination. However, if these alternates become unavailable, e.g., due to a major weather pattern such as a thunderstorm, or unanticipated military activity, a new airport will need to be identified. This unusual situation was explored during the Future Sky Safety research project [28] with ten aircrews (PF and PM) during a set of simulations in a full flight

simulator. The flight crew uses the procedures plus the Electronic Flight Bag (EFB) and other instruments/documentation sources to select between candidate airports, principally according to airport characteristics (so the particular aircraft type can land there) distance, fuel availability, and weather. Other factors to consider include ease of passengers continuing their journey without having to spend a night in the hotel (at the airline's cost), availability of a maintenance/services contract between the airline and the airport, as well as calculations concerning aircraft weight on arrival and runway length, and checking of NOTAMS related to the airport. In the Future Sky Safety project, these checks, as well as discussions between PF and PM about preferences based on local knowledge, and possible communications with the airline's Operations Control Centre, could easily take 12–15 min, during which time the aircraft continues flying, using up fuel, and cannot inform passengers where they are headed.

AI Assistant Concept: During the HAIKU project, a prototype AI algorithm embedded knowledge of all European commercial airports including real-time updates related to NOTAMS (Notice to Airmen: those affecting runway/taxiway work, unserviceable ground equipment, temporary obstacles, etc.), METeorological Aerodrome Reports (METAR), Significant Meteorological Information (SIGMET) and Pilot Reports (PIREP). The AS algorithm, based on hundreds of actual diversion cases had undergone human-supervised training to reflect all that pilots would typically consider, including their preferences, and hence give responses that pilots would typically choose. The processing time was 45 s instead of 12+ min to deliver three viable destinations, based on 12 factors. The PF and PM can review the ratings' explainability of the choices [29,30] if they wish, or they can simply select the first one or their preference of the three options. There is a degree of interaction and negotiation with the AI to decide the final destination, especially if the pilots decide to re-prioritise the 12 factors. This would possibly yield a different ordering of the airports suggested, or even trigger the identification of a different (but still viable) airport. Once the airport is selected, the AI uploads the destination and route into the FMS and changes heading. Because of this pilot–AI negotiation element, this AS sits between EASA Levels 2A and 2B.

Intended Benefits: The AS gives a much quicker response, saving ten minutes on average of fuel burn (including possibly heading in the wrong direction). It also ensures that all the key factors have been considered (in case the flight crew forget one) and can counter pilot selection biases. Its ratings of the factors are transparent to the crew. It considerably reduces the flight crew's cognitive load on what are principally information gathering and multi-parameter choice selection tasks.

AI High-Level Requirements: Airport and meteorological data must be up to date. A straightforward visual interface is required, either with easy textual query or Natural Language Processing (NLP) querying functionality. The latter would enable counterfactual reasoning, e.g., if the pilots ask "What about airport XYZ? Why isn't that in the list?"

3.2.4. Use Case 4: Safety Advisor

Current System: Today, when faults or system disturbances occur, if they are standardised and known there will be alerts and alarms for the flight crew, and the flight crew follows a checklist to determine what courses of action to take. However, sometimes faults occur that are less clear in their nature, origin and impacts. For these, the flight crew may need to carry out real-time diagnosis of what has happened, what it might lead to, and how to recover safely. Flight crew are trained in many different failure scenarios, but less so on complex technical failures (e.g., wiring connection failures, faulty sensor readings, electrical failures, etc.). Study of such scenarios during the Future Sky Safety research project [28] found such diagnosis very labour intensive for the PM in particular,

with much ‘head down’ time interrogating checklists, displays and instruments, and the Electronic Flight Bag (EFB). More generally, flight emergencies which are confusing to the flight crew remain a major risk category in aviation safety. Loss of Control in Flight (LOC-I) is EASA’s top risk category in European aviation (EASA refer to potential and actual LOC-I as ‘Aircraft Upset’). Relevant aviation accidents are Air France 447 (2009) [31], Air Asia 8501 (2014) [32] and the two B737 Max accidents (2018; 2019) [33]. Robust AI safety advice during such complex and confusing events would be a welcome protection from further catastrophe.

AI Assistance Concept: The idea of the AI-based Safety Advisor is to help the flight crew deal with a system disturbance event, whether time-pressured and dangerous, or not, gaining a quick and accurate diagnosis and resolution proposals. Early prototypes have been tested in aviation research scenarios in HAIKU (for startle response-see UC5) and HARVIS (Human Aircraft Roadmap for Virtual Intelligent System: for unstabilised approaches: [34]) projects, as well as ongoing post-HAIKU work. The future aim would be to cover a more extensive set of hazards or failure events, along with a broader-based small language model interface so that the flight crew can pose direct questions, with answers displayed on the cockpit Multi-Functional Display (MFD), and/or spoken. In the case of high urgency diagnoses (e.g., electrical faults causing sub-systems failures affecting flight stability, or even a cyber-attack affecting multiple systems), it is likely that flight crew will have to accept the AI’s diagnosis and resolution recommendation in order to stabilise flight conditions or avoid irrecoverable loss of control. The EASA Level ranges from 1B (decision support), to 2B due to the collaborative nature of the diagnosis and resolution determination phases, and the cases where the pilots accept the AI’s diagnosis due to time pressure, imminent danger and event complexity. In this latter case, the AI sets the goal/task and the flight crew executes it.

Intended Benefits: the aim is to support pilots when things go wrong in flight and the causes, consequences and required responses are not clear, helping them to understand the problem and gain rapid situation awareness of what is happening, along with the provision of guidance on how to rectify the fault or mitigate its consequences so that the flight can safely continue.

AI High-Level Requirements: Such an AS would need comprehensive development and testing to maximise accuracy. Collaborative dialogue should include the ability to pose counterfactuals (e.g., the pilots may have alternative diagnoses and want to know why they (or sub-elements) were not considered by the AI), and consider differential diagnoses (i.e., when symptoms match more than one cause). In order not to increase workload, a natural language interface would be preferable, i.e., pilots ask oral questions, and the AS responds either orally or in text and diagrams.

3.2.5. Use Case 5: Startle Response Support

Current System: In certain cases the flight crew may be subject to a ‘startle response’ event (e.g., see [35–37]). This is a psychophysiological phenomenon wherein the person affected suffers cognitive impairment for a short period of time (e.g., 20–30 s), for example in sudden unexpected events such as a lightning strike. In such cases flight crew may be unable to diagnose and react quickly to their situation, either because they are unfamiliar with the situation (it is beyond their experience and training), or because they are temporarily cognitively impaired.

AI Assistance Concept: In the HAIKU project a preliminary prototype of flight support was developed [38] in three components: physiological detection of startle response, support in overcoming the cognitive impairment, and directed situation awareness so that the flight can be brought under stable control. The intended application in HAIKU was for a

single pilot, rather than contemporary dual pilot configuration. The startle event was a simulated lightning strike, with a very loud bang and flash just outside the cockpit. The first startle support component relies on continuous physiological monitoring (e.g., heart rate, breathing rate, electrodermal activity, etc.), and is calibrated to the individual pilot (a baseline can be created during flight if necessary). The second element is a guided breathing system as recommended by EASA [37] to recover more quickly from cognitive impairment. This involved a slowly pulsing green ambient lighting on the cockpit dashboard, signifying a ‘breathe in’ portion of time {green on} and breathe out portion {green off}. The third element relies on active eye movement tracking, contrasting what the pilot is looking at (which instruments) in real time, compared to what they should be looking at according to Standard Operating Procedures (SOPs) and pilot flight recovery best practices. If there is a difference, this is highlighted on the cockpit displays as well as via oral ‘callouts’, e.g., “Vertical Speed”, telling them what to look at next. As soon as the pilot feels back in control, they can switch off the AI support, or the AS will disengage once the pilot’s eye track synchronises with its recommended instrument checking sequence.

AI High-Level Requirements: the physiological measurement equipment would need to be as unobtrusive as possible, and there are ethical considerations concerning pilot privacy of personal data (e.g., in relation to medical certification as being ‘fit to fly’). Similarly, eye movement tracking via special glasses might be unacceptable to some pilots. The directed attention approach needs to correspond to how most pilots would (i.e., should) react according to a range of situations, and presentation on the cockpit displays needs to be attention-directing without cluttering the displays or leading to added confusion.

The above are the five use cases, ranging from 1A to 2B in the EASA HAT classification system, selected for evaluation of B5T and CRM frameworks. The following section describes the methodology employed in the two expert reviews.

4. Method

4.1. Overall Design

This study adopts a twin-stream qualitative research design combining an expert review of the prevalence of the B5T model elements in the use cases, and practitioner-based evaluation of the potential impacts of the use cases on CRM. The methodology utilised is thus qualitative in nature and reliant on expert analysis of contextual use cases via two expert groups, the first comprising four Human Factors experts including one who is a B787 pilot, the second group comprising four airline pilots who are also involved in CRM training and development within their company. Without the use cases, any analysis risks remaining abstract, whereas the aim of the study is to prepare for real application of HAT in aviation transport, designated a safety-critical system according to both the EU Act on AI and EASA. Waiting for reliable quantitative data to examine teamworking with AI in realistic situations may be some years away, and therefore this case study approach is aimed at gaining a first understanding and estimation of the impacts of AI on teamworking. The research began in earnest in November 2025, with five B5T expert review meetings taking place between February and May 2026, and the CRM evaluation taking place during May 2026.

4.1.1. Specification of B5T and CRM Expert Review Research Questions

The focus of the B5T and CRM expert analyses of the five use cases was on identifying benefits of the AI-based system to the crew, and vulnerabilities caused by it, in normal and off-normal flight operations. This is generally referred to in the paper as ‘relevance’ of the two respective approaches to the identification of impacts on operational risk when working with AI-based systems, whether negative, e.g., disruption of situation awareness,

increase in workload, teamwork degradation, etc., or positive, e.g., increasing adaptability, reinforcing closed loop communications, etc. The two expert groups considered where the respective model elements helped the teamworking process, and where the models might need adjustment or augmentation to gain a better focus on the team–AI interaction.

The examination of the relevance of these two approaches for HAT assessment from a teamworking perspective was further elaborated into the following questions posed to both expert groups:

- What elements of either model are relevant in each use case?
- Are any relevant B5T/CRM elements missing?
- Are any elements missing important aspects or nuances?
- Are the impacts of the AS on B5T/CRM elements likely to be positive or negative?
- Are B5T and CRM suitable approaches for examining the teamworking impact of future AS and HAT concepts?

CRM Are new training strategies identifiable to enable flight crew to better manage AI-based only systems?

The first three lines of inquiry relate to the relevance, power and sensitivity of the two frameworks as applied to ASs and can be used to determine if detailed changes are advisable for either model. The fourth is of practical safety and design interest, and can result in additional research or safeguarding requirements for the AS. These four lines of enquiry converge into the fifth question, essentially asking whether the two models are appropriate for use when studying new AS and HAT concepts, perhaps with minor modification/augmentation, or whether more structural change is warranted, even to the extent of requiring a new framework or even a new paradigm. The final question is specific to CRM, as the introduction of AI-based systems into the cockpit may suggest new training requirements to ensure teamworking is not degraded by AI in the short or longer term.

4.1.2. Selection of HAT Use Cases

As noted in Section 3, the use cases cover EASA's 1A to 2B levels of AI autonomy, from augmenting information and aiding decision-making to task execution support and collaboration. UC5 is a single-pilot use case, thus creating more reliance on the AI, especially since it is assisting in an emergency event. The use cases, with the exception of UC4, have all been trialled with airline pilots, so that ways of working between human and AI have been evolved with, and tested by airline pilots in realistic simulations (UC4 is essentially a partial extension of UC5 across many more events). This level of contextualisation serves to 'ground' the deliberations of the two teams of experts (both teams contained pilot expertise).

However, since the five use-cases differ in task type, abnormality, time pressure, crew configuration, interface maturity and TRL, they cannot be used in a strict or classic experimental fashion, i.e., correlating certain findings by contrasting use-case observations according to level of AI autonomy, for example. Instead, they represent a use-case landscape that can be examined qualitatively by experts to draw insights and detect patterns within and across the different use cases. Ideally, if there were many such use cases a 'cleaner' set of differentiated use cases could have been selected. Unfortunately, a review of 27 European public-domain research projects on human–AI teaming [39] revealed only five eligible European research projects in the public domain that were cockpit-oriented, at TRL4 or above, and tested with licenced airline pilots. Of these, HARVIS is a precursor to UC5, and there was insufficient access to the operational details of SAFETEAM (<https://safeteamproject.eu>) (URL accessed on 23 September 2026) and DARWIN (<https://darwinai.eu>) (URL accessed on 23 September 2026) to generate detailed use case scenarios (the same applies to a recently finished study called LOKI (<https://www.dlr.de/en/me/research->

[and-transfer/projects-and-studies/loki](#)) (URL accessed on 23 September 2026). This left JARVIS and HAIKU as the only viable options available to the study authors. It is hoped that in future more use cases will become available.

4.2. Study 1: Theory-Driven Use Case Analysis (B5T)

4.2.1. Selection of Experts for B5T Analysis

Four experts were selected for the B5T analysis, all working for the same national aerospace research centre. Three were Human Factors experts working in the HAT area, familiar with the B5T model, one of whom had taken part in the HAIKU project (and knew UCs 3–5), another who had taken part in JARVIS project (and knew UCs 1 and 2 and a JARVIS correlate of UC3). The two HF experts, one of whom had detailed knowledge of HAIKU, the other of JARVIS, were not involved in the design/development of these projects' use-case systems. Rather they were involved in the Human Factors evaluations. As such they had no vested interest in the use-case AI systems nor their performance. Furthermore, this current study was not focused on which use-case or AI-based system performed best, instead only aiming to glean insights on AS impacts on teamworking-related operational risks, upon which neither JARVIS nor HAIKU had particularly focused. The fourth expert was an experienced senior airline captain trained and experienced in Human Factors, who was also a CRM trainer, and was also familiar with the B5T model. The team comprised one female and three males, aged 30 to 66 years, with an average of 30 years of experience (Human Factors/commercial flying). There was no 'team leader', and review meetings were run democratically.

4.2.2. Iterative Evaluation of B5T Elements in Each Use Case

The two experts who had previously worked on HAIKU and JARVIS described the use cases in detail to the other participants in the HF expert group, including discussions about which pilot (PF, PM or SPO) and pilot combination (e.g., PF and PM) would be involved in interactions with the AS function, and where AS outputs would appear in the cockpit interface (see Figure 3). A B787 was taken as the reference for all use cases, even if other aircraft models [A320 and an advanced cockpit interface for UCs 5 and 3] had been used in the original use case trials. Once the use cases were fully understood, each expert individually noted (using an Excel table), for each use case in turn, which elements were relevant for each use case, and then the group met to compare results, by comparing the Excel sheets and producing a combined one. Areas of agreement and disagreement were discussed, and it became apparent that a binary decision (yes/no) about element relevance was too broad. In subsequent meetings (five in total spread over three months), experts discussed their assessments of relevance until consensus was attained. The oral discussions were not recorded, rather the debates were captured in Excel and Word files updated and circulated after every meeting. Examples of the discussion points during this process are as follows:

- Initially for UC3, the three HF experts did not see it as relevant to *team leadership*. However, the 4th expert (professional pilot/HF expert) argued that, e.g., "*when AI assists in major re-routing finding an alternative airport to land (in case of emergency, thunderstorm, volcano, military activity, whatever) then it will influence team leadership (team direction and motivation), maybe overpowering "strong but wrong" leadership or help when both pilots aren't knowledgeable about which airport seems best.*" This argument was persuasive for the other three experts.
- In the first two meetings there was frequent disagreement over the elements *mutual performance monitoring* and *back-up behaviour*, as these concepts were seen as overlapping.

It was necessary to agree on how to discriminate the impact of these two elements, in particular in UCs 1 and 2 where most disagreement was occurring.

- There was initial disagreement on the element *mutual trust* in UC1. It was argued that having access to the same information would enhance the ability to crosscheck, and therefore interpersonal trust would be enhanced. After discussion this information symmetry was evaluated to reduce uncertainty and in fact *decrease* the need for interpersonal trust.
- The element *mutual performance monitoring* in UC3 was initially assessed as ‘relevant’. Later it was argued that not much changes with the introduction of the AS, i.e., mutual performance monitoring wouldn’t become better or worse compared to current operations. Its relevance was therefore downgraded.
- The fact that the AS would be checking aircraft status in UC5 was initially assessed as having impact on *mutual trust*. The argument was that pilot errors would become apparent and that this comparing of flight crew deliberations against own calculations was something akin to trust. Later discussion assessed this as performance monitoring from the AS. Furthermore, trust was deemed as a typically human–human construct, which involves a willingness to accept vulnerability based on the expectation that others are both willing and able to contribute to shared goals. It was agreed upon that this reciprocity does not exist in the crew–AS relationship.

There were mainly minor changes, plus a deepening of understanding concerning the nuances of the elements (e.g., what *mutual trust* means when talking about an AS). At the end of the fifth meeting the experts agreed no further changes were required, and ‘clean’ tables of results were finalised. The results of the B5T expert review are given in Section 5.1.

4.3. Study 2: CRM-Based Use-Case Analysis

4.3.1. Selection of Experts for CRM Analysis

For the CRM analysis team, four experienced airline pilots were selected who are part of their airline company’s CRM core team, which develops in-house CRM training. These subject matter experts comprised one female and three males (two captains and two first officers), aged 33 to 57 years, with an average of 23 years commercial flying experience. A CRM pilot who was involved in the B5T expert group was also there to ensure that any assumptions they made relating to the use cases team dynamics were concordant with the B5T expert team assumptions. Similarly, one of the first (B5T) expert team’s Human Factors experts also attended in case there were identifiable B5T impacts not discovered in the B5T analysis. These two experts functioned as facilitators only, not influencing the deliberations of the four other CRM experts, thus aiming to preserve independence between the two assessments and expert groups. However, given that there was not full independence between the two groups, findings may be best framed as project-informed propositions rather than results of independent validations. The Role Matrix in Table 4 below shows which experts were involved in each group.

Table 4. Expert role matrix for the two groups.

	Group 1 B5T (n = 4)				Group 2 CRM (n = 4)				
	HF1 Male	HF2 Male	HF3 Female	Pilot 1 Male	Pilot 2 Female	Pilot 3 Male	Pilot 4 Male	Pilot 5 Male	
B5T Group	x	x	x	x					
Prior involvement in HAIKU	x								
Prior involvement in JARVIS		x							

Table 4. Cont.

	Group 1 B5T (n = 4)				Group 2 CRM (n = 4)			
	HF1 Male	HF2 Male	HF3 Female	Pilot 1 Male	Pilot 2 Female	Pilot 3 Male	Pilot 4 Male	Pilot 5 Male
Facilitators for CRM Group			x	x				
CRM Group					x	x	x	x
Use Case Selector	x	x						
Data Recorder			x					

4.3.2. Evaluation of CRM Impacts

The approach for the CRM analysis employed a simplified Nominal Groups Technique (NGT) approach, a structured method for group face-to-face brainstorming that encourages contributions from everyone and facilitates quick agreement on the relative importance of issues, problems, or solutions [40]. NGT includes the benefits of in-depth discussion of a Focus Group, a face-to-face small-group technique in order to explore participants' perceptions of given topics, and the prioritising of results via a structured face-to-face group method achieving group consensus [41]. NGT can also avoid biases such as one expert dominating others in groups, by first allowing each expert to make their own individual assessment, then meeting to reach consensus, then allowing individuals to change their assessments afterwards if they wish to do so.

The four pilots were sent an information pack (a series of Powerpoint slides) explaining the study and the five use cases, including cockpit diagrams of where pilot–AI interaction would take place. Two core questions were posed to each individual pilot, for which they were asked to come with pre-prepared answers:

- Your opinion will be asked about how these five different use cases might influence CRM, as operationalised in the pilot core competencies.
- Furthermore, you will be asked what should be changed as a result of the introduction of such ASs in the CRM framework and/or CRM training of these competencies.

The four pilots then met during a day at an airport, with the two facilitators from the B5T group, and discussed their results and insights on the use cases one by one. An example of the collated output from the four experts, as presented to them in the joint session, is shown below in Figure 4.

The CRM experts aimed to reach consensus (full agreement) on the relations between the use cases and the pilot core competencies, which was achieved fairly readily. The rest of the session focused on the positive benefits of the use cases for CRM, and the potential downsides, and for the latter, possible mitigations (including training requirements). All results were captured in Excel sheets by one of the facilitators and then checked by the other one immediately after the meeting was finished. The CRM pilots were then sent the results and had one more opportunity to make changes (only minimal changes were made).

4.4. Assessment of Relevance/Utility of B5T and CRM for HAT Use Case Assessment

The same B5T team of experts first reviewed the models according to which elements were most important, simultaneously showing those that were not deemed as relevant. Relevance here relates primarily to operational risk. For example, if the AS could disrupt situation awareness, communications, increase workload, diminish team leadership or orientation, this puts the flight crew (and flight safety) at risk. B5T elements such as adaptability, mutual performance monitoring, back-up behaviour and mutual trust might

be expected to enhance operational risk management if the AS is strengthening them, but this needs to be assessed along with the risk of ‘downsides’ that could diminish the operational safety benefits arising from the AS. Such risk considerations are best judged by pilots themselves. This is the primary reason there was a pilot in group 1, and why group 2 consisted entirely of pilots experienced in operational CRM practices. Elements which might need adaptation in order to be more relevant to AI contexts, as well as potential missing elements were identified. Based on these two analyses of the results, a collective judgement was made of the operational relevance of both models, thus deeming whether they were suitable, even if only as a starting point, for the evaluation of the impacts on teamworking of future AS and HAT concepts.

Workload management	Situation awareness	Communication	Problem solving & Decision making	Leadership & Teamwork	(Application of Procedures)
Learning to prioritize tasks: How much time do you give your decision? Sometimes you have more time than others (e.g., plane is on fire)	Training on building individual SA: know which factors you always want to check and do not want to assume blindly. How do you keep your individual SA up to date when the answers are already being presented? Stay proactive with SA and threats along the way. Avoid over-reliance and complacency	Work on communication to promote shared SA: how do you keep everyone in the loop? Practice asking open-ended questions to each other. For example, how do you assess AI's judgment? What kind of risks do you see in the options that AI offers us?	Apply critical thinking very consciously: What am I looking at? What factors were taken into account in this decision? Is my current situation taken into account (e.g. problems with my flaps and runway length)? Continue training with/without AI, in various scenarios. Both practising wrong AI conclusion and an initially good advice that turns out to be inappropriate due to changing circumstances Awareness of the risks of taking the easy route and blindly following AI. Keep cross-checking.	Relates to 5.8 responsibility for decisions and actions: making people aware of their ultimate responsibility for the decision; AI just a tool Training in dealing with the outcomes of AI: how do you estimate its recommendations in value? Are you going to question the AI (dialogue) or discuss among each other (pilots)? Actively involve input of others, engaging others: How do you remove all doubt from the team? How do you keep everyone engaged?	KNO: AI literacy as a new competence: how reliable it is, how to interpret a confidence interval, etc. Existing OBs provide ample opportunities to include new skills, e.g., operates systems and equipment correctly, seeks accurate and adequate information (always remains skeptical), and validates the accuracy of information

Figure 4. Example excerpt from CRM expert group session for UC3 (prior to discussion).

4.5. Consolidation of Augmentation Requirements for B5T and CRM for HAT Applications

The specific changes, new elements or new CRM training requirements were consolidated by the HF expert group, to clarify the distinct differences between the two existing models and what is required to adapt them more closely to the consideration of future HAT systems.

5. Results

The impact analysis is in two parts, reviewing the five use cases through the ‘lenses’ of the Big 5 Teamwork model and then CRM.

5.1. Expert Analysis of B5T in Use Cases

Table 5 documents the expert insights into the relevance of the B5T elements for the five use cases. Results are then interpreted in terms of the research questions in Section 6.

5.2. Pilot Expert Review of CRM Impacts in Use Cases

The insights from the CRM expert review group are given in Table 6, use case by use case, for each of the pilot core competencies, and are discussed in terms of the research questions in Section 6. It may be noted that the B5T table of results is in B5T element sections, whereas the CRM table is in use case sections. This difference pertains to the

tendency of the first group to be more focused on the model, with the second group more focused on practical considerations of using the ASs. However, both tables distil useful insights for elements and use cases alike.

Table 5. Insights from expert review of five use cases using the B5T model.

B5T Element	Expert Observations and Contextual Insights
Team Leadership	Team leadership was not judged to be affected in UCs 1 and 2, unless the AS makes too many mistakes or is otherwise interfering with teamwork, in which case a decision must be made to ignore it or shut it down. E.g. in UC2, when flying from Initial Approach Fix to Final Approach Fix workload is very high; using the AS may interfere with other tasks in this flight phase. In UC3, the AS presents airports to the flight crew, and the captain decides which to select. The availability of three distinct solutions enables the captain to make a better-judged selection, enhancing leadership. In UC4, the AS can give the crew direction when they are unsure what is going on, allowing the captain to then take charge in leading the actions required. Complacency in decision-making (see ‘team orientation’ below) needs to be avoided by actively promoting critical thinking towards the Advisor, including whether to disregard its advice, or else shut it off. In UC5, the AS detects the startle reaction with the pilot, takes the lead in calming the pilot down and gives assistance in stabilising the aircraft. When assistance is no longer needed, the pilot takes the lead again and shuts down the AS. It is relevant to team leadership because it aims to put the pilot back in command.
Mutual Performance Monitoring	This element was seen as of highest relevance in UC4 and UC5. In UC5, initial monitoring is ‘one-way’ with the AS monitoring the physiological condition of the pilot; as the pilot is relying on the AS in a startle condition, the AS is not monitored. When the pilot has recovered from the startle condition and has regained enough cognitive resources, the pilot will begin monitoring the AS advice, and once the pilot has recovered, they can switch off the AS. In UC4, the AS can present its own diagnosis which the flight crew can then cross-check for themselves, to verify the diagnosis or identify alternatives. ‘Dialogue’ can ensue wherein the crew may postulate facts they believe the AS has missed or pose counterfactual questions. Such dialogue enables both the pilot(s) and the AS to reframe the problem and derive a new solution. The AS will also be checking aircraft status in real time, so that if pilot actions do not align with the AS’s predicted behaviour, this will be made apparent. Performance monitoring is therefore deemed mutual. In UC1, once a voice message is turned into text it remains available for both crew members. Even though pilots have different roles, the fact that it can be retrieved and reviewed means that both crew members have access to the same information. Thus, between pilots, performance monitoring concerning understanding ATC messages may not be necessary. The crew still listen to ATC calls and hence can monitor the AS in case it makes mistakes. However, as the AS is not monitoring crew performance, this factor is not considered ‘mutual’. In UC2, since information about RWY and TWY use is available to both team members earlier than normal, and for both of them at the same time, performance monitoring between pilots may be unnecessary. However, the crew must confirm or deny the AS’s prediction by listening to actual ATC clearances. As the AS presents the predicted routeing in front of the pilots on their screens, they must actively avoid confirmation bias (hearing what they want to hear). As the AS is not monitoring crew performance, this factor is not considered ‘mutual’. In UC3 this element was considered not relevant, although the explainability feature enables the crew to check the selection of proposed airports and monitor AS performance, and the PM may check various airports in parallel to the AS, and/or contact the Operations Control Centre for their advice. However, as with UC1 and 2, performance monitoring is not mutual.
Back-up Behaviour	This element had relevance to all five use cases. In UC1 the text version of the spoken ATC message helps ensure more reliable communication and thereby prevents workload rising due to errors. If the reliability of the AS is high in catching messages which are meant for their own aircraft, it can also offload the pilots because the constant scanning for their own callsign may not be needed (especially useful in high workload situations). In UC2 predicting the landing runway and taxiway routeing helps pilots in correctly programming their Flight Management System and building situation awareness. Especially in high workload situations, such as a runway change or a diversion to an unknown airport, it offloads the pilots because they do not have to spend time figuring out the taxi routeing after landing or take multiple taxi scenarios into account. The AS can also offload both pilots and ATC due to less clarification requests when ATC instructions concur with AS predictions. In UC3 the AS generates new alternate airports far quicker than the flight crew, off-loading this task from the flight crew. On the other hand, the back-up behaviour could lead to a temporary workload increase if negotiation with the AS is not successful. In UC4 the AS can be a useful check for pilots that they have gotten their initial diagnosis correct. If they have not, or are about to make a situation worse, or have completely the wrong mindset for what is actually happening, the AS could help them avoid misdiagnosis, confirmation bias, and tunnel vision, as well as partial solutions which can have unintended side-effects. On the other hand, instead of using the AS as a useful check on their own diagnosis, the crew could anchor their diagnosis on the recommendation of the AS and then gather information to agree with it, falling victim to confirmation bias [42]. Also, if the AS presents ambiguous (i.e., not explainable) information, this could frustrate the diagnosis and increase cognitive workload for the crew, making the AS’s support counterproductive. In UC5, the AS anticipates the needs of the pilot, assisting to regain flight stability, a calm condition, and to regain situation awareness. While workload is not shifted towards the AS, the workload is quickly manageable through the help of the AS.

Table 5. *Cont.*

B5T Element	Expert Observations and Contextual Insights
Adaptability	Adaptability is one of the more relevant elements, although in UC1 it was not judged relevant. However, in UC2 it is judged highly relevant. The information provided to the flight crew in an earlier stage of their flight will help them anticipate what will be coming shortly, making it easier for them to flexibly adjust to the instructions they receive from ATC. It is similarly of high relevance to UC3, the AS enabling the flight crew to more quickly adapt to a new situation, saving them valuable time and cognitive resources. Similarly, in UC4 the AS enables the crew to adapt to novel and complex time-pressured situations, possibly including events for which they were never trained. However, adaptability may be negatively impacted as the AS may be biased towards its training data and unable to think out of the box (e.g., for edge cases). It can also be biased because of reduced/faulty sensory input. If the crew does not realise this, readjusting strategy will be in the wrong direction. In UC5, which is entirely focused on sudden, unexpected events, the AS enables the pilot to quickly adapt to a new and urgent situation, saving them valuable time and cognitive resources.
Team Orientation	This element was less relevant to the selected use cases than most other elements, and in UC5 can be discounted as it is single-pilot operation, so there is no human team to orient. In contrast, for UC1, whereas RT is currently a team effort to make sense of ATC messages, e.g., with each pilot saying the clearance aloud, the AS may erode such practices and thereby reduce team orientation. This potential negative impact could be compensated by the visual (text) representation of the message available to each pilot. Similarly, the presentation of TWY routeing on the cockpit displays, visible to both pilots, could lead to less overt communication between the flight crew. This could have a slight negative impact on team orientation. In UC3 the AS may remind the flight crew of certain priorities that should drive their selection of alternate airport destination, e.g., cost to the company, availability of follow-on flights for passengers, potential runway contamination, etc. The AS may reduce the crew's initiative to search for other options or to consider other features, affecting team orientation negatively. However, if the flight crew takes the results of the AS as a starting point for decision-making on a diversion airport, and employs critical thinking, team orientation is unaffected. In UC4 the AS can let the crew know if they are focusing on the wrong priorities and help them determine a safer course of action and a better use of resources. There is however a danger of the crew becoming complacent, and instead of making joint diagnosis and each pilot asking the other about their diagnosis, simply letting the AS do the full diagnosis.
Shared Mental Models	This B5T element was judged to be one of the most important in relation to HAT systems. In UC1, the text message reinforces a common shared flight crew mental model of the clearance, but to achieve a definitive shared mental model (crew and AS), the AS would need to listen-in/monitor human responses and performance. Nevertheless, this reinforcement is seen as beneficial. However, pilots might be less inclined to listen to ATC and update their situation awareness (shared mental model) of other aircraft in their vicinity. Less understanding and projection of the current traffic situation can impact the crew when confronted with a sudden go-around, or a sudden delay, for example. Surprise reactions could occur, raising workload. Similarly, in UC2, pilots might become less inclined to actively process airport layout, standard taxi routeings, construction work in progress, etc. If AS predictions turn out to be wrong, pilot workload may increase significantly, as they will have to completely build up their SA to know where to taxi their aeroplane after landing. In UC3 the primary 'model' is geographical and temporal along with the factors used by the AS's algorithm, which was trained via hundreds of real diversion cases in a human-supervised learning approach, with expert pilots helping to train the model. This approach facilitates similarity between the AI model and the human one. In the cockpit, the AS advice and its rationale are available for the flight crew via the moving map and explainability displays. Because pilots do not have to search for information themselves, the presented information might not be processed actively by the pilots. This could diminish their own active mental model of the airports, and lead to issues if, for example, the situation worsens such that the newly selected destination airport also becomes unavailable. In UC4, the AS would be trained partly on Standard Operating Procedures (SOPs) and expert pilot responses to a range of emergencies, plus engineering maintenance manuals and procedures, and partly via using deep learning techniques for more complex failure scenarios where probabilistic analysis may underpin diagnosis. Advice from the first part is easier because it is based on models of how things work. The latter (deep learning) is more 'black box', and so AS explainability may be less straightforward. The real challenge here is therefore how to explain or represent the AI diagnosis and solution such that the crew can understand them in terms of their own mental models. In UC5, when startled, the pilot is in a confused state, and their mental model of the situation is no longer accurate or is degraded. The AS assists in regaining the mental model of the pilot by calming down the pilot and pointing out relevant instruments in the cockpit. The AS design challenge is that pilots do not all have an exactly homogeneous mental model, such that the AS's directed situation awareness may not closely follow the mental models of every pilot. This 'discordance' may trigger pilots to shut off the AS before they have sufficiently regained situation awareness.
Closed Loop Communication	In UC1 this element is highly pertinent, as the AS is aimed precisely at achieving more accurate and robust closed-loop communications and adherence to ATC instructions. Similarly, in UC2, TWY instructions by the Tower controllers are not always clear to a flight crew unfamiliar with the airport, so the AS will make such instructions appear clearer by giving visually displayed TWY guidance, e.g., illustrated on the cockpit displays. Similarly, in UC5 the AS indicates where to look (i.e., at which instruments, in sequence) through visual highlighting synchronised with oral callouts, thereby ensuring closed-loop (albeit one-way) communication. In UC3, the AS explainability concerning the factors assessed for each airport selected has been subject to pilot supervised training of the AS, thereby ensuring clear and precise terms for pilots. However, in negotiations, pilots will need to come up with the right prompts to get the desired information within a reasonably short time span. In contrast, in UC4 interpretation of AS outputs in terms of diagnosis and explanations could pose a problem if not sufficiently calibrated to pilots. Too much information will make it more complex, too little information will raise questions. The deep learning part of the UC4 AS may deliver its answers in a confident and unnuanced way that removes critical cues that could allow flight crew to calibrate the accuracy or reliability of its answers [42]. Conversely, communication might be affected positively if the AS reduces language barriers between the two humans on the flight deck.

Table 5. Cont.

B5T Element	Expert Observations and Contextual Insights
Mutual Trust	This element is most important in UC5, because in startle or surprise situations the pilot is highly vulnerable and has limited capacity to actively manage or evaluate system input. Although the AS can be manually disabled, the pilot may not be in a position to make this judgement during the event, making prior trust and training crucial for appropriate reliance. Shared access to radiotelephony (UC1) and routing information (UC2) increases information symmetry between crew members, enabling direct cross-checking, reducing uncertainty and thereby decreasing the need for interpersonal trust, but also potentially increasing over-reliance on the AS and reducing active monitoring. As a result, the crew becomes highly dependent on AS performance. AS failures in either use case may therefore degrade situation awareness, and so calibrated trust in the AS is critical. As the AS is not monitoring crew performance, trust cannot be considered ‘mutual’. Similarly, in UC3 shared pilot (PF and PM) access to alternate airport options increases information symmetry between crew members, reducing uncertainty and thereby decreasing the need for interpersonal trust. The crew must remain aware that the AS predictions could be wrong or insufficiently current (e.g., due to a rapidly evolving weather or military conflict situation). There is a risk of overreliance on the AS recommendations. Explainability is crucial to calibrate the crew’s trust in the system, to understand how the AS came to its recommendations and to decide whether that aligns with the crew’s priorities. Again, trust is not mutual. In UC4, by having a clearer diagnosis and agreeing on a plan of action, the flight crew can trust each other’s actions more, knowing they are ‘on the same page’. The crew may need to depend on the AS to support diagnosis of abnormal or unusual emergency situations, placing them in a highly vulnerable decision context, rendering trust critical, as both overreliance and under-reliance on the system’s output may compromise safety. Appropriate trust calibration is therefore required: the crew must be able to assess whether the system’s recommendations are reliable and relevant to the situation. Explainability becomes crucial in supporting this calibration, as it allows the crew to understand the basis of the diagnosis and better judge when to follow or disregard system advice. Again, trust is not mutual: although the AS can in theory point out faulty crew diagnoses, no AI trust is required.

Table 6. Insights from pilot expert group insights on CRM core competencies.

Use Case	Pilot Expert Group Insights on CRM Core Competencies
UC1–Voice to Text	Not surprisingly, the main benefit of this AS was seen in the core competency <i>communication</i> and the associated reduction in workload (<i>workload management</i>). These benefits may result from reduced miscommunication with ATC, fewer missed calls, better understanding of complex calls, no longer having to monitor your own callsign, and cognitive offloading radio frequency changes. Potential negative effects of this use case were expected due to possible overreliance. In the short term, this may lead to inattention and, over time, to an erosion of listening skills. If voice transmissions are no longer actively monitored, intonation may be missed, ATC errors may go unnoticed, and calls to other aircraft may no longer be processed. This last factor may cause a decrement in <i>situation awareness</i> of the general traffic situation. Situation awareness may also be negatively affected when attention is momentarily diverted from monitoring instruments to reading a text message. Another consideration is the possibility that the AS starts making errors, such as an incorrect transcription or uploading the wrong frequency. This may cause confusion for the pilots, increasing workload because they need to verify or clarify the information while simultaneously managing other tasks.
UC2–RWY & TWY Predictor	The benefits of the UC2 AS were mainly expected in <i>situation awareness</i> and <i>workload management</i>. An early prediction of the landing runway and subsequent taxi routing may focus the crew’s attention on just one runway–taxiway scenario, reducing the need to consider multiple alternatives and proactively prepare for them. Workload reductions are expected because the Flight Managements System (FMS) only has to be programmed once for the arrival and approach, and specific threats need to be assessed only once in the crew briefing. Currently, a runway change during arrival causes additional workload and distraction because the FMS and crew briefing need to be adapted. The feature of taxi route visualisation further supports situation awareness by helping pilots understand where they are and where they are going. This also reduces workload, outsourcing navigation to the AS. Of course, this outsourcing also has its downside. Pilots may become over-reliant on the system, pay less attention to information about work in progress or closed taxiways, and become less inclined to consider alternative runway–taxiway scenarios. Both may reduce situation awareness. This out-of-the-loop effect may lead to a sudden increase in workload when the predicted taxiway does not match ATC instructions and pilots must navigate independently.
UC3–Flight Re-Route Advisor	In UC3 the AS cognitively offloads the pilots during decision-making. The main benefits were expected in <i>workload management</i>, <i>situation awareness</i> and the subsequent <i>problem solving & decision-making</i> core competency. The AS quickly presents relevant information on candidate airports, gives three options, and ranks them accordingly to their associated risks and benefits. These are key steps in the decision-making process and are currently performed with limited decision support, increasing the risk that relevant factors are overlooked. The AS therefore supports a structured evaluation of the available options, reducing workload, enhancing situation awareness, and improving decision-making. However, this cognitive offloading was also expected to have a downside as pilots may become over-reliant on the AS. They may evaluate the provided information less critically or skip further individual or team discussions of the available options, both of which may reduce the team’s situation awareness and ultimately negatively affect decision quality. In the longer term, pilots may also lose proficiency in collaborative decision-making altogether, which can have serious consequences if the AS somehow fails. Another factor to consider is the <i>communication</i> with the AS. A lot of information is put together in presenting the factors considered for three airports, which potentially gives interpretation problems. Resolving these may require additional communication between the pilots or further interaction with the AS, potentially increasing workload.

Table 6. Cont.

Use Case	Pilot Expert Group Insights on CRM Core Competencies
UC4–Safety Advisor	In UC4 the AS provides the pilots with a quick diagnosis of potentially very complex problems and was expected to result in a reduction in <i>workload</i> , an increase in <i>situation awareness</i> , and improved <i>problem solving & decision-making</i> . The corollary of this support in relation to problem solving is again overreliance and cognitive skill loss. Pilots may accept the AS diagnosis without considering it critically and may lose their skills in working as a team to solve problems (e.g., sharing views and questioning each other about fact, option, risks and benefits). Another effect of this overreliance may be that pilots will not invest time in updating their technical knowledge (mental models) anymore, thereby further eroding their critical skills. Furthermore, possible pilot–AS conflict was expected when the AS provides unclear or incorrect information. Interpretation problems might occur, causing confusion. Together with the current inability to <i>communicate</i> with the AS (asking questions or adding sensory input the AS might have missed) this could increase workload, reduce situation awareness and deteriorate the problem-solving process).
UC5–Startle Response Support	In UC5 the AS aims first at calming the pilot down to free up cognitive resources (related to <i>workload management</i>) to come back in control, while pointing at the relevant instruments that require their attention, enhancing or restoring <i>situation awareness</i> , and increasing the ability for <i>problem solving and decision-making</i> . However, the AS could cause distraction or even confusion by the introduction of extra stimuli in the cockpit—resulting in the opposite of what is intended. An already overwhelmed pilot could become even more overwhelmed. Pilot–AS conflicts were seen as another source of negative impact, as incorrect detection of startle and errors when highlighting relevant instruments could decrease situation awareness and problem solving and decision-making capabilities. There may also be conflict with other procedures which demand immediate attention, e.g., for upset recovery, windshear or decompression.

6. Discussion

The aim of this paper was twofold:

- To assess which elements of the B5T model are relevant when applied to the HAT use cases in identifying potential AS impacts on teamworking performance, and to identify areas where the model may need updating to better accommodate such future operational AI-based systems.
- To determine the impact of HAT use cases on CRM and determine how the CRM model and associated training might need to evolve in relation to these use cases, to at least maintain the same level and quality of teamwork in the cockpit and ideally improve it.

6.1. Discussion of B5T Analysis Results

The B5T analysis results are discussed below in terms of the five lines of inquiry specified earlier. It may be noted that UCs 3–5 appear to call upon slightly more B5T elements than UCs 1–2, which require less cognitive resources (e.g., reading and comprehending ATC clearances and airport geography). Since the B5T elements are not scaled in any way, this is merely a qualitative observation. Nevertheless, UCs 3–5 all deal with non-nominal situations, so the need for more teaming can be expected.

What B5T elements are relevant across the use cases?

Adaptability is the B5T component that is mentioned most often as highly relevant across use cases (except for UC1). *Adaptability* is the intended benefit of most of these AI support systems. Examples include operations at large unfamiliar airports, unusual RWY/TWY configurations, major diversions, and sudden or unforeseen complex/rare events. In these contexts, AI support systems are intended to enhance human adaptability by providing a course of action repertoire and helping crews respond more effectively to non-routine conditions.

Similarly, *shared mental models* is relevant across most use cases, not surprising as most ASs are aimed at providing cognitive support to the flight crew by reinforcing and even extending the flight crew's situation awareness, especially in confusing and/or time-pressured scenarios when situation awareness may be diminished.

The relevance of *back-up behaviour* is again not surprising, as the AS in several of the use cases intends to reduce the workload by taking over tasks, thereby offloading the flight crew (especially UC3, 4 and 5).

Supporting *closed-loop communication* is the intended benefit of the ASs in UCs 1, 2 and 5, aimed at supporting accurate and clear pilot–ATC (UCs 1 and 2) and AI–pilot (UC5) communications.

Mutual performance monitoring is generally reinforced by the AI components, enhancing the shared situation awareness between flight crew members, as well as increasing the ability of one or other flight crew member to raise questions and identify issues.

Mutual trust was relevant across the use cases, mainly in terms of the crew’s reliance on AS capabilities and the need for appropriate trust calibration. Overreliance may result in insufficient monitoring of system outputs, whereas under-reliance may reduce the benefits of AI support. However, the reciprocal nature implied by the original B5T construct was largely absent, as trust in the AS–Crew direction was usually not applicable.

Team leadership and *team orientation* were the least prominent elements in the review, possibly because the humans remain in charge at all times (except potentially in UC4, and in UC5, temporarily). The main contributions to these two elements from the ASs in the use cases is firstly to support the captain in their leadership role, and secondly to ensure that the right priorities are being followed, including when to disregard/disable the AS.

Overall, therefore, the B5T model appears relevant to the AI use cases. In particular, *adaptability* and *shared mental models* appear to form a dominant dyad in three of the five cases.

Are any relevant teamworking elements missing?

One element that appears to be missing from B5T is *cognitive offloading*, defined as the delegation of cognitive tasks to external resources in order to reduce mental demand [43]. Cognitive offloading to an AS has both benefits and risks. By taking over routine or cognitively demanding tasks, the AS can free mental resources for higher-level decision-making. However, increasing reliance on the AS may also reduce the crew’s active cognitive engagement and, over time, diminish proficiency in these tasks. This may affect operational resilience if it reduces the crew’s preparedness for situations where human intervention is required, such as when the AS is unavailable, performs unreliably, or encounters situations outside its design assumptions or training dataset.

Cognitive offloading is related to back-up behaviour but reflects a challenge specific to HAT. Back-up behaviour refers to the ability to anticipate other team members’ needs through accurate knowledge about their responsibilities. In contrast, while an AS can take over tasks, it cannot take over responsibility. Offloading may reduce the crew’s visibility of a task and does not inherently provide redundancy, as the human may no longer retain the same level of proficiency required to perform the task independently. Thus, unlike human back-up behaviour, which can increase team resilience through complementary capabilities, cognitive offloading may engender a dependency on the continued availability and performance of the AS.

Are any B5T elements missing important aspects or nuances?

Mutual trust was identified as the most challenging B5T component to apply to HAT systems, as its relevance differs depending on the direction of the relationship. In human teams, trust involves a willingness to accept vulnerability based on the expectation that others are both willing and able to contribute to shared goals. This principle applies to the crew member–crew member relationship, where team members depend on each other and are jointly exposed to operational consequences. However, this reciprocity does not exist in the crew–AS relationship. Although ASs may monitor crew performance and provide feedback or backup support, they do not experience vulnerability, risk, or dependence on operational outcomes. The crew remains exposed to the consequences of AI-supported decisions, whereas the AS does not. Therefore, the crew–AS relationship is better characterised as reliance on AI capabilities rather than mutual trust. Similarly, the

AS–crew relationship is more appropriately described in terms of performance monitoring rather than trust.

This does not mean that trust becomes less relevant in HAT; rather, it changes in nature. In human teams, mutual trust enables effective coordination by allowing team members to delegate tasks, accept guidance, and rely on each other while maintaining appropriate monitoring. A similar balance, or calibration, is required in HAT. Pilots must be able to determine when an AS is reliable enough to use, when outputs require verification, and when independent judgement remains necessary. Both under-reliance and over-reliance can negatively affect performance: the former may prevent crews benefiting from available support, while the latter may reduce critical monitoring and awareness of potential system limitations. Therefore, in HAT contexts, the interpersonal concept of mutual trust may separate into two related constructs: mutual trust between human team members and *calibrated reliance* of humans on AI systems.

For *closed-loop communication*, the expert review identified a nuance for HAT not fully captured in the original B5T element, related to *explainability*. In human–human communications, information is conveyed not only through words, but also through intonation, emphasis, gestures, facial expressions, and a shared understanding of the situation. Such redundancy helps team members to quickly interpret both the content and intent of the message. This, together with the ability to rapidly explain why a suggested course of action is needed, makes human–human communication robust. ASs cannot yet rely on these implicit mechanisms. Instead, they must explicitly communicate the rationale underlying their recommendations in a manner that is concise, context-specific, and immediately interpretable. If the advice is generic and not contextualised in the local conditions of the event or disturbance as it unfolds, the pilots may have difficulty understanding or following such advice. What is needed, therefore, is clear and concise explainability, tailored to the current operational context, so that the flight crew or single pilot flying can see and interpret the advice quickly. If the closed-loop communication construct is to be augmented for HAT, this additional requirement could be described as *situational explainability*.

Similarly, the expert review proposes that *shared mental models* may need adaptation. ASs do not represent the world in the same way as humans. Rather than forming mental models, they largely maintain algorithmic and statistical representations derived from data. They compute, rather than think as humans do. For effective collaboration, it is therefore not necessary (nor practical) for humans and AI to share identical mental models, but for the crew to understand how the AS currently represents the situation and the basis upon which it produces a particular recommendation. This is illustrated in UC3, where the flight crew can inspect the factors underlying the proposed diversion airport, and similarly in UC4, where insight into the AS's reasoning would support decision-making in unfamiliar situations. We argue that the traditional notion of shared mental models should be extended towards *shared representation* [44]: a common representation of the current problem state, the information available to the AS, the rationale underlying its recommendations, and the uncertainty associated with its outputs. Such transparency enables the crew to anticipate the behaviour of the AS, coordinate their own actions accordingly, and compensate for the limited opportunities for natural communication with the AS (as discussed above). We therefore propose *shared representation* as a candidate for augmentation of the B5T framework for HAT.

Are the impacts of the AS on B5T elements likely to be positive or negative?

Whilst there are many potential benefits (e.g., enhancing crew adaptability), in a number of instances in the B5T expert review a potential gain in terms of back-up behaviour, mutual performance monitoring, team orientation, etc., could be neutralised by the loss

of a current crew function, e.g., vocalising and confirming ATC messages to each other. Care must be taken, therefore in ensuring that benefits from AS implementation are not erased by cancelling current good teamwork practices in the cockpit. A 'zero sum gain' in AS implementation should be avoided.

Is B5T a suitable approach for examining the teamworking impact of future AS and HAT concepts?

In summary, the result of the B5T expert review of the five use cases suggests that the B5T (a) does appear relevant, and (b) can be useful, leading to practical teamworking insights and the identification of associated operational risks/benefits, and (c) requires further exploration and potential augmentation in four areas: *cognitive offloading*, *calibrated reliance*, *shared representation* and *situational explainability*. It can therefore be a useful approach with which to examine future AS/HAT concepts, preferably with practitioner (pilot, controller, etc.) involvement in such a review process.

6.2. Discussion of CRM Analysis Results Across the Five Use Cases

Which CRM elements are relevant across the use cases?

Situation awareness was seen by the pilots as a prominent factor in all five use cases, perhaps not surprisingly since all five ASs provide new information to the pilots, which must be properly managed. The benefits could be substantial, facilitating the pilots in sense-making of radiotelephony (UC1), their planning (UC2), their decision-making (UC3&4) and focussing on the relevant cues needed for aircraft control (UC5). On the other hand, the drawbacks could also be substantial due to overreliance (all use cases), distraction (UC1&5), pilot preference bias (UC3), and confusion due to interpretation problems (UC1&4).

Workload management also appears in all use cases. If the ASs work as designed, they can reduce workload by preventing mistakes (UC1) and unnecessary changes (UC2), by saving time due to quick diagnosis (UC3&4), or by calming the pilot down (UC5). However, if things do not proceed as designed, workload is likely to increase. This could also be the case when out-of-the-loop problems occur (UC1&2), or when interpretation problems arise with given decision support advice (UC3&4).

Problem solving and decision-making was highly impacted in the UCs which were specifically intended to provide decision support (UC3&4). The competency was also relevant to UC5 but was less relevant in the other UCs.

Similarly, *communication* was highly implicated in the use case specifically designed to provide communication support (UC1), which showed positive and negative impact. It was also relevant in two other use cases, with possible impact on crew briefings (UC2) and the inability to ask for clarification (UC4). Furthermore, there was indirect impact caused by overreliance (UC3), which is primarily reflected in other competencies.

In contrast, *leadership and teamwork* was considered less relevant by the pilots when considering these five use cases, with some limited discussion related to the use cases designed for decision support (UC3&4).

Application of procedures was not considered or discussed much across the use cases, aside from the danger of erosion of discipline (UC1) and a possible disparity in procedures (SOPs), due to variability in the exact sequence in which pilots interrogate cockpit instruments during in-flight emergencies (UC5).

In summary, for these use cases the competencies can be split into three categories: 'highly relevant in all UCs' (*situation awareness* and *workload management*), 'moderately relevant' (*problem solving and decision-making* and *communication*), and 'low to medium relevance' (*leadership and teamwork* and *application of procedures*).

Are any relevant CRM elements missing?

Maintaining a critical mindset while using AI-based systems appears to be a missing element in the CRM framework when applied to these use cases. The CRM experts indicated that the danger of overreliance on an AI-based system might be more prevalent than with contemporary automation. Suggestions given by the AS may be persuasive just by their format, and a (relative) lack of knowledge or cognitive skill on the human side may give way to accepting suggestions without considering them critically. A critical use of AS should therefore be incorporated and trained in multiple competencies (e.g., *situation awareness* and *problem solving and decision-making*). It must cover knowledge on the limitations and risks of ASs (to reach an adequate mental model of the AS), an emphasis on the validation of information (by checking the accuracy of the source, by cross-checking with other sources, or by listening to your 'gut feeling'), and in general on maintaining *problem solving and decision-making* skills (e.g., by occasionally operating without the AS).

Pilot–AI conflict resolution is another element that appears to be missing, despite conflict management in general being an element of *leadership and teamwork*. The CRM experts indicated that a conflict may arise (a) between crew members on how to deal with the advice from the AS, (b) between crew member judgement and AS advice, and (c) between AS advice and other procedures. First, to handle these conflicts, guidelines should be in place. Second, pilots should be trained in these guidelines. Third, pilots should be made aware of their ultimate responsibility for decisions. Fourth, to prevent conflicts with other procedures, the information or advice given by an AS should be integrated in current (non-normal) procedures. And fifth, pilots should be trained to recognise and act when the AS makes errors.

Are any CRM elements missing important aspects or nuances?

As some of the use cases encompass communication with the AS (UC3, and possibly UC4) the pilot core competency *communication* should address aspects such as *formulating effective prompts* and the timing of communication with an AS. Information from the AS might raise more questions than answers if prompts are not formulated effectively. This could cause unwanted increase in workload. Also, without communication problems, workload is an aspect to be reckoned with when communicating with an AS. *Timing of human–AS communication* with an AS is therefore another aspect to be added to the competency. This will be even more important when working with ASs capable of collaboration (level 2B). Also, aspects such as communication of intent will then become relevant.

As far as the competency of *workload management* is concerned, pilots should be able to detect and *foresee unacceptable workload* caused by the AS and subsequently be able to reduce interaction to a minimum (when communication with the AS is the cause) or switch off the AS completely (when the AS is producing too many errors).

Are the impacts of the AS on CRM elements likely to be positive or negative?

As with the B5T analysis, a potential gain in CRM derived from AS usage could be neutralised by the loss of a current crew function. Overreliance is a danger not to be underestimated, as it not only has immediate effects on current crew function, but also in the long run. Immediate effects encompass accepting AS advice or information without further consideration instead of (cross-)checking the validity of information or involving the views of colleagues. In the long run, crew skills such as building up situation awareness from multiple sources can erode, as can problem solving and decision-making skills. This leaves the crew handicapped when the AS breaks down for some reason and therefore these skills need to be maintained in simulator training. Another way current crew function is impacted in the longer term is by affecting the crew's mental models. When relying on the AS too much, knowledge about underlying technical systems or aerodynamics may erode. If not maintained properly, this will again leave the crew handicapped when the AS

stops functioning as desired. Another effect of flawed mental models is that the crew may not be able to detect AI failures or other AI aberrant behaviour.

Is CRM a suitable approach for examining the teamworking impact of future AS and HAT concepts?

Overall, the result of the CRM expert review of the five use cases suggests that the CRM framework (a) does have relevance, and (b) can be useful, leading to practical teamworking insights and the identification of associated operational risks/benefits, and (c) may require augmentation in the pilot core competencies, *communication* and *workload management*, and should consider incorporating two new elements: *critical mindset while using AI-based systems*, and *pilot–AI conflict resolution*. CRM therefore appears to be a suitable and an appropriate perspective from which to examine the teamworking impacts of future AS/HAT concepts.

Potential CRM training for working with AI-based systems.

During the CRM group evaluation, a number of potential training avenues emerged in relation to these use cases:

- Pilots should be trained in discussing and questioning AS advice pilot-to-pilot, so that they can form a collective flight crew decision about whether to follow it or not.
- The foundation for such training is critical thinking, based on an understanding of how AIs work and can fail—including when the format of the AI advice can be overly persuasive—as well as human/team vulnerabilities when working with ASs.
- In the case where human–AI dialogue or natural language explainability is available, pilots should be trained in how to query the AS using precise prompts and/or counter-factual questioning, and how to evaluate the AS's answers. Such training should include negotiating with an AS, and when to override or switch off the AS.
- Such training should not be merely classroom or CBT-based but should involve real-time simulator training across a range of situations, including cases where the AS is incorrect or outside its training set, such that the pilots should decide to ignore its recommendation.
- Workload management with ASs should also be trained in simulators, including cases where the timing of the AS interaction can lead to workload spikes, and how to detect and avoid/prevent them.
- Additionally, such training must address pilot conflict resolution, e.g., where one pilot relies on the AS recommendation and the other does not, as such scenarios will inevitably arise.

Additionally, the pilot experts considered that a procedural solution to integrate critical use of ASs in already used workflows could be to adapt current decision-making models. For instance, FOR-DEC [45], a widely used acronym for decision-making in aviation could be adapted to *FORCDEC*. Facts, Options, Risks and benefits (the F-O-R, from FORDEC) can be presented through an AS (as in UC3) and may be accepted uncritically. By incorporating a critical thinking element, an extra 'C', a critical review of given information may be encouraged.

6.3. The Way Forward for Teamworking Evaluation of HAT Concepts

The expert review suggested that both approaches can identify potential advantages and vulnerabilities of the five ASs, which can inform HAT conceptualisation and design, as well as pilot training. The principal advantage of using these approaches to evaluate future HAT concepts is that normally such issues might not appear in use-case simulations and trials until much later in the AS development cycle (i.e., TRL 7–9). At such a late development stage, it could prove very difficult to alter the AS design and the associated HAT workflows. Identifying these issues early on is therefore a very useful capability of

this type of assessment, which can assist in developing more resilient concepts of operation, as well as increasing the likelihood of technology acceptance by future flight crews.

The alternative is to relegate such considerations to later development stages, e.g., TRL 7–9, on the assumption that training and CRM will cope with any downsides. However, there is a significant risk of undermining decades of proven performance of the dual-pilot system. The emergent teamworking impacts in AS conceptualisation must therefore not be downplayed.

Although the two approaches have significant overlap, they each appear to have their own merits, and in several cases have led to different (though non-conflicting) insights that should prove useful for safety and Human Factors assurance during the design and development stages, as well as informing operational training curricula. Until further study may occur using larger expert groups and other experimental methods, it may be advisable for those wishing to explore the impacts of AS and HAT on teamworking to apply both methods, as in this study. Since B5T is more theoretical, it may be most useful earlier in the design (TRL 3–6) when looking at Human Factors requirements for effective teamworking with ASs, with CRM employed later, once pilots can try out working with the AS in realistic simulators (TRL 5 onwards).

6.4. AI as Team Support, Not Team-Mate

Review of human and AI teamworking in actual use cases offers a relatively more concrete platform from which to consider the earlier raised question of whether AI is a tool or a teammate, even if such considerations can only reflect what is possible now, and so cannot fully resolve the issue. However, the perception of all experts in the two groups, across all the use cases, is that these ASs do not constitute ‘team-mates’, rather they are support tools, at most ‘Digital Assistants’.

This may evolve in the future as more advanced and collaborative (Level 2B) ASs are developed, but for now all five ASs are seen as tools, even if valuable and potentially critical. The human pilots remain in charge.

6.5. Limitations of the Study

There are three principal limitations of the study. The first is that the use cases only reach TRL6 at most, whereas CRM and teamwork effects may not emerge until later TRL stages where pilots can interact with the AS over more extended periods and with variable crew combinations in terms of experience range and age/generation/ AI familiarity. Therefore, more mature use cases scenarios are required for more in-depth study.

Second, for both reviews the number of Human Factors experts/CRM pilots was four. Further studies with larger samples of experts (e.g., 10) would be beneficial in validating/refuting/extending the findings in the study documented in this paper.

As noted earlier, some use cases were mixed, with a range of EASA AI autonomy levels (between 1A and 2B). As more use cases become available, more definitive studies will be possible using experiment-based rather than expert review-based paradigms.

6.6. Future Research Priorities

This study highlights four priority research avenues:

- i. Carry out similar studies with different AI use cases and a larger number of pilots, to confirm or elaborate the findings in the current study, preferably with pilots using the ASs in realistic simulations under standard crewing configurations, and applying situation awareness, workload and other measures to evaluate crew performance with and without AS assistance, also including instances of AI error.

- ii. Explore the identified candidate augmentations for the B5T model: cognitive offloading, calibrated reliance, shared representation and situational explainability;
- iii. For CRM, explore and develop the training avenues identified above, especially the development of a critical mindset and the potential utility of the proposed FORCDEC approach.
- iv. Develop and adapt the two approaches further so they can be used to support Safety and Human Factors Assurance functions associated with the future certification of AI-based systems.

7. Conclusions

Teamwork and good CRM are cornerstones of safe and efficient commercial flight operations. The future introduction of AI-based systems with varying degrees of autonomy into the cockpit must be safeguarded from the teamworking perspective, so that safety and efficiency are at least maintained, if not improved. This paper has reviewed such impacts via the ‘lenses’ of the B5T model and CRM, applied to five AI research prototypes at TRL 5–6, and has indeed found that a variety of potential impacts—positive and negative—are identifiable.

Flight crew teamworking with AI is likely to be a critical determinant of the success and safety of the integration of AI into cockpit operations, and both B5T and CRM have shown promise as approaches with which to identify impacts and develop safeguards, so as to realise AI’s benefits while mitigating its ‘downsides’.

The study suggests that further research and development is required for B5T, and for CRM training. For B5T the candidate research priorities are in the areas of cognitive offloading, calibrated reliance, shared representation and situational explainability. For CRM the priorities are in the critical use of AI-based systems, pilot–AI conflict resolution, as well as augmentation in the pilot core competencies communication and workload management. In particular, new training avenues for CRM with cockpit-integrated AI-based systems have been identified.

As future operational AI use cases develop beyond TRL6, this may prove the best period to validate and evaluate further teamworking impacts and the effectiveness of CRM adaptations. It will also present an opportunity to ‘sharpen’ both B5T and CRM approaches, or to develop new ones to better safeguard the critical TRL6–9 development and deployment phases, and realise tangible operational and safety benefits from AI-based systems in future manned flight systems.

Author Contributions: Conceptualization, E.K., B.K., J.v.R. and R.Z.; methodology, E.K., B.K., J.v.R. and R.Z.; formal analysis, E.K., B.K., J.v.R. and R.Z.; investigation, E.K. and J.v.R.; resources, B.K. and J.v.R.; data curation, B.K. and J.v.R.; writing—original draft preparation, E.K., B.K. and J.v.R.; writing—review and editing, B.K.; visualisation, B.K.; project administration, R.Z.; funding acquisition, R.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported through internal funding from the Royal Netherlands Aerospace Centre.

Informed Consent Statement: Informed consent was obtained from all participants involved, with full anonymity (no names taken and ‘Chatham House rules’ invoked), egalitarian discussion (‘ranks’ left out), no audio/video recordings, and voluntary participation.

Data Availability Statement: The main output contributions presented in this study are included in the article. Full detail has not been included due to its length and confidentiality conditions associated with the two expert groups. Further inquiries can however be directed to the corresponding author.

Acknowledgments: During the preparation of this manuscript, author(s) used ChatGPT (GPT-5.5) for language editing, including improvements to grammar, spelling, and readability. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AS	AI-based system
B5T	Big Five of Teaming model
CBTA	Competency-Based Training and Assessment
COM	Communication
CPDLC	Controller-Pilot Data-Linked Communication
CRM	Crew Resource Management
EASA	European Union Aviation Safety Agency
EBT	Evidence-Based Training
EFB	Electronic Flight Bag
FAA	Federal Aviation Administration
FMS	Flight Management System
FPA	Aeroplane Flight Path Management, automation (pilot core competency)
FPM	Aeroplane Flight Path Management, manual control (pilot core competency)
HAIRM	Human–AI Resource Management
HARVIS	Human Aircraft Roadmap for Virtual Intelligent System
HAT	Human–AI Teaming
HF	Human Factors
IATA	International Air Transport Association
KNO	Application of Knowledge (pilot core competency)
LoA	Level of Automation
LOC-I	Loss of Control in Flight
LTW	Leadership and Teamwork (pilot core competency)
METAR	Meteorological Aerodrome Report
MFD	Multi-Function Display
NGT	Nominal Groups Technique
NASA	National Aeronautics and Space Administration
NLP	Natural Language Processing
NOTAM	Notice to Airmen
OBs	Observable Behaviours
PF	Pilot Flying
PIREP	Pilot Report
PM	Pilot Monitoring
PRO	Application of Procedures and Compliance with Regulations (pilot core competency)
PSD	Problem Solving and Decision-Making (pilot core competency)
RWY	Runway
SAW	Situation Awareness and Management of Information (pilot core competency)
SIGMET	Significant Meteorological Information
SOP	Standard Operating Procedure
SPO	Single-Pilot Operations
TWY	Taxiway
UC	Use case
WLM	Workload Management (pilot core competency)

References

- Salas, E.; Sims, D.E.; Shawn Burke, C. Is There A “Big Five” in Teamwork? *Small Group Res.* **2005**, *36*, 555–599. [CrossRef]
- Helmreich, R.L.; Merritt, A.C.; Wilhelm, J.A. The Evolution of Crew Resource Management Training in Commercial Aviation. *Int. J. Aviat. Psychol.* **1999**, *9*, 19–32. [CrossRef] [PubMed]
- Berretta, S.; Tausch, A.; Ontrup, G.; Gilles, B.; Peifer, C.; Kluge, A. Defining Human-AI Teaming the Human-Centered Way: A Scoping Review and Network Analysis. *Front. Artif. Intell.* **2023**, *6*, 1250725. [CrossRef] [PubMed]
- Cooke, N.; Demir, M.; Huang, L. A Framework for Human-Autonomy Team Research. In *Engineering Psychology and Cognitive Ergonomics. Cognition and Design*; Harris, D., Li, W.-C., Eds.; Lecture Notes in Computer Science; Springer International Publishing: Cham, Switzerland, 2020; Volume 12187, pp. 134–146. ISBN 978-3-030-49182-6.
- Dubey, A.; Abhinav, K.; Jain, S.; Arora, V.; Puttaveerana, A. HACO: A Framework for Developing Human-AI Teaming. In *Proceedings of the 13th Innovations in Software Engineering Conference (Formerly Known as India Software Engineering Conference)*; ACM: Jabalpur, India, 2020; pp. 1–9.
- Lee, J.D.; See, K.A. Trust in Automation: Designing for Appropriate Reliance. *Hum. Factors* **2004**, *46*, 50–80. [CrossRef] [PubMed]
- Lyons, J.B.; Sycara, K.; Lewis, M.; Capiola, A. Human–Autonomy Teaming: Definitions, Debates, and Directions. *Front. Psychol.* **2021**, *12*, 589585. [CrossRef] [PubMed]
- Schelble, B.G.; Flathmann, C.; McNeese, N.J.; Freeman, G.; Mallick, R. Let’s Think Together! Assessing Shared Mental Models, Performance, and Trust in Human-Agent Teams. *Proc. ACM Hum.-Comput. Interact.* **2022**, *6*, 13. [CrossRef]
- Greenberg, A.M.; Marble, J.L. Foundational Concepts in Person-Machine Teaming. *Front. Phys.* **2023**, *10*, 1080132. [CrossRef]
- Kaliardos, W.N. Enough Fluff: Returning to Meaningful Perspectives on Automation. In *Proceedings of the 2022 IEEE/AIAA 41st Digital Avionics Systems Conference (DASC)*; IEEE: Portsmouth, VA, USA, 2022; pp. 1–9.
- Shively, R.J.; Lachter, J.; Koteskey, R.; Brandt, S.L. *Crew Resource Management for Automated Teammates (CRM-A)*; NASA Ames Research Center: Moffett Field, CA, USA, 2018.
- Salmon, P.M.; King, B.; Hall, D.; McLean, S.; Thompson, J.; Cooke, N.; Salas, E.; Loft, S.; Read, G.J.M. The Big Five Model of Teamwork and Human Autonomy Teams: A Scoping Review. *Appl. Ergon.* **2026**, *135*, 104761. [CrossRef] [PubMed]
- Salas, E.; Cooke, N.J.; Rosen, M.A. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Hum. Factors J. Hum. Factors Ergon. Soc.* **2008**, *50*, 540–547. [CrossRef] [PubMed]
- Helmreich, R.L.; Foushee, H.C. Why CRM? Empirical and Theoretical Bases of Human Factors Training. In *Crew Resource Management*; Academic Press: Cambridge, MA, USA, 2019; pp. 3–52.
- Pollitt, A.; Vlaskamp, D.; Blundell, J.; Landman, A. Whither CRM?—30 Years on: A Narrative Review and Position Paper on the Future of Aviation CRM Training. *Aerospace* **2026**, *13*, 500. [CrossRef]
- Salas, E.; Fowlkes, J.E.; Stout, R.J.; Milanovich, D.M.; Prince, C. Does CRM Training Improve Teamwork Skills in the Cockpit?: Two Evaluation Studies. *Hum. Factors J. Hum. Factors Ergon. Soc.* **1999**, *41*, 326–343. [CrossRef]
- Salas, E.; Burke, C.S.; Bowers, C.A.; Wilson, K.A. Team Training in the Skies: Does Crew Resource Management (CRM) Training Work? *Hum. Factors J. Hum. Factors Ergon. Soc.* **2001**, *43*, 641–674. [CrossRef] [PubMed]
- Terzioğlu, M. The Effects of Crew Resource Management on Flight Safety Culture: Corporate Crew Resource Management (CRM 7.0). *Aeronaut. J.* **2024**, *128*, 1743–1766. [CrossRef]
- IATA. *Competency-Based Training and Assessment (CBTA) Expansion within the Aviation System*; International Air Transport Association (IATA): Montreal, QC, Canada, 2024.
- IATA. *Data Report for Evidence-Based Training—Amendment 2021*; International Air Transport Association (IATA): Montreal, QC, Canada, 2021.
- EASA. *EBT Manual V2.3*; European Aviation Safety Agency (EASA): Cologne, Germany, 2025.
- HAIKU Project. HAIKU: Human-AI Teaming Knowledge and Understanding for Aviation Safety. Available online: <https://haikuproject.eu> (accessed on 23 September 2026).
- SESAR. Industrial Research Project: JARVIS—Just A Rather Very Intelligent System. Available online: <https://www.sesarju.eu/projects/JARVIS> (accessed on 23 September 2026).
- Cohen, M.C.; Mancenido, M.V.; Chiou, E.K.; Cooke, N.J. Teamness and Trust in AI-Enabled Decision Support Systems: Current Challenges and Future Directions. In *Proceedings of the CEUR Workshop Proceedings, Munich, Germany, 26 June 2023*; Volume 3456, pp. 175–187.
- FAA. *Roadmap for Artificial Intelligence Safety Assurance*; Federal Aviation Administration (FAA): Washington, DC, USA, 2024.
- EASA. *EASA Concept Paper: Guidance for Level 1 & 2 Machine-Learning Applications*; European Aviation Safety Agency (EASA): Cologne, Germany, 2024.
- Kirwan, B. Human Factors Requirements for Human-AI Teaming in Aviation. *Future Transp.* **2025**, *5*, 42. [CrossRef]
- CORDIS. Future Sky Safety. Available online: <https://cordis.europa.eu/project/id/640597> (accessed on 23 September 2026).
- Gunning, D. Explainable Artificial Intelligence (XAI). *Sci. Robot.* **2019**, *4*, eaay7120. [CrossRef] [PubMed]

30. Venditti, R.; Minaskan, N.; Biliri, E.; Villegas, M.; Kirwan, B.; Westin, C.; Basjuka, J.; Pozzi, S. Construal Level Theory (CLT) for Designing Explanation Interfaces in Operational Contexts. In Proceedings of the 13th International Conference on Human Interaction & Emerging Technologies: Artificial Intelligence & Future Applications, Costa Del Sol, Spain, 22 April 2025; Volume 161, pp. 268–276.
31. BEA. *Accident to the Airbus A330 Registered F-GZCP Operated by Air France Occured on 06/01/2009 in the Atlantic Ocean/Flight AF447 Rio Paris*; Bureau d’Enquetes et d’Analyses Pour la Sécurité de L’aviation Civile: Le Bourget Cedex, France, 2012.
32. KNKT. *Aircraft Accident Investigation Report: PT. Indonesia Air Asia Airbus A320-216; PK-AXC Karimata Strait*; Komite Nasional Keselamatan Transportasi: Central Jakarta, Indonesia, 2014.
33. Dias, M.d.O.; Teles, A.; Lopes, R.d.O.A. Could BOING 737 MAX Crashes Be Avoided? Factors That Undermined Project Safety. *Glob. Sci. J. GSJ* **2020**, *8*, 187–196.
34. Duchevet, A.; Imbert, J.-P.; De La Hogue, T.; Ferreira, A.; Moens, L.; Colomer, A.; Cantero, J.; Bejarano, C.; Rodríguez Vázquez, A.L. HARVIS: A Digital Assistant Based on Cognitive Computing for Non-Stabilized Approaches in Single Pilot Operations. In Proceedings of the 25th EURO Working Group on Transportation Meeting (EWGT 2022), Bratislava, Slovakia, 14–16 September 2022; pp. 253–261.
35. Field, J.N.; Boland, E.J.; van Rooij, J.M.; Mohrmann, J.F.W.; Smeltink, J.W. *Startle Effect Management*; European Aviation Safety Agency (EASA): Amsterdam, The Netherlands, 2018.
36. Vlaskamp, D.; Landman, A.; Van Rooij, J.; Li, W.-C.; Blundell, J. Airline Pilots’ Perceived Operational Benefit of a Startle and Surprise Management Method: A Qualitative Study. In *Proceedings of the 2nd International Conference on Cognitive Aircraft Systems*; SCITEPRESS—Science and Technology Publications: Toulouse, France, 2024; pp. 29–34.
37. Vlaskamp, D.; Landman, A.; Van Rooij, J.; Blundell, J. Recovery from Startle and Surprise: A Survey of Airline Pilots’ Operational Experience Using a Startle and Surprise Management Method. *Int. J. Ind. Ergon.* **2025**, *107*, 103733. [[CrossRef](#)]
38. Duchevet, A.; Vo, D.-B.; Peyruqueou, V.; De-La-Hogue, T.; Garcia, J.; Causse, M.; Imbert, J.-P. FOCUS: An Intelligent Startle Management Assistant for Maximizing Pilot Resilience. In *Proceedings of the 2nd International Conference on Cognitive Aircraft Systems*; SCITEPRESS—Science and Technology Publications: Toulouse, France, 2024; pp. 5–12.
39. Kirwan, B.; Noren, A. A Review of 27 European Aviation Human-AI Teaming Research Projects: What Have We Learned, What Have We yet to Learn? Presented at the Human Computer Interaction (HCI) International 2026, Montreal, QC, Canada, 26–31 July 2026.
40. McMillan, S.S.; King, M.; Tully, M.P. How to Use the Nominal Group and Delphi Techniques. *Int. J. Clin. Pharm.* **2016**, *38*, 655–662. [[CrossRef](#)] [[PubMed](#)]
41. Varga-Atkins, T.; McIsaac, J.; Willis, I. Focus Group Meets Nominal Group Technique: An Effective Combination for Student Evaluation? *Innov. Educ. Teach. Int.* **2017**, *54*, 289–300. [[CrossRef](#)]
42. Endsley, M.R. Ironies of Artificial Intelligence. *Ergonomics* **2023**, *66*, 1656–1668. [[CrossRef](#)] [[PubMed](#)]
43. Chirayath, G.; Premamalini, K.; Joseph, J. Cognitive Offloading or Cognitive Overload? How AI Alters the Mental Architecture of Coping. *Front. Psychol.* **2025**, *16*, 1699320. [[CrossRef](#)] [[PubMed](#)]
44. Christoffersen, K.; Woods, D.D. How To Make Automated Systems Team Players. In *Advances in Human Performance and Cognitive Engineering Research*; Elsevier: Amsterdam, The Netherlands, 2002; Volume 2, pp. 1–12. ISBN 978-0-7623-0748-7.
45. Hörmann, H.J. FOR-DEC—A Prescriptive Model for Aeronautical Decision Making. In Proceedings of the 21st Conference of the European Association for Aviation Psychology, Dublin, Ireland, 2 March 1995; Volume 3.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.