

Article

# Human Factors Requirements for Human-AI Teaming in Aviation

Barry Kirwan 

EUROCONTROL, EIH, Bois-des Bordes, F-91222 Bretigny sur Orge, France; barry@kirwanhf.com

**Abstract:** The advent of Artificial Intelligence in the cockpit and the air traffic control centre in the coming decade could mark a step-change improvement in aviation safety, or else could usher in a flush of ‘AI-induced’ accidents. Given that contemporary AI has well-known weaknesses, from data biases and edge or corner effects, to outright ‘hallucinations’, in the mid-term AI will almost certainly be partnered with human expertise, its outputs monitored and tempered by human judgement. This is already enshrined in the EU Act on AI, with adherence to principles of human agency and oversight required in safety-critical domains such as aviation. However, such sound policies and principles are unlikely to be enough. Human interactions with current automation in the cockpit or air traffic control tower require extensive requirements, methods, and validations to ensure a robust (accident-free) partnership. Since AI will inevitably push the boundaries of traditional human-automation interaction, there is a need to revisit Human Factors to meet the challenges of future human-AI interaction design. This paper briefly reviews the types of AI and ‘Intelligent Agents’ along with their associated levels of AI autonomy being considered for future aviation applications. It then reviews the evolution of Human Factors to identify the critical areas where Human Factors can aid future human-AI teaming performance and safety, to generate a detailed requirements set organised for Human AI Teaming design. The resultant requirements set comprises eight Human Factors areas, from Human-Centred Design to Organisational Readiness, and 165 detailed requirements, and has been applied to three AI-based Intelligent Agent prototypes (two cockpit, one air traffic control tower). These early applications suggest that the new requirements set is scalable to different design maturity levels and different levels of AI autonomy, and acceptable as an approach to Human-AI Teaming design teams.



Academic Editor: Lynnette Dray

Received: 13 December 2024

Revised: 16 March 2025

Accepted: 21 March 2025

Published: 5 April 2025

**Citation:** Kirwan, B. Human Factors Requirements for Human-AI Teaming in Aviation. *Future Transp.* **2025**, *5*, 42. <https://doi.org/10.3390/futuretransp5020042>

**Copyright:** © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** aviation; human-AI teaming; intelligent agents; human factors requirements

## 1. Artificial Intelligence in Aviation

Artificial Intelligence is beginning to appear in civil aviation, principally via Machine Learning and Deep Learning approaches [1] in a wide variety of applications, including, for example, flight operations and unmanned aerial vehicles [2], weather prediction [3], and numerous improvements in air traffic management [4]. Such augmentations to system performance largely enhance efficiency and safety of operations. There is, however, a broad scope for the increased uptake of AI in cockpit and air traffic control settings, in the guise of future ‘Intelligent Agents’ that could assist pilots and air traffic controllers interactively and dynamically in real-time flight operations. This would rely on Machine Learning approaches rather than Generative AI models such as ChatGPT [5] and other Large Language Models, as the latter are currently ruled out by the EU Act on AI [6] for safety-critical systems due to the problem of hallucinations [7,8]. But Intelligent Agents

would afford more human-machine interaction, lending more autonomy to the IA to take on tasks, as well as giving advice during challenging and time-critical flight upsets.

Assistance by AI at such a level is known as Human AI Teaming (HAT: also Human Autonomy Teaming and Human Machine Teaming) [9]. Early examples of HAT prototypes include cockpit support for startle response [10], determining safe alternate airports due to severe weather degradation [3], delivering air traffic control sector workload prognoses [11], cockpit management of unstable approaches [12], single-pilot operations [13], and ATC support for landing/arrival sequencing [14].

It is useful to put HAT into context by considering actual examples of use cases. This paper is based on research under the auspices of the Horizon Europe HAIKU project (<https://haikuproject.eu/> accessed on 20 March 2025), which is exploring six futuristic HAT use cases—two cockpit, two air traffic, and two airport—with varying levels of AI autonomy. Three of the use cases are more team-oriented in nature (the other three are more similar to machine learning support for an airport, use of a chatbot by passengers, and an early concept study of AI controlling drone traffic with very little human interaction), as outlined below and illustrated thematically in Figure 1 (all six use cases are shown):

1. UC1—a cockpit AI to help a single pilot recover from a sudden event that induces ‘startle response’. Startle response is when a pilot in the cockpit is startled by a sudden, unexpected event in or outside the cockpit, leading to a temporary disruption of cognitive functioning, usually lasting approximately 20 s. [10]. The AI directs the pilot concerning instruments to focus on in order to resolve the emergency situation. Although the AI supports and directs the pilot, the pilot remains in charge throughout.
2. UC2—a cockpit AI to help flight crew re-route an aircraft to a new airport destination due to deteriorating weather or airport closure, for example, taking into account a large number of factors (e.g., category of aircraft and runway length; fuel available and distance to airport; connections for passengers, etc.). The flight crew remain in charge, but communicate/negotiate with the AI to derive the optimal solution.
3. UC4—a digital assistant for remote tower operations, to alleviate the tower controller’s workload by carrying out repetitive tasks. The tower controller monitors the situation and intervenes if there is a deviation from normal (e.g., a go-around situation, or an aircraft that fails to vacate the runway). The controller is in charge, but the AI can take certain actions unless the controller vetoes them.

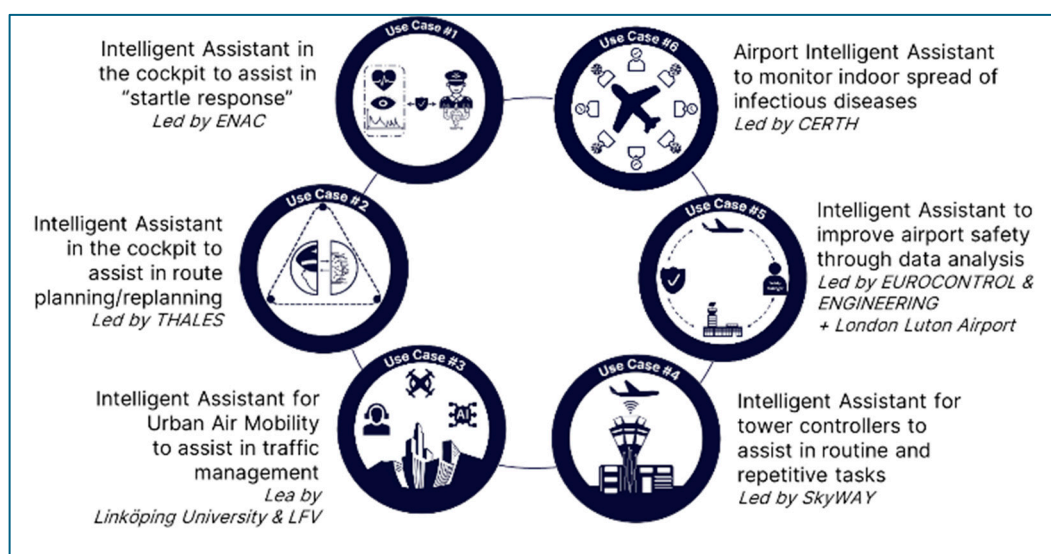


Figure 1. HAIKU HAT Use Cases.

These use cases are in relatively early design stages, but AI prototypes have been developed allowing human-in-the-loop simulations to explore human-AI teaming in realistic aviation scenarios with licensed pilots and controllers as participants. These use cases therefore serve as useful ‘testbeds’ for Human Factors and HAT approaches.

The fundamental question arising for such future HAT concepts, and the principal focus of this paper, is how to ensure that such HAT systems will be safe and operationally effective if implemented, given that most existing Human Factors assurance approaches have not been designed or developed to deal with AI and HAT.

Currently, new cockpit or ATM system designs are subject to Human Factors design requirements e.g., EASA’s CS25.1302 for cockpit Human Factors design ([https://www.easa.europa.eu/sites/default/files/dfu/CS-25\\_Amdt%203\\_19.09.07\\_Consolidated%20version.pdf](https://www.easa.europa.eu/sites/default/files/dfu/CS-25_Amdt%203_19.09.07_Consolidated%20version.pdf) (accessed on 20 March 2025)), and SESAR Human Performance Assurance Process (<https://www.sesarju.eu/sites/default/files/documents/transversal/SESAR%202020-%20-%20Human%20Performance%20Assessment%20Guidance.pdf> (accessed on 20 March 2025)) guidance for air traffic towers and en route control centres. Such requirements systems have evolved over many decades of civil aviation, encompassing pilot and air traffic controller (ATCO) experience and insight, incident and accident experience, and Human Factors research. However, AI is developing at a rapid pace and could have profound implications for aviation systems and human-machine interaction. It will be hard for a traditional approach of gaining experience to keep up with AI developments. The last time there was a radical change in human-machine arrangements—namely the introduction of glass cockpits into the industry—it initially led to a spate of ‘automation-assisted accidents’, Appendix 1 in [15]. The key question is therefore how to prepare for AI advances and ensure that near-future Human-AI Teaming concepts of operation are both effective and safe.

Aviation regulators such as the European Union Aviation Safety Agency (EASA), as well as regulators in other safety-critical domains such as Oil and Gas [16], are well aware of this predicament, and EASA in particular have issued advance guidance for aviation ‘Human-AI Teaming’ scenarios [17,18], depending on the level of autonomy of the AI under consideration for implementation into flight operations. Such guidance is both timely and welcome and sets safeguards or ‘guardrails’ for designers and developers of future aviation HAT system concepts.

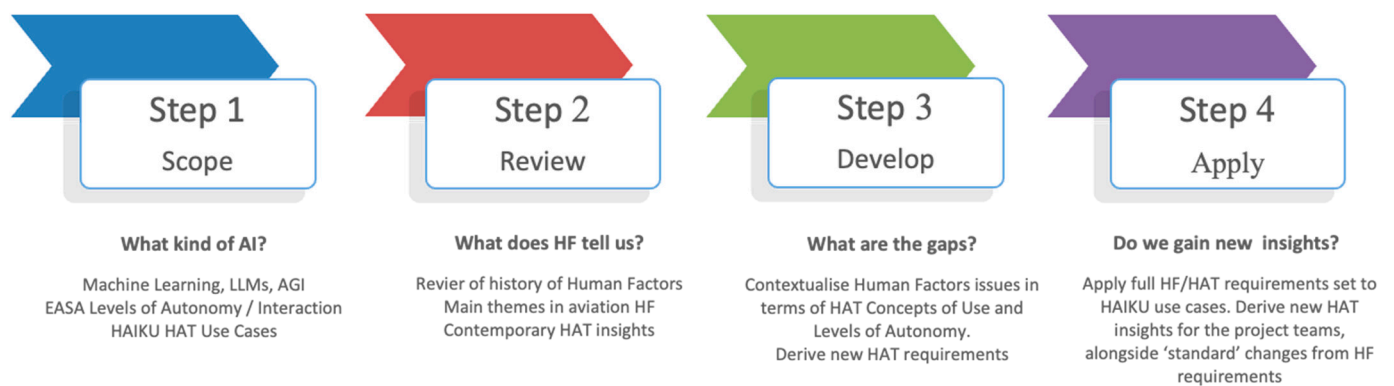
In parallel with the issuance of EASA’s guidance, the HAIKU project has been developing and testing its own set of guidance material for future aviation HAT developers. HAIKU’s mission, that of aiding designers in developing human-centric HAT systems, leads to a broader scope than that of a regulator, for example, also focusing on organisational practices, and staff competence and well-being factors that can affect not only technology acceptance, but indirectly system performance and the ability to detect AI errors [19]. This broader scope leads to a wider range of Human Factors requirements than a regulatory set. This paper therefore documents the process for the development of the HAIKU HAT requirements and shows both their application and added value via insights gained from three HAIKU HAT use cases.

## 2. Research Questions and Approach

The development of a preliminary set of new Human Factors HAT requirements for aviation applications has been guided by four over-arching questions, shown as four steps in Figure 2:

1. What type of AI and Human-AI Teaming characteristics are likely in future aviation concepts?

2. What does the existing body of Human Factors knowledge suggest we should focus on for HAT systems?
3. What are the Human-AI Teaming requirements arising from the gap between what we have currently in HF requirements systems, and the challenges of future HAT concepts of operation?
4. Are the new HAT requirements fit for purpose, i.e., can they be used by project teams to identify new system design insights to safeguard and optimise human-AI team performance?



**Figure 2.** Overview of HAT Requirements Development and Testing Approaches.

The first step seeks to sharpen the focus on the type of AI being considered, namely, Machine Learning (ML) systems developed to produce Intelligent Assistants/ Agents to support the air traffic controller in an air traffic control centre or tower, or pilots in the cockpit, whether for current dual-pilot operations or potential future single-pilot operations. The ML focus is because currently GenAI/ Large Language Models or LLMs are not allowed for safety-critical systems according to EU law, and Artificial General Intelligence does not yet exist [20,21]. The AI focus of interest is further actualised by considering the type of interaction with operational end users, in particular, the level of autonomy of the AI (the relative degree of control by human and AI elements). Given that the focus of this paper is aviation, the EASA guidance on levels of automation is used as the principal framework to consider levels of AI assistance and style of interaction and work-sharing between AI and human elements. These levels are exemplified by three HAIKU use cases, concretising the concepts of operation of future AI-based systems.

The second step involves a review of the existing knowledge base of Human Factors from its outset 7 decades ago until recent early studies of Human AI Teaming (HAT). After an early focus on physical ergonomics, much of the remaining focus of Human Factors has been on people interacting with automation. AI is likely to change this picture, however, as it can have a level of autonomy not seen in current aviation systems. The historical review, although inevitably somewhat subjective according to the author's HF experience over the past 4 decades, allows consideration of how some of the formative theories of Human Factors might have addressed (and could still address) the potential HAT scenarios afforded by AI-based systems.

Step 3 concerns contextualisation, essentially rendering the theoretical Human Factors considerations into practical requirements that can be tested in use case applications. Existing industrial HF requirements are not theoretical, since they are used as design support as well as in certification systems. New HAT requirements must therefore also be contextual so they can be applied to actual (existing or design-stage) aviation AI-based systems and system elements, verifiable via system performance evidence. HF Requirements therefore embody Human Factors theory but are contextualised according to workplace equipment

and modes of interaction with operational end users (in this case, pilots and air traffic controllers). Step 3 essentially considers the pertinent aspects from the HF review in the context of future HAT concepts of operation. The result is a set of new HAT (HF) requirements couched in language designers can work with.

The fourth, final step sees the application of the new requirements to the HAT use cases, to see if they result in novel insights and design improvements that are convincing to the respective project teams. If this is so, then the new set of HAT requirements can serve as a preliminary approach for other future HAT systems evaluations, the requirements set to be updated as user experience and HF research evolve with respect to HAT development and implementation.

### 3. Step 1: Scoping the HAT Requirements

#### 3.1. What Kind of AI?

Artificial Intelligence (AI) has a relatively long history, as illustrated in the lower half of Figure 3, dating back to the 1950s [22,23] and arguably even back to the 19th Century via Charles Babbage’s counting machines [24]. AI initially had the goal of replicating certain human cognitive abilities, albeit with more reliability. Hence, the fundamental notion was one of computation, of ‘crunching the numbers’, which humans could do given enough time, but rarely error-free. As computing power grew, it soon surpassed what humans could do even given infinite time, and with the advent of deep learning [1], it could sometimes derive novel solutions to problems that we would never have thought of [21]. Thus, the goal of AI shifted from replicating human cognitive capabilities to surpassing them.

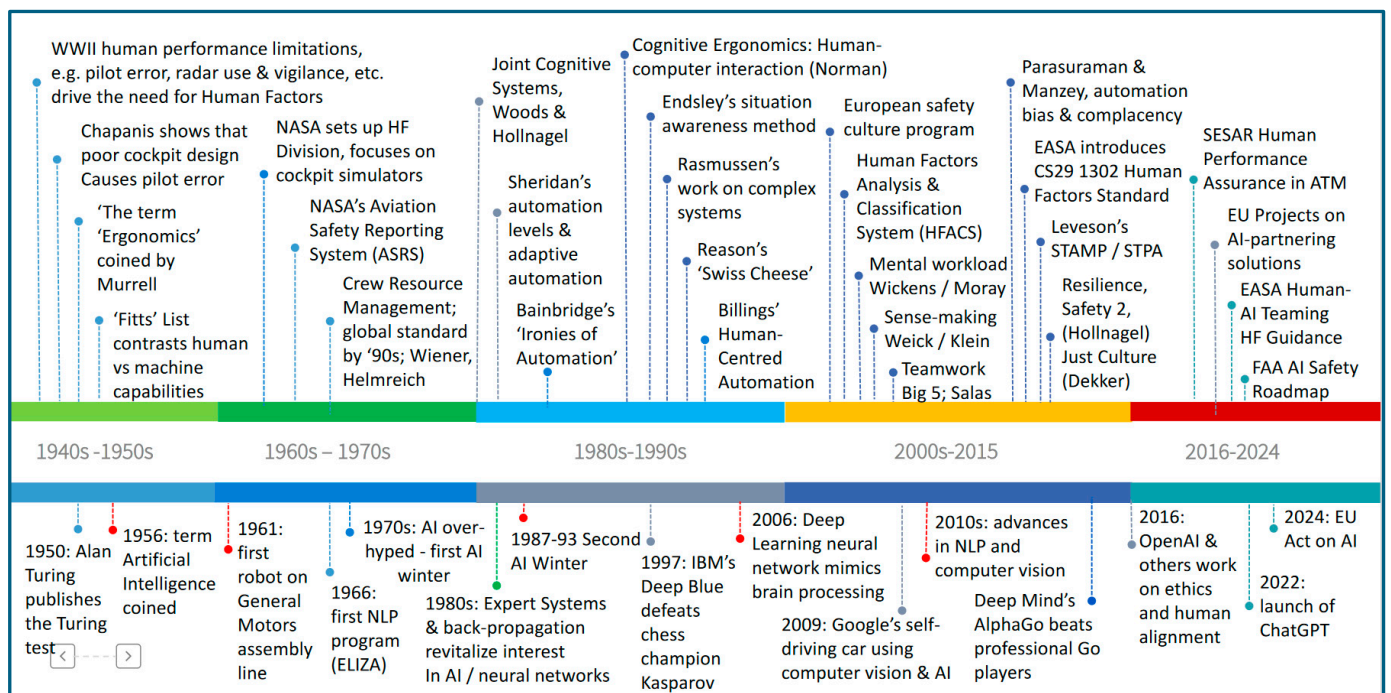


Figure 3. Timeline of Parallel Evolution of AI (lower section) and Human Factors (upper section).

The AI timeline in the lower part of Figure 1 shows that many inventions or initiatives believed generally to be relatively recent, have in fact been around for a long time, including robots, natural language processing, neural networks, computer vision, and even self-driving cars. However, many of these were prototypes, and did not go ‘mainstream’ until recently (robot assembly for car manufacturing being a notable exception) [25].

As Figure 1 also shows, AI has already experienced two ‘AI winters’, where its promise vastly exceeded its delivery, leading to an investment and technological cliff edge, so that AI largely disappeared for a while from the public eye [26,27]. But as computing power increased by orders of magnitude, and as Machine Learning approaches began to show their worth and earn their keep, AI once again caught the public’s attention, and has attracted both brilliant minds and eye-watering levels of investment. The public release of ChatGPT [5] and other Generative AI (hereafter called GenAI) systems in 2022 transformed AI from being a technophile, jargon-imbued subject to being a workplace and even household commodity. Even if most people have little idea of how AI works, they know that it can do things for them, whether helping their research, finishing or finessing a report, or providing a diagnosis of their health symptoms. Though many have experienced firsthand the errors and biases of GenAI systems, they accept the trade-off between its accessibility and instant power to answer questions and requests, against the occasional inaccuracies or plausible fabrications called hallucinations [8].

A definition of Artificial Intelligence pertinent to this paper is as follows [28]:

*“...the broad suite of technologies that can match or surpass human capabilities, particularly those involving cognition.”*

Most AI systems today represent what is known as Narrow AI [29], namely AI-based systems and services focused on a specific domain such as aviation, typically using Machine Learning (ML). In ‘normal’ programming, e.g., for automation, a machine is programmed exactly what to do and, given stable inputs, that is exactly what it will do. The code may be complex, but is completely explainable, at least to a software analyst or data scientist. Narrow AI supports humans in their analysis, decision-making, and other tasks. In cases where tasks are well-specified and predictable, it can execute its functions without human intervention and with minimal supervision.

Machine Learning can develop models and predictions that perhaps, given enough time, humans could do. Deep Learning is different and can come up with solutions humans likely would never think of. That said, Deep Learning typically uses artificial neural networks, themselves inspired by the way human brains work. Deep Learning is used for some of the more complex human cognitive processes we take for granted, such as natural language processing [30] and image recognition [31], and also for tackling complex problems such as finding cures for intractable diseases [32] (see [9,21,25,33–35] for a general summary of contemporary AI and HAT application areas).

The hallmark of Generative AI (GenAI) tools, such as ChatGPT, Google Gemini, DALL-E, and DeepSeek-R1 are that they can create new content that is often indistinguishable from human-generated content. They utilise deep learning neural networks trained on vast data sets and natural language processing to render interaction with human users smoother. Large Language Models (LLMs) like ChatGPT can respond to a human user to any query. Whether the response is a valid or correct one is another matter [36].

The point about ML, Deep Learning, and GenAI is that these systems are still ‘crunching the numbers’, and in the case of LLMs they are generally predicting the next word in a sequence. Nowhere is there understanding, or a mind, or thought. They may be very useful, but they are all essentially ‘idiot savants’ [28]. The problem is that particularly with advanced LLMs, it can feel to the user as if they are interacting with a person (known as anthropomorphising or personifying AI) [37,38]. This can matter in a safety-critical environment and is returned to at the end of Step 2.

Artificial General Intelligence (AGI) does not yet exist but is predicted to emerge in the coming decades, e.g., by 2041 [20]. It would effectively comprise a mind capable of independent reasoning and could therefore in theory attain sentience. AGI would be able to set its own objective functions (goals), and its intelligence could grow very rapidly to

eclipse that of human beings [21]. While the step from GenAI to AGI may seem small, given the way LLMs such as ChatGPT can summarise vast swathes of knowledge, in practice the step is significant. For the rest of this paper, therefore, AGI is ignored, as it may never be realised.

What could exist in a relatively short time frame are ‘Intelligent Assistants’ (IAs) or ‘Digital Colleagues’, as envisaged by the concept of Human-AI Teaming, also called Human Autonomy Teaming, both using the same acronym HAT or else Human Machine Teaming [9,11,13,17–19,34,35]. The essential nature of HAT and in particular Intelligent Assistants (IAs), which differentiates it from conventional and contemporary automation, is the idea of an IA having a degree of autonomy/agency such that it can have intent, form goals and decisions, and execute such decisions, in collaboration with one or more human agents. Such IAs therefore could, in the coming decade, collaborate with and even converse with pilots and air traffic controllers in operational settings. These IAs are still Narrow AI (ML, including Deep Learning) with or without Natural Language Processing capability. They would not be GenAI, at least not now in the current legal framework in Europe given the EU Act on AI, and would certainly not constitute AGI. Early examples of HAT prototypes have already been cited earlier in the introduction [10–14] and are further expanded in Section 3.3.

First, it is necessary to further define the characteristics of intelligent agents, as meta-reviews have noted the construct confusion currently existing in the domain of Human-AI Teaming [34,35]. This is settled pragmatically by consideration of the six EASA HAT categories, as they define the level of AI (or IA) autonomy at each level.

### 3.2. AI Levels of Autonomy—the European Aviation Regulatory Perspective

Given that Intelligent Assistants (IAs) are intended to interact and collaborate with humans in aviation contexts, there is clearly an increase in their autonomy compared to automation simply presenting information and warnings. This shift in degree of autonomy affects the relationship between the human and the AI-based automation in two ways. First, the information or advice, or even executive action, can be based on calculations that are opaque to the end users (e.g., pilots), because the level of complexity and transparency of how AIs derive their answers means that no amount of theoretical training for pilots will enable them to follow the IA’s ‘reasoning’, unless an additional layer of ‘explainability’ is afforded to the pilot by the AI-based automation. The pilot must therefore come to trust the IA, or its advice will be rejected. Second, the role of the pilot is affected, because currently the pilot is always ready (i.e., trained) to take over in case the automation (e.g., automatic landing) fails. The fundamental notion of collaboration suggests an interdependence; control becomes to a greater or lesser degree shared between human and AI.

It is therefore useful to map the degree of IA autonomy, and autonomy sharing between human and IA. Increasing levels of autonomy can be represented on a scale, and the most influential scale in aviation currently is that provided by the European Union Aviation Safety Agency (EASA). Recent guidance on Human-AI Teaming (HAT) from the EASA [17,18] envisages six categories of future Human-AI partnerships:

- 1A—Machine learning support (already existing today)
- 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 running the show, 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.

It has been argued that AI innovation, for all its benefits, is essentially ‘just more automation’ supporting the human operator [37]. The critical threshold in AI autonomy where this may no longer hold appears to be between category 2A and 2B [39,40], since this is different from what we have today in civil aviation cockpits. The distinction between 2A and 2B is clarified by EASA as follows, including what each one is and is not [40]:

- Cooperation Level 2A: cooperation is a process in which the AI-based system works to help the end user accomplish his or her own goal. The AI-based system works according to a predefined task-allocation pattern with informative feedback to the end user on the decisions and/or actions implementation. The cooperation process follows a directive approach. Cooperation does not imply a shared situation awareness between the end user and the AI-based system. Communication is not a paramount capability for cooperation.
- Collaboration Level 2B: collaboration is a process in which the end user and the AI-based system work together and jointly 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.

There are currently no [AI-based] civil aviation systems that autonomously share tasks with front-line users (e.g., pilots) and can negotiate, make trade-offs, change priorities, and initiate and execute tasks under their own initiative. Even for ‘lesser’ autonomy levels such as 1B to 2A, and also for 3A, there remains the opacity issue, wherein the AI is often akin to a ‘black box’, and as such may surprise, confound, or confuse the end users, because most end users will not be able to follow the computations underpinning AI advice. This leads to the *raison d’être* of this paper, namely that AI and IAs represent something novel, and as such, conventional automation interface design and Human Factors approaches may not be sufficient to assure the safe use and acceptability of such systems. At the very least, issues such as roles and responsibilities, trust in automation, and situation awareness will have additional significant nuances that may require augmentation of existing approaches. Furthermore, new issues such as operational explainability of AI systems may require completely new approaches or design requirements [34,35].

### 3.3. Example Intelligent Agent Use Cases

Three aviation HAT Intelligent Agent use cases are presented below, from the HAIKU project. The first use case (UC1) addresses support to a single pilot in the event of startle response, wherein a sudden unexpected serious event (e.g., a lightning strike) can cause ‘startle’, leading to diminished cognitive performance for a short period of time (e.g., 20 s) [10]. The FOCUS (Flight Operational Companion for Unexpected Situations) IA supports the pilot firstly by detecting startle via various psycho-physiological sensors (breathing, heart rate, skin conductance etc.) analysed by a trained AI, using an AI technique known as Extreme Gradient Boosting (see [1]). Second, it analyses where the pilot is looking compared to where the pilot should be looking given the situation, and if different, FOCUS highlights the relevant parameter on the cockpit displays (see Figure 4). This is effectively directed situation awareness. It is based on Standard Operating Procedures (SOPs) for emergency events (e.g., lightning strike), coupled with sensors related to dynamic flight parameters (which together define for the IA what is happening and what needs to be done), compared to where the pilot is looking (via eye tracking), and what the pilot is doing (all in real time). The IA may therefore highlight certain displays or display segments in the cockpit (e.g., the Primary Flight Display, or PFD) to the single pilot, e.g., the



Vertical Speed display. Once the pilot has looked at the display or display element, FOCUS then considers what is next. If the pilot looks at the correct next display, there is no further need for highlighting.



**Figure 4.** Images from HAIKU Use Case 1 simulation (FOCUS simulation development at Ecole Nationale Aviation Civile (ENAC) Labs, Toulouse, France) [10], showing the cockpit simulator (top right), a pilot in the cockpit being supported after startle (top left), an extract from the eye tracking analysis of key parameters the pilot may need to focus on (bottom right), and highlighted directed situation awareness (vertical speed) on the primary flight display (PFD) (bottom left).

FOCUS is there to help the pilot regain situation awareness and stabilise the aircraft. Once the pilot feels back in control, they can cancel the automation, and carry on flying without the IA.

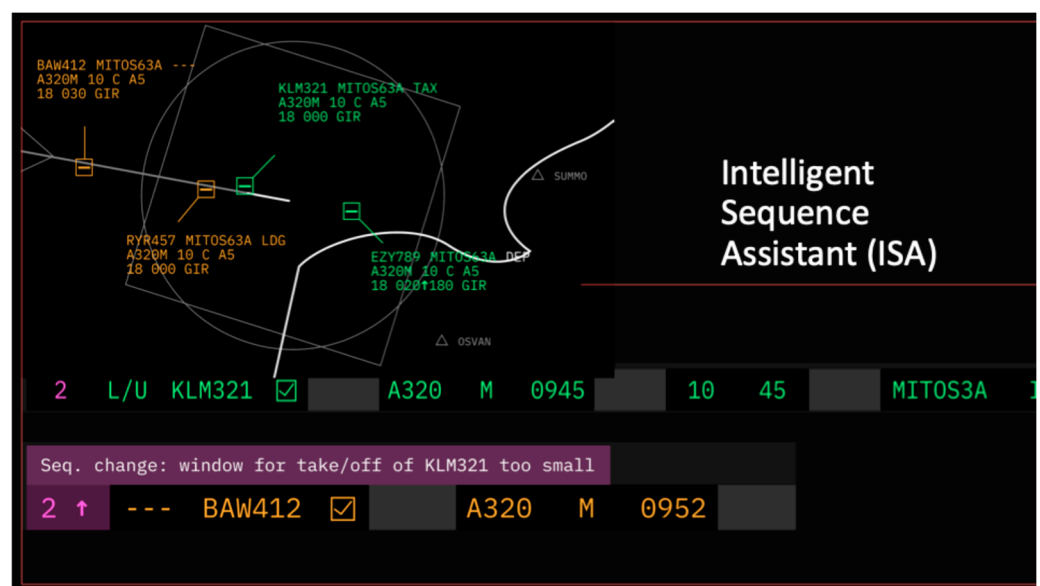
Two real-time simulations have been carried out with airline pilots using FOCUS in a startle scenario (lightning strike). The high-level results show that the FOCUS system aided pilots suffering a startle response in quickly regaining situation awareness and control.

A second HAIKU use case (UC2) concerns advising a two-pilot crew in case of a major diversion, e.g., due to a developing weather pattern (e.g., major thunderstorm), rendering the original destination airport and the pilot's back-up alternate airport inaccessible. Such re-routing while en route can take the flight crew 12–15 min to find a suitable airport according to a number of factors including remaining fuel, runway characteristics for the aircraft and its passenger manifest, whether the airline has a maintenance or ground handling contract with the airport, and how easy it will be for the passengers to reach their final destination from the newly chosen airport. An AI-based assistant has been developed called COMBI (using a data-based supervised learning AI ML approach), which can identify up to three airports within reach. The IA's calculation time takes less than a minute instead of 12–15 min. The flight crew can then select one of the airports and can also query COMBI's selection according to a number of parameters; the IA therefore has a degree of operational explainability. There is also a degree of negotiation between pilots and IA; hence, in EASA terms, the system is category 2A/2B. The cockpit simulator platform for this use case is shown in Figure 5.



**Figure 5.** Simulator platform for COMBI IA (Figure shows the simulator platform for the COMBI AI experiments (Thales Avionics, Bordeaux, France)).

A third HAIKU use case is for an air traffic control tower. In this use case, an Intelligent Sequence Assistant (ISA), based on a neural network, is being developed to support and enhance decision-making for Air Traffic Controllers. ISA optimises runway utilisation in single-runway airports, providing real-time sequencing suggestions for arriving and departing aircraft. ISA computes the ordered sequence of aircraft that will use the runway, the order displayed on the tower controller’s Human Machine Interface (HMI) via numbers placed on the electronic strips of each aircraft, e.g., Figure 6 in the controller’s ‘bay management area’. If an event (e.g., in the figure the BAW is flying faster than expected) triggers a resequencing, ISA updates the sequence in real time, and the results are displayed on the HMI to the controller. ISA also provides explainability on-demand. For example, in Figure 6, ISA signals that the take-off ‘window’ for the KLM is now too small due to the BAW’s increased speed, and so the BAW will land prior to the KLM take-off and takes position ‘2’ in the strip (as indicated by the upward arrow in the bottom left hand corner of Figure 6).



**Figure 6.** ISA in use on the control tower controller’s HMI (ISA under development at Skyway ANSP, Spain).

The real-time assistance provided by ISA ensures timely and accurate forecast updates, allowing Tower Air Traffic Controllers (ATCOs) to manage the traffic flow more efficiently,

with more 'look-ahead time' than currently, as ISA can see further 'upstream'. The benefits are improved decision-making, enhanced runway utilisation, increased operational efficiency, and a safer and more streamlined air traffic flow that reduces the need for 'go-arounds'. In EASA terms, this use case is HAT category 2A, as the AI gives advice, and if the ATCO does nothing, after a certain short period of time ISA will implement the change.

These three use cases form the testbed for the application and evaluation of the HAT Human Factors Requirements developed in the next two sections.

## 4. Review of Human Factors

In order to determine what is needed from Human Factors for future HAT systems, it is useful to briefly review Human Factors as a whole, over its 70-year history, to see how it has developed into its current capability set, in order to better understand what needs to be added to accommodate new AI-based systems and, in particular, Intelligent Assistants. As noted earlier, such a review is inevitably somewhat subjective in terms of the choices of key historical themes or 'waypoints' in the evolution of Human Factors. The choices are based on 4 decades of experience in three industrial sectors (aviation, nuclear power, and petrochemical), including contributions to the development of Human Factors requirements in each of those industries, as well as applying them to actual nuclear plants, air traffic control systems, and oil rig human interface designs.

### 4.1. Key Waypoints in Human Factors

Human Factors and Ergonomics have always focused largely on human work and the workplace. As highlighted in the upper half of Figure 3, Human Factors started out with a focus on physical ergonomics and on the layout of cockpit instruments and controls, for example. By the 1980s and 1990s the focus had shifted to cognitive ergonomics, in parallel with the fact that many work situations involved computing and automation. As work complexity grew, constructs such as situation awareness and mental workload came to the fore, and as automation became the norm, there was a corresponding need to consider complacency and bias, as well as a move from a focus on human error to system resilience. Arguably, these developments over time have prepared Human Factors for the next major step change in human-machine interaction associated with AI, HAT, and IAs.

This section therefore details key milestones and themes in Human Factors over the past 7 decades that have helped application domains such as aviation achieve and sustain very high levels of safety. These are the principal focal areas and capabilities of Human Factors that need to be revisited (or even resurrected), reviewed, and updated in the context of AI, in order to form the basis of a Human Factors Requirements set for AI-based systems. Each milestone is outlined below, with one or more key references, along with the implications for Human Factors assurance of future Human-AI Teaming systems in aviation.

#### 4.1.1. Fitts' List

One of the first landmark achievements was the development of a contrasting list of what machines are good at, versus what humans are good at. This was developed by Paul Fitts and is perennially known as Fitts' List [41,42]. Notably at the time, most of the cognitive 'heavy lifting', including pattern recognition and interpretation of 'noisy' data, was left to the human. A more recent analysis [43] showed that the distinction between people and machine's relative capabilities has shifted, or at least blurred, and is likely to blur further given ongoing advances in AI (both ML and GenAI). It would seem timely to review the relative strengths of human and AI (including LLMs), as otherwise allocation of tasks will revert to the 'left-over principle' (see Ironies of Automation below), in which

humans are allocated tasks and functions the AI can't easily do, whether or not the human can do them or take over from the AI when it fails.

#### 4.1.2. Aviation Safety Reporting System

Human Factors was given an early boost when NASA created a Human Factors group to contribute to the Apollo space missions. At around that time, the Aviation Safety Reporting System (ASRS) was set up by the FAA and NASA (<https://asrs.arc.nasa.gov/> (accessed on 20 March 2025)). ASRS collected data on safety-related events (e.g., mistakes made by pilots), guaranteeing pilots immunity (with a few exceptions) from prosecution if they reported. ASRS provided a constant wellspring of viable information about what was not working in the cockpit or on the ground. This enabled a continuous feedback loop that has run in parallel with the inexorable increase of the role of technology in the cockpit.

If (when) AI begins to be used operationally in the cockpit or ATC Ops Room or Control Tower, ASRS and equivalent systems (e.g., ECCAIRS in Europe [<https://aviationreporting.eu/en>] (accessed on 20 March 2025)) would also likely need additional categories to capture AI-related characteristics of events, and errors that arise during Human-AI Teaming operations (see HFACS, later).

#### 4.1.3. Crew Resource Management

Crew Resource Management [44] (originally called Cockpit Resource Management), was to become one of the major contributions of Human Factors to aviation safety. CRM gave flight crews the means to become more resilient against human errors and teamwork problems, and complemented ongoing work to make the cockpit design itself less error-prone. The need for CRM grew from the world's worst civil aviation air disaster at Tenerife airport in 1977 but was also linked to the United Airlines Flight 173 air crash in 1978 [45]. Both accidents highlighted team and communication errors, and the need for specific training on leadership, decision-making, and communication on the flight deck, and with air traffic control (which in Europe later developed its own version of CRM called Team Resource Management or TRM [46]). CRM focuses on proper use of the available human resources in operational teams and has spread to other domains such as maritime via Bridge Resources Management (BRM) or Maritime Resources management (MRM) [47]. CRM remains strong today, is currently in its sixth generation, and continues to be a mainstay in aviation Human Factors and aviation safety. If AI-based systems begin to play a role in either cockpit or ATC workplaces, or significantly support flight crew/ATCOs, potential impacts on CRM/TRM need to be understood and safeguards put in place. This will be particularly the case with Intelligent Assistants at EASA category 2A or 2B, where the flight crew and the IA are sharing tasks and even negotiating over how best to achieve goals.

#### 4.1.4. Human-Centred Design and Human Computer Interaction

The 1980s is generally considered to be the epoch wherein Human Factors and Ergonomics became more 'cognitive' in their approach. Norman's landmark work on design [48] reflected the ubiquitous rise of computer usage in aviation and other industrial domains and ushered in a lasting focus on human-computer interaction (HCI) and the benefits of human-centred design (HCD) [49]. It also contributed to the development of the broader and still flourishing field of Usability [50]. Human-AI interaction, whether via keyboard, speech, or other media, will be critical for the safe and effective introduction of AI-based systems into operational aviation contexts. A superordinate Human-Centred Design approach to the development and validation of Human-AI Teaming solutions would facilitate the principles of both human agency and human oversight [51]. Many HCI and Usability requirements will probably still apply to AI-based interfaces and interaction, with the proviso that new requirements may well be needed for a human-AI speech interface via

natural language processing (NLP) as well as operational explainability (OpXAI) to help the end user understand the AI's decision-making processes [52]. Operational explainability is poised to become a major new area for Human Factors research.

#### 4.1.5. Joint Cognitive Systems

Joint Cognitive Systems (JCS) [53] and Cognitive Systems Engineering (CSE) [54] introduced the concept of a cognitive system as an adaptive system that functions using knowledge about itself and the environment in the planning and modification of actions. CSE sought to side-step the man-vs machine paradigm and consider the cognitive output of both working together. Key to this is the idea of a mental model of the automation (or AI) and how it works and delivers its output. With its focus on the joint cognitive picture emerging from the agents in the system (human/machine), it paved the way for Situation Awareness (see later). Its focus on examining work in situ, rather than in the lab, and its focus on more naturalistic examination of decision-making scenarios (aka naturalistic decision-making [55]), also paved the way for Safety II and Resilience movements in Human Factors (see later). Given that JCS/CSE are fundamentally about cooperation, unsurprisingly they also drew together a range of disciplines, forming a hybrid community comprising Computing/Data Science, Systems Engineering and Systems Thinking, Human Factors, Neuro-Science, and Social Sciences (including Psychology). Such a hybrid community approach would probably benefit current Human-AI Teaming design and development.

#### 4.1.6. Ironies of Automation

At around the same time, Bainbridge [56] published her seminal 'Ironies of Automation' article, which highlighted some of the key dilemmas of human-automation pairing that still exist today and are relevant to human-AI teaming. As an example, as automation increases, human work can require exhausting monitoring tasks, so that rather than needing less training, operators need to be trained more to be ready for the rare but crucial interventions. Bainbridge and her colleagues also stated the case that automation is often given precedence, while humans are left to do the things that automation cannot do, including stepping in when the automation can no longer deal with the current conditions, or simply fails. The importance of the Ironies is that they act as checks and balances for system designers, with warnings about certain design philosophies and pathways that tend not to work, leading instead to endemic system performance problems and drawbacks. Endsley [57] has already begun the process of updating the Ironies for HAT systems.

#### 4.1.7. Levels of Automation and Adaptive Automation

Sheridan was one of the key proponents of 'levels of automation', which can be seen as an alternative or complement to Fitts' List. His 10 levels of automation [58,59] run from fully manual to fully automated:

1. The computer offers no assistance, human must take all decisions and actions
2. The computer offers a complete set of decision/action alternatives, or
3. Narrows the selection down to a few, or
4. Suggests one alternative, and
5. Executes that suggestion if the human approves, or
6. Allows the human a restricted veto time before automatic execution
7. Executes automatically, then necessarily informs the human, and
8. Informs the human only if asked, or
9. Informs the human only if it, the computer, decides to
10. The computer decides everything, acts autonomously, ignores the human

Sheridan's work also contributed to the notion of adaptive automation [60], wherein, for example, the automation could step in when (or ideally, before) the human became overloaded in a work situation. The earlier-mentioned case study on startle response [10] is effectively AI-supported adaptive automation, detecting startle and then directing the pilot's attention to key display components to stabilise the aircraft. Adaptive automation suggests the need for a number of requirements, including how to safeguard the role and expertise of the human, and how to switch from human to AI and back again as required. Adaptive automation also raises ethical issues in terms of data protection, e.g., where AI components such as neural networks use real-time human performance data (EEG, heart rate, galvanic skin response, etc.) as inputs to determine when to take over. Pilots may have concerns over the measurement and recording of such data, as it could reflect on their medical fitness for duty, a pilot licensing requirement.

#### 4.1.8. Situation Awareness, Mental Workload, and Sense-Making

Prior to Situation Awareness as a Human Factors construct, there were (and still remain) related concepts such as vigilance and attention [61], born from the early study of WWII radar display operators (and later, air traffic controllers) and the ability to detect signals such as incoming missiles or aircraft, given that the signal-to-noise ratio was low, and given both time-on-task and fatigue. Whereas attention and vigilance can be considered to be states related to alertness or arousal, situation awareness tends to be more contextualised, e.g., awareness of elements in the environment, such as aircraft in a sector of airspace.

Situation Awareness (SA) therefore focused less on the state of cognitive arousal, and more on the detail of what the human is aware of and what it means for the current and future operation. Indeed, the principal SA method, the Situation Awareness Global Assessment Technique (SAGAT, [62]) focused on three time frames: past, current, and future. As complexity rose in human-plus-automation environments such as cockpits and air traffic control centres, the question became how well the human understood what was going on, what was going to happen in the near future, and how to (re)act. The next question became how much the pilot or air traffic controller could assimilate from their controls and displays, in both short time frames and longer durations. This led to the study of mental workload (MWL), considering the task demands vs. human capabilities and capacities, including problems of overload and underload [61,63].

SA and Mental Workload signalled a deepening of the focus on cognitive activities, relating them in a measurable fashion to the actual cognitive work and operational context, delivering viable metrics that could be measured in simulations or in situ. Both SA and MWL have become mainstays in the design, validation, and operation of human-operated aviation systems. They are useful constructs because they help designers determine what the flight crew need to know (and see/hear), and when, and in what sequence and priority order, and when users may become overloaded. However, the perception of what is happening (and going to happen) to an object may be insufficient when the AI is processing its inputs; the human may need to understand what is behind the AI output.

More recently, this delving into mental processes has focused on sense-making, which is the way people make sense out of their experience in the world [64]. Sense-making deals with the human need to comprehend, often via the exploration of information. In relation to AI, sense-making helps to test the plausibility of an AI's explanations as well as anomalous outputs or event characteristics. Sense-making is a useful construct particularly in conditions where uncertainty is high and not all signals may be present, or where some signals may be erroneous. Unfortunately, for such situations, as can happen in flight upsets (also known as loss of control in flight), there is no straightforward associated metric

to determine how easily flight crew will be able to make sense of a particular scenario. Instead, realistic simulations are carried out with pilots undergoing abnormal events, and SA measurements and post-simulation debriefs, as well as safety and aircraft performance measures, are arguably the best way to determine the safety of the cockpit design or air traffic control system.

There is a very real danger that AI systems, which tend to be 'black boxes', can undermine the human crew's situation awareness, both in terms of what is going on, and of what the AI is doing or attempting to do. A critical question therefore becomes how to develop an interface and interaction means so that the AI and the human can remain 'on the same page'.

Additionally, there will need to be operational explainability (OpXAI), so that the human crews can determine (i.e., make sense of) why a course of action has been recommended (or taken) by the AI. Such explainability needs to be in an operational context in language that crews can follow (as opposed to data analytic explainability, which refers more to how to trace an AI's outputs to its internal architecture, data sources, and algorithmic processes).

Workload will become more nuanced with human-AI teaming partnerships, as there may be more periods of underload followed by intense workload periods, especially if the AI is suddenly unable to function.

Realistic and real-time simulations with human crews as occur now, both for flight crew and air traffic controllers, must continue. These will become human-plus-AI simulations. The human crews need to see how AI works in realistic contexts, so they can gain trust in it. They also need to see it when it fails or becomes unavailable, so that they can recover from such scenarios. Low SA, plus a sudden spike in mental workload due to loss of the AI support, and an inability of the AI to explain its recommendations, could well be a recipe for disaster.

#### 4.1.9. Rasmussen and Reason—Complex Systems, Swiss Cheese, and Accident Aetiology

Rasmussen's work on complex systems and safety, underpinned by his Skill, Rule and Knowledge-Based Behaviour hierarchy [65], had a big impact in the 1980s and 1990s in many high-risk industry domains. It may be worth revisiting this model as there are undoubtedly skill sets that may be lost (this already happens with automation), and rules (e.g., Standard Operating Procedures [SOPs] in cockpits) may become more fluid as AI-based support systems find ever-new ways of optimising operations in real time. Perhaps most interesting will be the area of knowledge-based behaviour (KBB), namely, having to consider what is going on in a situation based on a fundamental understanding of how the system works and responds to external/internal perturbations. KBB incorporates not only factual knowledge (also called declarative knowledge) but also experience amassed over years of operating a system (e.g., an aircraft) in a wide range of conditions.

KBB can be supported by Ecological Interface Design (EID) [66], resulting in high-level displays to monitor critical functions or safety parameters, as were developed in the nuclear power domain to avoid misdiagnoses of nuclear emergencies. This is an attempt to take a complex system's inputs and outputs and make sense of them. However, the problem with AI is that its complexity may be unfathomable for humans, at least in reasonable and normal operational timescales. This suggests a need for display approaches that make the complex system's workings and output relatable, backed up by explainability function that can at least approximate what the AI has done and why.

Since AI operational explainability can probably never be completely trustworthy (it is an approximation, and few will be capable of understanding what goes on 'under the hood' of an AI), an interface that affords the pilot or other aviation worker an effective

system safety overview, unfiltered by AI, would seem a sensible precaution. Moreover, since diagnosis of an aircraft emergency may become a shared human-AI process, the question is one of how to ensure that no human ‘tunnel vision’ or misdiagnosis by an AI leads to catastrophe. A deeper question becomes how to visualise AI performance so that it is evident to the human when the AI is operating outside its knowledge base or has limited statistical confidence in its prognosis.

Reason’s so-called Swiss Cheese model of accident aetiology [67], although dated, is still in regular use today. It proposes that accidents occur via vulnerabilities in a succession of barriers (e.g., organisation, preconditions, [un]safe acts, and defences). The vulnerabilities are like the holes in Swiss cheese, and the larger or more proliferated they are, the easier it is for them to ‘line up’ and for an accident to occur. What is interesting is that AI could in theory affect all these layers, either increasing or decreasing the size and quantity of the holes. It could also reduce the independence between each barrier, so that in reality, a system has fewer barriers before an accident occurs.

It would be useful to consider how AI could affect the Swiss Cheese layers differentially, e.g., use of LLMs at the organisational and preconditions layers, and AI-based tools at the unsafe acts and defences layers. Such a layered model also leads to the question of how different AIs will interface with one another. We already talk of Human-AI Teaming, but there will also be Human-AI-AI-Human and Human-AI-Human-AI variants before long, potentially allowing for problems to propagate unchecked across traditional ‘defence-in-depth’ boundaries.

#### 4.1.10. Human-Centred Automation

In the 1990s, after a series of ‘automation-assisted aviation accidents’ following the introduction of glass cockpits, Billings [15] developed the concept of Human-Centred Automation. This tradition has generally persisted in aviation ever since. The nine core principles of HCA are as follows:

- i. The human must be in command
- ii. To command effectively, the human must be involved
- iii. To be involved, the human must be informed
- iv. The human must be able to monitor the automated system
- v. Automated systems must be predictable
- vi. Automated systems must be able to monitor the human
- vii. Each element of the system must have knowledge of the others’ intent
- viii. Functions should be automated only if there is good reason to do so
- ix. Automation should be designed to be simple to train, learn, and operate

These are the ‘headline’ principles, but there are many others in this watershed research carried out by Billings and others, e.g., ‘automation should not be allowed to fail silently’. Such principles (i.e., all of them, not only the top nine) could be revisited for AI, as already several of them are in danger of not being upheld, e.g., some AI-based systems may not be predictable; reciprocal knowledge of the others’ intent may prove difficult to achieve in practice.

#### 4.1.11. HFACS (and NASA–HFACS and SHIELD)

The Human Factors Analysis and Classification System (HFACS) approach [68] was developed for the US Navy, and has proven popular, often leading to ‘variants’ in other domains including space (NASA–HFACS [69]) and maritime (the SHIELD taxonomy, [70]). HFACS and equivalent taxonomies of human error and its causes/contributory factors have been in use for decades to determine how and why an error occurred, and how to prevent its recurrence. HFACS-like systems do not address only the surface factors



(what happened), but deeper causes, including Human Factors elements, supervisory practices, and organisational and cultural factors. In this respect, HFACS embodies a Swiss Cheese approach.

As AI-based systems are introduced to support flight crews and air traffic controllers in their daily operations, taxonomies such as HFACS will likely need updating, adding new terms linked to human-AI interactions. Although some blanket terms already exist, such as complacency and over-trust with respect to automation, these are probably not nuanced enough to capture the full extent of the transactional relationships that will exist between human crews and AI support systems, especially as those systems become more advanced and even executive (i.e., not requiring human oversight).

Additionally, as the AI is seen more as an 'agent', consideration must be given to the ways in which it, too, can fail. Already, as noted above, some general failure modes have been considered, such as data biasing, hallucinations, edge and corner cases, etc., but there are likely many more, some of which may be subtle and hard or even impossible to detect. The scenario in which 'data forensics' is required to understand why an AI suggested something that contributed to an accident, is probably not far in the future.

#### 4.1.12. Safety Culture

Safety culture, namely, the priority given to safety by the organisation, from the CEO down to the front-line and support workers, originated in the nuclear power industry following the Chernobyl disaster [71]. It has been applied to aviation most notably since the Uberlingen midair collision in 2002 and mentioned frequently in other aviation accidents. Today, more than 30 European air traffic management organisations have undergone safety culture surveys [72]. Commercial aviation (air traffic control and commercial flight crews) is generally seen as having a positive safety culture.

As noted in a companion paper to this one [73], the introduction of AI into operational aviation systems could aid or degrade safety culture in aviation. In particular, the concern is that human personnel may delegate some of their safety responsibility to the AI, especially if the AI is taking more of an executive role. There is therefore a critical need for a new safety culture approach to monitor what is happening to safety culture as AI is introduced into the cockpit and ATC Ops room or Tower.

#### 4.1.13. Teamwork and the Big Five

According to the 'Big Five' theory of Teamwork [74], the core components of teamwork include team leadership, mutual performance monitoring, back-up behaviour, adaptability, and team orientation, which collectively lead to team effectiveness. These components are predicated on shared mental models, mutual trust, and closed-loop communication. Similar to Crew Resource Management, Teamwork could be significantly disrupted by the introduction of IAs. However, AI could also act as a diverse and potentially more comprehensive 'mental model' backup system for the pilots, one that is instantly available and adaptable when a sudden unexpected situation occurs. This contrasts with today, where it can take humans a short time to re-adapt, time that can be crucial in an emergency scenario in-flight (one of the principal reasons there are always two pilots in the cockpits of commercial airliners). Probably key for Human-AI Teamwork will be trust and closed-loop communication. The latter will likely entail short, succinct, and contextual explainability provided by the AI, whether in procedural or natural language, and/or visually via displays. Team orientation will be an issue for AI-based systems that aim to carry out and execute certain tasks. The very existence of such agents suggests the need for specialised training for leadership of a human-AI team.

#### 4.1.14. Bias and Complacency

Parasuraman [75] considered the potential end states of automation as implemented in aviation systems. He defined four (non-exclusive) categories: use, misuse (over-reliance), disuse (disengagement), and abuse (poor allocation decisions between human and machine). He found the primary factors influencing these end states to be trust, mental workload, risk, automation reliability and consistency, and knowing the state of the automation. One of the more worrying biases was the withdrawal of attention from cross-checking the automation and considering contradictory evidence, summarised as 'looking but not seeing'. The effects such as complacency (not checking) and automation bias (over-trust) are not easy to fix, e.g., via training, and are apparently prevalent in both experienced or naïve (i.e., new) users [76].

As with several other major study areas of Human Factors, the area of automation bias (especially complacency) needs to be revisited, for several reasons. The first is that cross-checking AI is likely to be more complex, as the way the AI works will itself be more complex and sometimes not open to scrutiny (either non-explainable or unfathomable for humans). The availability and salience of contradictory evidence is highly pertinent to the human capacity of acting as a back-up to the AI, given the well-known biases of humans, including representativeness, availability, anchoring and confirmation biases [77,78] (AIs, especially LLMs, are also not immune to biases). The human may want to know why a particular course of action was suggested and others were ignored. Ways to show such 'alternates' therefore need to be considered, possibly including the trade-offs the AI has made, or data it has ignored as outliers or irrelevant. Similarly, knowing the state of the AI will also be important; not simply whether it is 'on' or 'off', but its confidence level given the situation at hand compared to the data it was trained on. There is also the question of prior experience: existing pilots can compare their experience to what the AI is suggesting, whereas new pilots (in the future) who have never known a system without AI support, may not have such 'unfiltered' prior experience.

#### 4.1.15. SHELL, STAMP/STPA, HAZOP, and FRAM

There are a number of means of analysing the risks associated with human-machine systems. A general thematic framework systems approach model is provided by SHELL [79], which considers the software (procedures and rules), hardware, environment, and 'liveware' (humans and teams), and how they can interact to either yield safe or unsafe outcomes. Approaches such as the Systems Theoretic Accident Model and Processes (STAMP) framework and its derivative Systems Theoretic Process Analysis (STPA), both developed by Leveson [80], deliver a systematic and formalised analytic approach for identifying human-related hazards and potential mitigations. The Hazard and Operability Study (HAZOP) approach developed in the chemical industry in the 1970s by Kletz [81], though arguably less structured, is still an insightful and comparatively agile approach to identifying human-related hazards in complex systems, including those with AI [14,82]. The Functional Resonance Analysis Method [83] has as core premises that most of the time, most things go right, and that there is often a large gap between 'work as imagined/designed' and 'work as done' in the real operational system (this perspective on risk known colloquially as Safety II [84]).

As a safety-critical industry, aviation must carry out hazard assessment to identify potential hazards and develop appropriate mitigations. At the moment, all of the methods above (and others) can be used to identify hazards that can occur in human-AI systems. The problem is that we are missing two inputs. The first is a model of how the AI can fail, or exhibit aberrant behaviour, or suggest inaccurate/biased resolutions or advice. We already know some of the answers, in terms of hallucinations, edge cases, corner cases,

biased data usage, etc. [85]. But these are generic. What we need is a way to determine when these and other AI ‘failure modes’ are likely, given the type of ML/LLM being used, its data, and the operational context it is being applied to. It is likely that a taxonomy of AI failure mechanisms will develop as more experience is gained with AI-based systems, but it would be preferable not to have to learn the hard way.

The second unknown, or ‘barely known’, relates to ‘human+AI’ failure modes, i.e., the likely failure types when people are using and interacting with various types of AI tools. We already have ‘complacency’, but this is a catch-all term, and knowing when it is likely or not, as a function of the human-AI system design, is unpredictable. This is problematic: how can a safety-critical system be certified if complacency is a likely user characteristic and the AI can fail? There is a need for greater understanding of the evolution of human-AI inter-relationships. This may entail longer-term study of human-AI working partnerships, perhaps in extended simulations (lasting months rather than days or weeks), effectively constituting a safe ‘sandbox’ in which to see emergent behaviour of both human and AI, moving beyond ‘work as imagined’ and ‘work as done’, to ‘work as AI-assisted’.

#### 4.1.16. Just Culture and AI

Since the implementation of ASRS, aviation has generally been seen as having an effective reporting culture, enabling it to be an informed culture, learning from events to continually improve. The reporting culture is predicated on a Just Culture [86], in which the aviation system prefers to learn from mistakes than blame people for them, as long as there is no reckless behaviour or intention to cause damage or harm. In European aviation, this principle is enshrined in law (Regulation (EU) No 376/2014 of the European Parliament and of the Council of 3 April 2014). But there is a potential double-bind in the future for human agents in aviation [87]. If they are advised by an AI to do something and it results in an accident, they may be asked by a court of law to justify why they did not recognise the advice was faulty. Conversely, if they choose not to follow AI advice and there is an accident, they will be asked why they did not follow the AI’s advice. Responsibility and justice are human constructs, and an AI, even if its future role is that of an executive agent in an operational aviation system, cannot be prosecuted in a court of law, and prosecuting its developers will likely be fraught with legal complexities. For example, what judge or jury will have sufficient AI literacy to understand what really happened in such an accident, and how will they overcome hindsight and other biases (e.g., ‘the pilot should have known better’) in forming their deliberations?

Just Culture could be a major deal-breaker for unions and professional associations who feel their members (pilots, controllers, others) are at risk of being prosecuted in an area with little legal precedent, and given juridical framework variations according to country, and potentially serious criminal charges. Hypothetical test cases should be carried out in legal sandboxes to anticipate legal argumentation and potential outcomes in this area, and more generally to expand the Just Culture ‘playbook’ [88]

#### 4.1.17. HF Requirements Systems—EASA CS25.1302, SESAR HP, SAFEMODE, FAA

Over the decades there have been various Human Factors requirements and assurance approaches. Of particular note in Europe is EASA Certification Standard (CS) 25.1302 ([https://www.easa.europa.eu/sites/default/files/dfu/CS-25\\_Amdt%203\\_19.09.07\\_Consolidated%20version.pdf](https://www.easa.europa.eu/sites/default/files/dfu/CS-25_Amdt%203_19.09.07_Consolidated%20version.pdf) (accessed on 20 March 2025)), which is concerned with controls and displays in cockpits, and contains detailed guidance on all aspects of information usage including display design, situation awareness, workload, alarms, etc. Essentially, all controls and displays need to be fit for human purpose whether in normal, degraded mode, or emergency situations. Whilst there is no European-wide regulatory equivalent for Air

Traffic Management (ATM) systems in Europe, there is comparable guidance available via the SESAR (Single European Sky ATM Research) programme and its Human Performance Assessment Process (HPAP) (<https://www.sesarju.eu/sites/default/files/documents/transversal/SESAR%202020%20-%20Human%20Performance%20Assessment%20Guidance.pdf> (accessed on 20 March 2025)). This approach breaks down four high level areas—roles and responsibilities, the human-machine interface, teams and communications, and transition to operations—into detailed and measurable requirements in an argument-based structure. The SESAR HPAP allows a Human Factors ‘case’ to be built for a new system design or change, showing where the design complies with Human Factors principles, and where it does not (or does not need to). More recently, an EU research project called SAFEMODE (<https://www.safemodeproject.eu/> (accessed on 20 March 2025)) has developed a Human Factors assurance platform called HF Compass, which refers to the HPAP but also highlights more than 20 tools and techniques that can be used to provide evidence that Human Factors has been assured for a new system. Very recently, EASA (<https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2> (accessed on 20 March 2025)) has provided preliminary guidance on the Human Factors assurances required for AI-based systems in aviation, in the form of regulatory requirements. The Federal Aviation Agency (FAA) has also recently released its own Roadmap ([https://www.faa.gov/aircraft/air\\_cert/step/roadmap\\_for\\_AI\\_safety\\_assurance](https://www.faa.gov/aircraft/air_cert/step/roadmap_for_AI_safety_assurance) (accessed on 20 March 2025)) for safety assurance of AI-based systems in aviation, though it is currently focused more on safety than Human Factors.

The existing guidance from CS25 1302, the SESAR HPAP, and the EASA guidance on Human Factors aspects of human-AI Teaming systems, all offer excellent foundations for an approach focused on Human-AI systems. Section 5 accordingly presents a Human Factors Assurance system for Human-AI Teaming systems.

#### 4.2. Contemporary Human Factors and AI Perspectives

Having reviewed the historical landmarks in Human Factors and their implications for Human-AI systems, this subsection briefly reviews a sample of the more recent emerging Human Factors literature on Human-AI Teaming, focusing on a model of HAT and the issues of personification (anthropomorphism) of AI and emotion-mimicking AI.

##### 4.2.1. HACO—A Human-AI Teaming Taxonomy

In a recent paper on Human AI Collaboration (HACO), a Human-AI Teaming (HAT) taxonomy has been usefully mapped out [89] and its key insights relevant to this paper are outlined below.

The HACO concept has six core tenets: context awareness, goal awareness, effective communication, pro-activeness, predictability, and observability. Taken as a whole, one way to summarise these tenets is that they all ensure that the team members, including the AI, are ‘on the same page’ with what is happening and what to do about it. In practice, this should mean that the workflow remains smooth, without surprises, significant breaks or disruptions, or conflicts. As with human teams, it does not require perfect understanding of one another, but an appreciation of individual behavioural norms (including style of interaction), pace of work, skill sets and capabilities, and limitations.

Context awareness is possibly a more useful (and less anthropomorphic) label than situation awareness when applied to intelligent agents. It seeks to establish what the AI was responding to within its inputs and data sets. However, this is not always simply the superficial data in the environment (e.g., weather patterns that might affect flight route), but the way the AI will use statistical and other algorithms to interpret such data according to the task and goals at hand. The design challenge is to maintain the usability adage of

‘what you see is what you get’, whether this is achieved via natural language dialogue or (more likely) visual media that can decrypt the AI’s computational processes in meaningful and quickly ‘graspable’ ways for the end user.

Goal awareness (a precursor to goal alignment) is a higher-level attribute, related to context awareness, and ensures that human and AI goals are aligned. This becomes important in work arrangement scenarios where the AI is able to modify the goal hierarchy, and where goals may dynamically shift during a scenario (e.g., for an emergency, as conditions worsen or become more stable). It can also be important where there are a mixture of safety and other goals, some of which may conflict with safety. Similarly, as for contextual awareness, a design challenge is how to ensure the human is aware of the goals the AI is working towards as they change and evolve.

Effective Communication can occur via various modalities, from natural language and even gestures, to digital displays and procedural textual responses to human prompts. Communication for AI tools or systems below EASA’s category 2B may not need to be that advanced, though for 1B and 2A there may require explainability if the AI’s task or output is sufficiently complex. For human AI collaboration at the 2B level, there will likely not only need to be communication, but a degree of rapport, so that the AI is communicating in terms and contexts familiar to the human. This implies a shared understanding of the local operational environment conditions and practices. For example, this can refer not only to a specific aircraft type such as an A320 or B737, but how those aircraft types are fine-tuned by the airlines using them, along with their Standard Operating Procedures and day-to-day working practices.

Proactiveness links again to EASA’s 2B and above categorisations, whereby the AI can initiate its own tasks and shift or re-prioritise goals, giving the AI a degree of autonomy and *agency* (since it can act under its own initiative). However, in contrast to the distinctive categorisations of EASA the HACO authors [89] suggest the concept of ‘sliding autonomy’, wherein the human (or the system) determines the level of autonomy. This is of interest in aviation in cases where flight crew or air traffic controllers, for example, may be overwhelmed by a temporary surge in tasks or traffic, respectively, and may wish to ‘hand-off’ certain tasks (especially low-level ones) to an AI. This is effectively an update of the Adaptive Automation concept.

The three HAIKU use cases are close to this notion of sliding or rather ‘flexible’ autonomy. In UC1 (startle response) the pilot can initiate and stop the support at any time, depending on their situation awareness and the degree to which they feel in control. In UC2 (re-routing support) the pilots have many options in how to use the automation, whether to find their own solution, accept the first one offered by the IA, or delve deeper into the rationale behind the three airports offered. In UC4, the controllers stated that they would likely use ISA in high workload situations, depending on their workload capacity and the complexity of the traffic situation. This partial use was also to avoid ‘skill-fade’.

Predictability can refer both to the degree to which the pace of work of an AI is understood, trackable, and manageable, to the degree to which an AI might ‘surprise’ the humans in the team via its outputs. The former probably requires some kind of overview display to know what the AI is doing and the progress on its tasks or goals. The latter will depend on the amount of human-AI training afforded prior to teamworking in real operational settings.

Observability refers to the transparency of the progress of the AI agent when resolving a problem, and can be linked to explainability, though in practice the AI’s workings might be routinely monitored via a dashboard or other visualisation rather than a stream of textual explanations, with the user able to pose questions as needed.

Three additional HAT attributes are mentioned in [89] and are worth reiterating here:

- Directing attention to critical features, suggestions and warnings during an emergency or complex work situation. This could be of particular benefit in flight upset conditions in aircraft suffering major disturbances, as in UC1.
- Calibrated Trust wherein the humans learn when to trust and when to ignore the suggestions or override the decisions of the AI, an over-riding concern in UC2.
- Adaptability to the tempo and state of the team functioning, as in the controllers' approach to using ISA in UC4.

With respect to Adaptability, one of the hallmarks of an effective team, it may be useful to resurrect High Reliability Organisation (HRO) theory [90]. All five of the pillars of HROs—preoccupation with failure, commitment to resilience, reluctance to simplify explanations, deference to expertise, and sensitivity to operations—could well apply to Human-AI Teams. HRO theory leans towards collective mindfulness of the team, and here is where it is necessary to consider and contrast how humans think 'in the moment', compared to how an AI might build up its own situation representation or context assessment.

#### 4.2.2. AI Anthropomorphism and Emotional AI

Human-AI Teaming can be considered to be an anthropomorphic term [37], suggesting that the AI is a team player, denoting human qualities to a machine. Anthropomorphism relates to the identity we assign an AI system, and hence the degree of agency we accord it. The more we 'personify' an IA, the more the danger of delegating responsibility to it, or of surrendering authority to its logic and databases, or of basically second-guessing ourselves rather than the AI. In this vein, it is worth recalling that AI systems are created by people (data scientists), and there are many human choices, not to mention selecting algorithms by trial and error, that go into development of AI systems, e.g., choice of training, validation and test datasets, choice of hyperparameters, selection of most appropriate algorithm, etc. [1].

The recent FAA Roadmap on AI [38] is explicitly against the personification of AI, and given that AI has no sentience and will likely not do for some time, it is right to avoid AI personification, especially in safety-critical operations. A key practical consideration, however, is whether treating future AI systems as a team member could enhance overall team performance. This is as yet unknown, but it leads to a second question of whether we can tell the difference between a human and an AI. In a recent study of 'emotional attachment' to AI as team members [91], most participants could tell the difference between an AI and a human based on their interactions with both. Another study [92] examined human trust in AIs as a function of the perception of the AI's identity. The study found that AI 'teammate' performance matters, whereas AI identity does not. The study authors cautioned against using deceit to pretend an AI is a human. Deception about AI teammate's identity (pretending it is a human) did not improve overall performance and led to less acceptance of AI solutions. Knowing it is an AI actually improved overall performance.

In a non-safety-critical domain study [93], two key acceptance parameters for emotional AI were found to be the potential to erode existing social cohesion in the team, and authority (impacts on team members' status quo). As with other studies in this area, the authors found people judged machines by outcomes. A further study [94] found that monitoring people's behavior and emotional activity (speech, gestures, facial expressions, and physiological reactions), even if supposedly for health and well-being, can be seen as intrusive. Such monitoring activities can be for good reasons, such as detecting and combatting stress (e.g., UC1), fatigue and boredom monitoring, and for error avoidance, and of course, productivity. Nevertheless, people may be uncomfortable with this level of personal intrusion of their behavior, bodies, and personal data.

Overall, therefore, preliminary indications are that aviation needs neither anthropomorphic (personified) nor emotional AI. What matters is the effectiveness of the AI in the execution of its tasks.

## 5. Human Factors Requirements for Human-AI Systems

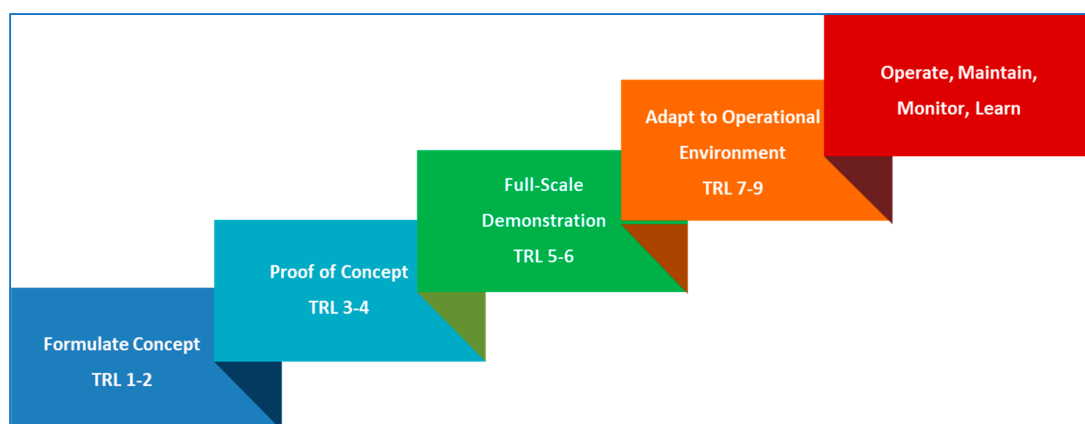
Having reviewed types of AI, levels of AI autonomy, and examples of prototype Intelligent Agents, as well as Human Factors themes relevant to AI-based systems in aviation, this section (Step 3) outlines the development of a preliminary set of Human Factors requirements for future aviation HAT systems. A Human Factors Requirements System for HAT systems should satisfy several requirements of its own:

1. It must capture the key Human Factors areas of concern with Human-AI systems.
2. It must specify these requirements in ways that are answerable and justifiable via evidence.
3. It must accommodate the various forms of AI that humans may need to interact with in safety-critical systems (note—this currently excludes LLMs) both now and in the medium future, including ML and Intelligent Agents or Assistants (EASA's Categories 1A through to 3A).
4. It must be capable of working at different stages of design maturity of the Human-AI system, from early concept through to deployment into operations.

Accordingly, in this section, the issue of how Human Factors can inform the different stages in the design, development, and deployment life cycle of new HAT systems is considered first. Second, a framework or overlying architecture for HAT is presented, to group the various requirements thematically. This will aid in their implementation in programming and resourcing Human Factors integration into future HAT systems. Third, a sample of the new requirements themselves, contextualised in HAT and autonomy-sharing terms in such a way that they can be applied meaningfully by HAT system developers, is presented. These are then applied and evaluated in Section 6 (Step 4).

### 5.1. HAT Requirements and a Design Life-Cycle Framework

In aviation, both in Europe and North America, the NASA system of Technology Readiness Levels (<https://www.nasa.gov/directorates/somd/space-communications-navigation-program/technology-readiness-levels/> (accessed on 20 March 2025)) (TRLs) is the most commonly used framework for aviation system design life cycle maturity and is illustrated in simplified form in Figure 7.



**Figure 7.** Technology Readiness Levels.

Following on from the principles of Human-Centred Design, Human Factors needs to be integrated from the early design stages. Otherwise, the design of the AI will be a

poor 'fit' to the human end user, and the only significant degree of freedom left to optimise HAT performance will be training. Training end users to operate a poorly designed system is a poor design strategy. Therefore, Human Factors requirements should ideally be applied from the early concept stages onwards, until the system is operational. The critical design stages, however, in terms of integration of Human Factors to deliver optimal system performance, are TRLs 3-6, as these stages determine how the human-machine (AI) interactions will take place.

Typically, TRLs 1-2 are concerned with early concept exploration. For Human-AI System design, at this TRL usually the most important considerations focus on Roles and Responsibilities, namely who (or what) will be in charge when the AI is operating. Decisions about sense-making, such as how shared situation awareness will be established, and primary human-AI interaction modes, e.g., 'conventional' (keyboard, mouse, touchscreen, etc.) or more advanced (speech, gesture recognition) can be made, as well as whether the AI will be used by a single person or a team. It is at this stage that key design choices are made concerning the adoption of a Human-Centred Design approach, including the use of end users in informing the early 'foundational' concept.

TRLs 3-4 add a lot more detail, fleshing out the concept's architecture, and gaining a picture of what the AI will look and feel like to interact with, via early prototypes and walk-throughs of human-AI interaction scenarios. The area of sense-making is crucial here, and communication and teamworking aspects will become clearer.

At TRLs 5-6 models and prototypes are developed iteratively until a full-scope demonstration is completed and tested in a realistic simulation with operational end users. This period of design and development sees many issues 'nailed down' and solidified into the design and operational concept (CONOPS), having been validated via robust testing with human end users. Risk studies focus on errors and failures seen, as well as those that could conceivably occur once the system is in operation.

TRLs 7-9 prepare the concept for deployment in real-world settings. This is the time to consider in earnest the ramifications of entering a human-AI system into an operational organisation, with requirements for competencies and training of end users, as well as socio-technical considerations such as staffing, user acceptance, ethics, and well-being. These latter issues can determine whether the system is accepted and used to its full extent.

After TRL 9 the system is in operational use. Since most AI systems are learning systems, monitoring—particularly in the first 6 to 12 months—will be a critical determinant of sustainable system performance.

## 5.2. The Human Factors HAT Requirements Set Architecture

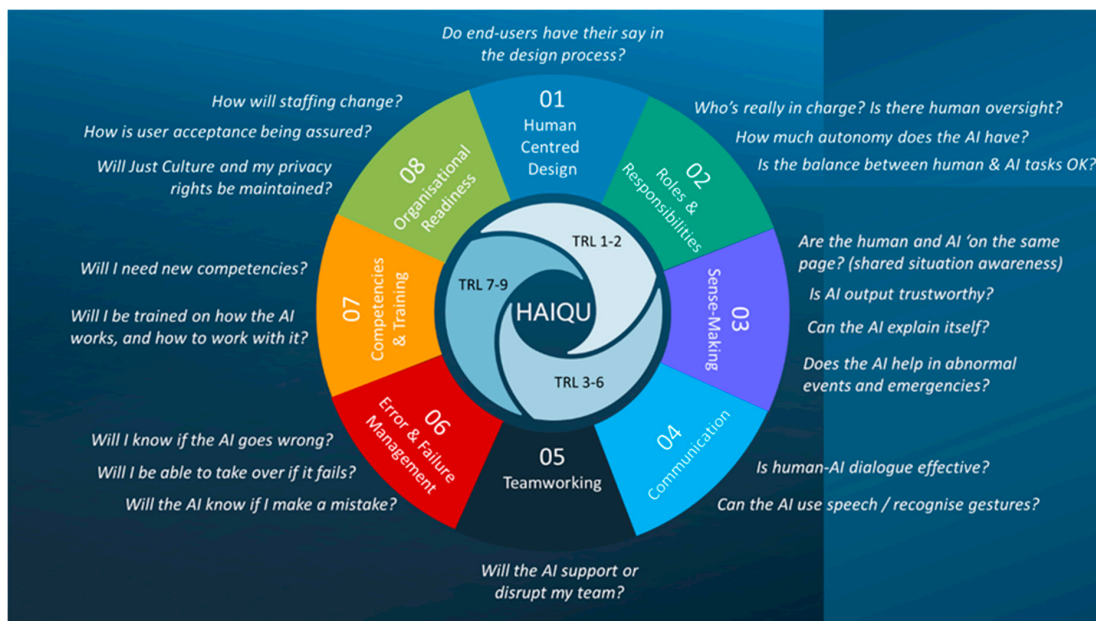
Based on the literature review in Section 4, including the SESAR HPAP and the recent EASA Guidance on Human Factors for AI-based systems, as well as the EU Act on AI, eight overall areas have been identified, as shown in Figure 8 and outlined below.

1. Human-Centred Design—this is an over-arching Human Factors area, aimed at ensuring the HAT is developed with the human end-user in mind, seeking their involvement in every design stage.
2. Roles and responsibilities—this area is crucial if the intent is to have a powerful productive human-AI partnership, and helps ensure the human retains both agency and 'final authority' of the HAT system's output. It is also a reminder that only humans can have responsibility—an AI, no matter how sophisticated, is computer code. It also aims to ensure the end user still has a viable and rewarding role.
3. Sense-Making—this is where shared situation awareness, operational explainability, and human-AI interaction sit, and as such has the largest number of requirements. Arguably, this area could be entitled (shared) situation awareness, but sense-making



includes not only what is happening and going to happen, but why it is happening, and why the AI makes certain assessments and predictions.

4. Communication—this area will no doubt evolve as HATs incorporate natural language processing (NLP), whether using pilot/ATCO ‘procedural’ phraseology or natural language.
5. Teamworking—this is possibly the area in most urgent need of research for HAT, in terms of how such teamworking should function in the future. For now, the requirements are largely based on existing knowledge and practices.
6. Errors and Failure Management—the requirements here focus on identification of AI ‘aberrant behaviour’ and the subsequent ability of human end users to detect, correct, or ‘step in’ to recover the system safely.
7. Competencies and Training—these requirements are typically applied once the design is fully formalised, tested, and stable (TRL 7 onwards). The requirements for preparing end users to work with and manage AI-based systems will not be ‘business as usual’; new training approaches and practices will almost certainly be required (e.g., pilots and controllers who participated in UC1, 2 and 4 simulations stated they would want specialised training).
8. Organisational Readiness—the final phase of integration into an operational system is critical if the system is to be accepted and used by its intended user population. In design integration, it is easy to fall at this last fence. Impacts on staffing levels and levels of pay, concerns of staff and unions, as well as ethical and well-being issues are key considerations at this stage to ensure a smooth HAT-system integration. This is therefore where socio-technical considerations come to the fore.



**Figure 8.** Human Factors Requirements System Architecture.

This eight-area architecture goes beyond the SESAR HPAP four-area structure (Roles and Responsibilities, Interface Design, Teamworking and Transition) and has added new sub-areas including trustworthiness, AI autonomy, operational explainability, speech recognition and human-AI dialogue, Just Culture, etc. Many of the requirements from EASA’s CS 25 1302 are evident at the specific requirements level but have been re-oriented or augmented to focus on AI and HAT. The original requirements set was written before the most recent EASA HAT guidance release, and since that time alignment efforts have been made. The resultant HF Requirements set is significantly larger than the EASA one, as it

deals with a number of areas and sub-areas outside EASA's current focus (e.g., related to roles and responsibilities organisational readiness, competencies, human agency, etc.). The detailed structure is as follows, showing all subcategories and the number of requirements in each of the total of 17 categories and subcategories (165 requirements in total, of which 51 are common requirements with EASA's guidance):

1. Human-Centred Design (5)
2. Roles and Responsibilities
  - a. Human and AI Autonomy (7)
  - b. Balance of Human and AI Tasks (13)
  - c. Human Oversight (10)
3. Sense-Making
  - (a) Shared Situation Awareness (12)
  - (b) Trustworthy Information (12)
  - (c) Explainability (13)
  - (d) Abnormal Events, Degraded Modes and Emergencies (7)
4. Communication
  - a. Human-AI Dialogue (10)
  - b. Speech and Gestures (11)
5. Teamworking (11)
6. Errors and Failure Management (13)
7. Competencies and Training
  - a. New Competencies (9)
  - b. New Training Needs (9)
8. Organisational Readiness
  - a. Staffing (8)
  - b. User Acceptance (9)
  - c. Ethics and Wellbeing (8)

The relevance of the Human Factors areas to the different TRL are also highlighted in Figure 8, to help the user adapt their evaluation approach. As illustrated in the figure, the application of Human Factors areas to TRLs are not always clear cut, so some degree of judgement must be used. For many projects, in early TRLs entire areas may be deemed 'TBD' (to be decided later) or even N/A (not applicable—for now) and returned to in later design stages. For example, many research projects will not consider Human Factors areas 7 and 8, although it is worth considering impacts on staffing early, since this can be a 'deal-breaker' for any project.

### 5.3. Detailed HAT Requirements

Table 1 shows a selection of the HAT Human Factors requirements developed and applied in the HAIKU project. The selection highlights newer requirements, those of particular relevance to HAT and Intelligent Agents. The full set of requirements also includes 'conventional' requirements (e.g., avoid clutter in the human-machine interface displays) that already exist in contemporary HF Requirements systems (EASA, SESAR, FAA, etc.), since these must still be applied to HAT systems to ensure usability etc. The requirements in Table 1 focus on the first six areas of the HAT architecture in Figure 8, as this is where there has been most application with the use cases, which are mainly TRL 3-6, such that training and competencies (area 7) and organisational readiness (area 8) considerations have yet to occur.

**Table 1.** Human-AI Teaming Human Factors Requirements (extract).

New HAT Requirement	HUMAN FACTORS AREA	Origin HF Area
1. Are end user opinions helping to inform and validate the design concept, as part of an integrated project team including product owner, data scientists, safety, security, Human Factors and operational expertise?	HUMAN-CENTRED DESIGN	Joint Cognitive Systems/ Cognitive Systems Engineering/ Human-Centred Automation (HCA)
2. Are end users involved in any hazard identification exercises (e.g., HAZOP, STPA, FRAM etc.)?		SHELL, STAMP, HAZOP &FRAM
3. What is the overall level of autonomy—is the human still in charge?		
4. If the level of autonomy changes dynamically, has it been determined when/why it changes?		Levels of automation/adaptive automation
5. If there is task-switching, is it controlled by the human, by the AI, or a mixture of both?		HCA
6. In the case of AI control, can the human reject a task?		HCD
7. If the AI cedes control to the human unexpectedly, is there enough information for the human to safely take control?	ROLES and RESPONSIBILITIES ➤ Human & AI Autonomy ➤ Balance of Human/AI Tasks ➤ Human Oversight	Situation Awareness HCA Ironies of Automation/AI
8. Does the human have a mental model of how the AI performs each task?		Joint Cognitive Systems/ Cognitive Systems Engineering/ Situation Awareness HACO—goal awareness and predictability
9. Can the human monitor and adjust the AI’s goal formulation/prioritisation?		HCA
10. If human-AI negotiation is possible, does the human make the final call?		Fitts List
11. Can the AI detect poor decision-making by the user and offer alternatives?		HACO goal awareness
12. Can the human retain a strategic overview of the tasks and system performance/safety?		HCA, Complex Systems, SA, Sense-Making, Fitts List HACO goal alignment
13. Does the AI build its own situation representation?	SENSE-MAKINGShared Situation Awareness	Sense-Making
14. Is the AI’s situation representation made accessible to the end user, via visualisation and/or dialogue?		JCS, HCD, HCA
15. Does the AI-human interface reinforce the end user’s situation awareness, so that human and AI can remain ‘on the same page’?		Situation Awareness (SA) Ecological Interface Design HACO

**Table 1.** *Cont.*

New HAT Requirement	HUMAN FACTORS AREA	Origin HF Area
16. For complex situations, does the AI offer a diagnosis and rationale, along with the problem’s root cause and solution, and a prediction of likely operational consequences?		Complex Systems (KBB) SA (current/predictive) Ironies of Automation/AI
17. Is it made clear to the end user when the alerting situation raised by the AI is resolved, or if actions taken are not resolving the threat?		HCD, HCA SA & SM
18. Is the information or decision provided in a timely fashion for the human to consider and ‘weigh’ it before acting or accepting it?		JCS/CSE, HCA HCD
19. Is the information/decision offered accompanied by uncertainty estimates upon request?		Sense-Making, HCA Bias and Complacency
20. Can the human modify the AI’s parameters to explore alternative courses of action?	Trustworthy Information	JCS/CSE, UCD, HCA, HACO Sense-Making Complex Systems/KBB
21. If the AI is making trade-offs, are these made visible to the human?		Ironies of Automation/ AI, HCA
22. Can the end user alter key parameters the AI is optimising?		Sense-Making, JCS/CSE Situation Awareness, HCA
23. Are end users aware of the data sources the AI uses to build its situation representation		SA and SM Bias & Complacency
24. Can the human query the information/decision via an explainability function?		Sense-Making
25. Does explainability detail how the advice was derived, in end user (operational) terms?		HRO theory: sensitivity to operations, JCS, Sense-Making, SA
26. Is explainability multi-levelled, based on different levels of abstraction, including context of the AI’s goals and ‘reasoning’ available to the end user, any historical perspective underlying the AI’s reasoning and key data sources it has accessed, so the user can fully judge its appropriateness to the situation, and progressively determine how far (or when) to trust the AI?	Explainability	Complex Systems/KBB HRO–reluctance to simplify explanations Big 5 (shared mental models) Sense-Making, SA
27. Can the AI explain both its current goal and longer-term strategy if it has one?		Sense-Making Big 5 shared mental model
28. Can the human view both data that were used and data that were ignored by the AI, e.g., anomalies or outliers?		HCD/HCA Ironies of Automation/AI

**Table 1.** *Cont.*

New HAT Requirement	HUMAN FACTORS AREA	Origin HF Area
29. Does the level of human workload enable the human to remain proactive rather than reactive, except for short periods?	Abnormal Events and Emergencies	Mental Workload Ironies of Automation/ AI
30. Can the human retain situation awareness in emergencies/abnormal events, including what the AI is doing (and why)?		Complex Systems/KBB Ironies of Automation/ AI
31. Is it clear to both human and AI which tasks are safety-critical, and when safety is being threatened?		Safety Culture
32. If natural language is used, how is context-sensitivity ensured, so that misunderstandings are avoided?	COMMUNICATIONS	Big 5 (closed-loop comms; shared mental models) HCA, HCD, UCD HACO Emotional Mimicry Non-Personification of AI
33. Does the AI communication mode, whether oral or audio-visual, avoid the use of emotional mimicry (i.e., mimicking human emotions)?		
34. Are humans made aware of what human performance aspects the AI may be monitoring or recording, e.g., speech (to detect stress or fatigue), gestures, psychophysiological parameters (EEG, heart rate measures, skin conductance, eye movements, etc.)?		
35. Are the boundaries of the AI’s language capabilities and limitations made known to the user?		
36. Is the dynamically evolving AI-derived situation representation communicated to the entire team to ensure coherent team situation awareness?		
37. Does the design optimise Human AI team resource management and minimise team functioning errors?	Teamwork	CRM/TRM/BRM The Big 5 SA and Sense-Making Complex Systems/KBB
38. Are there sufficient skilled human crew to operate or recover or stabilise the system (i.e., the aircraft, air traffic situation, etc.) in case of AI failure or erroneous behaviour?	ERROR and FAILURE MANAGEMENT	CRM/TRM/BRM The Big 5 SA and Sense-Making Complex Systems/KBB
39. Is the AI robust against edge and corner cases, data bias, and data poisoning?		HAZOP/STPA/FRAM
40. Is the human trained on AI error modes and how to verify AI results?		Bias and Complacency SA and SM
41. Has the human seen examples of AI incorrect information/advice in simulation training?		Complex Systems/KBB JCS/CSE
42. Are there sufficient ‘unfiltered’ (non-AI) displays of critical functions to allow the human to verify independently true system status?		ASRS, HFACS, SHIELD Swiss Cheese
43. Are logs available for post-failure analysis, to know what happened and avoid recurrence of the failure?		Accident Aetiology

Table 1 also indicates the principal Human Factors themes that have influenced the various HAT requirements. As with conventional HF Requirements sets, there is often not a one-to-one relationship between a theoretical strand of Human Factors and an industrial HF requirement. Rather, the HAT requirements are a reflection of certain HF theoretical bases, contextualised via the use cases available, integrated with insights from the recently emerging HAT literature.

## 6. Application of Human Factors Requirements to Three HAT Prototypes

The aim of the HAT HF requirements questions is to help the designers, developers and Human Factors specialists realise and deliver a highly usable and safe Human-AI system, one that end users can use effectively and will want to use (avoiding misuse, disuse, and abuse, as mentioned earlier). The HF requirements set was applied to the three HAIKU use cases described earlier, between December 2023–January 2025, typically taking a full day with the respective design teams. An extract from the evaluations is shown in Table 2 for a sample of the questions from the first four Human Factors Areas.

For UC1, at several points during the HF Requirements review with the design team, potential improvements were derived for the AI system, in terms of how it is visualised and used by the pilot. In total, 27 issues were raised for further consideration by the design team, in terms of potential changes to the design or further tests to be carried out in the second set of simulations (Val2). A number of requirements were deemed ‘not applicable’ to the concept being evaluated, and many of the requirements from the last two areas (Competencies and Training; Organisational Readiness) were not answerable at this design stage.

The Human Factors requirements evaluation process for UC1 has resulted in a number of refinements to the HAT design, including: AI display aspects related to the ‘on/off’ status of the AI support; use of aural SA directional guidance; application of workload measures; consideration of how to better maintain a strategic overview during the emergency; pilot trust issues with the AI; consideration of the utility of personalisation of the AI to individual pilots; consideration of potential interference of the AI support with other alerts during an emergency; and use of HAZOP for identification of potential failure modes and recovery/mitigation measures. More generally, it raised issues that could be tested in Val2.

In all three use cases, the requirements evaluation process was found by the participating design teams to be useful and even provocative in elaborating and refining the concept: e.g., in UC2 on Roles and Responsibilities and Sense-Making; in UC4 on the detail of Human-AI Teaming interaction as well as Error and Failure Management, and in a further airport use case (UC5 in HAIKU, EASA category 1A/1B, at TRL 7) on Sense-Making (including the usability and explainability of the system) and Organisational Readiness (e.g., staff allocation to the AI, new competencies and training requirements, etc.). The product design teams found the approach to be straightforward and added value to the design process.

**Table 2.** Example responses to HF Requirements Questions from HAIKU Use Cases.

Human Factors Requirement Question	Y/N N/A TBD	Justification
Human-Centred Design		
Are licensed end users participating in design exercises such as focus groups, scenario-based testing, prototyping, and simulation (e.g., ranging from desk-top simulation to full scope simulation)?	Y	For UC1, ENAC pilots and commercial airline pilots are involved in Val1 (five pilots) and Val2 (12 pilots) real-time simulation exercises in a static A320 cockpit simulator. For UC2, 12 pilots participated in simulations. For UC4 a number of controllers have participated in design activities and real-time simulations in Skyway’s Tower simulator in Madrid.
Are end user opinions helping to inform and validate the design concept, as part of an integrated project team including product owner, data scientists, safety, security, Human Factors, and operational expertise?	Y	Pilots are involved in UC1, and the product owner is a pilot. Additionally, there are Data Science and Human Factors experts in the design team. Security is outside the scope of HAIKU, and UC1 is TRL4-5. For UC2, a Thales test pilot is involved with the design team. For UC4, the product owner is a Tower air traffic supervisor.
Are end users involved in any hazard identification exercises (e.g., HAZOP, STPA, FRAM etc.)?	Y	All three use cases have undertaken HAZOPs with end users (pilots and controllers) who have experienced the AI in simulations participating in the HAZOP process.
Roles and Responsibilities		
Are there any new roles, or suppressed roles?	Y	For UC1 this is a Single Pilot Operation (SPO) concept study, so one flight crew member is no longer in the cockpit. There are no staffing or role impacts on UC2 or UC4.
What is the level of autonomy—is the human still in charge?	Y	The end user remains in charge in all three use cases.
Does this level of autonomy change dynamically? Who/what determines when it changes?	TBD	In UC1 the IA is triggered by detection of pilot startle; the pilot can activate/deactivate the IA at any time. In UC2 the IA is triggered by certain circumstances, and the pilots can ignore it if they choose to do so. For UC4 the IA suggests changes when required, and if the ATCO does nothing the change of landing/take-off sequence will be automatically implemented. The ATCO can also switch the IA on and off.
Are the new/residual human roles consistent, and seen as meaningful by the intended users?	Y	Yes, for pilots in UC1 the AI is like a clever flight director or attention director, but the pilot remains in control. For UC2 the IA’s advice on three airports is very quick, with the supervised training still being fine-tuned to ensure the recommendations fit with pilots’ expertise and preferences. For UC4 the advice is seen as useful, giving them forewarning of arrivals/departures pressures.

Table 2. Cont.

Human Factors Requirement Question	Y/N N/A TBD	Justification
Sense-Making		
Is the interaction medium appropriate for the task, e.g., keyboard, touchscreen, voice, and even gesture recognition?	Y/ TBD	Startle and SA support colour-coding was appreciated in Val 1. Supporting displays on the Electronic Flight Bag (EFB) were not used due to the emergency nature of the event. Red was seen as too strong. Voice was suggested to back up the visual direction of SA (this has since been implemented). Changes have been tested in Val2 experiment (still under analysis). For UC2 the touchscreen display is seen as appropriate, and for UC4 display elements have been integrated into their normal radar displays.
Does the AI build its own situation representation?	Y	Yes, for UC1 coming from the aircraft data-bus, and from the pilot's attentional behaviour (eye-tracking). Context is also from the SOPs (Standard Operating Procedures) for the events. For UC2 extensive details of all European airports plus dynamic weather information, plus aircraft characteristics and passenger manifests, as well as remaining fuel, altitude, etc., are used to compute optimum alternates. For UC4 it is computing times to land and separation distances for a specified single-runway airport (Alicante in Spain), with a database of tens of thousands of landings in varied conditions to render predictions accurate.
Is the AI's situation representation made accessible to the end user, via visualisation and/or dialogue?	Y	For UC1, the EFB (electronic flight bag) to the captain's left summarises the AI's situation assessment. For UC2 the results of the three airports selected are shown on the moving map display and in an icon-based display, with a further explainability layer accessible to the flight crew. For UC4 only the output is shown, with a single line of explainability (usually, this is enough), due to the short timescale for accepting or rejecting the advice.
Does the AI-human interface reinforce the end user's situation awareness, so that human and AI can remain 'on the same page'?	Y	Pilots in UC1 felt it helped their SA and speed of regaining a situational picture. In UC2 the icon display and explainability layer unpack the AI's computation, showing which factors were prioritized. This helps when the pilots are unfamiliar with the airports available. For UC4 the display is clear and sharpens their own SA and 'look-ahead' time (Level 3 SA).
Can the human modify the AI's parameters to explore alternative courses of action?	N/ TBD	In UC1, no. The pilot can follow an alternative course of action, though there is no 'interaction' on this with the AI, as it is an emergency. In UC2 the flight crew can modify the goal priorities of the IA. In UC4 the ATCOs cannot modify the parameters.



Table 2. Cont.

Human Factors Requirement Question	Y/N N/A TBD	Justification
Is at least some operational explainability possible, rather than the AI being a 'black box'?	Y	In UC1, explainability is via the EFB. However, due to the very short response times in a loss of control in flight scenario, pilots had little time for explainability in the two simulations. This could differ in a scenario where the event was less clear cut, e.g., electronics failures, bus-bar failures, automation malfunction, etc. For UC2, there is a high degree of explainability. In UC4, the explainability needs are basic (aircraft on approach coming in too fast/too slow, etc.) and are deemed sufficient.
Does the AI possess the ability to detect human errors or misjudgement and notify them or directly correct them?	Y	In UC1, the AI is intended to detect temporary performance decrement due to startle, and to guide the pilot, but does not go as far as correcting his/her action. However, the dynamic highlighting of key instruments, along with callouts (e.g., "vertical speed!") could be considered a form of error correction. UC2 has the potential to aid error detection, e.g., failing to consider one of the variables in airport selection, since all the key parameters are used and displayed by the IA. UC4 has the potential to correct errors of judgment and memory failures/omissions or vigilance failures, or at the least alert controllers to something they have overlooked or misjudged.
Does the level of human workload enable the human to remain proactive rather than reactive, except for short periods?	Y	For the types of sudden scenarios in UC1, the pilot is in reactive mode. The answer 'yes' is given as it is a short period, and the aim is to reduce cognitive stress, giving them more 'headspace' to deal with the event. For UC2, it is a very clear 'yes', since it will save them at least 10 min and considerable work for the PNF (pilot not flying/first officer). For UC4, the ATCOs suggested its primary benefit may be when workload increases.
Can the human detect errors by the AI and intercede accordingly?	Y	In all three simulations, pilots noticed if something was incorrect, usually due to an error in the IA's database. Such errors should largely be eradicated by TRL 9. No edge cases or hallucinations nor alignment errors have arisen so far.
Does the human trust the AI, but not over-trust it? Is the human taught how to recognise AI malfunction or bad judgement?	Y	Pilots and controllers did not over-trust the IAs, knowing they were prototypes. They were not taught how to recognize aberrant IA behaviour. A general comment, though, from many end users, is that if such tools are to be implemented in the cockpit or ATC tower, they must be trustworthy and highly reliable; one serious mistake would irrevocably break trust and lead to non-use of the IA.

Additionally, the requirements evaluation process aided the enrichment of the collective understanding of the HAT design teams, and clarified the IA's intended operational modes of use, as well as integrating Human Factors into the HAT design. There is an intention to apply the requirements set to one final Urban Air Mobility use case (TRL 1-2, EASA category 3A), to see how useful the requirements are at such an early design stage. All results of these applications will be published in the final documentation for the HAIKU project at the end of 2025 [95].

## 7. Conclusions

A provisional HAT Human Factors requirements system comprising eight Human Factors areas and 165 detailed requirements questions has been developed and applied to three contemporary aviation HAT research use cases. The application of the requirements approach has been found to be both relevant and useful by the use case owners and design teams and has resulted in enhancements and refinements to the HAT system designs. In this sense, the approach appears to be fit for purpose, and scalable to projects of differing levels of AI autonomy and design maturity.

However, this is but a preliminary step. There is much research to do and experience to be gained with HAT systems and the integration of Intelligent Agents into safety-critical aviation operational settings. Certain Human Factors areas stand out as priorities for research, e.g., Human-IA Teamworking arrangements, sensemaking, shared situation awareness and operational explainability, AI error and failure management, and how to train end users to work effectively with IAs. As more HAT prototypes are developed and tested, experience of how best to design HATs will accrue, and the requirements can evolve in tandem, even adding new HF areas or sub-areas where advisable.

As a final conclusion, all the HAIKU use cases are augmenting human performance and capabilities, rather than seeking to supplant human capabilities in aviation systems. This is not only to maintain human agency and final authority, but also because it appears to be the best way to optimize overall aviation transportation system safety and performance. Human-AI Teaming therefore appears a useful and potentially valuable research and development avenue to pursue.

**Funding:** This publication is based on work performed in the HAIKU Project which has received funding from the European Union's Horizon Europe research and innovation program, under Grant Agreement no 101075332. Any dissemination reflects the authors' view only, and the European Commission is not responsible for any use that may be made of information it contains.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects in the studies cited.

**Data Availability Statement:** Data not available due to privacy restrictions. More data may be released by subsequent papers from the HAIKU project.

**Acknowledgments:** The author would like to acknowledge the support of the ENAC UC1 Team, in particular Jean-Paul Imbert and Alexandre Ducheve, as well as Roberto Venditti and Nikolas Giampaolo of Deep Blue and Miguel Villegas Sanchez of Skyway for their support on UC4, Jaime Diaz-Pineda (Thales Avionics), Ricardo Reis and Anais Villani (Embraer), Théodore Letouze and Charles Dormoy (Bordeaux University: ENSC and CATIE respectively) for their support on UC2, and Ryan Elliott and Liam Bolger of London Luton Airport for their support on UC5.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Burkov, A. *The Hundred Page Machine Learning Book*; Andriy Burkov: Quebec City, QC, Canada, 2019; ISBN 199957950X.
2. Morgan, G.; Grabowski, M. Human machine teaming in Mobile Miniaturized Aviation Logistics systems in safety-critical settings. *J. Saf. Sustain.* 2025, in press. [CrossRef]
3. Dalmau-Codina, R.; Gawinowski, G. Learning with Confidence the Likelihood of Flight Diversion Due to Adverse Weather at Destination. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 5615–5624. [CrossRef]
4. European Commission. *CORDIS Results Pack on AI In Air Traffic Management: A Thematic Collection of Innovative EU-Funded Research Results*; European Commission: Luxembourg, 2022. Available online: <https://www.sesarju.eu/node/4254> (accessed on 20 March 2025).
5. OpenAI. GPT-4 Technical Report. ArXiv Preprint abs/2303.08774. 2023. Available online: <https://arxiv.org/abs/2303.08774> (accessed on 20 March 2025).
6. European Parliament. EU AI Act: First Regulation on Artificial Intelligence. 2023. Available online: <https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence> (accessed on 20 March 2025).
7. Weinberg, J.; Goldhardt, J.; Patterson, S.; Kepros, J. Assessment of accuracy of an early artificial intelligence large language model at summarizing medical literature: ChatGPT 3.5 vs. ChatGPT 4.0. *J. Med. Artif. Intell.* **2024**, *7*, 33. [CrossRef]
8. Huang, L.; Yu, W.; Ma, W.; Zhong, W.; Feng, Z.; Wang, H.; Chen, Q.; Peng, W.; Feng, J.; Qin, B.; et al. A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Question. *ACM Trans. Inf. Syst.* **2024**, *43*, 1–55. [CrossRef]
9. Bjurling, O.; Müller, H.; Burgén, J.; Bouvet, C.J.; Berberian, B. Enabling Human-Autonomy Teaming in Aviation: A Framework to Address Human Factors in Digital Assistants Design. *J. Phys. Conf. Ser.* **2024**, *2716*, 012076. [CrossRef]
10. Duchevet, A.; Dong-Bach, V.; 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 ICCAS 2024, Toulouse, France, 16–17 May 2024.
11. SAFETEAM EU Project. 2023. Available online: <https://safeteamproject.eu/1186> (accessed on 20 March 2025).
12. Duchevet, A.; Imbert, J.-P.; De La Hogue, T.; Ferreirab, A.; Moensb, L.; Colomerc, A.; Canteroc, J.; Bejaranod, 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 34th Conference of the European Association for Aviation Psychology, Athens, Transportation Research Procedia, Gibraltar, UK, 26–30 September 2022; Volume 66, pp. 253–261. [CrossRef]
13. Minaskan, N.; Alban-Dormoy, C.; Pagani, A.; Andre, J.-M.; Stricker, D. Human Intelligent Machine Teaming in Single Pilot Operation: A Case Study. In *Augmented Cognition*; Bd. 13310; Springer International Publishing: Cham, Switzerland, 2022; pp. 348–360.
14. Kirwan, B.; Venditti, R.; Giampaolo, N.; Villegas Sanchez, M. A Human Centric Design Approach for Future Human-AI Teams in Aviation. In *Human Interactions and Emerging Technologies, Proceedings of the IHET 2024, Venice, Italy, 26–28 August 2024*; Ahram, T., Casarotto, L., Costa, P., Eds.; AHFE International: Orlando, FL, USA, 2024. [CrossRef]
15. Billings, C.E. *Human-Centred Aviation Automation: Principles and guidelines (Report No. NASA-TM-110381)*; National Aeronautics and Space Administration: Washington, DC, USA, 1996. Available online: <https://ntrs.nasa.gov/citations/19960016374> (accessed on 20 March 2025).
16. Bergh, L.I.; Teigen, K.S. AI safety: A regulatory perspective. In Proceedings of the 2024 4th International Conference on Applied Artificial Intelligence (ICAPAI) | 979-8-3503-4976-4/24/\$31.00 ©2024 IEEE | , Halden, Norway, 16 April 2024. [CrossRef]
17. EASA. EASA Concept Paper: First Usable Guidance for level 1 & 2 Machine Learning Applications. February 2023. Available online: <https://www.easa.europa.eu/en/newsroom-and-events/news/easa-artificial-intelligence-roadmap-20-published> (accessed on 20 March 2025).
18. EASA. Artificial Intelligence Concept Paper Issue 2. Guidance for Level 1 & 2 Machine-Learning Applications. 2024. Available online: <https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-concept-paper-issue-2> (accessed on 20 March 2025).
19. Naikar, N.; Brady, A.; Moy, G.M.; Kwok, H.W. Designing human-AI systems for complex settings: Ideas from distributed, joint and self-organising perspectives of sociotechnical systems and cognitive work analysis. *Ergonomics* **2023**, *55*, 1669–1694. [CrossRef] [PubMed]
20. Macey-Dare, R. How Soon is Now? Predicting the Expected Arrival Date of AGI- Artificial General Intelligence. 1204. 2023. Available online: [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4496418](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4496418) (accessed on 20 March 2025).
21. Lamb, H.; Levy, J.; Quigley, C. *Simply Artificial Intelligence*. In *DK Simply Books*; Penguin: London, UK, 2023.
22. Turing, A.M.; Copeland, B.J. *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life Plus The Secrets of Enigma*; Oxford University Press: Oxford, UK, 2004.
23. Turing, A.M. Computing machinery and intelligence. *Mind* **1950**, *49*, 433–460. [CrossRef]
24. Subrata, D. *It Began with Babbage: The Genesis of Computer Science*; Oxford University Press: Oxford, UK, 2014; p. 22, ISBN 978-0-19-930943-6.
25. Mueller, J.P.; Massaron, L.; Diamond, S. *Artificial Intelligence for Dummies*, 3rd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2024.

26. Lighthill, J. Artificial Intelligence: A General Survey. In *Artificial Intelligence—A Paper Symposium*; UK Science Research Council: Swindon, UK, 1973. Available online: [http://www.chilton-computing.org.uk/inf/literature/reports/lighthill\\_report/p001.htm](http://www.chilton-computing.org.uk/inf/literature/reports/lighthill_report/p001.htm) (accessed on 20 March 2025).
27. Available online: <https://www.perplexity.ai/page/a-historical-overview-of-ai-wi-A8daV1D9Qr2STQ6tgLEOtg> (accessed on 20 March 2025).
28. DeCanio, S. Robots and Humans—Complements or substitutes? *J. Macroecon.* **2016**, *49*, 280–291. [CrossRef]
29. Defoe, A. *AI Governance » A Research Agenda*; Future of Humanity Institute: Oxford, UK, 2017. Available online: <https://www.fhi.ox.ac.uk/wp-content/uploads/GovAI-Agenda.pdf> (accessed on 20 March 2025).
30. Dubey, P.; Dubey, P.; Hitesh, G. Enhancing sentiment analysis through deep layer integration with long short-term memory networks. *Int. J. Electr. Comput. Eng.* **2025**, *15*, 949–957. [CrossRef]
31. Luan, H.; Yang, K.; Hu, T.; Hu, J.; Liu, S.; Li, R.; He, J.; Yan, R.; Guo, X.; Qian, X.; et al. Review of deep learning-based pathological image classification: From task-specific models to foundation models. *Future Gener. Comput. Syst.* **2025**, *164*, 107578. [CrossRef]
32. Elazab, A.; Wang, C.; Abdelaziz, M.; Zhang, J.; Gu, J.; Gorriz, J.M.; Zhang, Y.; Chang, C. Alzheimer’s disease diagnosis from single and multimodal data using machine and deep learning models: Achievements and future directions. *Expert Syst. Appl.* **2024**, *255*, 124780. [CrossRef]
33. Lopes, N.M.; Aparicio, M.; Neves, F.T. Challenges and Prospects of Artificial Intelligence in Aviation: Bibliometric Study, Data Science and Management. *J. Clean Prod.* **2016**, *112*, 521–531. [CrossRef]
34. O’Neill, T.; McNeese, N.J.; Barron, A.; Schelble, B.G. Human- autonomy teaming: A review and analysis of the empirical literature. *Hum. Fact.* **2022**, *64*, 904–938. [CrossRef]
35. 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]
36. Hicks, M.T.; Humphries, J.; Slater, J. ChatGPT is bullshit. *Ethics Inf. Technol.* **2024**, *26*, 38. [CrossRef]
37. Kaliardos, W. *Enough Fluff: Returning to Meaningful Perspectives on Automation*; FAA, US Department of Transportation: Washington, DC, USA, 2023. Available online: <https://rosap.ntl.bts.gov/view/dot/64829> (accessed on 20 March 2025).
38. FAA. Roadmap for Artificial Intelligence Safety. Assurance. 2024. Available online: [https://www.faa.gov/aircraft/air\\_cert/step/roadmap\\_for\\_AI\\_safety\\_assurance](https://www.faa.gov/aircraft/air_cert/step/roadmap_for_AI_safety_assurance) (accessed on 20 March 2025).
39. Kirwan, B. 2B or not 2B? The AI Challenge to Civil Aviation Human Factors. In *Contemporary Ergonomics and Human Factors 2024*; Golightly, D., Balfe, N., Charles, R., Eds.; Chartered Institute of Ergonomics & Human Factors: Kenilworth, UK, 2024; pp. 36–44, ISBN 978-1-9996527-6-0.
40. Kilner, A.; Pelchen-Medwed, R.; Soudain, G.; Labatut, M.; Denis, C. Exploring Cooperation and Collaboration in Human AI Teaming (HAT)—EASA AI Concept paper V2.0. In Proceedings of the 14th European Association for Aviation Psychology Conference EAAP 35, Thessaloniki, Greece, 8–11 October 2024.
41. Fitts, P.M. *Human Engineering for an Effective Air-Navigation and Traffic-Control System*; Ohio State University: Columbus, OH, USA, 1951. Available online: <https://psycnet.apa.org/record/1952-01751-000> (accessed on 20 March 2025).
42. Chapanis, A. On the allocation of functions between men and machines. *Occup. Psychol.* **1965**, *39*, 1–11.
43. De Winter, J.C.F.; Hancock, P.A. Reflections on the 1951 Fitts list: Do humans believe now that machines surpass them? Paper presented at 6th International Conference on Applied Human Factors and Ergonomics (AHFE 2015) and the Affiliated Conferences, AHFE 2015. *Procedia Manuf.* **2015**, *3*, 5334–5341. [CrossRef]
44. 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]
45. Available online: [https://en.wikipedia.org/wiki/Crew\\_resource\\_management](https://en.wikipedia.org/wiki/Crew_resource_management) (accessed on 20 March 2025).
46. Available online: <https://skybrary.aero/articles/team-resource-management-trm> (accessed on 20 March 2025).
47. Available online: [https://en.wikipedia.org/wiki/Maritime\\_resource\\_management](https://en.wikipedia.org/wiki/Maritime_resource_management) (accessed on 20 March 2025).
48. Norman, D.A. *The Design of Everyday Things (Revised and Expanded Editions ed.)*; The MIT Press: Cambridge, MA, USA; London, UK, 2013; ISBN 978-0-262-52567-1.
49. Norman, D.A. *User Centered System Design: New Perspectives on Human-Computer Interaction*; CRC: Boca Raton, FL, USA, 1986; ISBN 978-0-89859-872-8.
50. Schneiderman, B.; Plaisant, C. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*; Addison-Wesley: Boston, MA, USA, 2010; ISBN 9780321537355.
51. Schneiderman, B. *Human-Centred AI*; Oxford University Press: Oxford, UK, 2022.
52. Gunning, D.; Aha, D. DARPA’s Explainable Artificial Intelligence (XAI) Program. *AI Mag.* **2019**, *40*, 44–58. [CrossRef]
53. Hollnagel, E.; Woods, D.D. Cognitive Systems Engineering: New wine in new bottles. *Int. J. Man-Mach. Stud.* **1983**, *18*, 583–600. [CrossRef]
54. Hollnagel, E.; Woods, D.D. *Joint Cognitive Systems: Foundations of Cognitive Systems Engineering*; CRC Press: Boca Raton, FL, USA, 2005.
55. Edwin, H. How a Cockpit Remembers Its Speeds. *Cogn. Sci.* **1995**, *19*, 265–288.
56. Bainbridge, L. Ironies of automation. *Automatica* **1983**, *19*, 775–779. [CrossRef]

57. Endsley, M.R. Ironies of artificial intelligence. *Ergonomics* **2023**, *66*, 11.
58. Sheridan, T.B.; Verplank, W.L. *Human and Computer Control of Undersea Teleoperators*; Department of Mechanical Engineering, MIT: Cambridge, MA, USA, 1978.
59. Sheridan, T.B. Automation, authority and angst—Revisited. In *Human Factors Society, Proceedings of the Human Factors Society 35th Annual Meeting, San Francisco, CA, USA, 2–6 September 1991*; Human Factors & Ergonomics Society Press: New York, NY, USA, 1991; pp. 18–26.
60. Parasuraman, R.; Sheridan, T.B.; Wickens, C.D. A Model for Types and Levels of Human Interaction with Automation. *IEEE Trans. Syst. Man Cybern.—Part A Syst. Hum.* **2000**, *30*, 286–297. [[CrossRef](#)]
61. Wickens, C.; Hollands, J.G.; Banbury, S.; Parasuraman, R. *Engineering Psychology and Human Performance*; Taylor and Francis: London, UK, 2012.
62. Endsley, M.R. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors* **1995**, *37*, 32–64. [[CrossRef](#)]
63. Moray, N. (Ed.) Mental workload—Its theory and measurement. In *NATO Conference Series III on Human Factors, Greece, 1979*; Springer: New York, NY, USA, 2013; ISBN 9781475708851.
64. Klein, G.; Moon, B.; Hoffman, R. Making Sense of Sensemaking 1: Alternative Perspectives. *IEEE Intell. Syst.* **2006**, *21*, 70–73. [[CrossRef](#)]
65. Rasmussen, J. Skills, rules, knowledge; signals, signs, and symbols; and other distinctions in human performance models. *IEEE Trans. Syst. Man Cybern.* **1983**, *3*, 257–266.
66. Vicente, K.J.; Rasmussen, J. Ecological interface design: Theoretical foundations. *IEEE Trans. Syst. Man Cybern.* **1992**, *22*, 589–606.
67. Reason, J.T. *Human Error*; Cambridge University Press: Cambridge, UK, 1990.
68. Shappell, S.A.; Wiegmann, D.A. *The Human Factors Analysis and Classification System—HFACS*; DOT/FAA/AM-00/7; U.S. Department of Transportation: Washington, DC, USA, 2000.
69. Dillinger, T.; Kiriokos, N. *NASA Office of Safety and Mission Assurance Human Factors Handbook: Procedural Guidance and Tools*; NASA/SP-2019-220204; National Aeronautics and Space Administration (NASA): Washington, DC, USA, 2019.
70. Stroeve, S.; Kirwan, B.; Turan, O.; Kurt, R.E.; van Doorn, B.; Save, L.; Jonk, P.; Navas de Maya, B.; Kilner, A.; Verhoeven, R.; et al. SHIELD Human Factors Taxonomy and Database for Learning from Aviation and Maritime Safety Occurrences. *Safety* **2023**, *9*, 14. [[CrossRef](#)]
71. IAEA. *Safety Culture*; Safety Series No. 75-INSAG-4; International Atomic Energy Agency: Vienna, Austria, 1991.
72. Kirwan, B.; Shorrocks, S.T. A view from elsewhere: Safety culture in European air traffic management. In *Patient Safety Culture*; Waterson, P., Ed.; Ashgate: Aldershot, UK, 2015; pp. 349–370.
73. Kirwan, B. The Impact of Artificial Intelligence on Future Aviation Safety Culture. *Future Transp.* **2024**, *4*, 349–379. [[CrossRef](#)]
74. Salas, E.; Sims, D.E.; Burke, C.S. Is There a “Big Five” in Teamwork? *Small Group Res.* **2005**, *36*, 555–599. [[CrossRef](#)]
75. Parasuraman, R. Humans and Automation: Use, Misuse, Disuse, and Abuse. *Hum. Factors* **1997**, *39*, 230–253.
76. Parasuraman, R.; Manzey, D.H. Complacency and bias in human use of automation: An attentional integration. *Hum. Factors* **2012**, *52*, 3. [[CrossRef](#)]
77. Tversky, A.; Kahneman, D. Judgment under Uncertainty: Heuristics and Biases. *Science* **1974**, *185*, 1124–1131. [[PubMed](#)]
78. Klayman, J. Varieties of Confirmation Bias. *Psychol. Learn. Motiv.* **1995**, *32*, 385–418. [[CrossRef](#)]
79. Hawkins, F.H. *Human Factors in Flight*, 2nd ed.; Orlandy, H.W., Ed.; Routledge: London, UK, 1993. [[CrossRef](#)]
80. Leveson, N.G. Safety Analysis in Early Concept Development and Requirements Generation. In *Paper Presented at the 28th Annual INCOSE International Symposium*; Wiley: Hoboken, NJ, USA, 2018. Available online: <https://hdl.handle.net/1721.1/126541> (accessed on 20 March 2025).
81. Kletz, T. *HAZOP and HAZAN: Identifying and Assessing Process Industry Hazards*, 4th ed.; Taylor and Francis: Oxfordshire, UK, 1999.
82. Single, J.J.; Schmidt, J.; Denecke, J. Computer-Aided Hazop: Ontologies and Ai for Hazard Identification and Propagation. *Comput. Aided Chem. Eng.* **2020**, *48*, 1783–1788. [[CrossRef](#)]
83. Hollnagel, E. *FRAM: The Functional Resonance Analysis Method. Modelling Complex Socio-Technical Systems*; CRC Press: Boca Raton, FL, USA, 2012. [[CrossRef](#)]
84. Hollnagel, E. A Tale of Two Safeties. *Int. J. Nucl. Saf. Simul.* **2013**, *4*, 1–9. Available online: [https://www.erikhollnagel.com/A\\_tale\\_of\\_two\\_safeties.pdf](https://www.erikhollnagel.com/A_tale_of_two_safeties.pdf) (accessed on 20 March 2025).
85. Kumar, R.S.S.; Snover, J.; O’Brien, D.; Albert, K.; Viljoen, S. *Failure Modes in Machine Learning*; Microsoft Corporation & Berkman Klein Center for Internet and Society at Harvard University: Cambridge, MA, USA, 2019.
86. Available online: <https://skybrary.aero/articles/just-culture> (accessed on 20 March 2025).
87. Franchina, F. Artificial Intelligence and the Just Culture Principle. *Hindsight* **35**, November, pp. 39–42. EUROCONTROL, rue de la Fusée 96, B-1130, Brussels. 2023. Available online: <https://skybrary.aero/articles/hindsight-35> (accessed on 20 March 2025).
88. MARC Baumgartner; Malakis, S. Just Culture and Artificial Intelligence: Do We Need to Expand the Just Culture Playbook? *Hindsight* **35**, November, pp.43-45. EUROCONTROL, Rue de la Fusee 96, B-1130 Brussels. 2023. Available online: <https://skybrary.aero/articles/hindsight-35> (accessed on 20 March 2025).

89. Dubey, A.; Kumar, A.; Jain, S.; Arora, V.; Puttaveerana, A. HACO: A Framework for Developing Human-AI Teaming. In Proceedings of the 13th Innovations in Software Engineering Conference on Formerly Known as India Software Engineering Conference, Kurukshetra, India, 20–22 February 2025; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–9. [[CrossRef](#)]
90. Rochlin, G.I. Reliable Organizations: Present Research and Future Directions. *J. Contingencies Crisis Manag.* **1996**, *4*, 55–59, ISSN 1468-5973. [[CrossRef](#)]
91. Schecter, A.; Hohenstein, J.; Larson, L.; Harris, A.; Hou, T.; Lee, W.; Lauharatanahirun, N.; DeChurch, L.; Contractor, N.; Jung, M. Vero: An accessible method for studying human-AI teamwork. *Comput. Hum. Behav.* **2023**, *141*, 107606.
92. Zhang, G.; Chong, L.; Kotovsky, K.; Cagan, J. Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Comput. Hum. Behav.* **2023**, *139*, 107536.
93. Ho, M.-T.; Mantello, P. An analytical framework for studying attitude towards emotional AI: The three-pronged approach. *MethodsX* **2023**, *10*, 102149. [[PubMed](#)]
94. Lees, M.J.; Johnstone, M.C. Implementing safety features of Industry 4.0 without compromising safety culture. *Int. Fed. Autom. Control (IFAC) Pap. Online* **2021**, *54*, 680–685.
95. HAIKU EU Project (2023-25). Available online: <https://cordis.europa.eu/project/id/101075332> (accessed on 20 March 2025).

**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.