HAIQU - A Human Factors Requirements App for Human-AI Teams in Aviation

Venditti, R.[1], Pozzi, S.,[1] Frau, G.,[1] Salam, R.,[1] Imbert, J-P.,[2] Duchevet, A.[2] and Kirwan, B[3]

[1]Deep Blue, [2]ENAC, [3]EUROCONTROL

SUMMARY

Contemporary Human Factors requirements sets, such as those embodied in EASA CS29.1302 for cockpit design, and the SESAR Human Performance Assessment Process (HPAP) requirements for air traffic management systems, are unlikely to be sufficient to account for Operational Explainability (OpXAI), shared situation awareness, and other elements associated with proposed Human-AI Teaming systems. Whilst the European Union Aviation Safety Agency (EASA) has provided new guidance on a number of these areas, their focus is largely on safety, with less attention on other areas such as Roles and Responsibilities, Competencies and Training, and Organisational Readiness - all concerned with systems integration.

HAIQU (Human-AI Teaming Questionnaire) is a freely available Web App developed in the Horizon Europe HAIKU project. The app aims to make standards and regulations accessible and user-friendly for design teams looking to integrate AI capabilities into safety critical applications. HAIQU contains 180 requirements in eight Human Factors areas (Human-Centred Design, Roles & Responsibilities, Sense-Making, Communication, Teamworking, Error and Failure Management, Competencies and Training, and Organisational Readiness). The App is sensitive to different design maturity levels and AI autonomy levels, consistent with EASA's classification of Human-AI Teaming arrangements.

This paper firstly situates the requirements set in terms of current EU and EASA regulations, as well as contemporary Human Factors guidance sources for cockpit and air traffic management systems. It then focuses on showing how HAIQU works and can serve as an aid to Product or Design Teams managing the integration of Human Factors into future AI-based systems, using as illustration a use case in cockpit Human-AI Teaming.

KEYWORDS

Aviation, AI, Human-AI Teaming, Human Factors Requirements, Design

Background

The next decade is likely to see the introduction of AI-based Intelligent Assistants (IAs) in operational aviation contexts, whether to augment pilot, controller, and airport operatives' capabilities, or to support new concepts such as single pilot operations and urban air mobility (city-wide drone and sky-taxi air traffic control)[1]. Whilst some argue that AI is 'just more automation' (e.g. Kaliardos, 2023), others consider that future AI could herald a more radical change in human machine collaboration. AI has the potential to drastically impact several sectors, improving different key performance areas, i.e. safety, efficiency, predictability. However, it also presents a

[1] See https://haikuproject.eu and https://safeteamproject.eu

new set of challenges such as algorithmic bias, lack of transparency, over-reliance and unclear human-machine responsibility boundaries.

The European Union's Artificial Intelligence Act (AI Act: European Parliament, 2023[2]) introduces a proportionate risk-based approach to AI regulation, which imposes a gradual scheme of requirements and obligations depending on the level of risk posed to health, safety and fundamental rights. This approach assigns regulatory requirements to four distinct risk categories: unacceptable, high, limited and minimal (see Figure 1). This hierarchical structure follows the principle that higher risks warrant stricter requirements and obligations, aiming to balance innovation with the protection of fundamental rights and values.

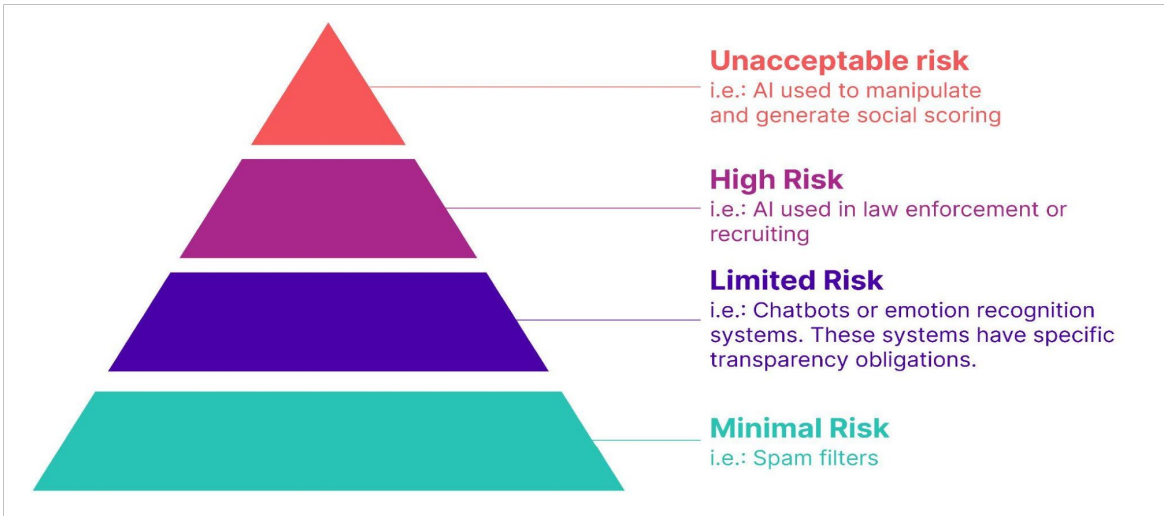

Figure 1: A Proportionate Risk-based approach to AI (adapted from European Parliament, 2023)

The Need for a Requirement Tool

The EU AI Act requirements complement other aviation sector-specific regulations, such as those established by the European Union Aviation Safety Agency (EASA, 2023; 2024). EASA has developed guidance on Human Factors for Human-AI Teaming systems, encompassing operational explainability, natural language processing interfaces, and traditional aspects like design and human-machine interaction, bringing new requirements to the scene. However, EASA's focus is safety, whereas Human Factors (HF) considers in more depth other performance elements such as roles and responsibilities, competencies, and wellbeing, that can have indirect impacts on safety.

Some of these factors are highlighted in the European Act on AI, but not crystallised into measurable requirements. HF requirements for AI systems therefore remain somewhat scattered across different sources and often lack practical implementation guidance for designers. Additionally, while regulatory entities effectively communicate principles and requirements through declarative statements ("the designer shall"), they rarely address the procedural aspects ("how to").

The HAIQU (Human-AI Questionnaire) Web App addresses this gap by translating regulatory requirements into actionable questions for design teams, helping them improve their AI systems and meet compliance standards. HAIQU has been developed in the context of the Horizon Europe funded HAIKU (Human-AI teaming Knowledge and Understanding for aviation safety) project

(https://haikuproject.eu). HAIKU aims to explore Human-AI Teaming via six aviation use cases involving AI prototypes (2 cockpit, one ATM, two airport and one urban air mobility)

2 https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2021)698792

The HAIQU concept

The HAIQU team has derived and synthesized Human Factors requirements by reviewing multiple regulatory sources, including EASA's regulatory requirements for Human-AI interactive systems in aviation, as well as requirements arising from the EU Act on AI, and Human Factors guidance from platforms such as SESAR[3], EASA regulation CS25/1302[4], and the Human Factors Compass[5]. These requirements, expressed as statements, were then translated into a set of questions for Product or Design Teams to evaluate their AI design against the HF requirements relevant to both its maturity level and the level of AI autonomy in the Human-AI Teaming system concept.

As an example of how HAIQU renders declarative requirements into more achievable ones contextualised in Human Factors capabilities, in the EU AI Act it is stated that AI-based systems should be human-centred (AI Act: European Parliament, Recital 6, 2023), implying the need for human involvement in the design process. However, it provides no specific guidance on how to achieve this. The related HAIQU question is "*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)?*" With such a question the Product Team can immediately tell if the requirement is satisfied, and if not, they know what needs to be done.

Similarly, in the EASA guidance there is an operational explainability requirement (EXP 18) that "*The training and instructions available for the end user should include procedures for handling possible outputs of the ODD [Operational Design Domain[6]] monitoring and output confidence monitoring*." This is translated and expanded in HAIQU as follows:

i. "*Are users trained to recognise and take corrective action on strange or erroneous AI outputs?*"

ii. "*Does advice offered come with an indication of its confidence or uncertainty?"*

iii. *"Has the end user seen examples of AI incorrect information / advice in simulation training?"*

At a simple level these questions can be answered 'yes' or 'no'. If 'yes', evidence should be available in terms of training schedules and records [(i) and (iii) above], and of the presence of confidence estimation parameters in the AI (ii). This expansion of the requirements can give the Product Team more latitude, e.g. if confidence estimates are not available (as these can be difficult to generate with some AI systems) or are deemed inadvisable, training and testing of training in simulators [(i) and (iii)] offer viable alternatives. This points to one of the aims of the HAIKU project, namely to be flexible and proportionate, and to give the Product or Design Team options. This degree of flexibility and 'scalability' is important, as the HAIQU questionnaire is by its nature one for self-reflection by the Team – there is no set of magic or secret answers, though often the 'ideal' answer is implied in the question. Nor is it intended as a regulatory compliance tool (though it may be used to increase the likelihood of regulatory compliance). Rather, it is for the team to see

3 SESAR Human Performance Guidelines https://www.sesarju.eu/sites/default/files/documents/transversal/SESAR%202020%20%20Human%20Performance%20Assessment%20Guidance.pdf

4 EASA CS25 1302 https://www.easa.europa.eu/en/search?keys=CS25%201302&f[0]=origin:EASA+Pro

5 Human Factors Compass SAFEMODE Project https://www.safemodeproject.eu/EhuridGuidedPaths.aspx

6 Operational Design Domain defines the specific operating conditions under which an aviation system (or a part of it) is designed to function safely and effectively.

where they are strong on assuring high human-AI team performance with their system, and where more work may be advisable.

The resulting 180 questions (for more on their derivation see Kirwan, 2025) effectively guide teams into the exploration of eight key areas, *Human-Centred Design, Roles & Responsibilities, Sensemaking, Communication, Teamworking, Error & Failure Management, Competencies & Training,* and *Organisational Readiness*. An overview of the HAIQU architecture is given in Figure 2 (Kirwan, op cit).

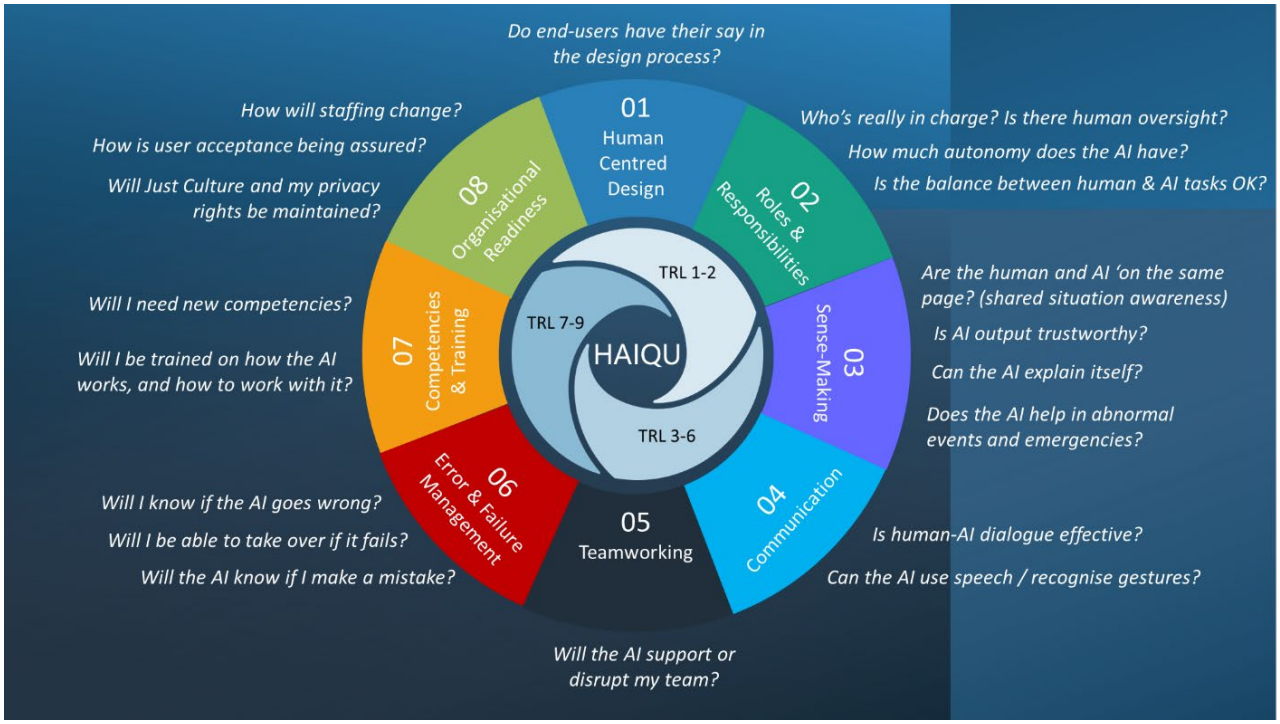

Figure 2: HAIQU Eight Overarching Human Factors Areas (from Kirwan, 2025)

How HAIQU Looks and Works

The questions are embedded in a collaborative web-platform (https://haiqu.eu/). The landing page is shown in Figure 3. Users can sign in to the platform, register their system, and begin the questionnaire immediately. HAIQU supports iterative development by enabling assessments at various stages of design maturity. While some questions may need to be revisited later in the development cycle, early identification of potential issues allows for timely adjustments before designs become too rigid to modify. By the time a system approaches deployment, teams will have comprehensively addressed all relevant requirements, establishing a solid foundation for operating their Human-AI system.

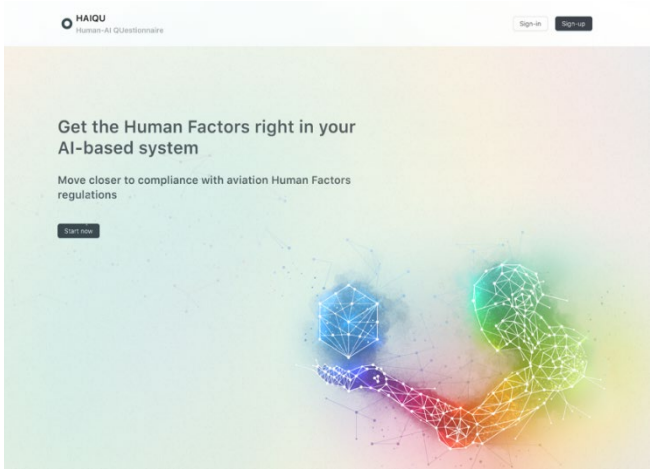

Figure 3: HAIQU Landing Page

HAIQU Features

To achieve its goal and concretely help Product and Design Teams, the HAIQU app offers several key features designed to support diverse user needs.

Progress tracking and saving: HAIQU is comprehensive, and it may not be possible to complete all relevant questions in one sitting. Some questions may be non-applicable, some will be quick to answer, and others may take more time and require deeper consideration and evidence gathering. For this reason, the tool saves the status of the questionnaire, giving the possibility for users to continue when they see fit.

Interaction with questions: Users can quickly evaluate requirements using four response options (see Figure 4): 'Yes' (requirement met), 'No' (requirement not met), 'TBD' (to be addressed later), or 'N/A' (not applicable). The interface prioritises quick navigation and fast interaction, allowing users to navigate smoothly between different assessment areas.

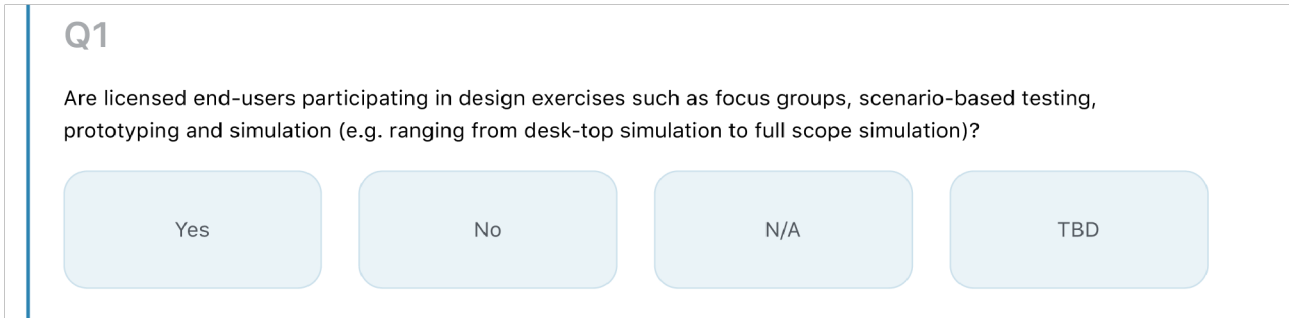

Figure 4: Example of a HAIQU Question

Action tracking: When requirements are marked as "No" or "TBD", the system prompts teams to document specific actions needed for compliance. This feature helps teams develop concrete implementation plans, with all actions visible to the entire team through a shared workspace.

Real-time monitoring: A dynamic dashboard provides real-time visibility of the evaluation progress across all eight areas or selected subsets relevant to the project. The system records all responses, including supporting evidence, justifications and identified actions. This tracking enables teams to identify areas needing attention. For instance, if a project shows strong compliance in Human Centred Design but reveals gaps in Error & Failure Management, teams can prioritise improving their error management strategies.

As an example, Figure 5 shows the HAIQU status of a 'project-in-progress' called FOCUS (Flight Operational Companion for Unexpected Situations: Duchevet et al, 2024), which is a cockpit-based research prototype Human-AI Teaming tool aimed at helping pilots in sudden in-flight emergencies.

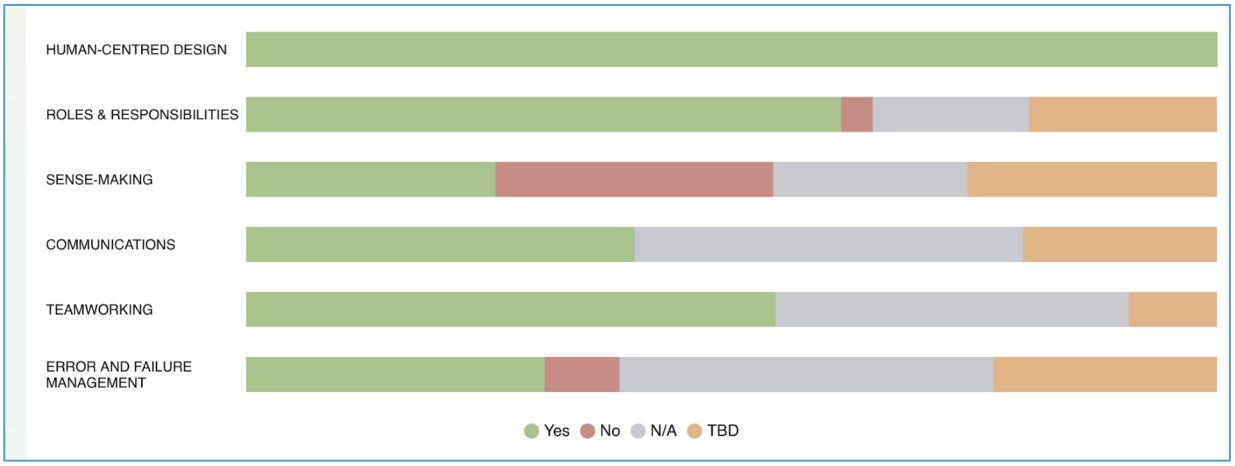

Figure 5: Example of Intermediate HAIQU Results for a Human-AI Teaming Project

This figure shows that the project is doing well in certain areas, whereas more work is needed in others. Two other parts of the Dashboard are shown in Figure 6. The left part of Figure 6 shows the status of answers to the relevant questions. For FOCUS, 120 out of 180 are deemed relevant based on the level of maturity of the concept (it is a research prototype that can be used in a real-time simulation) and the level of AI autonomy and functionality (it is EASA Category 2A and does not have AI natural language processing). The questions answered 'Yes' can point to substantiative evidence, those answered 'No' or 'N/A' (not applicable) need to have an associated justification, and those deemed TBD (To be done) need a task or work plan.

The right-hand side of Figure 6 offers a 'helicopter view' and is a spider chart showing the overall 'health' of the project when considered against the eight HAIQU Human Factors areas. In this case, due to the maturity level of the research project, *Competencies & Training* and *Organisational Readiness* are not yet assessable, so are set to zero. *Human Centred Design* is fully addressed, and *Teaming* and *Communications* are in progress, while *Sense-Making* and *Error and Failure Management* are probably the next priorities in the development process. The Dashboard can be reviewed at key stages (e.g. after a major simulation with end-users) to keep an eye on progress.

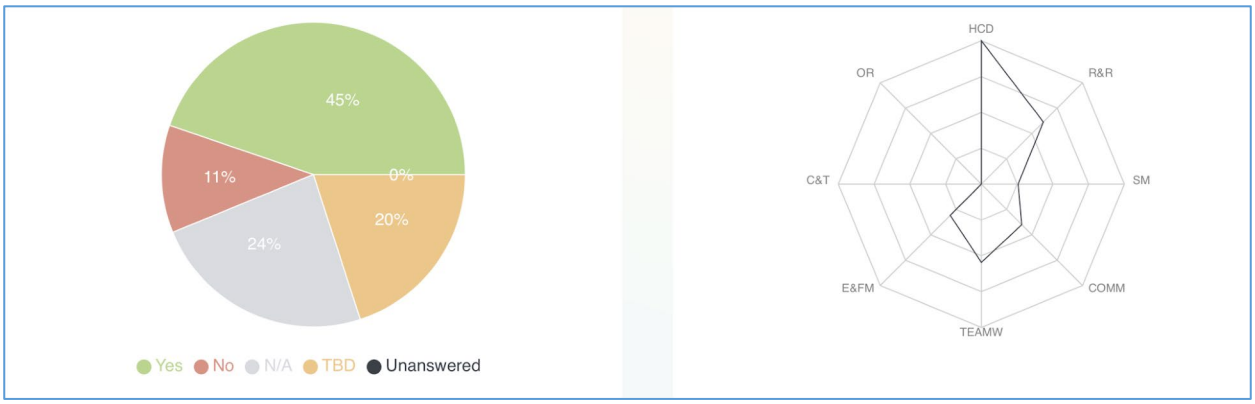

Figure 6: Key HAIQU Dashboard Elements

Figure 7 shows example 'TBD' tasks for FOCUS related to 'Sense-Making', showing the original HAIQU question and the task plan to resolve it.

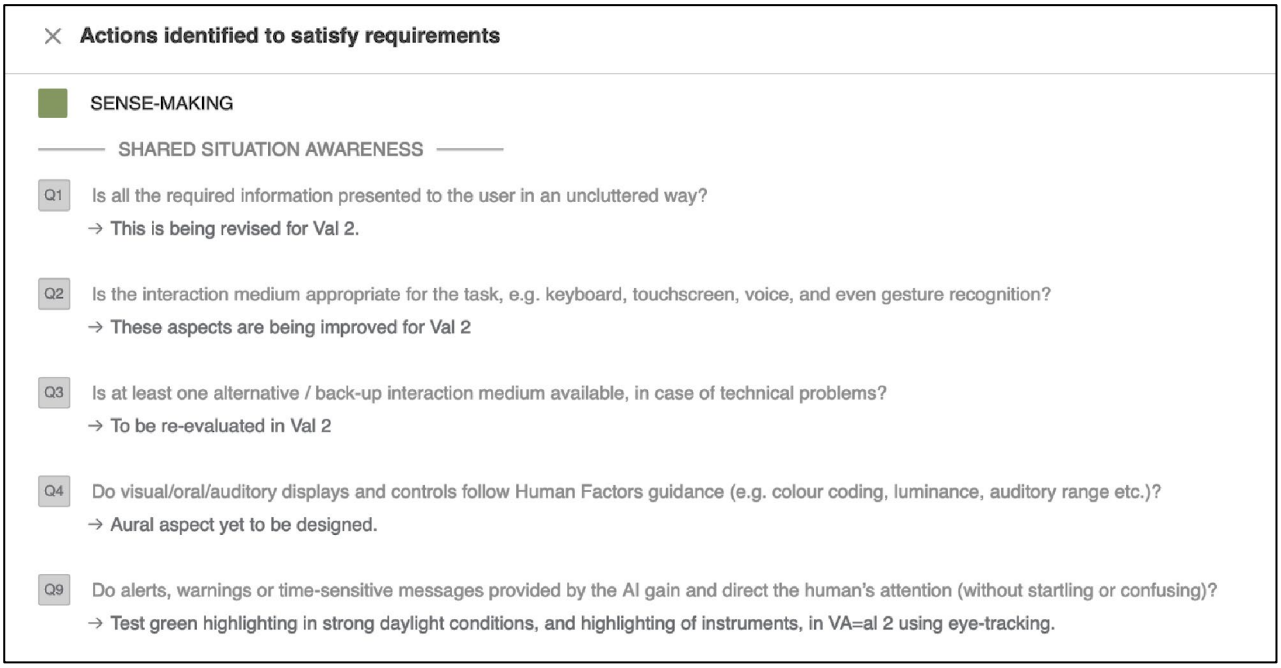

Figure 7: Example of residual tasks to be carried out to satisfy requirements.

In Figure 7, 'Val 2' refers to a second real-time simulation with pilot end-users, carried out and being analysed at the time of writing this paper. Many of the TBD issues have been resolved or moved closer to resolution. For example, an aural component has been designed and implemented (Q2; Q3; Q4) and appears to be working better for pilots (who are accustomed to aural alerts/instructions during emergencies). Clutter (Q1) has been alleviated, and the examination of the adequacy of the visibility/salience of the green illumination in the cockpit (Q9) is under analysis.

Dual assessment mode: To accommodate different project needs, HAIQU offers two modes: Guided and User-driven. When a system is added to the app, the app asks several high-level questions about the system in the "Scoping" section. These questions investigate the AI system's maturity level (from early research to deployment-ready), EASA classification (1A to 3B), and intended capabilities (ranging from text interaction to complex gesture recognition). Then, users are prompted to choose between the Guided or the User-driven approach. If users pick the Guided approach, the app intelligently filters questions based on the answers provided in the Scoping and only shows relevant questions for their system. This step significantly cuts down the time required for the questionnaire (e.g. for FOCUS it reduced the requirements set from 180 to 120), streamlining the whole process to only relevant areas and questions. This is a quicker albeit not as deep approach. Alternatively, the User-driven approach allows users to freely explore any Human Factors area or question they deem relevant to their project, and they are free to go as in-depth as they like. Note that if the initial scoping assessment is carried out when the project is a research project, which subsequently becomes a 'real' project intended for full operational implementation, the scoping will have to be reassessed at the appropriate stage in the project. This has occurred for one of the HAIKU use cases.

Sharing for cross functional collaboration: Given the broad scope of expertise required (one person cannot answer all the questions, as they cover a range of areas or expertise), the platform is built for collaborative use by cross-functional teams, including product owners, experienced (and ideally currently licensed) end-users, AI specialists, Human Factors and safety experts. Individual users can also "share" an assessment with other users in only a few minutes, so that questions can be addressed collaboratively by users who are not in the same building, for example. That said, in its application so far (four applications), HAIQU evaluations benefit from having face-to-face sessions,

as otherwise some details are often left unsaid, or else people do not always speak up when joining meetings remotely. Also, for one use case in particular, the HAIKU session spawned significant discussion on the operational concept of the Human-AI Teaming prototype itself, and the entire team found this discussion very productive, and for some of the junior members, very instructive.

Conclusions

HAIQU has already received positive feedback from the HAIKU project teams who have used it. The platform has been successfully tested with four distinct use cases: two cockpit applications, one air traffic management (ATM) application and one airport-based application. The platform has proven effective in challenging design teams to deepen their consideration of human-centric aspects of their systems. For instance, the ATM use case revealed that, while the team had addressed most requirements, they lacked robust safeguarding strategies for managing "unscripted" user interactions with the AI system. The interaction with HAIQU prompted designers to take a more detailed look at error management, making them think about edge cases they had initially overlooked.

While HAIQU has initially been developed for aviation applications, its core “formula”- translating regulatory requirements into actionable questions for design teams - has broader potential. This approach could be valuable across various safety-critical sectors, including nuclear power, oil and gas, defence and healthcare, where human-AI interaction must be carefully managed within strict regulatory frameworks.

Funding & Disclaimer

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.

References

Duchevet, A., Dong-Bach V., Peyruqueou, V., De-La-Hogue, T., Garcia, J., Causse, M. and Imbert, J-P. (2024) FOCUS: An Intelligent Startle Management Assistant for Maximizing Pilot Resilience. ICCAS 2024, Toulouse, France.

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

EASA (2024) EASA Artificial Intelligence Concept Paper Issue 2. April. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligenceconcept-paper-issue-2

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

European Parliament (2023) EU AI Act: Recital 6. https://artificialintelligenceact.eu/recital/6/

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

Kirwan, B. (2025: preprint) Human Factors Requirements for Aviation Human-AI Teaming. Future Transportation. doi:10.20944 https://www.preprints.org/manuscript/202501.0974/v1