SafeTeam

D5.1: Bow-Tie Analysis and Alternative Risk Assessment Frameworks

Document number	D5.1
Document title	**Bow-Tie Analysis and Alternative Risk Assessment Frameworks**
Version	1.0
Work package	WP5
Edition date	20.06.2025
Responsible unit	**CAAi**
Dissemination level	PU
Project acronym	SafeTeam
Grant	101069877
Call	Safe, Resilient Transport and Smart Mobility services for passengers and goods (HORIZON-CL5-2021-D6-01)
Topic	HORIZON-CL5-2021-D6-01-13: Safe automation and human factors in aviation – intelligent integration and assistance

This project has been co-funded by the European Union under Grant Agreement 101069877

Co-funded by the European Union

SafeTeam Consortium

	Innaxis (INX)
	Agencia Estatal de Seguridad Aérea (AESA)
	Technische Universität München (TUM)
	DataBeacon
	ONERA
	Rise Research Institutes of Sweden AB (RISE)
	PEGASUS HAVA TASIMACILIGI ANONIM SIRKETI (PEGASUS)
	CAA INTERNATIONAL LIMITED (CAAi)

Document change record

Version	Date	Status
0.1	30/04/2025	Drat submission
1.0	10/06/2025	Amendments Addressing Consortium Feedback
1.1	20/06/2025	Final version for submission

Table of Contents

1 Scope

The aviation industry is increasingly exploring the potential of machine learning (ML) and concepts of Artificial Intelligence (AI) to enhance flight safety. Use cases such as predicting unstable approaches during aircraft landings, digital assistance for area ATC, and drone risks assessment are concepts being explored.

Project SafeTeam is a consortium of organisations across Europe established to develop these use cases as early adopters to understand the challenges of implementation for more automated digital assistance. SafeTeam aims to progress on the human factor's aspects on the use of digital assistants for aviation, including a deeper understanding on the technology and processes that will facilitate the adoption of more autonomous tools with integration into operations and enhancing human cognitive abilities. The project also aims to investigate approval and certification issues regarding aspects related to the human ability to operate more sophisticated automated tools in order to develop the explainability of larger concepts of AI operations.

The objectives of regulatory involvement within project SafeTeam are to review at a high level, the development of the consortium cases aimed to address the following objectives:

- Facilitate a discussion on human-centric approach to automation and its integration into a wide spectrum of air traffic operations.
- Propose potential methodologies for the assessment and monitoring of the system performance, with a core focus on safety, to help the consortium progress the topic of Human-Machine cooperation.
- Through education of regulatory frameworks and methods, progress the development of Digital Assistants for aviation operations in support of human performance for all development, testing, validation, and verification phases.
- Comment on the potential regulatory requirements for automation tools to address safety critical elements.

2 Risk Assessment Frameworks

2.1 Regulatory approach to risk assessments

Risk assessments are a fundamental tool for aviation safety oversight. They provide a structured approach to identifying threats, evaluating the likelihood and severity of hazardous events, and determining whether risks can be accepted within the boundaries of the regulatory frameworks in place to prevent incidents. Through this process, regulators are able to make informed decisions about the safety of new technologies, operational practices, and system changes.

Historically, established methods such as Bowtie Modelling have served the aviation system well. These approaches are most effective where systems behave in predictable ways, hazards are well understood, and human operators remain central to system control.

The increasing introduction of machine learning, artificial intelligence and higher levels of automation introduces new types of risk. These technologies challenge many of the core assumptions behind traditional assessment techniques, particularly in the following areas:

- Behavioural uncertainty: Developed systems may evolve in ways that are not easily predictable or fully testable in advance.
- Lack of transparency: Complex models can make it difficult to explain or justify decision making. This increases difficulty to identify hidden failure paths.
- Variable outcomes: Some systems do not produce consistent outputs, even when inputs appear identical. Systems may behave differently to the same risk that may occur multiple times.
- Reduced human control: Automation may shift the human role from active control to passive oversight, with implications for situational awareness and response time. While risk assessments currently account for the primary barrier for safety to be a human, furthermore legal and ethical questions are being raised to machine systems taking on the primary safety decision making.

Project SafeTeam aims to research, from a regulatory perspective, the development of safe integration of advanced automation technologies. To achieve this, the CAA conducted reviews with the SafeTeam consortium on whether existing risk assessments remain suitable for the new types of systems we are beginning to see, or whether a different approach is needed. The primary review consisted on bowtie modelling as a known standard for risk identification.

The input for regulatory research for automation applications is vital and aim to be structured around:

- Strengthening regulatory confidence: Regulators must be confident in the safety cases presented. This is including those involving systems that cannot be fully assessed through conventional means.
- Ensuring proportionality: Identifying correct methods of assessment may allow us to focus on the specific characteristics of novel technologies without applying unnecessarily restrictive or inappropriate barriers.
- Supporting transparency and trust: Improved assessment approaches can help address legal and ethical concerns around the explainability of artificial intelligence and support clearer communication with both industry and the public.
- Responding to new forms of risk: Some risks associated with artificial intelligence and automation may only emerge through interaction with complex environments. This calls for assessment methods that allow for dynamic behaviour, such as simulation-based analysis or continuous monitoring.

By reviewing and evolving our approach to risk assessment, we aim to support innovation while maintaining the highest standards of safety. Project SafeTeam aims to contribute to ensuring regulatory frameworks remains both robust and adaptable as the aviation sector continues to transform.

2.2 Bowtie Overview

The Bowtie Model consists of different elements that build up the safety risk picture. The safety risk picture revolves around the hazard (something in, around or part of an organisation or activity which has the potential to cause damage or harm) and the top event (the release or loss of control over a hazard known as the undesired system state).

Consideration is then turned to the threats (a possible direct cause for the top event), consequences (results of the top event directly ending in loss or damage) and the controls (any measure taken which acts against some undesirable force or intention).

The controls can be populated on either side of the model showing:

Left hand side of the model	Right hand side of the model
Preventative measures which eliminate the threat entirely or prevent the threat from causing the top event recovery	Measures which reduce the likelihood of the consequence owing to the top event being "live" or mitigate the severity of the consequence

The Bowtie Model explores the escalation factors (the reasoning to why a control may be defeated or less effective) of all controls allowing the allocation of escalation factor controls. These prevent the escalation factors having an impact on the prevention or recovery controls. Further attributes, such as control effectiveness or criticality can be allocated to the Bowtie Model to evaluate the safety risk picture as part of an effective Safety management System (SMS).

Bowtie assessments produce a visual tool which effectively depicts risk, providing an opportunity to identify and assess the key safety barriers either in place or the ones lacking, between a safety event and an unsafe outcome.

Bowtie models are a key component of Performance Based Regulation (PBR) and support:

- An enhanced, graphic representation of risk.
- A balanced and cross domain risk overview for the whole aviation system between internal and external stakeholders (including third party risks and exposure).
- An increased awareness and understanding of the safety risk leading to the 'Key Risk Areas
- The comprehensive and wide-ranging practical guidance material for safety risk management at an operational and regulatory level.
- An identification of critical risk controls and an assessment of their effectiveness.
- An identification of Safety Performance Indicators (SPIs) to monitor performance of risk control.

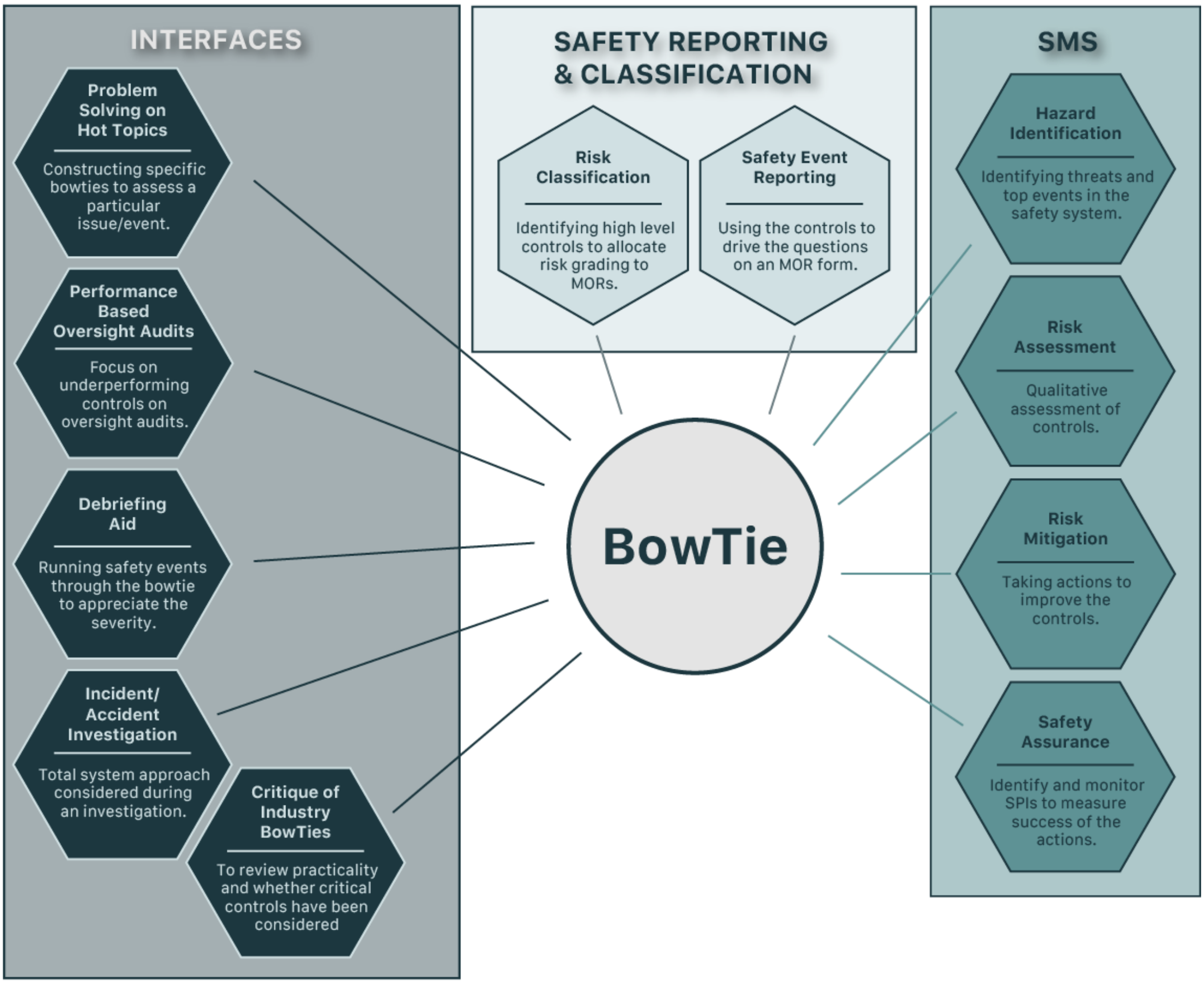

Figure 1: Variety of tasks that can be represented through Bowtie Assessments

Figure 2: Visual representation of Bowtie Model structure

2.2.1 Project SafeTeam – CAA Bowtie Workshop

A two-day workshop was planned to support WP5 (see Agenda in Annex I). The workshop was held on the 16th and 17th of September 2024 at the UK Civil Aviation Authority (CAA) offices at Gatwick Airport, UK and virtually for consortium partners unable to attend in person. It brought together twenty-one consortium members to examine the suitability of the Bowtie methodology for risk assessments in high autonomy driven aviation use cases.

The 1st day focused on establishing a shared understanding of Bowtie analysis, including its core components such as hazards, top events, threats, barriers, consequences, and escalation factors, as well as its practical applications and digital tools. Through interactive exercises and running example scenarios with the consortium, participants gained insight into the method's strengths in illustrating linear cause and effect relationships and barrier logic.

When delving deeper into the question of applying more advanced automation systems to this method of risk assessment, Bowtie analysis was found to be limited in capturing the dynamic, probabilistic, and emergent properties of AI technologies, particularly in relation to learning systems and complex decision-making frameworks.

The 2nd day introduced Human Factors and explored their integration into Bowtie structures. It became apparent that incorporating sociotechnical considerations, especially those involving human and machine interaction, added layers of complexity that the traditional Bowtie framework was not well equipped to accommodate. Group discussions and activities reinforced the conclusion that while Bowtie can serve as a useful tool for conventional hazards, it lacks the flexibility required for systems of systems thinking and the iterative nature of modern AI safety assessments. The workshop therefore concluded that Bowtie is not an appropriate technique for assessing risks associated with higher levels of automation. Alternative approaches, which can better suit to dynamic modelling, may offer a more comprehensive basis for evaluating these use cases.

2.3 Alternative Risk Assessment Frameworks

When assessing the right form of assessment for ML systems in the development stage, regulators typically do not prescribe a specific assessment method but instead evaluate whether the chosen framework effectively addresses critical safety challenges and known gaps related to a policy framework. This ensures that there is a visible demonstration of a robust understanding of the system, its risks, and the mitigations in place that can drop the risk of a hazard down to a level deemed "As Low as Reasonably Practicable" (ALARP).

As expressed through UK CAA CAP 1059: "a risk may be described as ALARP if you have taken all reasonable action to mitigate the risk and the cost (in terms of time, effort and money) of taking further action would be 'grossly disproportionate' to any further reduction in the level of risk". Reducing a risk to ALARP does not mean that the risk has been eliminated as some level of risk remains; however, the organisation has accepted the remaining level of risk.

With regarding to project SafeTeam, while the 3 Use Cases presented in the previous work packages show progression from discovery to testing phases, the technology developed in the project time frame is not yet mature enough for operational deployment, that is aligned with the targeted TRL in an Innovation Action. Accordingly, and following the scope of the research and the targeted maturity level, the risk assessments cannot yet be concluded to determine which methodology addresses most of the key gaps and challenges regarding the ML systems. The following section therefore presents an overview of existing risk assessment frameworks that could be applied at a higher maturity level stage, including an analysis of their applicability. While the 3 Use Cases under research present an opportunity for digital assistance in aviation, each of them might opt for different assessment frameworks.

Continued research, testing, and collaboration with stakeholders are essential to unlock its full potential and ensure its safe integration into aviation operations.

2.3.1 Analysis on Alternative Frameworks

This section aims to discuss some of the alternative methodologies to Bowties and what they can cover in addressing safety assessments. This is up to organisations to determine a "best fit" framework for further discussions at a regulatory level in any future deployment phase.

Assessment Framework	Analysis
Systems Theoretic Accident Model and Process/Analysis (STAMP/STPA)	**What it is:** STAMP/STPA is a technique for the development of a safety assessment that helps anticipate hazardous scenarios namely focussing on causes from Software and Automation based systems. **How it works:** STAMP/STPA focuses on "control actions" and how unsafe controls can cause hazardous events. This is identified differently from linear cause effect chains as seen in bowties. **What it identifies:** STAMP/STPA can help capture emergent behaviour of ML systems based on human-machine interactions and dynamics of a system that may be deemed unsafe or unknown at current stages of maturity. **Noted potential use cases:** Flight deck automation, ATC Automation, Decision making tools concerning early AI adoption.
Functional Resonance Analysis Method (FRAM)	**What it is:** FRAM looks at how different parts of a system, for example, people, machines, or ML models would normally work and how small changes in behaviours can add up and lead to potential threats. FRAM is not just about looking for failures but asks how things usually go right and what happens when that normal behaviour changes. For reviews on ML systems, FRAM can show how changes in data, model decisions, or how people use the system can combine in unexpected ways to create safety risks. **How it works:** FRAM methodology focuses on demonstrating how a process is done through multiple functions and activities. Identifying where ML components are involved in a process can help understand where key challenges may be to address the safety critical responsibility of the ML systems. **What it identifies**: FRAM focuses on variability and does not correlate linear error to a potential failure of an ML system. **Noted potential use cases**: ML tools that help decision making for landing and approach.
Goal Structuring Notation (GSN)	**What it is:** GSN helps build a structured safety argument that a ML system is safe for a specific use case. **How it works:** By providing links to the argument statement with evidence of safe operations and justifications for a method of operations, this provides a rationale for trusting ML systems for more complex decision-based ops. **What it identifies:** GSN can provide clear and robust evidence of safe practices concerning ML systems that can be linked to a safety statement for a particular use case. **Notable potential use cases**: Collision avoidance safety assurances.

Bayesian Belief Network (BBN)	**What it is:** BBN is a model which looks to represent variables through nodes and edge links. Edge links represent probabilistic relationships and each node being a noted variable. **How it works:** A BBN will model "casual relationships" of a system, not just correlations this is useful when evaluating how certain conditions like sensor failure affect outcomes such as incorrect predictions of an ML system. **What it identifies**: BBNs can identify decision making that based on data or incomplete data, allowing for a more dynamic approach as a risk assessment methodology for ML and software-based systems. **Notable potential use cases:** ATC Automation and, Predictive maintenance regimes.

2.3.2 Ethical and Sociotechnical Assessments

While alternative frameworks are available to break down how a ML system can operate safely, it is key to understand the safety critical applications where decision making responsibilities are called into question between human and machine. Therefore, examination of more human factors-based and sociotechnical natures is also utilised.

As project SafeTeam mainly focuses on human- AI teaming development, the UK CAA Horizon Scanning Team conducted further research into frameworks that specifically target this topic. This analysis is purely from a regulatory strategy perspective and is open to dialogue as to further ways organisations can explore capturing safety risk with regard to human factors.

There are several assessment tools and methodologies that can be employed to evaluate the risks, ethical considerations, and sociotechnical impact of these systems.

These assessments are intended to be complementary and were discussed as such by the consortium. They are introduced to provide a human factors perspective, offering additional insight in an area that remains relatively unfamiliar to regulatory authorities for policy framework development. By integrating physical, technical, and sociotechnical considerations, the approach enables a broader and more comprehensive understanding of risk. It is important to note that this is offered as a suggestion to support improved safety practices and culture, rather than as a prescriptive requirement.

Assessment Framework	Analysis
Failure Mode and Effect Analysis (FMEA)	**What it is:** FMEA is a systematic method for evaluating potential failure modes within a given system and determining their causes and effects. It aims to prioritize failures based on their severity, likelihood, and detectability. **What it identifies**: Critical failure modes in ML algorithms, hardware, and human interactions, along with the potential consequences of failures in decision-making, predictions, or actions. It also assesses the likelihood of these failures occurring and the detectability of such failures. **Benefits**: Identifies and mitigates potential failure points in a system before they occur. FMEA also prioritises actions based on the severity and impact of the failure modes. This approach encourages proactive safety and risk management. **SafeTeam Context:** FMEA can be used to identify possible failure scenarios for the developed ML systems for ATC optimisation and unstable approach detection. It allows to understand when the system might make incorrect predictions (e.g., predicting false unstable approach), and develop corrective actions to mitigate risks that may lead to a safety critical event.

Fault Tree Analysis (FTA)	**What it is**: FTA is a top-down, deductive method for analysing the causes of system failures. It focuses on identifying root causes for potential failures by being visualised in a "tree" diagram that branches out from the failure to the contributing factors. **What it identifies:** The root causes of system failures, potential breakdowns in the interaction between ML models, sensors, and other system components, and areas where multiple system failures could converge, leading to a major safety issue. **Benefits:** Helps to understand complex ML systems and their interdependencies, identifying vulnerabilities that could lead to catastrophic failures and therefore safety critical events. FTA also provides a structured approach to tracing causes and effects in safety critical parts of an ML systems. **SafeTeam Context**: FTA could be used to analyse the causes of an ML failure that impacts flight safety. Where detections are given or optimisation for en-route traffic is suggested, FTA allows to trace false readings back to system design, data quality, or any human-machine interaction issues.
Human Error Assessment and Reduction (HEAR)	**What it is:** The HEAR methodology is designed to evaluate human error and its potential consequences in systems involving human machine interactions. It aims to identify how errors can be mitigated through system design, training, and operational processes. **What it identifies:** Potential human errors arising from the use of ML systems, particularly in interactions with pilots, ATC or maintenance staff. **Benefits:** This tool often addresses areas where human decision making might conflict with or depend on automated ML predictions. HEAR can lead to developers to suggest improvements in user interface, training, and system design to minimise human error. The HEAR is human factors cantered and often leads to training development or improvement of user interfaces. **SafeTeam Context:** HEAR can help assess the likelihood of errors in interactions between flight crew and ML systems, ensuring that systems are intuitive, transparent, and easy to understand. A tool for predicting unstable approaches, there may be a conflict of pilot’s decision to agree with a detection. HEAR can help justify the nature of a decision taken by active pilots based on competency and suggest better practices or adjustments for system developers to better integrate and ML tools.

Machine Learning Impact assessment (MLIA)	**What it is:** The MLIA is an assessment tool designed specifically to evaluate the impacts of deploying ML systems in various sectors. It focuses on the potential risks, societal impacts, and compliance with ethical and regulatory standards in the nominated region of operations in the world. **What it identifies**: This approach identifies both the technical and non-technical risks associated with ML deployments, ensuring that the system is aligned with legal, ethical, and regulatory requirements such as ICAO DOC 10151. **Benefits**: MLIA helps users of ML systems to gain a better understanding of the broader implications of introducing ML systems into complex environments, especially in the context to aviation, it allows for more informed decision-making and risk management at a holistic level rather than technical. **SafeTeam Context:** MLIA can help identify potential risks and societal impacts related to the use cases presented and developed for aviation deployment. While other methods that have been analysed here focus aviation context, a larger broader overview of ML system impacts to an industry helps provide assurance to regulators that developers have an understand of the larger safety challenges and complexities the industry operates in. This analysis is to be matured once more key testing is concluded for the use cases.

2.3.3 Regulatory Approach to Assessment of Safety Cases

Eventually, once innovative Use Cases reach a level of maturity to undergo assessment by regulators, the proposed application to a change in operation with ML systems will be reviewed by regulatory SME's (Subject Matter Experts). The UK CAA publication CAP 1801: Assessment of Change Safety Cases provides a structured approach for evaluating safety cases associated with changes to operational systems. These changes can be to incorporate new systems or new responsibilities of personnel. This document is primarily used to ensure that modifications to existing systems maintain safety standards, through a thorough review of potential risks and safety impacts.

For the ongoing project developing ML systems for predicting unstable approaches and optimising en-route ATC operations, CAP 1801 can be leveraged as part of the risk assessment process phase once the outputs of project SafeTeam are at a maturity level to undergo regulatory assessment. CAP 1801 methodology aligns with evaluating safety risks when introducing advanced automation technologies, particularly of those within the remit of Safe Team's research and maturity as seen.

By applying the principles outlined in CAP 1801, the design and implementation phases of the automation system can integrate a robust risk management framework. This ensures that both safety and regulatory compliance are considered throughout the development of these innovative technologies, particularly in areas that directly impact flight safety and air traffic control operations.

In this context, CAP 1801 could support the identification and mitigation of potential hazards linked to the ML system's operational integration, as well as guide the development of safety cases for new functionalities, ensuring their safe and effective deployment.

It is important to note, that while CAP 1801 provides a generic overview to a regulatory approach to safety assessment for safety cases, this document is a guidance and would be used in line with regulations and guidance of the specific areas of application (e.g., Flight Operations). Currently the applications of more ML systems and autonomy are under review by the regulator with the launch of the UK CAA AI strategy. Specific regulatory updates to capability policy are yet to mature and this document will not be able to comment on those implications for Use Cases that have been presented by project SafeTeam.

3 Recommendations

Upon the review and consultation with the consortium for project SafeTeam, the UK CAA can provide high level recommendations for further maturity of the use cases that have been in development.

As the UK CAA's AI strategy continues to evolve, and with the outputs from Project SafeTeam requiring further testing and validation, the recommendations set out here have not been formally reviewed or endorsed by CAA Subject Matter Experts (SMEs) and may be subject to change. While CAA SMEs have provided input into project SafeTeam to support early regulatory thinking on safety-critical elements, it has been concluded that it would be premature to establish a regulatory position at this stage. A formal review by the relevant regulatory experts will take place closer to the point of live operational trials once greater system maturity has been achieved. The recommendations regarding work package 5.1 are as follows:

- Human Factors Research: Continue development to examine how human factors such as pilot confidence, decision-making processes for ATC, and training can influence the ML system’s reliability. Additionally, exploring the relationship between pilot self-assessments of approach stability and ML predictions could provide valuable insights for enhancing trust in the system and improving operational integration. This will ensure future discussions at a regulatory level can be more specific rather. Researching into system operations from a technical level is the next mature phase from a strategy review of the project.

- Alternative Risk Assessment Frameworks: Given the probabilistic and dynamic nature of machine learning systems, alternative approaches to traditional bowtie analysis should be considered. Risk assessment methodologies should be designed to reflect the inherent uncertainties of predictive algorithms, supporting more adaptive and nuanced safety assessments. Work package 5 as a whole aims to give further guidance on various assessment frameworks, relevant CAA frameworks that tackle regulatory complexities and more human factors-based scenarios. Future activity should review the alternative assessment methodologies and discover which ones identify key risks for mitigation activities. This will allow maturity of regulatory conversation for approval of more complex live trials.

Annex

SafeTeam/ CAA Bowtie Workshop Agenda

Location: CAA Aviation House Gatwick or Online (Microsoft Teams)

Date: 16-17th September 2024

Time: 9:00-16:00

Facilitator: UK Civil Aviation Authority

Agenda Items

Day 1: Bowtie Fundamentals

9:00 AM - 9:30 AM: Welcome and Introduction

- Introduction to the workshop objectives and structure.
- Overview of the consortium's goals in AI for aviation.
- Brief participant introductions and expectations.

9:30 AM - 10:30 AM: Introduction to Bowtie Methodology

- Understanding Bowtie fundamentals: Overview of Bowtie methodologies.
- Components: Hazards, top events, threats, consequences, barriers, and escalation factors.
- The Bowtie application importance

10:30 AM - 10:45 AM: Morning Break

10:45 AM - 12:00 PM: Practical Application of Bowtie Analysis

- Case Study: Example of Bowtie application in aviation.
- Interactive Exercise: Designing your own bowtie.

12:00 PM - 1:00 PM: Lunch Break

1:00 PM - 2:30 PM: Applying Bowtie to AI Use Cases in Aviation

- Identifying potential hazards and risks in AI-based aviation use cases.
- Group Activity: Participants review SafeTeam AI use case and apply Bowtie ideologies to risk assessment diagram.
- Discussion: Common threats and barriers in AI, focusing on regulatory compliance.

2:30 PM - 2:45 PM: Afternoon Break

2:45 PM - 3:45 PM: Advanced Bowtie Techniques

- Integrating Bowtie analysis with other risk management tools
- Digital tools and software for Bowtie analysis.

3:45 PM - 4:00 PM: Wrap-up and Preparation for Day 2

- Review of key takeaways from Day 1.

- Overview of the Day 2 agenda
- Any other business

Day 2: Human Factors in AI Use Cases for Aviation

9:00 AM - 9:30 AM: Recap of Day 1 and Introduction to Human Factors

- Quick recap of Bowtie fundamentals discussed on Day 1.
- Introduction to Human Factors: What they are and why they matter.

9:30 AM - 10:30 AM: Human Factors in Aviation Safety

- Overview of Human Factors principles.
- Case Studies: Examples of Human Factors in aviation incidents and their relevance to AI.

10:30 AM - 10:45 AM: Morning Break

10:45 AM - 12:00 PM: Identifying Human Factors in AI Use Cases

- Discussion: Potential Human Factors issues in AI applications in SafeTeam applications.
- Group Activity: Participants analyse a SafeTeam use case to identify potential Human Factors risks.

12:00 PM - 1:00 PM: Lunch Break

1:00 PM - 2:30 PM: Integrating Human Factors into Bowtie Analysis

- Linking Human Factors to Bowtie diagrams: Where and how to include Human Factors considerations for Bowtie risk assessments.

2:30 PM - 2:45 PM: Afternoon Break

2:45 PM - 3:45 PM: Regulatory Considerations and Readiness

- Overview of aviation regulations related to AI and Human Factors.
- Discussion: Assessing technology readiness of SafeTeam use cases and compliance using Bowtie and Human Factors insights.

3:45 PM - 4:00 PM: Final Wrap-up and Next Steps

- Review of the workshop’s key outcomes.
- Discussing next steps for the consortium in applying these methods to their AI use cases and reporting.
- Q&A and Closing Remarks.