Federal Aviation Administration

Roadmap for Artificial Intelligence Safety Assurance

Version I

Roadmap for Artificial Intelligence Safety Assurance

Federal Aviation Administration

Foreword

The recent acceleration in the development of artificial intelligence provides new opportunities to leverage the technology to support a safe aviation system, while posing new risks if not appropriately qualified and used. In the face of these challenges and opportunities, we have developed this roadmap to explain our approach to developing methods to assure the safety of the technology and introduce it for safety. The guiding principles described in the roadmap are fundamental to how we are approaching this new technology, leveraging the extensive safety experience of the aviation community with related technologies such as complex systems, software, and automation.

The development of this roadmap was informed by a series of technical interchanges with industry and other aviation regulators. It lays out a strategy to pursue both the safety of AI and the use of AI for safety. It exists within the larger ecosystem for the federal use of regulation of AI, as addressed in *Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* [1].

The challenge we face for the 21st century is how to deliver increasing levels of service while maintaining the safest, most efficient aerospace system in the world. Technology and innovation present tremendous opportunities and challenges for sustainability of the aerospace system. We look forward to working with the aviation industry, other aviation regulators, and US government agencies in executing and refining this roadmap.

David H. Boulter
Federal Aviation Administration
Associate Administrator
Aviation Safety

[1] President of the United States, Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, issued on Oct. 30, 2023.

Executive Summary

Artificial Intelligence (AI) has generated significant interest in aviation, from offline applications to process control to on-aircraft autonomy. Prior to its utilization in aviation, this technology must demonstrate its safety. However, it presents a new challenge for systems that achieve performance and capability by learning rather than design due to the absence of engineering principles that guide the traditional engineering design process.

The primary purpose of this document is to provide a path to assuring the safety of AI in aircraft and related systems for inflight operations. While considering the importance of the safety assurance of AI, it became apparent that there are significant opportunities to use AI for safety and the scope of the roadmap was expanded to introduce this topic.

The objectives of this roadmap are to establish the guiding principles for assuring the safety of artificial intelligence in aviation and to establish priorities and plans for its safe introduction into aviation. The treatment of the ethical use of AI is outside the scope of this roadmap.

The document introduces a set of principles (valued concepts that are independent of time, resources, and situation) to guide the development of methods for the safety assurance of AI in aircraft and aircraft operations, including:

- **Work Within the Aviation Ecosystem**: Use existing aviation safety requirements. Introduce AI within this structured, disciplined, and risk-managed ecosystem.
- **Focus on Safety Assurance and Safety Enhancements**: Center AI development in aviation on safety of AI and using AI for safety, targeting improvements in safety, and ensuring ethical considerations, particularly those that may affect safety, are addressed.
- **Avoid Personification**: Treat AI as a tool, not a human. Emphasize clear responsibility assignment and avoid human-centric language to maintain a clear understanding of AI's role and limitations in aviation.
- **Differentiate between Learned and Learning AI**[2]: Establish distinct safety assurance methodologies for learned (static) AI and learning (dynamic) AI, understanding the difference their respective operational and safety implications.
- **Take an Incremental Approach**: Implement AI in aviation incrementally, learning and adapting safety assurance methods based on real-world application and experience.
- **Leverage the Safety Continuum**: Utilize the safety continuum, starting with lower-risk applications to gain experience and inform broader applications and safety methods.

[2] Practitioners of AI use a variety of terms to describe these two concepts, and common terms are not universally accepted. This roadmap uses these terms to differentiate a critical characteristic for safety – whether a given algorithm is under the direct control of the designer (learned AI) or if it adapts in the operational environment (learning AI). Some experts refer to these concepts as off-line and on-line learning, which are not adopted here due to potential confusion with the learning while being connected to the Internet.

- **Leverage Industry Consensus Standards**: Adopt industry consensus standards for AI safety assurance in aviation, as appropriate, promoting global harmonization and adapting to technological changes while aligning with the principles in the roadmap.

This roadmap identifies next steps in five areas to enable safety enhancements and the safe use of AI:

- **Collaboration**: Establish ongoing partnerships with industry, government, the International Civil Aviation Organization (ICAO), and other appropriate domestic and international stakeholders to share insights and develop harmonized global AI safety assurance methods, as appropriate.
- **FAA Workforce Readiness**: Enhance knowledge and experience in the Federal Aviation Administration (FAA) workforce to effectively oversee the design and integration of AI technologies in aviation.
- **Assure the Safety of AI**: Adapt and develop safety assurance methods specifically for AI, utilizing conventional frameworks and developing new approaches tailored to AI's unique challenges. Ensure any AI system put into an aircraft is safe.
- **Use AI for Safety**: Leverage AI to improve the fidelity and effectiveness of safety lifecycle processes, from monitoring to system development and testing.
- **Aviation Safety Research**: Conduct focused research to develop and evaluate methods for assuring the safety of AI systems and using AI to enhance overall aviation safety.

This is a living document that the FAA plans to update periodically to both reflect progress in safety assurance and adapt to rapidly evolving AI technology.

Introduction

AI is a machine-based system that can, for a given set of human-defined objectives, make estimations, predictions, recommendations, or decisions influencing real or virtual environments. AI systems use data inputs provided by machines, sensors, or humans to perceive real and virtual environments, conceptualize such perceptions into model inference (through analysis in an automated manner), and use this model inference to formulate options for information or action.[1]

The Research, Engineering, and Development Advisory Committee (REDAC) in 2022 recommended that FAA establish a roadmap for AI to alleviate uncertainty in the industry with a clear direction of how this innovative technology can be used in airborne applications. The industry lacks a method for the safety assurance of AI. Safety assurance encompasses all the activities and artifacts to provide sufficient justification that the risks are acceptable.

The FAA held a series of technical interchange meetings to listen to a variety of industry opinions and priorities throughout 2023 and 2024. Aircraft and avionics designers are seeking to incorporate AI technology into their designs safely, with an initial focus on the implementations of complex algorithms and image processing tasks. System designers are faced with developing new methods to assure the safety of their system using AI and coordinate those methods with the FAA. This provides an agile approach to enable innovation, but the uncertainty regarding acceptable methods creates barrier to market adoption.

Through the technical interchanges, the broader potential for AI in aviation became apparent. The aviation community has invested in data analytics to monitor data from real-world operations and identify precursor events and conditions to then adopt risk mitigations before an accident or incident occurs. AI has already been used in limited areas for curating data for analysis and using this data to estimate risk. For example, the latest standards for the Aircraft Collision Avoidance System (ACAS) replaced scenario-based algorithm design of the Traffic Collision Avoidance System (TCAS v7.1 and earlier) with weighted risk models developed through machine learning (ACAS X). There is opportunity for AI to provide an additional perspective when analyzing system risks and mitigations. Recognizing these opportunities, this roadmap also addresses the use of AI as a tool to improve safety assurance.

National Context for Safe and Secure AI

This roadmap is being developed within a broader, evolving national framework established for the safe, secure, and trustworthy development and use of AI, including in appropriate cases its adoption and regulation across the federal government.

On October 30, 2023, President Biden issued Executive Order 14110, *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, which indicates the "highest urgency on governing the development and use of AI safely and responsibly" and therefore "compels the United States to lead in this moment for the sake of our security, economy, and society."

This latest Executive Order supplements Executive Orders 13859[3], *Maintain American Leadership in Artificial Intelligence,* and 13960[4], *Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government,* recognizing that "continued American leadership in AI is of paramount importance to maintaining the economic and national security of the United States," but "the ongoing adoption and acceptance of AI will depend significantly on public trust." This roadmap addresses a narrow but important intersection between national policies relating to AI and the aviation regulatory context for safety. The aviation community should participate, and make use of, standards and practices in the general technical community, and leverage those developments or develop aviation-specific standards to fit within the unique regulatory regime of aviation, as appropriate.

In addition to this *Roadmap for AI Safety Assurance* that focuses on aircraft and related systems, FAA is developing an *FAA Artificial Intelligence Strategy* to address the use of AI for a wider variety of applications and objectives. The FAA engages in diverse activities as it serves in the roles of an aviation and commercial space regulator and air navigation service provider and engages in activities necessary to support those missions spanning recruiting, training, facilities, maintenance, and more. AI has the potential to add tremendous value in identifying safety risks and increasing operational and business process efficiency. Because of this diversity, AI use cases are driven by mission need. The FAA learns from these use cases, and best practices from the use cases are centralized and shared.

---

[3] President of the United States, Executive Order 13859: Maintaining American Leadership in Artificial Intelligence, issued on Feb. 11, 2019.

[4] Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 85 Fed. Reg. 78939 (Dec. 10, 2023).

Safety of AI

The application of AI faces the challenge of determining how to assure the safe operation of an AI system which was not traditionally designed, but instead learned how to perform its task. Conventional aviation safety assurance techniques assume that a designer can explain every aspect of the system design, but such explanations are not readily extendable to AI. The aviation regulatory system is also reliant on the system designer assuming responsibility for their design, so that an AI developer is responsible for the AI system performance and characteristics and must be able to provide assurance that it is acceptable.

Guiding Principles for Safety Assurance

The following principles will guide the development of requirements and methods for the safety assurance of AI in aircraft and aircraft operations. These principles were informed through the technical outreach to industry and are based on aviation experience and best practices of safety, standards, and oversight and history of collaboration with industry on the introduction of new technologies.

- Work Within the Aviation Ecosystem
- Focus on Safety Assurance and Safety Enhancements
- Avoid Personification
- Differentiate Between Learned and Learning AI
- Take an Incremental Approach
- Leverage the Safety Continuum
- Leverage Industry Consensus Standards

Work within the Aviation Ecosystem

Principle: Use existing civil aviation safety requirements, processes, and methods to introduce AI, except where they are found to be inadequate.

The general field of AI has developed terms and concepts within that field that are distinct from the terms and concepts in aviation. It is paramount to recognize that the tremendous safety record in aviation has been achieved through assignment of responsibilities, a disciplined systems engineering process, thorough testing and analysis, and management of any risks while in-service. Aviation is unique in this regard, and it is appropriate to place AI within the aviation context rather than aviation within an AI context.

AI is an enabling technology that interfaces with several significant safety assurance characteristics, such as the key factors shown in Figure 1.

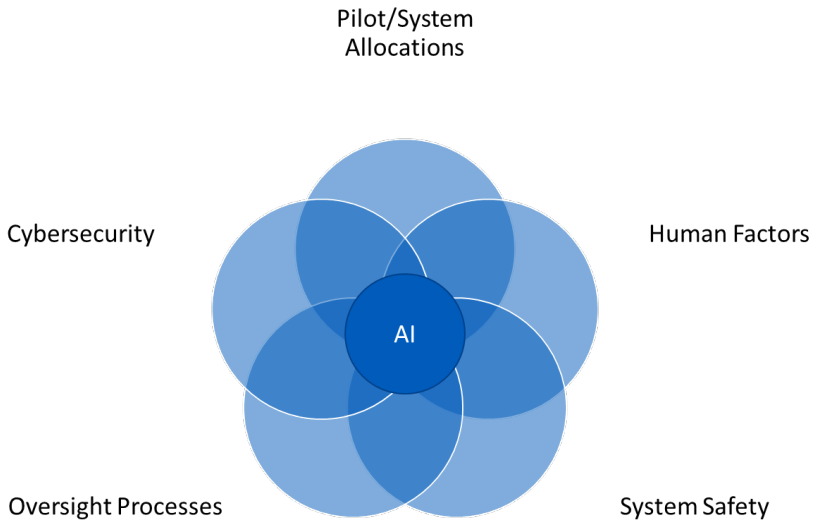

Figure 1 AI within the Aviation Ecosystem

This principle leads to several key conclusions. First, pre-existing requirements and methods for evaluating the safety implications of any system are addressed in FAA regulations, FAA policies, and FAA-accepted industry standards for system safety assessments. The existing regulations for aircraft systems and equipment (e.g., subpart F for 14 Code of Federal Regulation (CFR) parts 23, 25, 27, 29) are performance-based regulations that ensure any system performs its intended function and does not introduce an unacceptable hazard. Industry standards provide appropriate methods for determining the safety criticality of a particular subsystem or algorithm and designate a design assurance level.

Additionally, aviation regulations and guidance already address automation and the role of the pilot and other crewmembers. While AI is frequently considered a tool to develop or provide more advanced levels of automation, there is already considerable experience in human factors design principles, evaluation, and training in the aviation context that should be applied. Regulations and guidance continue to improve with experience and as new automation capabilities become feasible.

Issues associated with human-automation integration should be addressed as human factors and automation issues, and not as AI issues, unless the use of AI introduces risks that might not be present with other types of automation.

Focus on Safety Assurance and Safety Enhancements

Principle: Address safety within the regulatory framework.

This roadmap is intended only to address safety assurance of AI in aviation, and selected applications where the introduction of AI has the potential to improve safety. This scope is aligned to FAA authority to promote safe flight of civil aircraft.

This roadmap does not address societal aspects with the use of AI which are outside of the FAA's authority. There may be some common considerations where the safety of an AI application can be impacted by biases in training data, such as a pilot-health monitoring system that works more effectively for some ethnicities than others. These issues are addressed within the scope of safety assurance, in that the designer of such a system must show that the system performs its function across the entire community of pilots without unfair advantages.

Avoid Personification

Principle: Treat AI as an algorithm or computer, not as a human.

With AI technologies, it is common for developers and the media to portray AI as machines that interact like humans. People frequently refer to simulated assistants by name, such as Siri or Alexa, personifying the software that is responding to prompts. While this may serve as an effective marketing tool, it is not conducive to assuring safe operation of these complex systems in aviation.

Personifying AI can erode safety by creating ambiguity on the assignment of responsibility for safe operation. As certain operations, traditionally accomplished by people, are instead accomplished by automation, responsibility shifts from the human operator to the system designer. The system designer must delineate the responsibilities that are assigned to human beings as compared to the requirements that are assigned to systems and tools and must do so in a manner consistent with applicable aviation regulatory requirements and international standards. The responsibility for systems to meet their requirements rests with the system designer and AI developer, not the AI itself.

Aviation experience with complex automation and human factors has highlighted the importance for the human operator, the pilot, to have a solid understanding of the modes, operation, and malfunction of the automation. Personifying AI applications suggests that they have human-like capabilities and potentially unexpected behavior. This contributes to the false impression that the modes, operation, and malfunction would be that of a human, and that AI is an entity which can be responsible. While the normal operation may be intended to automate something that can be performed by a human, the modes and malfunctions are notably different. The safety of future operations depends on the pilot understanding that a system containing AI is just a system and not another human with whom they can reason or negotiate.

For these reasons, this roadmap avoids the use of human-centric language when referring to AI. For example, AI cannot be a part of crew-resource management (CRM) but can affect crew responsibilities. AI cannot be a copilot but can perform autopilot functions and affect how a pilot

performs their duties. AI may have a degree of control authority over specific flight functions but is not accountable for anything; the designer and maintainer of the AI are accountable unless that responsibility has been allocated elsewhere by applicable law. This roadmap also refers to the safety assurance of AI as a responsibility of the designer, unless allocated elsewhere by applicable law, and avoids mentioning trust in AI.

Differentiate Between Learned AI and Learning AI

Principle: Distinguish between the safety assurance methodology for learned AI implementations and learning AI implementations.

A key characteristic of AI in the context of safety assurance is that AI systems often gain significant capabilities through learning processes, in addition to what is explicitly programmed or designed. The use of AI can be further divided between implementations of fixed and deterministic algorithms, and implementations that continue to change through learning in the operational environment.

- **Learned AI Implementation**: AI implementations that are static in the operating environment. These implementations encompass Machine Learning (ML) algorithms that are developed through offline training, and the specific learned AI is subjected to design time safety assurance processes prior to its use in an operational context.
- **Learning AI Implementation**: AI implementations that are dynamic in the operating environment. These implementations involve the incorporation of learning mechanisms into the AI system itself, so it is not possible to qualify each version: instead, the safety assurance strategy will need to address the learning itself.

The safety assurance for a learned AI implementation can be performed as part of the system design and validation. Once completed, the AI implementation is accepted, and the in-service monitoring of the AI implementation is part of the continuous operational safety (COS) program for the aircraft. A system designer may record in-service operational data to further train the AI implementation and deploy an updated version, but each new version is subjected to safety assurance. The developer of a learned AI implementation must consider the range of operating conditions that will be encountered throughout the product lifecycle as part of the initial safety assurance. Changes in that environment to depart from the range of operating conditions initially considered may impact the performance or functionality of the AI and would be addressed under the COS program in the same way they would be for traditionally designed systems.

Learned AI implementations with frequent offline updates pose an opportunity and a challenge. The collection of in-service data provides a means to monitor how well the deployed product is working to quickly detect in-service deficiencies and inform how urgently they should be corrected. The frequency of updating the version is constrained by the need to provide safety assurance of the update. This is an important consideration in the development of safety

assurance methods, which should describe sufficient and proportional strategies to assure the safety of a revision.

In contrast, a system that continues to learn in the operating environment must build its safety assurance into the operating environment or include safety assurance as part of the process of learning. Given the difficulty of this task, the learning systems are posing safety concerns if assurance are not properly addressed, and therefore are seen as ongoing research topics. Learning AI implementations may adapt in a manner that degrades performance, ultimately weakening their original safety profile. Cases in which a system learns anomalous, ungeneralizable, or inaccurate information will require new mitigation strategies. Learning systems may necessitate new regulations to assure the continued safety of the evolving system, as for active monitoring of performance or recurrent certification.

Differentiating between these two classes of AI implementations is important in developing strategies to assure their safety, and in educating users on the expected modes, operation, and malfunction of the automation.

Take an Incremental Approach

Principle: Adopt an incremental approach to the introduction of AI in aviation safety assurance, updating with experience.

With evolving AI technology, the types of evidence necessary to assure safety are not predetermined, largely due to the diverse nature of the AI discipline and the wide range of applications. Given the lack of experience and uniformity in AI safety assurance, a single guideline defining one type of evidence that is necessary for all AI applications might not be feasible, usable, or appropriate. Instead, FAA promotes an incremental approach where individual companies and FAA can gain experience on specific projects.

The FAA will host technical exchanges for the community to share experiences and lessons-learned regarding assurance concepts and methods in a non-binding and open forum. The experience from real projects keeps the safety assurance criteria scaled to real-world applications. For example, FAA is working with companies on avionics applications of AI in aircraft, using project-specific issue papers that can be tailored to each project.

Leverage the Safety Continuum

Principle: Gain experience in lower-risk applications.

The Safety Continuum refers to the spectrum of risk levels acceptable to society. The highest level of safety is expected for scheduled passenger service, transitioning to a lower safety threshold for research experimental flights and drone operations. We can gain experience with AI in experimental aircraft, without trying to provide the assurance that would be required for that AI to be used in scheduled passenger air carrier operations. Small uncrewed aircraft also provide

ideal vehicles and operations to gain early experience which can be used to further inform future versions of this roadmap. The experience that is gained can inform safety assurance methods relevant to other applications and safety objectives.

The Safety Continuum can be particularly valuable when used in conjunction with the Design Assurance Levels (DALs) that indicate the safety criticality of a function within the aircraft and operational context. DALs are assigned through the system safety assessment process based on the potential risks of failure or malfunction. DALs indicate the necessary stringency and scrutiny to assure an avionics subsystem from a safety perspective. This measurement depends on the failure conditions of those components and the risks associated with each type of failure. DALs[5] are divided into five different levels, which correlate to the five hazard classifications in transport category airplanes: catastrophic, hazardous, major, minor, and no safety effect. For aircraft operations at the other end of the Safety Continuum, such as small, uncrewed aircraft, a design assurance level of D may be appropriate for systems that could lead to loss of control. For aircraft operations it may be appropriate to allow aircraft functions that have a failure classification of Minor as appropriate to implement AI system functions taking an incremental approach to allowing higher critical functions over time. We can build our experience on lower criticality functions before applying them to higher criticality functions. This enables the aviation community to gain experience with systems that would demand a very high level of rigor by focusing on the lower levels first and building up the DALs based on experience.

Another motivation for leveraging the continuum is to appropriately scale the assurance expectations to the risk. The safety assurance method must be sufficient for the risks involved but should not impose undue burdens which increase cost and present a barrier to gaining experience at lower criticality levels. This experience is necessary to inform how to address safety assurance at higher levels.

Leverage Industry Consensus Standards

Principle: Use industry consensus standards that provide acceptable means of compliance for aviation safety assurance, as appropriate.

In developing and applying safety assurance methods to AI, FAA will participate in developing industry consensus standards and utilize them as they are completed and applicable to aviation safety assurance, as appropriate. Industry standards can promote global regulatory harmonization if widely adopted and can be agile in adapting to the changing technology. The aviation community should consider the standards developed for other applications outside of aviation and develop aviation-specific standards that reflect the principles outlined in this roadmap, as appropriate.

---

[5] DO-178C: Software Considerations in Airborne Systems and Equipment Certification, RTCA, Inc., December 2011.

Existing Aviation Applications

AI is already being introduced into various aspects of an aircraft lifecycle and operation. AI assurance guidelines must address the breadth of AI applications and resulting risks.

- **Design**: AI may be applied in the design activities and artifact generation processes. AI can be used to generate other AI models, as a tool in model-based software development. AI classification models can be used to perform labeling of input data where similar data are organized into a group, and the collective characteristic of such group is used as label for the data in that group. Distributions of the input data can be used in specification of requirements and corresponding response in design. Documentation supporting the certification processes that currently exist can also be auto generated through similar applications.
- **Production**: Robots are actively utilized in manufacturing and vision systems are in place as part of a quality control (QC) framework. AI can extend the QC framework through in-depth analysis of production faults and line failures to identify outliers and statistical distributions on parts tolerances.
- **Operations**: AI-based functions can support flight operations, including dispatch, training and training simulators, and scenario prediction. AI applications can also aid in document generation such as training manuals and Safety Risk Management (SRM) support.
- **Maintenance**: Many maintenance concepts already employ a digital twin, whereby a model of the component is represented in a virtual context that receives real inputs. When the performance of the real system varies from the digital twin, a call for preventive maintenance can be made. An AI application can conduct a thorough comparison, considering multiple factors beyond predefined triggers, particularly when the system is allowed to learn over time.
- **Retirement Phase**: AI can enhance various processes. It can aid in predictive maintenance, optimizing the decommissioning and recycling of aircraft parts, and managing assets efficiently. AI also aids in analyzing historical operational data for future design insights, ensuring compliance with regulatory requirements, and improving safety through detailed analysis. Additionally, decommissioned aircraft and components find use in AI-driven pilot training simulators and educational tools, contributing to the training of aviation professionals.

Continued operational safety programs rely on safety metrics and key performance indicators that may benefit from an AI application to identify trends that are unrecognizable by a human observer without additional analysis. Such safety monitoring systems would serve in real time, as appropriate, to provide a safety status that can be used within, or for oversight over, an operation.

Application of Principles to Aircraft Certification Safety Assurance of Learned AI Implementations

This section illustrates how the guiding principles affect the safety assurance strategy for on-aircraft learned AI implementations. An aircraft with a standard airworthiness certificate must be

certified again to incorporate AI. The FAA has relevant regulations for systems and equipment, which require that the aircraft designer show that the system or equipment performs its intended function and that any failures are mitigated to an acceptable level. These performance-based regulations were used when the software was introduced into the type design and are expected to be adequate for the safety of learned AI implementations. Manufacturers must comply with relevant performance-based regulations and demonstrate through rigorous testing and analysis that the software meets specified safety and performance standards. Continuous oversight and post-certification evaluations are conducted to ensure compliance and address any emerging risks or changes in technology.

In the context of system design and safety assurance, Figure 2 shows the engineering design process that starts with aircraft requirements, followed by system requirements, item requirements, item design, item verification, system verification, and aircraft verification. The questions that must be addressed are (1) What is unique to AI/ML that cannot be addressed with existing regulations and guidance? and (2) Should existing regulations and guidance be modified to address AI/ML-related items or will new regulations and guidance or special conditions need to be developed for this purpose?

The unique characteristic of AI is that the implementation is learned, rather than designed. The design process starts with the aircraft requirements, from which system and item requirements are developed based on the aircraft system and item designs. In traditional systems engineering, the designer can derive the requirements at the lower level and validate those requirements to the next higher level through rational explanation and traceability. When AI is used to develop a learned AI implementation, those links are broken; the designer cannot derive the requirements that directly describe the AI implementation and cannot validate them by showing that they provide coverage to the next higher level. The lower-level requirements are the learned AI implementation (an algorithm). For example, it may be a large neural network (NN) with thousands of weights performing basic arithmetic functions that cannot be traced to higher level requirements.

All AI algorithms are implemented in hardware and software, which is reflected in the item design. Hardware implementations include the design of a specialized chip, or the configuration of a specialized chip to function according to an AI algorithm. Conventional engineering methods of development for the item hardware and software itself, shown in Figure 2, can be applied but may need to be adapted to reflect that they can only trace the requirements to the definition of the AI implementation itself, and not the higher-level requirements.

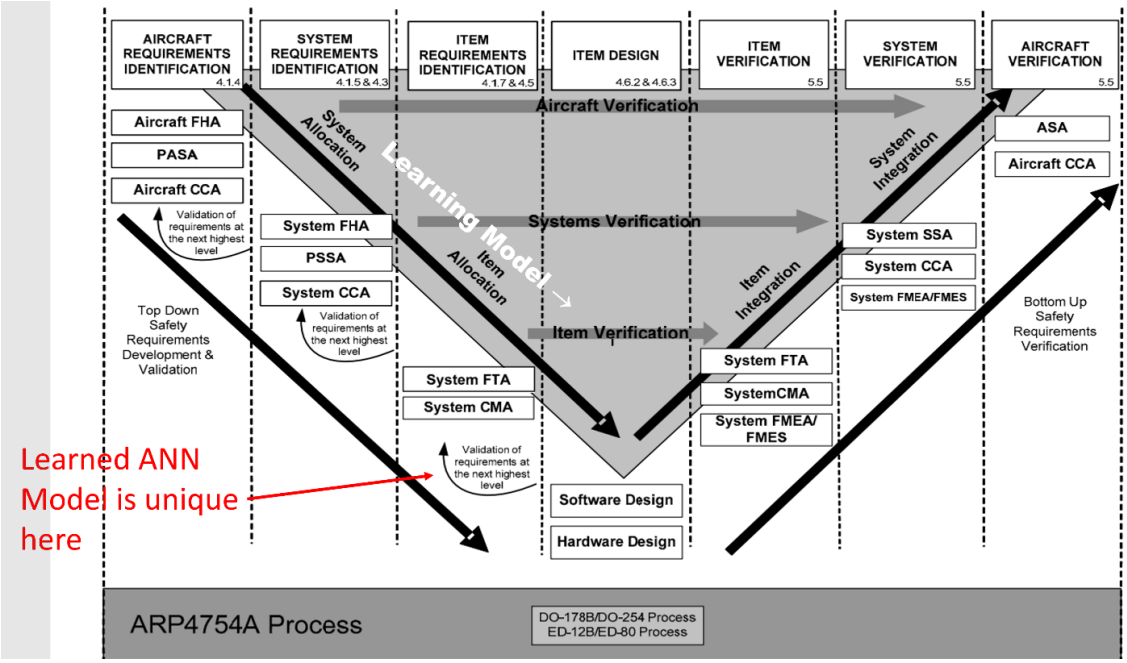

Figure 2 Development of AI Algorithm May Require Additional Adaptation (at Item Level)

ARP4754A[6] and ARP4761[7] describe the system safety methodology to determine a DAL for a given item. The DAL indicates the extent to which that item can pose a hazard to aviation and differentiates the degree of rigor that must be applied when assuring the safety of the item. This method is already used for software and complex hardware and can be applied to learned AI implementations to provide a starting point for addressing the safety assurance of AI. However, existing software and complex hardware guidance is not adequate to address learned AI implementations. For example, the guidance for the traceability of software or complex hardware cannot be extended to an AI system, where there is not any additional information outside of the input-output context that can explain the internal dynamics of an AI system developed through data.

Resolving this issue will necessitate new assurance methods and may include establishing criteria for choosing suitable formal methods and testing tools that are compatible with the stochastic nature of AI and providing a framework for the traceability and validation of AI-based software to ensure it meets the high safety standards required in aviation. The properties of training datasets and the nature of NN weights are often inscrutable to human review or may deviate significantly from the properties typically analyzed in design assurance. Therefore, validation and verification methods in DO-254[8] fall short when applied to AI. Another challenge is

[6] ARP4754A: Guidelines for Development of Civil Aircraft and Systems, SAE International, December 2010.

[7] ARP4761: Guidelines and Methods for Conducting the Safety Assessment Process on Civil Airborne Systems and Equipment, SAE International, December 1996.

[8] RTCA DO-254: Design Assurance Guidance for Airborne Electronic Hardware, RTCA, Inc., April 2000

performing the Failure Mode and Effects Analysis (FMEA), given that the anomalous behavior of the AI implementation may not be identified, or even identifiable.

This may be addressed through extensive stress testing for AI systems at item level (for V&V of the functionality), and additional test at both the system and aircraft levels (for failure analysis). This procedure begins at the item level and includes critical software and hardware testing, integral to a thorough and complete safety evaluation.

AI for Safety

Existing FAA regulations and guidance assign responsibility and define expectations throughout the safety lifecycle as depicted in Figure 3. Those documents did not contemplate and do not specifically address the use of AI. However, since those documents are general in nature, they can be applied to the use of AI. AI-generated content can be inaccurate, misleading, or entirely false, or may incorporate copyrighted material or other intellectual property for which an appropriate license or other necessary permission has not been obtained. Companies are responsible for their products, their compliance, and the demonstration of compliance, regardless of how those artifacts are created.

The application of AI in the tools and processes for assuring the safety of aircraft and operations may introduce significant improvements. In considering the use of AI to improve safety as a research thrust in the long-term strategy, the objective is to identify those aspects of the safety lifecycle that are most likely to benefit from the emerging AI technologies and focus FAA attention in those areas. Figure 3 shows various example applications of AI in the lifecycle of an aircraft-by-aircraft designers, manufacturers, operators, and the regulator.

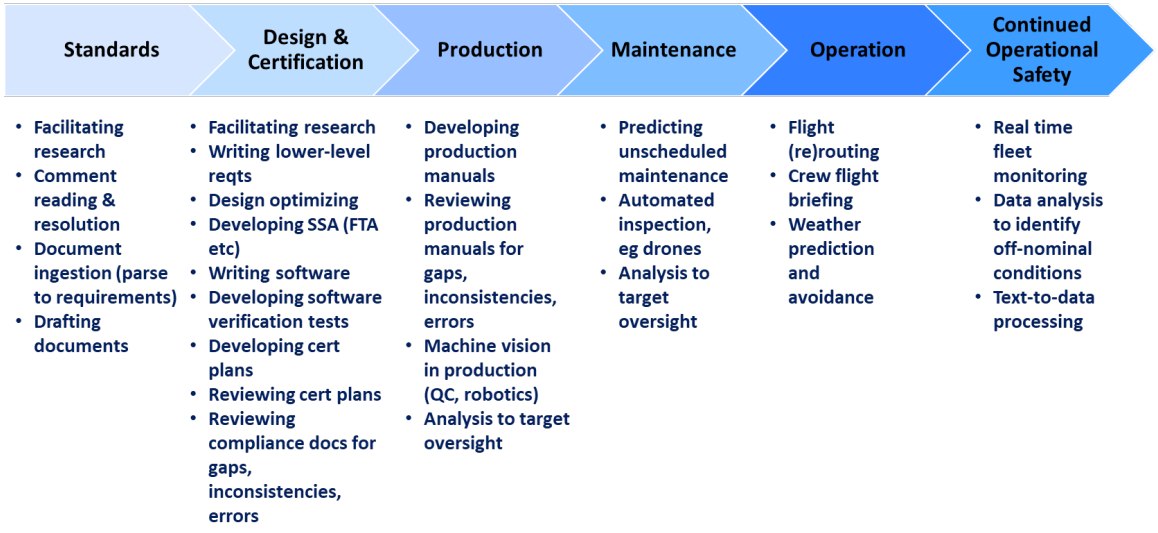

Figure 3 Example Use Cases for AI in Aircraft Safety Lifecycle

AI, in the context of data analytics for large data sets, can identify details that would have been lost or missed in traditional analysis methods. This is done through data mining where data representing special cases or having off-nominal conditions can be discovered with data clustering techniques, and data with similar characteristics are grouped into subsets that are analyzed separately to both avoid inaccuracies and identify hidden information that can reveal outliers, unexpected scenarios, or oversights.

AI may also improve safety in other contexts, such as in improving the quality of pilot training, qualification and currency, or air traffic management and associated services. The use of AI to improve the efficient use of airspace is outside the scope of this roadmap.

AI Safety Assurance Roadmap

AI is a rapidly emerging set of technologies that demands a flexible strategy that can adapt as experience is gained and capabilities expand. Through that evolution, the principles in previous sections will guide the adoption of specific products and decisions.

This roadmap identifies **key actions** to enable the safety enhancements and safe use of AI:

- Collaboration
- FAA Workforce Readiness
- Assuring the Safety of AI
- Use of AI in Safety Lifecycle
- Aviation Safety Research

These key actions are addressed below, but do not dictate specifics in the framework of certification. Section 6 provides the anticipated milestones along the road to AI safety assurance.

Collaboration

Ongoing collaboration is essential to sharing experience and insights, identifying new opportunities and risks, avoiding redundant activities, and keeping pace with the broader technical community. The FAA will continue to convene aviation technical interchange meetings that are open to industry and participate in the development of industry standards. The FAA is also working with foreign civil aviation authorities to share ideas and experience with a goal of harmonizing AI safety assurance methods.

The FAA collaborates within the US government, particularly with other organizations with aviation or other transportation responsibility. This includes the National Aeronautics and Space Administration (NASA), the Department of Defense (DOD), the modal administrations within the Department of Transportation, and the Highly Automated System Safety Center of Excellence.

The aviation community should also participate within the broader AI community, particularly in the development of standards that may have utility in aviation. For example, National Institute of Standards and Technology (NIST) has developed a risk management framework and is developing additional standards concerning the broader scope of general use of AI specified by the Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence issued on October 2023.

FAA Workforce Readiness

The FAA's ability to oversee the safe introduction of a new technology is dependent on factors that include the knowledge and experience of the workforce. Because AI is a new and rapidly

expanding field, FAA cannot rely on its conventional aviation experience. The FAA has hired a Chief Scientific and Technical Advisor for AI/ML, along with several certification specialists with AI experience. Additionally, a Community of Practice for FAA staff will start in 2024 to foster continuous learning and provide peer support to specialists facing AI-related safety assurance projects. A companion series of technical interchange meetings with industry have begun and will continue.

With respect to formal training, FAA continues to leverage external courses for engineers who initially need training. The FAA will leverage training development for government and evaluate if specific training for safety assurance within FAA becomes necessary.

Assuring the Safety of AI

Motivated by the guiding principle to reuse the existing aviation safety oversight framework, FAA expects to apply the conventional systems safety, human factors, and software standards to systems that incorporate AI, to the greatest extent possible. An important consideration is the application of artificial neural network (ANN) models, which are inspired by the understanding of the structure of biological brains, using interconnected nodes to process and transmit information in complex patterns. The validation of a learned AI system must be addressed by supplemental guidance: How can someone assure that a learned ANN model performs as expected, and that any unexpected deviations are mitigated in the integrated system?

A conceptual framework for addressing this question could be seen in the form of Overarching Properties (OPs)[9]. These are generic concepts for assuring system safety[10] that can be used as a strategy to explore alternative methods of assuring software based on assurance cases. The conceptual framework of the OPs is developed in three parts: intent, correctness, and innocuity. From this framework, well-defined arguments can be constructed to support establishing these three properties with a degree of rigor commensurate with the risk consequence of the application.

The development of methods of compliance for learned AI systems can be broken down into a two-step strategy. The first step is a general engineering view of a system developed through Machine Learning as a model, which is just a mathematical function. The second is the implementation of the model, which is entirely traditional. The machine learning process is illustrated in Figure 4.

---

[9] C. Michael Holloway, The Overarching Properties & Overarching Properties Related Arguments, NASA, July 2022

[10] This research work was conducted by FAA in collaboration with NASA and the industry.

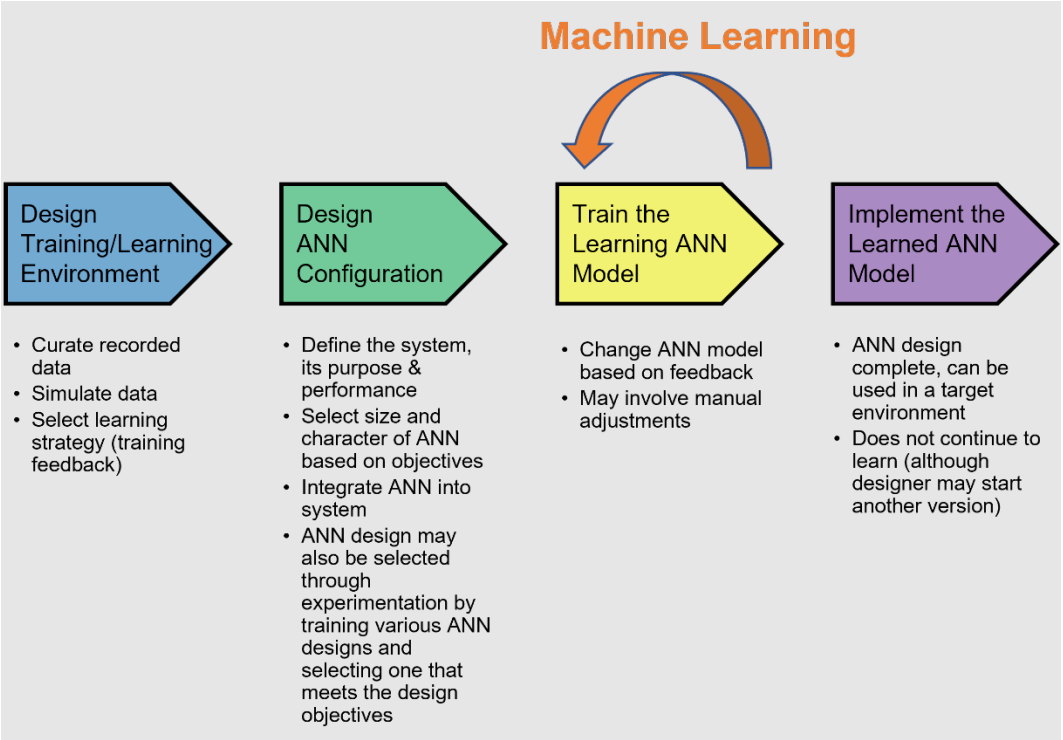

Figure 4 Machine Learning Process

The strategy will be utilized through a combination of experience, consensus standards, guidance material, and research to address learned AI systems.

- **Experience:** Based on the safety continuum and incremental approach outlined in section 3, FAA encourages companies to *initiate projects with lower criticality AI systems* to enable the industry and FAA to gain experience with developing, integrating, testing, and assuring AI. Project-specific issue papers provide a flexible means to develop tailored criteria to a specific project, while the final special conditions document the resultant criteria to inform future projects and common criteria, as appropriate. Companies are encouraged to work with FAA early in their product development and involve FAA to build common experiences upon which standard methods of compliance can be established.
- **Consensus Standards:** The FAA is also supporting the development of acceptable *voluntary consensus standards* for the means to provide safety assurance of AI. This includes the application, and possible extension, of existing standards for systems (SAE ARP/4754 and RTCA/DO-254), software (RTCA DO-178C/EUROCAE ED-12), software tool qualification (RTCA DO-330/EUROCAE ED-215), and model-based design (RTCA/DO-331/EUROCAE ED-218). The FAA is also supporting the development of new standards for AI in SAE G-34 and anticipates additional specific standards may be appropriate for different types of AI learning. There is also considerable development on safety and trust in AI outside of aviation (e.g., NIST),

and FAA will participate in and leverage the use of those standards to the extent that they can be integrated into the aviation ecosystem.

- **Guidance Material:** The FAA will develop guidance material based on the experience and consensus standards described above, as appropriate. As such, it is expected to follow an iterative and progressive approach, starting with guidance for lower-criticality applications and culminating in higher safety-critical aircraft systems. The timing of the guidance will be dictated by the pace of experience and consensus standards and continued technological developments. The initial guidance will address the use of pre-existing guidance and address the unique aspects of low criticality learned ANN models.
- **Research:** The FAA will use the technical interchange forum to promote the sharing of research throughout the aviation community. The FAA is funding research into potential means of compliance, spanning from low-criticality applications to safety-critical systems. The research is described in section 5.4 of this roadmap.

Safety assurance methods for learning AI is a lower priority. Further research is needed before focused activities can begin for such systems.

The FAA is also pursuing global harmonization with other civil aviation authorities as appropriate and is working directly with leading AI aviation industry applicants to identify common criteria for safety assurance of AI and address differences as they are discovered.

Harmonized safety assurance methods for high criticality learned AI systems might be possible in three to five years with focused and deliberate research and collaboration between industry and government. From the guiding principle of taking an incremental approach, there are specific AI applications from the aviation industry that can provide a clearer understanding of what is involved in the safety assurance process. It is reasonable to expect consideration of three specific scenarios commonly seen in the aviation community where AI is being used: (1) AI as an efficient implementation of the solution to a mathematical equation (or set of equations) that might be unsolvable or too tedious to solve in real-time, (2) AI as a monitoring tool to recognize characteristics of an image, and (3) AI as a perception tool to understand the surrounding workspace.

From the guiding principle of using existing regulations as much as possible, there are several objectives of studying the three specific AI applications above. One is to identify the gaps in the existing safety assurance process that is in place for software design and implementation where they cannot handle AI systems. Next is to provide specific instances of how each of the three AI applications is addressed. Last is to present the preliminary findings to the Community of Practice and encourage similar applications that can lead to the generalization of the findings into regulatory standards.

The safety assurance of learning AI systems is likely to remain in the discovery readiness level for more than three years. The challenges in assuring the safety of a system that continues to change while in service are extensive and are a lower priority in the short term as described above. The research thrust of using AI to assure and improve the safety of all systems in aviation, is a primary direction for research aimed at demonstrating feasibility in three to six years, as illustrated in Figure 5.

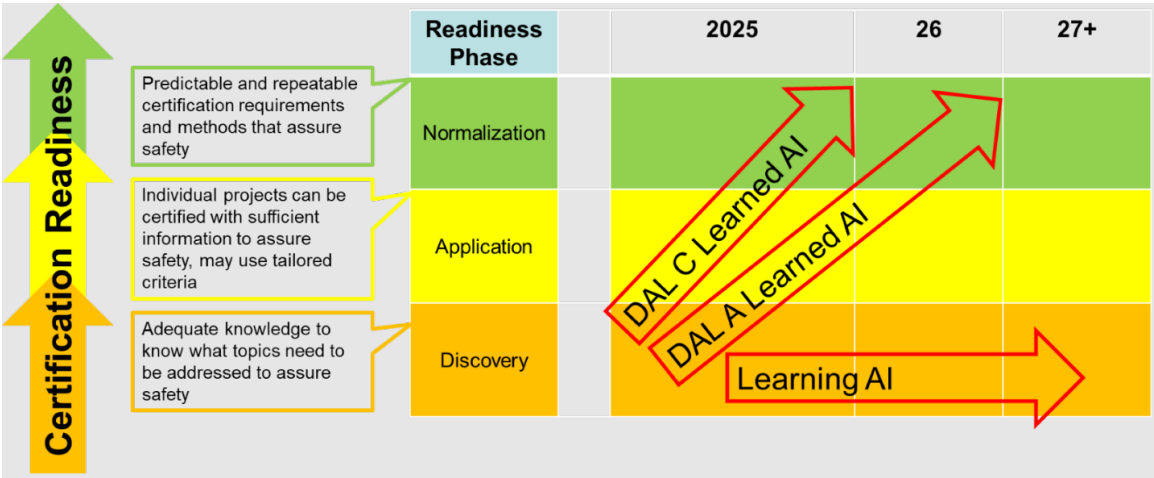

Figure 5 Certification Readiness

Use of AI in the Safety Lifecycle

The aviation community has achieved tremendous advances in safety by monitoring for and mitigating precursor events that could contribute to an accident or incident. The highest initial priority for the use of AI in the safety lifecycle is also the most mature: using AI to curate, process, and analyze data to identify precursors, anomalous events, and risk patterns. The FAA has plans to test and publish guidance on the use of AI in systems monitoring.

A second priority is the potential use of AI to improve the fidelity and completeness of software development, software testing, and requirements traceability. LLM, trained on the rules and expectations of industry standards for these activities, may provide robust tools to automate those activities with a higher degree of assurance than when they are accomplished manually. However, it is possible for the LLM to generate spurious or hallucinatory material or fail to comply with some criteria. Recognizing that the existing aviation systems are vulnerable to human error in these activities, research into the potential of AI for these uses is a high priority.

In the longer term, it may also be possible for AI to automate the review of compliance reports and data, including system safety assessments. Modern aircraft are far too complex for a single human to read and recall all aspects of a single aircraft's design, and aircraft safety is dependent on a team of people with a variety of expertise working together effectively. As LLM AI matures, it may become possible for a single instance of an AI to review all material relating to an aircraft

and identify gaps or discrepancies in the associated compliance and safety reports. This is an important area for research, paced by the development and applications of LLM.

Aviation Safety Research

The two fundamental research objectives are consistent with the scope of this roadmap: (1) Safety of AI: To identify and evaluate the effectiveness of methods to assure the safety of AI systems, and (2) AI for Safety: To use AI to assure and improve safety.

The research on AI techniques and tools is well-established in industry and academia and encompasses safety and security. Focused research on the assurance of AI applications in safety-critical aviation systems is limited. The FAA, NASA, and DOD have conducted some research in this area.

The FAA will continue to collaborate with other government agencies and industry stakeholders to conduct research and develop recommendations that serve as guidance for AI systems' assurance. The objective is to identify and evaluate various methods for safety assurance. The FAA is collaborating with NASA on the exploration of an alternate assurance process utilizing OPs as a conceptual framework. This approach shows promise when traditional compliance methods cannot be readily applied to emerging technologies such as AI. Ongoing research is being conducted in collaboration with industry partners to develop case studies that demonstrate the practical application of OPs.

Additional research is needed to provide insight for the development of guidance and training material. These initiatives include exploring efficient computational means of compliance, establishing test cases for safety analysis, gaining a deeper understanding of learning-based generalization and rigor, evaluating approaches to scale from low-criticality advisory systems to safety-critical systems, and exploring the explainability of AI systems. By actively pursuing these research endeavors, the aviation community can advance safety assurance methods and provide a predictable path to safely expand the use of AI in aviation.

The FAA has already concluded several research initiatives in AI. Recognizing both the pace of AI developments and the scale of industry investment in AI, both internal and external to aviation, it is essential that the aviation community focus aviation safety research on the topics that are uniquely related to AI and for which research can be concluded in time to inform decisions, regulations, or guidance, as appropriate. The FAA has identified several areas where research would be informative:

- **Aberrant Behavior Review**: The largest challenge for AI is in assuring innocuity, as described in section 5.3. Research should be conducted and sustained into the types of aberrant or unexpected behavior of learned AI systems to enable system designers to account for their potential manifestation in their systems and their safety mitigations.

- **Safety Assurance Using Formal/Numerical Methods**: This research should assess the effectiveness of potential methods to analyze the performance of AI systems during training and subsequent qualification. The method should provide performance indicators that can be used for assessing the safety of an integrated complex system, in which AI is a component.
- **Safety Assurance Using Systems and Testing Methods**: This research should assess the effectiveness of potential methods to constrain and test the performance of learned AI implementations in integrated systems. For example, can a designer adequately describe the environment in which the system will operate, and can the designer define adequate performance requirements and constraints to assure the integrated system is safe?
- **Use of AI in Data Analytics for Risk Identification**: This research should assess the effectiveness in using Data Mining to identify precursors to anomalies.
- **Use of Software Runtime Assurance in AI**: The commonly used idea of continuously monitoring the value of the output of a system to restrict it to below a preset threshold value, originally developed to assure the internal stability of an integrated system, may be extendable to assure the effectiveness and/or safety of learned AI systems. By investigating a Runtime Assurance (RTA) system, the questions that must be answered are: (1) Was the evaluation model designed correctly? (2) Why is the secondary system used as a backup (instead of as a primary system) when it can guarantee correct/safe performance? and (3) What are the confidence levels for accepting that the secondary system can guarantee correct/safe performance? To address these questions, it is strategic to leverage the partnership with the industry to get full cooperation to gain sufficient information about the RTA System, and to leverage the interest of the research institutions to press on developing specific methods that can provide supporting evidence to answer these questions.
- **Use of AI in Software Verification**: An interesting research topic for using AI to generate test data with the intention of verifying the functional requirements is in progress. In this endeavor, each requirement must be clearly defined with specific verification methods, and AI can be used to generate comprehensive test plans and test data to confirm that the requirement is satisfied. Where a property cannot be shown through testing, AI may be able to show a potential vulnerability by explaining the conditions under which it occurs, for example, a system deadlock during run-time due to design errors.
- **Use of AI in System Safety Reviews**: A System Safety Review will generally reveal the “expected” safety performance in anticipated scenarios that are represented in specific test cases. However, AI can be used in the unsupervised learning mode to identify what might have been overlooked by humans who must manually analyze the test data. In this approach, data mining is used to identify precursors to unexpected failure that can help humans in the review loop to evaluate the consequence of these unexpected failures.

Expected Milestones and Timeline

This section summarizes the planned and proposed activities that support the primary purpose of the roadmap, to provide a path to assuring the safety of AI in aircraft and related systems for inflight operations and to explore the use of AI for safety. These activities are planned with the current focus on addressing Machine Learning, a main component of AI that is currently being used in the development of many aviation applications that are under review at the FAA.

Development of Training Materials and Recruitment for Workforce Readiness

	Description	**Schedule (CY)**
AI Overview	The FAA and industry should provide access to AI introductory courses for the workforce to understand the potential and limitations of AI.	Ongoing
Introduction to AI Safety Assurance	The FAA should provide training to FAA and designee workforce regarding the safety vulnerabilities of AI in aircraft supersystems, the generic issue paper and how FAA is managing the risk.	3Q2025
AI Safety Assurance Methods	The Standard Organization(s) should provide training regarding safety assurance methods and how they can be demonstrated once such methods are established.	TBD
Recruit AI Experts	The office of Aviation Safety will recruit expertise in artificial intelligence.	Ongoing

AI in the Aircraft Supersystem

	Description	Schedule (CY)
Industry-initiated certification projects	FAA and individual applicants coordinate through project-specific issue papers to gain experience in safety assurance methods in a realistic context.	Ongoing
Policy Statement: Introduction of AI in Aircraft Certification	Policy statement to inform applicants that plan to use AI in their systems, or in the development of documents, that they should disclose that use and discuss a certification path with the FAA early in the development phase. FAA will be directly involved in AI projects due to novel or unusual nature.	4Q2024 (Draft)
Policy Memo: Considerations for AI/Machine Learning Used in the Development of Airborne Systems and Equipment	Policy memo to provide a list of considerations to be used as a starting point for discussion between the FAA and the applicant and documented in an issue paper. The policy memo will also include information about FAA involvement. The policy memo will be updated to include additional information as needed.	4Q2024 (Draft)
Certification Position Paper(s) (CPPs)	One or more CPPs to document stable policy related to specific use cases of AI/ML. CPPs will be based on experience with several project specific issue papers that can be generalized to address a certain use case.	1Q2026
Advisory Circular(s) (ACs)	One or more ACs developed from stable policy and harmonized with other certification authorities.	TBD
Standards		
SAE G34/EUROCAE (Industry)	Industry-developed standard for learned AI safety assurance.	TBD
Proposed (Industry)	Industry-developed standard[11] for black-box algorithm validation (applicable to some	TBD

[11] The proposal can be done at Technical Exchange Meetings and through communications with all interested Standard Organizations, the decision of who will accept that role is with individual Standard Organizations.

	learned AI where black-box strategy is adequate)	

Research

	Description	Schedule (CY)
Safety Assurance of AI		
AI Assurance Technical Exchange	Convene interested industry to share information, ideas and experience relating to AI safety assurance.	Ongoing
Aberrant Behavior	Aggregate the types of aberrant or unexpected behavior of learned AI systems to enable system designers to account for their potential manifestation in their systems.	4Q2025
Systems and Testing Strategy	Assess the effectiveness of black box system-level validation techniques and system-level requirement definition.	4Q2025
Numerical Methods	Assess the effectiveness of potential methods to analyze the performance of AI systems during training and subsequent qualification as it relates to high-level system requirements.	2Q2026
Training-based Methods	Assess the Advantages and Disadvantages between Different Training Methods and Identify the Appropriate Scenario for Each Training Method	4Q2024
AI in Safety Lifecycle		
Data Analytics	Assessment of Existing Methods and Applicability to Safety Lifecycle	4Q2024
Risk Modeling	Assessment of the Effectiveness of Machine Learning for Modeling Risk	4Q2025
Software runtime assurance	Evaluate the Effectiveness of Digital Twin Concept to Monitor AI Performance during Operation	4Q2025
Software verification	Evaluate Proposed Standards for Verification of AI Algorithms Implemented in Software Form	TBD
System safety reviews	Identify Areas in the System Safety Review (SSR) Process that Can Benefit from Data Analytics (for analysis) and Large Language Model (for report generation)	TBD

AI Roadmap is a Living Document

This document reflects a point-in-time snapshot of a fast-paced and evolving technology. Changes to this roadmap will be implemented based on experience, standards development, and research. It is important to recognize that the maturation of the technology is occurring in a much broader community than aviation, and following the further developments outside of aviation is as critical as advancing the activities outlined in this roadmap within the aviation community.