European Union Aviation Safety Agency

ARTIFICIAL INTELLIGENCE ROADMAP 2.0

Human-centric approach to AI in aviation

May 2023 | Version 2.0

easa.europa.eu/ai

Disclaimer

This document and all information contained or referred to herein are provided for information purposes only. It reflects knowledge that was current at the time that the document was generated. Whilst every care has been taken in preparing the content of the report to avoid errors, the Agency makes no warranty as to the accuracy, completeness or currency of the content. The Agency shall not be liable for any kind of damages or other claims or demands incurred as a result of incorrect, insufficient or invalid data, or arising out of or in connection with the use, copying or display of the content, to the extent permitted by European and national laws. The information contained in the document should not be construed as legal advice.

Photocredits

istock

Copyright

Contents

A. Foreword

The aviation industry has always been at the forefront of technological innovation, constantly pushing the boundaries to make air transport safer, more efficient and more accessible. From the earliest pioneer flights to the present day, aviation has undergone many technological revolutions, each contributing to the evolution of safer air travel.

The latest revolution is the rise of artificial intelligence (AI) and its potential to transform the world of aviation. AI allows us to create intelligent systems that can provide advanced assistance solutions to human end users, optimise aircraft performance, improve air traffic management and in turn enhance safety in ways that were previously unimaginable. However, the deployment of AI in aviation also poses new challenges and questions that need to be addressed.

As the European Union Aviation Safety Agency (EASA), we are committed to ensuring that the aviation industry benefits from the potential of AI while maintaining the highest standards of safety, security and environmental protection.

To achieve this, EASA has worked over the past 3 years with all stakeholders in the aviation industry. EASA's first two AI concept papers have paved the way for the approval and deployment of safety-related AI systems for end-user support (pilots, ATCOs, etc.) and are already being applied to certification projects through special conditions.

However, this is not all and a series of issues related to the use of AI in aviation still need to be addressed, such as:

- How to establish public confidence in AI-enabled aviation products?
- How to prepare for the certification and approval of advanced automation?
- How to integrate the ethical dimension of AI (transparency, non-discrimination, fairness, etc.) into oversight processes?
- What additional processes, methods and standards do we need to develop to unlock the potential of AI to further improve the current level of air transport safety?

The EASA AI Roadmap 2.0 is an update of the EASA vision for addressing the challenges and opportunities of AI in aviation, intended to serve as a basis for continued discussion with the Agency stakeholders. It is a living document, which will be amended regularly, augmented, deepened, improved through discussions and exchanges of views, but also, practical work on AI development in which the Agency is already engaged.

It builds further on the central notion of ***trustworthiness of AI*** and identifies high-level objectives to be met and actions to be taken to respond to the above questions. Moreover, it addresses a number of challenges that the Agency will have to meet; for instance, in developing new staff competency and processes that will contribute to the overall EU strategy and initiatives on AI.

We acknowledge that implementing this Roadmap is a complex and challenging task that requires collaboration and coordination between all stakeholders in the aviation industry. Only by working together can we ensure that AI technologies are deployed in a way that benefits the industry and the flying public.

Patrick KY
Executive Director
European Union Aviation Safety Agency

B. Introduction

AI is being adopted widely and rapidly, including in the aviation domain. While the concept of AI has been in existence since the 1950s, its development has significantly accelerated in the last decade due to three concurrent factors:

- Capacity to collect and store massive amounts of data;
- Increase in computing power; and
- Development of increasingly powerful algorithms and architectures.

AI systems are already integrated in everyday technologies like smartphones and personal assistants, and we can see that the aviation system is already affected by this technological revolution.

As concerns the aviation sector, AI not only affects the products and services provided by the industry; it also triggers the rise of new business models. This affects most of the domains under the mandate of the Agency, and its core processes (certification, rulemaking, organisation approvals, and standardisation) are impacted. This in turn affects the competency framework of Agency staff.

Beyond this, the liability, ethical, social and societal dimension of AI should also be considered.

In October 2018, the Agency had set up an internal task force on AI, with a view to developing an AI Roadmap 1.0 [1] that identified for all affected domains of the Agency:

- the key opportunities and challenges created by the introduction of AI in aviation;
- how this may impact the Agency in terms of organisation, processes, and regulations; and
- the actions that the Agency should undertake to meet those challenges.

The work of the AI task force resulted in the publication of EASA AI Roadmap 1.0 in February 2020. The implementation of the initial plan has already generated two major deliverables: the Concept Paper 'First usable guidance for Level 1 ML applications' [2] in December 2021, and the new proposed revision of that document, namely the 'Guidance for Level 1 & 2 ML applications' [3] that was released for public consultation in February 2023.

The AI activity has now evolved to a cross-domain AI Programme in which numerous specialists throughout the Agency are involved. This structure enhances the capacity of EASA to develop an all-encompassing strategy in relation to AI and addresses the scope as widened by this updated Roadmap.

The scope and plan developed in version 1.0 remained valid for the elapsed 3 years. It now requires an update based on the push from the technological development in the field of AI and the updated perspectives from the aviation industry.

The purpose of this AI Roadmap 2.0 is not only to communicate on the Agency vision for the deployment of AI in the aviation domain, but also to further serve as a basis for interaction with its stakeholders on this topic. In this perspective, this document is further intended as a dynamic document, which will be revised, improved and enriched with time as the Agency will gain experience on AI developments and stakeholders will provide their input and share their vision with the Agency.

C. What is AI?

AI is a relatively old field of computer science that encompasses several techniques and covers a wide spectrum of applications. AI is a broad term and its definition has evolved as technology developed. For this version 2.0 of this AI Roadmap, EASA has moved to the even wider-spectrum definition from the 'Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence' (EU Artificial Intelligence Act) [4], that is ***'technology that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with'***.

In line with Annex I to the Proposal for an EU AI Act, AI techniques and approaches can be divided in ***machine learning approaches*** (also known as data-driven AI), ***logic- and knowledge-based approaches*** (also known as symbolic AI) and ***statistical approaches***.

When drafting the first version of this AI Roadmap, the breakthrough was linked with ***machine learning (ML)*** and, in particular, ***deep learning (DL)***. Even if the use of learning solutions remains predominant in the applications and use cases received from the aviation industry, it turns out that meeting the high safety standards brough by current aviation regulations pushes certain applicants towards a renewed set of ***knowledge-based AI***. This is one of the main drivers for EASA to update this document.

Moreover, it is important to note that those different AI approaches may be used in combination (also known as ***hybrid AI***), which is also considered to fall within the scope of this Roadmap.

Consequently, the EASA AI Roadmap has been extended to encompass all techniques and approaches described in the following figure.

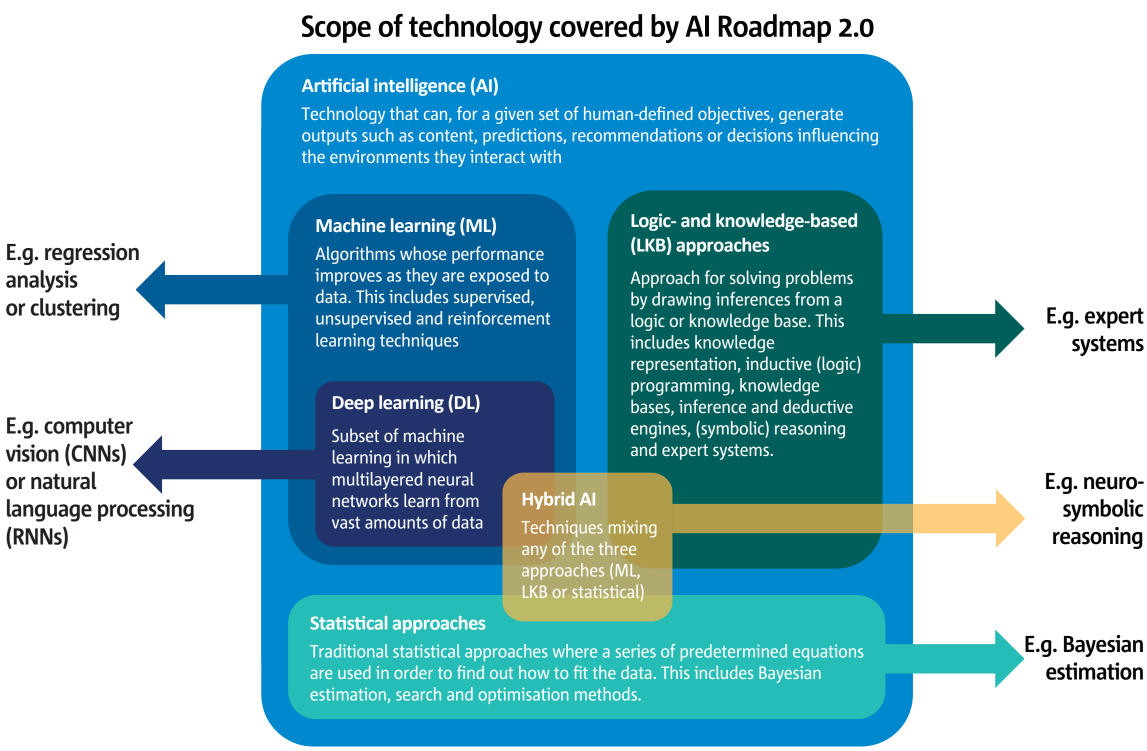


Figure 1: AI taxonomy in this Roadmap

D. The EU AI strategy

Massive investments around AI and data as the new gold

AI is a strategic technology that can improve among others healthcare, energy, transport, resources, finance or justice. Through the Horizon Europe and Digital Europe programmes, the Commission plans to invest 1 billion EUR per year in AI. It will mobilise additional investments from the private sector and the Member States in order to reach an annual investment volume of 20 billion EUR over the next decade, mainly to support research, health, transport, and common data spaces, which are the AI 'fuel'.

This is in addition to the 2.16 billion EUR per year already allocated in the frame of the Horizon Europe programme (cluster 5), which has paved the way to the SESAR3 JU and to the Clean Aviation JU. Both Joint Undertakings contribute to shaping the future aviation system with more reliance on new technologies as they oversee a large number of projects related to the deployment of AI in aviation.

EU strategy for AI at a global level

Note: Part of this section was co-constructed by humans using ChatGPT (OpenAI GPT-3) to illustrate the growing of capability of AI; however, still requiring some human oversight. The final text has been reviewed and corrected by human readers.

The European Union (EU) has developed a comprehensive strategy for AI at a global level, which is designed to ensure that AI is developed and used in a way that is human-centric, trustworthy and safe.

The Coordinated Plan on Artificial Intelligence 2021 Review[1] indicates that the EU's strategy for AI is focused on accelerating investments, acting on AI strategies and programmes and aligning policies.

The strategy involves four main actions:

- enabling conditions for AI development and uptake in the EU, including establishing a regulatory framework that promotes innovation while ensuring safety and ethics;
- making the EU the place where excellence thrives from the lab to market, fostering an ecosystem that supports start-ups and SMEs in AI development;
- ensuring that AI works for people and is a force for good in society, with a focus on promoting diversity, transparency, and accountability;
- building excellence in AI through strategic leadership in high-impact sectors, such as healthcare, mobility, and energy, by fostering collaboration between industry, academia, and public institutions, and ensuring access to data and computing power.

Overall, the EU is committed to a human-centric approach to AI that respects fundamental rights and values, promotes inclusion and diversity, and supports sustainable and responsible innovation.

EU AI regulations: the EU AI Act

The EU AI Act [4] is a regulatory proposal that was published by the European Commission in April 2021. It aims to create a harmonised framework for the development and use of AI across the European Union.

The proposed regulation distinguishes AI systems into three categories according to the level of risk they pose: (i) unacceptable risk, (ii) high risk, and (iii) low or minimal risk.

1 https://ec.europa.eu/newsroom/dae/redirection/document/75787

The key provisions of the EU AI Act include:

- Risk assessment: AI systems will be categorised based on their risk level, and high-risk systems will require a conformity assessment before they can be placed on the market or put into service.
- Ban on unacceptable AI: The regulation prohibits certain AI practices that are considered to be unacceptable and contrary to EU values, such as AI systems that manipulate human behaviour, exploit vulnerabilities of specific groups, or use subliminal techniques.
- Transparency and traceability: The regulation requires that AI systems be designed in a way that ensures transparency and traceability, so that users can understand how decisions are made and what data is used to train the system.
- Data governance: The regulation aims to ensure that data used to train AI systems is of high quality and unbiased, and that privacy and data protection rules are respected.
- Human oversight: The regulation requires that AI systems have appropriate human oversight and control, especially in high-risk situations such as in healthcare or public services.

The EU AI Act is still in the proposal stage and will need to be approved by the European Parliament and Council before it becomes law. However, when adopted, it will have a significant impact on the development and use of AI across the EU, ensuring that AI is developed and used in a way that is safe.

As reflected in Article 81 of the proposed EU AI Act, the anticipated impact on aviation regulations should consist in ensuring that the requirements set out in Title III, Chapter 2 of the AI Act is accounted for in future implementing and delegated acts related to airworthiness, ATM/ANS and unmanned aircraft. EASA AI Concept Papers [2] and [3] were built to ensure full compatibility with the EU AI Act [4] Title III, Chapter 2 for all domains under the remit of EASA. Complete traceability will be detailed upon final publication of the EU AI regulations.

Other AI-related regulations and directives are considered by EASA in the preparation of the rulemaking concept that is described in Chapter H, in particular the proposed EU Data Act [4] and EU AI Liability Directive [6].

AI trustworthiness as a key driver for an ethical AI

Most probably more than any technological fundamental evolutions so far, AI raises major ethical questions. A European ethical approach to AI is central to strengthen citizens' trust in the digital development and aims at building a competitive advantage for European companies. Only if AI is developed and used in a way that respects widely shared ethical values, can it be considered trustworthy. Therefore, there is a need for ethical guidelines that build on the existing regulatory framework.

In June 2018, the Commission set up a High-Level Expert Group on Artificial Intelligence (AI HLEG)[2], the general objective of which was to support the implementation of the European strategy on AI. This includes the elaboration of recommendations on future-related policy development and on ethical, legal and societal issues related to AI, including socio-economic challenges. In March 2019, the AI HLEG proposed the following seven key requirements for trustworthy AI, which were published in its report on Ethics Guidelines on Artificial Intelligence [5]:

2 https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence

7 key ethical requirements for trustworthy AI

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

Figure 2: Overview of the ethical guidelines from the EC HLEG

The guidelines developed by AI HLEG are non-binding and as such do not create any new legal obligations. The EASA strategy embraces this approach from an aviation perspective.

Finally, the AI trustworthiness concept is considered as a key enabler of the societal acceptance of AI (see Chapter G).

Other relevant initiatives for the development of AI in Europe

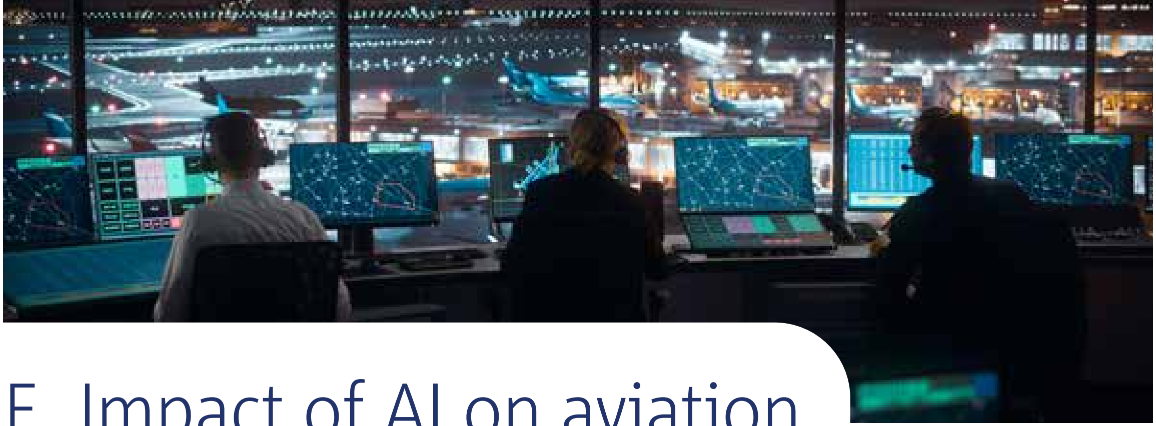
The DIGITALEUROPE[3] programme will be crucial to make AI available to small/medium-sized enterprises across all Member States, through innovation hubs, data spaces, testing/experimentation and training programmes. Its budget is 9.2 billion EUR for 2021-2027. DIGITALEUROPE is an international non-profit association whose membership includes 41 national trade associations from across Europe as well as 102 corporations that are global leaders in their field of activity. Its mission is to shape a business, policy and regulatory environment in Europe that best nurtures and supports digital technology industries.

The European High-Performance Computing Joint Undertaking (EuroHPC) will develop the next generation of supercomputers because computing capacity is essential for processing data and training AI, and Europe needs to master the full digital value chain. The ongoing partnership with Member States and industry on microelectronic components and systems (ECSEL) as well as the European Processor Initiative will contribute to the development of low-power processor technology for trustworthy and secure high-performance edge computing.

On the software side, the European Commission also proposes to develop common 'European libraries' of algorithms that would be accessible to all. Like data, AI algorithms are a key governance instrument to ensure the independence of EU industry from the 'AI mega-players'.

As for the work on ethical guidelines for AI, all these initiatives build on close cooperation of all concerned stakeholders, Member States, industry, societal actors and citizens. Overall, Europe's approach to AI shows how economic competitiveness and societal trust must start from the same fundamental values and mutually reinforce each other.

3 https://www.digitaleurope.org/



E. Impact of AI on aviation

Multiple domains of the aviation sector will be impacted by this emerging technology. The air transport system is facing new challenges: increase in air traffic volumes, more stringent environmental standards, growing complexity of systems, greater focus on competitiveness, for which AI could provide opportunities. This section first provides a general overview of the anticipated impact on each aviation domain in terms of use cases. It then provides an overview of the Rulemaking strategy anticipated by EASA including a discussion on the specificities for each domain.

1. Aircraft design and operation

AI, and more specifically the ML field of AI, is bringing an enormous potential for developing applications that would not have been possible with the development techniques that have been used so far.

On the one hand, the breakthrough of DL has brought about a wide range of applications that could benefit aviation; in particular, computer vision and natural language processing (NLP) bring new perspectives when dealing with perception applications; also, time-series analysis can increase the capabilities to make sense out of sensor data. In aviation, these types of application could open the door to solutions such as high-resolution camera-based traffic detection or assistance in ATC communication through speech-to-text capability. On the other hand, the hybridisation of DL solutions with logic- and knowledge-based approaches, open a promising path to efficient decision-making support, in view of enhanced virtual assistance to the pilot.

AI is a clear enabler for a wide range of applications in support of aircraft design and operations. AI may assist the crew by advising on routine tasks (e.g. flight profile optimisation) or providing enhanced advice on aircraft management issues or flight tactical nature, helping the crew to take decisions in particular in high-workload circumstances (e.g. go around or diversion). AI may also support the crew by anticipating and preventing some critical situations according to the operational context and the crew health situation (e.g. stress, health, etc.). Level 1 AI applications are already under the process of certification in the general aviation domain thanks to the first special condition on Trustworthiness of Machine Learning based Systems that was published in April 2022. AI could also be used in nearly any application that implies mathematical optimisation problems, removing the need for analysis of all possible combinations of associated parameter values and logical conditions. Typical applications of ML could be flight control laws optimisation, sensor calibration, fuel tank quantity evaluation, icing detection and many more to come. Furthermore, AI could also be used to embed complex models on board aircraft systems, for instance by using surrogate models that are more memory and processing efficient.

Moreover, AI brings new opportunities for advanced automation applications involving human-AI teaming. However, ML approaches alone may not always provide the necessary level of assurance in AI-based systems when it comes to automatic decisions (Level 2 AI and above); therefore, the aviation industry is also investigating the use of hybrid AI solutions in an attempt to overcome possible shortcomings of ML solutions.

In a wider perspective, the most discussed application of AI is autonomous flight. At the point we are in this quest, it becomes however clear that current available technology does not match the anticipated levels of adaptability, rationality and management of uncertainty that aviation products should be meeting to reach autonomy. Nevertheless, the drone market paves the way towards advanced automation and we can see the emergence of new business models striving for the creation of air taxi systems to respond to the demand for urban air mobility. Such vehicles will inevitably have to rely on systems to enable complex decisions, e.g. to ensure the safe flight and landing or to manage the separation between air vehicles with reduced distances compared to current ATM practices. This is where AI comes into play: to enable advanced automation, very powerful AI models will be necessary to process and use the huge amount of data generated by the embedded sensors and by the machine-to-machine communications to support flights without human intervention.

Moreover, AI could also be used for improving the design processes. For instance, ML-based tools could be developed to support engineering judgement in the selection of relevant sets of non-regression tests. AI/ML can also provide a solution for the modelling of physical phenomena (e.g. through the use of surrogate models) and could be used for facilitating design space exploration and optimising qualification processes that rely on physical phenomena demonstration (e.g. EMI, EMC, HIRF).

To ensure safe operations, crew training is another essential consideration. The use of AI gives rise to adaptive training solutions, where ML could enhance the effectiveness of training activities by leveraging the large amount of data collected during training and operations.

2. Aircraft production and maintenance

Production and maintenance (including component logistics) are domains where digitalisation is likely to affect processes and business models significantly.

With digitalisation, the amount of data handled by production and maintenance organisations is steadily growing and with this, the need to rely upon AI to handle this data is also increasing. Among the trends to be mentioned are the development of surrogate models and digital twins in the manufacturing industry, the introduction of internet of things (IoT) in the production chains and the development of predictive maintenance where the vast amount of data and the need to identify low signals will most certainly require the use of AI.

Nowadays, engine manufacturers do not really sell engines and spare parts anymore, but rather flight hours. This paradigm shift implies that, to avoid penalties for delays, engine dispatch reliability and safety are part of the same concept. AI-based predictive maintenance, fuelled by an enormous amount of fleet data, allows to anticipate failures and provide preventive remedies.

Industry key players have already recognised the value of predictive maintenance. For instance, Airbus' Aircraft Maintenance Analysis (Airman), used by more than a hundred customers, constantly monitors health and transmits faults or warning messages to ground control, providing rapid access to maintenance documents and troubleshooting steps prioritised by likelihood of success.

Certain university research estimate that predictive maintenance can increase aircraft availability by up to 35 %[4].

3. Environment

Among the multiple applications of AI, the optimisation of trajectories is one example of how AI can help reducing carbon emissions.

Beyond this, AI gives an unprecedented opportunity for the Agency to improve its capability to deal with environmental protection; for instance, regarding impact assessments.

Assessing the environmental impacts of aviation, such as noise around airports or in-flight engine emissions, is a data and computation-intensive activity that has significantly evolved over the past decades together with machine capabilities. Based on data sets available to the Agency (global weather data, flight data recorder (FDR) information, worldwide radar (ADS-B) flight trajectory data, etc.), ML algorithms could be developed to assess the fuel consumption of virtually any flight. This would allow the Agency to perform its impact assessments in a more effective and continuously improving manner.

4 PREDICTIVE & DETECTIVE MAINTENANCE: EFFECTIVE TOOLS IN THE MANAGEMENT OF AERONAUTICAL PRODUCTS. José Cândido de Almeida Júnior, Rogerio Botelho Parra FUMEC University — 31st Congress of the International Council of the Aeronautical Sciences — Belo Horizonte, Brazil; September 09-14, 2018

4. Air traffic management

The ATM/ANS domain foresees important deployments of AI/ML applications. AI-enabled assistants have already been introduced in operations to support air traffic controllers (air traffic controllers (ATCOs), flow management positions (FMPs)) or other end users of the ATM/ANS domain. Just to mention one example, such assistants can improve the 4D trajectory predictions, thus improving the quality and accuracy of any local or network situation assessment (early detection of hotspots).

By analysing data on weather patterns, sectors configurations, air traffic congestions, and other factors, AI/ML models could support the optimisation of flight routes to reduce flight time, fuel consumption, and, ultimately, costs. Such an optimisation would then lead to a more efficient air traffic management system, reducing delays and increasing the capacity of air travel.

AI/ML applications could help ATCOs make more informed decisions more quickly, potentially reducing delays and improving safety. ML models could be integrated into decision support tools providing real-time guidance to ATCOs. These support tools could provide recommendations based on the current situation, including information on potential conflicts, and suggest the best course of action for resolving conflicts, or alternatively propose solutions that the ATCO is familiar with. Some research groups under SESAR 3 projects, industry players or air navigation service providers (ANSPs) are investigating applications that will make use of reinforcement learning for the purpose of conflict detection and resolution (CDR).

5. Aerodromes

In the aerodromes domain, applications of AI/ML could be envisaged both for the airside and terminal-based operations.

On the airside, the following use cases of AI/ML related to aerodrome safety are anticipated:

- Detection of foreign object debris (FOD) on the runway: FOD prevention and the inspection of movement area for the presence of FOD is a core activity of aerodrome operators. The application of ML in FOD detection has the potential to make current systems more reliable.

- Avian radars: at airports, the prevention of bird strikes to aircraft is an ongoing challenge. Avian radars can track the exact flight paths of both flocks and individual birds. ML solutions could support automatic detection and logging of hundreds of birds simultaneously, including their size, speed, direction, and flight path, thereby creating situational awareness and allowing for a better response by bird control staff.

- UAS detection systems: the surroundings of aerodromes may be affected by the unlawful use of unmanned aircraft, representing a hazard to aircraft landing and taking off. Today’s technology-based counter-UAS solutions are mostly multi-sensor-based, as no single technology is sufficient to support the system to perform satisfactorily. The improvement of such technologies with ML appears to be the logical evolution.

Inside the terminal and in relation to passenger services and passenger management, AI has manifold application areas. For example, AI is integrated with airport security systems such as screening, perimeter security and surveillance since these will enable the aerodrome operator to improve the safety and security of the passengers. Furthermore, border control and police forces use facial recognition and millimetre-wave technologies to scan people walking through a portable security gate. ML techniques are used to automatically analyse data for threats, including explosives and firearms, while ignoring non-dangerous items — for example, keys and belt buckles — that users may be carrying. In addition, ML techniques are used by customs to detect prohibited or restricted items in luggage.

6. Drones, U-space and innovative air mobility

The integration of manned and unmanned aircraft, while ensuring safe sharing of the airspace between airspace users, and ultimately the implementation of advanced U-space services will only be possible with high levels of automation and use of disruptive technologies like AI/ML.

The early implementation of AI/ML solutions will be essential to widely enable complex drone operations in environments which are fast-evolving and where stringent requirements apply, such as urban areas or congested control tower regions (CTR). For instance, AI/ML solutions could allow for a dynamic and fast reaction (e.g. autonomous change of trajectory) to sudden changes in the operational environment (e.g. encounters or threats, implementation of dynamic airspace reconfiguration/restriction). AI/ML solutions will play a crucial role in enabling safe conduct of the drone operations also in case of a contingency/emergency situation — for instance, in detecting obstacles (e.g. cranes), detecting or predicting icing conditions, or determining the risk on the ground (e.g. presence of public on a pre-planned landing site).

Similarly, efficient implementation of U-space, to cope with large numbers of drones simultaneously operating in the same volume of airspace will not be possible using traditional approaches. AI/ML will be a key enabler to satisfy the related required performance requirements.

More particularly, the AI/ML solutions will be a technical prerequisite for:

- detect and avoid (DAA) solutions, in particular by relying on the performance of AI/ML solutions in analysing data acquired from radar or camera-based systems;
- adaptive deconfliction, by dynamically predicting the risk of encountering an intruder along the flight path and adjusting in advance the drones' trajectory to ensure continuous separation in space and time;
- autonomous localisation/navigation (without GPS) solutions to reap the benefits from AI/ML techniques; for instance, by improving and simplifying current positioning sensors, data aggregation and the overall performance of the functions.

7. Cybersecurity

The cybersecurity domain encompasses three main elements:

- The system/organisation which has vulnerabilities that lead to the risk of being exploited, causing thus operational impacts;
- The threat (e.g. a malware) which could cause harm to a system or organisation by exploiting its vulnerabilities; and
- The security control/countermeasure introduced by the defender, which mitigates one or more security risks.

Emergence of the use of AI will affect all three elements.

- With AI, the system improves its effectiveness, but may also encompass new kinds of vulnerabilities to cyberattacks. These new types of vulnerabilities need to be better understood (e.g. data poisoning) and specific security controls (technical or organisational) for them need to be defined.
- On the threats side, nowadays, malware are already mutating (i.e. adapting their behaviour depending on the running environment). Moreover, researchers have demonstrated the feasibility of a new class of AI-powered malware (e.g. DeepLocker). Using AI for cyberattacks will certainly improve the efficiency of the threats by developing the ability of circumventing the conventional rule-based detection systems and ultimately making cyberattacks adaptive and autonomous. AI-powered attacks may be soon deployed and it is essential that adequate countermeasures be identified.

- On the defender side, we shall also consider the opportunity of introducing AI in countermeasures and security controls to improve their effectiveness. To this extent, we may benefit from AI technology for the automatic detection and patching of systems' vulnerabilities (prevention), as well as for the identification of threats on a behavioural basis (detection).

The scope of the AI Concept Paper guidance related to information security is for now limited to addressing the first point of the above list. A future version of this guidance may extend to the third point, considering to a certain extent the second one as well from a defender perspective.

8. Safety risk management

Data science is a specialised domain that combines multiple areas such as statistics, mathematics, intelligent data capture techniques, data cleansing, mining and programming to prepare and align big sets of data for intelligent analysis to extract insights, pattern and information. Data science is quite a challenging area due to the complexities involved in combining and applying different methods, algorithms, and complex programming techniques to perform intelligent or predictive analysis from large volumes and large varieties of data.

The emergence of the use of AI will affect many aspects of the data science technology, mainly for the data analysis part, such as the identification of complex data correlation (pattern discovery). Applied to the EASA context, AI technology will empower the safety intelligence by, for instance, improving the vulnerability discovery capabilities.

Generally, in the domain of EASA safety intelligence and management, AI is a key enabler to support:

- emerging risks detection;
- risk classification of occurrences; and
- Safety Risk Portfolio design and prioritisation of safety issues.

Looking more specifically at the application of AI to the EASA Data 4 Safety (D4S) project, ML could provide solutions to deal with D4S data (i.e. large sources of operational data like flight data, safety reports, weather data) and traffic data (very large volumes, variable and complex data silos, numerous potential quality issues) both during the collection/preparation of the data sets and the analysis steps.

AI can provide solutions to infer knowledge through:

- understanding data (e.g. risk modelling) thanks to the ad hoc analysis of large amounts of historical data;
- identifying hidden correlations in the data, between the different silos of data, fully leveraging on data fusion;
- vulnerability discovery; and
- anomaly detection thanks to the analysis of the data flowing into the Big Data architecture incrementally and the detection of any unusual evolutions (anomalies).

In the longer term, we can anticipate that AI will be a solution to deal with real-time data flows and enable real-time risk management.

F. Common AI challenges in aviation

From the analysis of the impact presented in Chapter E, a set of common challenges has been identified.

The scope of the first version of the EASA AI Roadmap focused on ML and DL techniques and included a set of initial challenges that needed to be addressed by the guidance development. While the initial list of challenges is still valid, is has been updated to address the new scope set in this updated AI Roadmap 2.0. Here is an overview of the main challenges that will require further consideration and guidance development in the future versions of the EASA AI Concept Paper:

— ***Adapting assurance frameworks to cover the specificities of identified AI techniques and address development errors in AI-based systems and their constituents***

  The development of AI-based systems and constituents puts emphasis on other parts of the process than the traditional development assurance; namely, the definition of metrics for performance verification, the knowledge and data management as a proxy for requirements traceability, the selection of model architecture and algorithms, etc. AI assurance methodologies and processes for learning approaches (refer to the W-shaped learning assurance process in the EASA AI Concept Paper [3]) will require further generalisation and extension to address the other AI approaches.

— ***Difficulties in keeping a comprehensive description of the intended behaviour and creating a framework for knowledge and data management***

  Development assurance principles have been confirmed to remain adapted to the capture, validation and verification of the intended behaviour at a higher level, and at item level, to perform the implementation of the AI/ML models into hardware and software.

  Nevertheless, when it comes to AI design processes, the behaviour is contained in the knowledge bases and data sets that are used to design the AI constituents and models. It is thus more challenging to maintain a traceability link with higher-level requirements and to ensure the completeness and correctness of the knowledge bases and data sets. The quality of the knowledge bases and data sets has a great importance, as incomplete or incorrect data can influence the behaviour of the training model. In this respect, there is an underlying challenge in capturing the operational design domain (ODD), especially in high-dimensional input spaces.

— ***Coping with limits in predictability and explainability of the AI application behaviour, considering the inference mechanisms and the model complexity***

  Even if an AI model may be deterministic from a mathematical perspective (e.g. fixed weights in an NN), for any new input at inference-time, the output will depend on the correlation between that input with the data set that was used for the design process. This and the inherent complexity of AI models can lead to unpredictable outputs that may be difficult to explain. This is often confused with a lack of determinism of the model, so it is clearer to speak about unpredictability of the AI application.

  On the one hand, there is a need to address the challenge of reaching appropriate generalisation guarantees for ML models. On the other hand, there is a need to increase the capability of making more understandable the conditions that led to a given output, further investigating the notion of 'explainability of AI' (both in the development and in the operational phases).

— ***Managing shared operational authority in novel types of human-AI teaming (HAT)***

With the evolution of the AI-based systems towards Level 2 AI applications, there is a need for guidance on how to effectively introduce and use this concept of human-AI teaming (HAT). This involves in particular the challenges of the shared understanding of goals and decision-making processes between the human and AI-based systems.

— ***Elaborating pertinent guarantees on stability and robustness of AI models and on the absence of 'unintended behaviour' in AI applications***

Due to the statistical nature of data-driven and statistical applications, they are subject to variability in their output for small variations on their input (that may even be imperceptible by a human). There is a need to investigate new methods to verify the stability and robustness of AI applications, as well as to evaluate the completeness of the verification. The use of formal methods as providing suitable verification means is further investigated. Such methods may as well be a means to compensate for the lack of coverage analyses, while observing closely and attentively the underlying scalability issues presented by these methods.

— ***Addressing model bias and variance trade-off in the various steps of the AI assurance processes***

AI solutions are subject to bias and variance (see [3] for more details), which can compromise the integrity of their outputs.

From EASA perspective, one of the most challenging aspects when collecting, preparing or using knowledge bases and data sets, is therefore the capability to identify, detect and finally mitigate adequately any bias or variance — introduced at any time during the 'knowledge and data management' or the 'model design' processes — that could have a negative impact on safety.

— ***Dealing with adaptive learning processes***

Real-time learning in operations is a parameter that is expected to introduce a great deal of complexity in the capability to provide assurance on the ever-changing software. This is incompatible with current certification processes and would require large changes in the current regulations and guidance. At this stage, this is considered a complex issue that requires the introduction of a strong limitation for safety-related systems in the aviation domain.

— ***Embedding AI models in safety-related systems***

This challenge includes at least two important considerations. First, to manage optimisation between the trained model and the model embedded on the target platform, so as to ensure that the semantics of the inference model is preserved. Second, to investigate the need for additional guidance for the qualification of different types of hardware (e.g. including dedicated accelerators) that are necessary to execute large ML models.

G. AI trustworthiness concept

The common thread that creates a continuity between the first version of the EASA AI Roadmap, and its deliverables, and this second version is the concept of ***AI trustworthiness.***

This document identifies four building blocks that are considered essential for creating a framework for AI trustworthiness and for enabling readiness for use of AI in aviation. All four building blocks constitute the backbone of this EASA AI roadmap. The objectives related to each of the building blocks are further developed in the successive versions of the EASA AI Concept Paper and are organised along the line of the end-to-end life cycle introduced at the end of this Chapter.

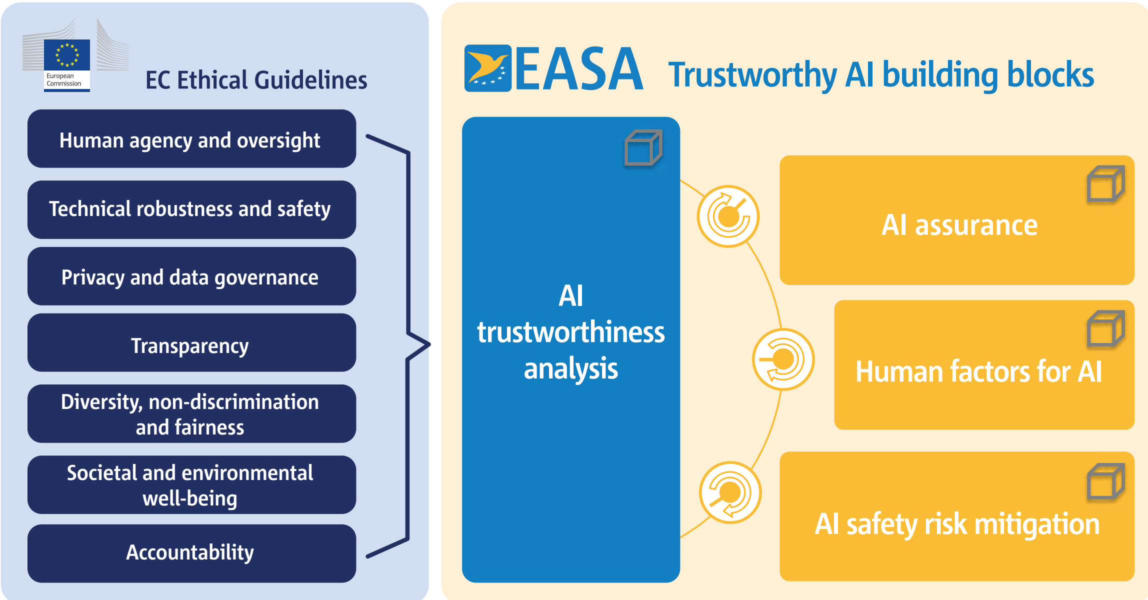


Figure 3: EASA AI trustworthiness building blocks

1. AI trustworthiness analysis

General approach for the trustworthiness analysis building block

The ***AI trustworthiness analysis*** starts with a ***characterisation of the AI application*** and encompasses the ***safety assessment*** and ***security assessment*** that are key elements of the trustworthiness analysis concept. It also includes an ***ethics-based assessment*** that creates an interface with the EU Ethical Guidelines developed by the EU Commission (EU High-Level Expert Group on AI, 2019). All four assessments (i.e. characterisation of the AI, safety, security and ethics-based) are important prerequisites for the development of any system developed with or embedding AI, and are not only preliminary steps but also integral processes towards approval of safety-related AI. Moreover, the AI trustworthiness analysis serves as a gate to the three other technical building blocks.

Focus on human oversight

To focus on one important aspect of the future EU regulations on AI, the AI trustworthiness analysis addresses the level of 'human oversight' in the AI application.

The first version of the EASA AI Roadmap identified three general AI levels. This scheme had been proposed based on prognostics from industry regarding three main types of AI applications, namely human assistance (Level 1 AI), human-AI teaming (Level 2 AI) and advanced automation (Level 3 AI).

This classification scheme has been refined while developing AI guidance, with an additional split for level 2 AI, based on the human-AI teaming (HAT) guidance and presented in the EASA AI Concept Paper [3]. Level 3 will surely require further adaptations when EASA develops Issue 03 of its AI Concept Paper. One distinction between the anticipated levels 3A and 3B could be the potential loss of human oversight, e.g. when losing the communication and control link with a drone. In this context, a further split of level 3 will probably be necessary to account for the likely emergence of autonomous products in the longer term, which may require an update of the EU AI Act Article 14 before being enabled.

The resulting refinement of the three scenarios is reflected in the following figure:

Level 1 AI: assistance to human	**Level 2 AI: human-AI teaming**	**Level 3 AI: advanced automation**
• Level 1A: Human augmentation • Level 1B: Human cognitive assistance in decision-making and action selection	• Level 2A: Human and AI-based system cooperation • Level 2B: Human and AI-based system collaboration	• Level 3A: The AI-based system performs decisions and actions that are overridable by the human. • Level 3B: The AI-based system performs non-overridable decisions and actions (e.g. to support safety upon loss of human oversight).

Figure 4: Classification of AI applications

It is also important to note that the proportionality and modulation of the AI guidance is driven by both the level of AI (as an output of the characterisation of the AI application) and the criticality of the application (as an output of the safety and security assessments). The trustworthiness analysis is always necessary and should be performed in its full spectrum for any application. For the other three building blocks, the potentiometers represented in Figure 3 illustrate that the depth of guidance could be adapted depending on the application.

2. AI assurance concept

The ***AI assurance*** building block is intended to address the AI-specific guidance pertaining to the AI-based system. It encompasses two major topics. Firstly, ***learning assurance*** covers the paradigm shift from programming to learning, as the existing development assurance methods are not adapted to cover learning processes specific to AI/ML. Secondly, ***development (respectively post-ops) explainability*** deals with the capability to provide human users (e.g. developers, auditors) with understandable, reliable and relevant information with the appropriate level of details on how an AI/ML application produces (respectively has produced) its results. This building block also includes the data recording capabilities, addressing two specific operational and post-operational purposes: on the one hand the continuous monitoring of the safety of the AI-based system and on the other hand the support to incident or accident investigation.

3. Human factors for AI

The ***human factors for AI*** building block introduces the necessary guidance to account for the specific human factors needs linked with the introduction of AI. Among other aspects, ***AI operational explainability*** deals with the capability to provide the human end users with understandable, reliable and relevant information with the appropriate level of details and with appropriate timing on how an AI/ML application produces its results. This block has also introduced the concept of ***human-AI teaming*** to ensure adequate cooperation or collaboration between human end users and AI-based systems (see [3] for more details on the human-AI teaming concept) to achieve certain goals.

4. AI safety risk mitigation

The ***AI safety risk mitigation*** building block considers that we may not always be able to open the 'AI black box' to the extent required and that the associated residual risk may need to be addressed to deal with the inherent uncertainty of AI.

This concept covers a wide range of solutions and may need to be further split in several blocks in the future.

5. AI application life cycle overview

Another important finding made during the first 3 years of development of guidance for safety-related AI applications, is that the life cycle process for AI-based systems has a larger scope than the one considered for traditional systems development.

For instance, adequate data management as part of the learning assurance process requires a number of requirements on the data collection and data governance phases upstream, at least for the highest level of criticality of applications.

Also, the continuous safety management process that is introduced in the EASA AI Concept Paper Issue 02 creates a set of additional requirements on the operational phase of the product life cycle which goes beyond existing requirements (and may trigger the need for re-training of AI/ML models).

Finally, additional requirements on the users' (e.g. engineers) and end users' (e.g. pilots, ATCOs) training phases are anticipated through the requirements for aviation organisations contained in Section C.6 of the EASA AI Concept Paper Issue 02. This is another extension of the life cycle process as represented on the following figure.

The guidance developed in the EASA AI Concept Papers already addresses most of the phases identified in Figure 5: AI application life cycle process, except for the knowledge and data governance phase, which will require further consideration in Issue 03.

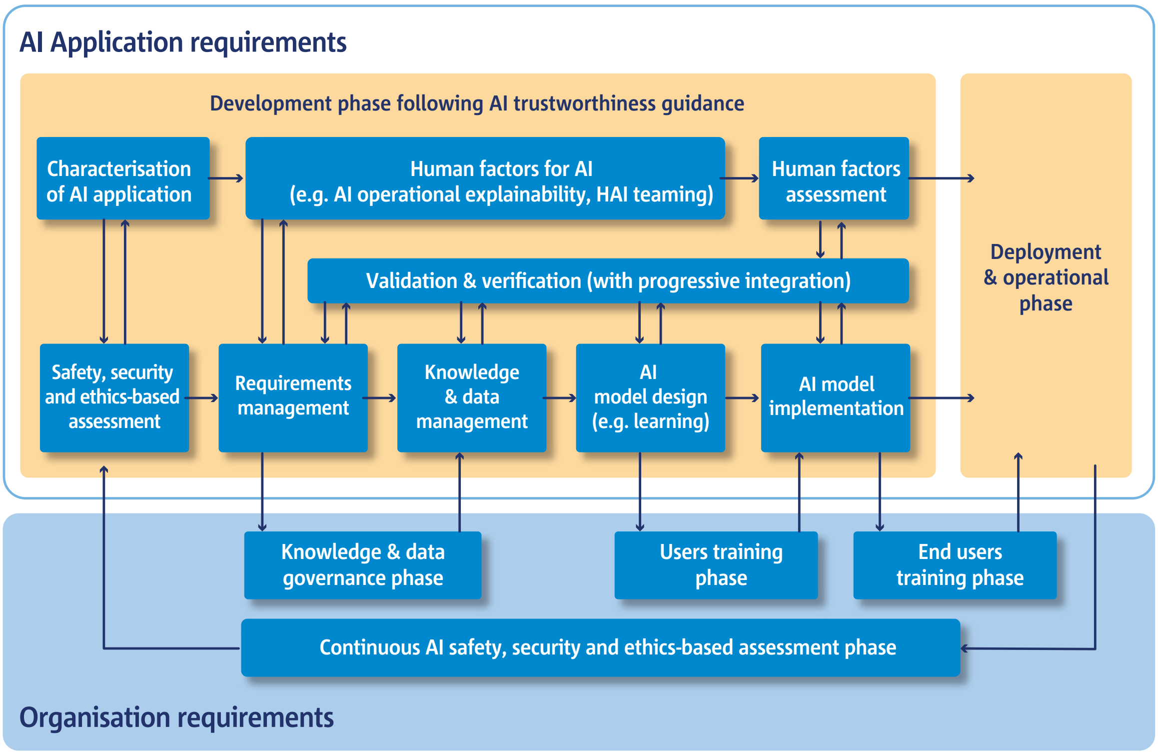


Figure 5: AI application life cycle process

H. Rulemaking concept for AI

The analysis of the anticipated impact of AI on the various domains (see Chapter E) has been initiated in the EASA AI Concept Papers and indicates that the statement of issue is largely shared across domains. However, the first guidance development shows in addition the need to account for certain domain specificities.

This calls for a mixed rulemaking approach, involving on the one hand cross-domain rules (horizontal) and, on the other hand, domain-specific rules (vertical).

This is anticipated to be developed in two steps:

- Step 1: development of a transversal Part-AI to contain the three major provisions anticipated in the EASA Concept paper guidance: requirements for authorities (Part-AI.AR), requirements for organisations (Part-AI.OR) and requirements linked to AI trustworthiness (Part-AI.TR) as described in Chapter G. In addition, some acceptable means of compliance (AMC) and guidance material (GM) are anticipated to account for the 'anticipated MOC' currently captured in the EASA AI Concept Papers, to reference or complement where necessary applicable industry standards.
- Step 2: analysis, per domain, of those requirements that are domain-specific and those that need to be complemented to provide an adequate regulatory basis for deploying the new Part-AI.

This rulemaking approach will also address the need to account for the EU AI Act [4] Title III, Chapter 2. As the needs in terms of rulemaking activities are already identified in the European Plan for Aviation Safety (EPAS) 2023-2025 Volume I Section 3.4.1 'Artificial intelligence (AI) in the Aviation Programme', a detailed rulemaking plan is anticipated to be developed in the next revision of the EPAS Volume II.

The regulatory structure is anticipated as follows:

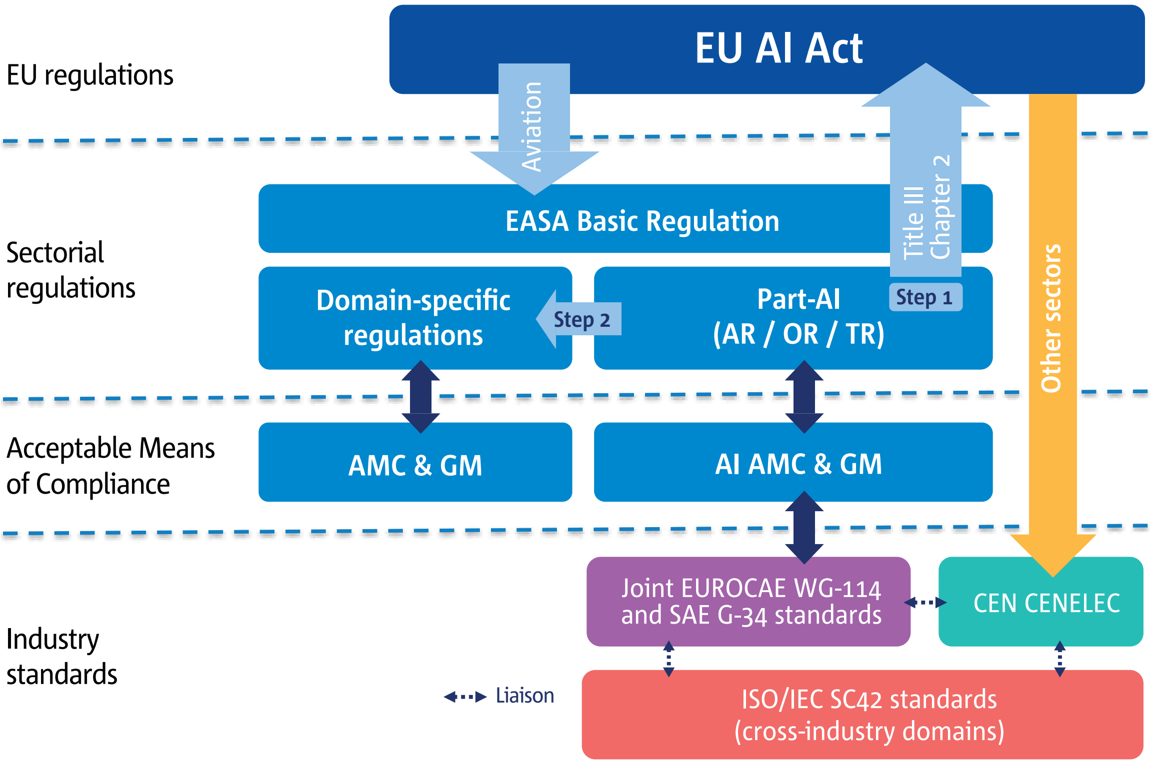


Figure 6: Anticipated regulatory structure for AI

I. Other challenges for EASA posed by the introduction of AI in aviation

1. Staff competency

As the Agency's core safety functions (certification, rules and standards development, approval of organisations) will be impacted by the introduction of AI, it is of utmost importance to ensure that the Agency's personnel will get the right level of AI expertise to carry out their tasks.

Contrary to industry personnel, the Agency staff is not directly exposed to or involved in the development of AI. This poses the risk of a knowledge gap between EASA and industry experts, which could be detrimental to the fulfilment of the EASA core safety functions.

In order to overcome this risk, not only is it necessary to train EASA personnel on AI but also to expose them to AI/ML practices as soon as possible.

A comprehensive training plan has been initiated and will be further deployed to ensure peer-to-peer discussions with our industry stakeholders and other regulatory authorities.

Partnership agreements with the industry, as explained in Section H.3 should also be the vehicle for practical exposure to AI.

2. Research

The need to support the EU Research agenda has been reinforced in the EASA Basic Regulation (see in particular Recital 58[5] and Article 86[6])

Recommendations from the EU Commission on AI indicate that Europe must aim for scientific AI leadership. Both purpose-driven and fundamental research in all aspects of AI must be secured in order to promote AI that is trustworthy and to address relevant scientific, ethical, sociocultural and industrial challenges. This necessitates a European research community that can unite through strong collaboration, and that can join forces with industry and society at large to build on European research strengths and enhance Europe's well-being.

From an aviation perspective, EASA identifies the key research topics and setting up research priorities, among others for AI, in collaboration with all research stakeholders, other authorities and industry partners.

EASA has identified a number of key research projects in which to participate. EASA has also launched, under the Horizon Europe contribution agreement, a comprehensive research project (machine learning application approval (MLEAP[7])) to enable the identification of anticipated MOC for challenging objectives of the learning assurance framework.

5 (58) On the basis of its technical expertise, the Agency should assist the Commission in the definition of research policy and in the implementation of Union research programmes. It should be allowed to conduct research which is immediately needed and to participate in ad hoc research projects under the Union Framework Programme for Research and Innovation or other Union and non-Union private or public funding programmes.

6 Article 86 Research and innovation. The Agency shall assist the Commission and the Member States in identifying key research themes in the field of civil aviation to contribute to ensuring consistency and coordination between publicly funded research and development and policies falling within the scope of this Regulation. [...]

7 https://www.easa.europa.eu/en/research-projects/machine-learning-application-approval

3. Support to industry

A growing number of industry projects will include AI/ML in the near future. It is essential for the Agency to be in a position to actively support those projects.

In order to best support industry projects, the Agency will have to meet four challenges:

- to adapt the regulatory framework to allow for the early implementation of innovation;
- to overcome the knowledge asymmetry with industry;
- to integrate 'new entrants' (i.e. those new stakeholders not necessarily having the aviation safety culture) in the aviation community;
- to adapt processes to new technologies, including certification of AI.

Working in partnership with industry is at the centre of the Agency Strategy on Innovation. In fact, it is the only possible way for the Agency to meet the four above challenges.

Innovation partnership contracts (IPCs) and memoranda of cooperation/understanding (MoC/MoU) on innovation are the tools that EASA is using to collaborate on innovation with the industry.

IPCs aim at supporting innovative industry projects at the conceptual phase. On one hand, they allow the industry to benefit from the Agency technical expertise and aviation safety culture, which is particularly relevant to new entrants. On the other hand, they give the Agency the opportunity to learn from new technologies at an early stage of development, and thus to identify possible regulatory gaps and safety challenges.

MoC/MoU on innovation are broader partnership agreements, intended to facilitate the exchange of information on innovation between key industry players and the Agency. They allow the Agency to be kept aware of the most advanced developments on innovation but also to launch joint initiatives (training, workshops, and exchange of staff) that will further enhance the Agency understanding and vision of upcoming developments.

Support to industry also involves the follow-up of and participation in standards development activities. EASA actively participates in the joint EUROCAE/SAE WG-114/G-34 and closely monitors the progress of other committees on AI, in particular ISO/IEC SC42 and CEN CENELEC JTC21, with the ambition to ensure, as far as practicable, consistency between the EASA AI guidance and the AI-specific standards, including a shared understanding on general concepts and terminologies.

It is an Agency priority to intensify those partnerships with the industry regarding both AI as well as other technologies.

4. Impact on EASA processes

For a public administration authority like EASA the potential of AI is twofold. Firstly, AI has an economic potential as it can improve efficiency by automating routine tasks to support staff, optimising rational decision-making and in a wider horizon automating certain processes. Secondly, AI may support higher public service quality.

The examples below are based on actual cases already implemented in administrations or private organisations:

- Automatic text correction in forms and documents
- Analysis of letters and CVs for the recruitment process
- Analysis of contracts and legal documents
- Automatic legal checking in rulemaking (consistency, readability, legal certainty)

- Data-based organisation and standardisation auditing (POA, DOA, MOA)
- Assessment of the data provided by stakeholders in the certification and organisation approval processes.

The deployment of concrete AI use cases at Agency level will be assessed and managed as part of EASA's own digital transformation programme, in line with the EASA Integrated Management Standards (IMS), also taking into account the role they may play in facilitating and supporting the ongoing digitalisation of the EU aviation industry — in line with the Agency's related programme.

J. Time frame

The deployment of learning processes in projects for civil aircraft certification has already started. The Agency received the first project applications making limited use of AI/ML solutions.

The timeline proposed in version 1.0 of this Roadmap was based on an initial forecast from the major players in the specific domain of large commercial air transport and was already anticipated to require further adjustments. It offered the benefit of triggering an aggressive initial EASA AI Roadmap timeline, enabling development of early guidance in support of the first applications. This was achieved through the release of the two first EASA Concept Papers [2] and [3]. Feedback from stakeholders has been integrated in the new timeline as shown on Figure 7.

In the deployment of AI solutions, most of the industry players are envisaging first assistance scenarios, corresponding to Level 1 AI applications. This step is foreseen for 2025 in this document; however, it could even happen earlier considering the timeline of current applications.

Subsequent development will consist in the gradual ramp up to more automated solutions (Level 2 AI). This includes the deployment of progressively more automated solutions to assist the pilot flying in extended minimal crew operations (eMCO) and single-pilot operations (SPO) in large commercial air transport, or, in the ATM domain, for CDR through the use of virtual co-controllers. This is currently foreseen to happen around 2035; however, this is not a formal target but rather a prediction based on current prognostics from the major aviation stakeholders.

The next steps are advanced automation with human supervision (Level 3A AI) and ultimately without human oversight (Level 3B AI). These steps will likely take place in a progressive way between 2035 and 2050. We notice in certain aviation domains a push to get to the last step more quickly, especially from the drones industry; nevertheless, for CAT operations, a stepped approach will be observed.

The ultimate step towards autonomy (for which a further split in Level 3 AI will likely have to be made) is most probably to be shifted even after 2050, considering the current state of the art in AI technologies.

With an initial focus on examples from the domains of commercial air transport and ATM/UTM, the timeline associated with the three steps described above could be:

- First step: human assistance/augmentation (2023-2025+)
- Second step: human-AI teaming (2025-2035+)
- Third step: advanced automation and autonomous AI (2035-2050+)

Other domains, like innovative aerial services, maintenance or aerodromes are anticipated to develop AI solutions in parallel within a similar time frame and will benefit from the deliverables of this Roadmap.

The deliverables of the AI Roadmap are as far as possible aligned with the various industry roadmaps. The initial phase has started with the publication of the first usable guidance for Level 1 AI/ML (human assistance/augmentation) in 2021, followed by guidance on Level 2 (human-AI collaboration) in 2023. Guidance on Level 3 (advanced automation) will be produced for consultation in 2025. A consolidation phase will follow, which foresees the finalisation of the EASA AI/ML policy by 2028.

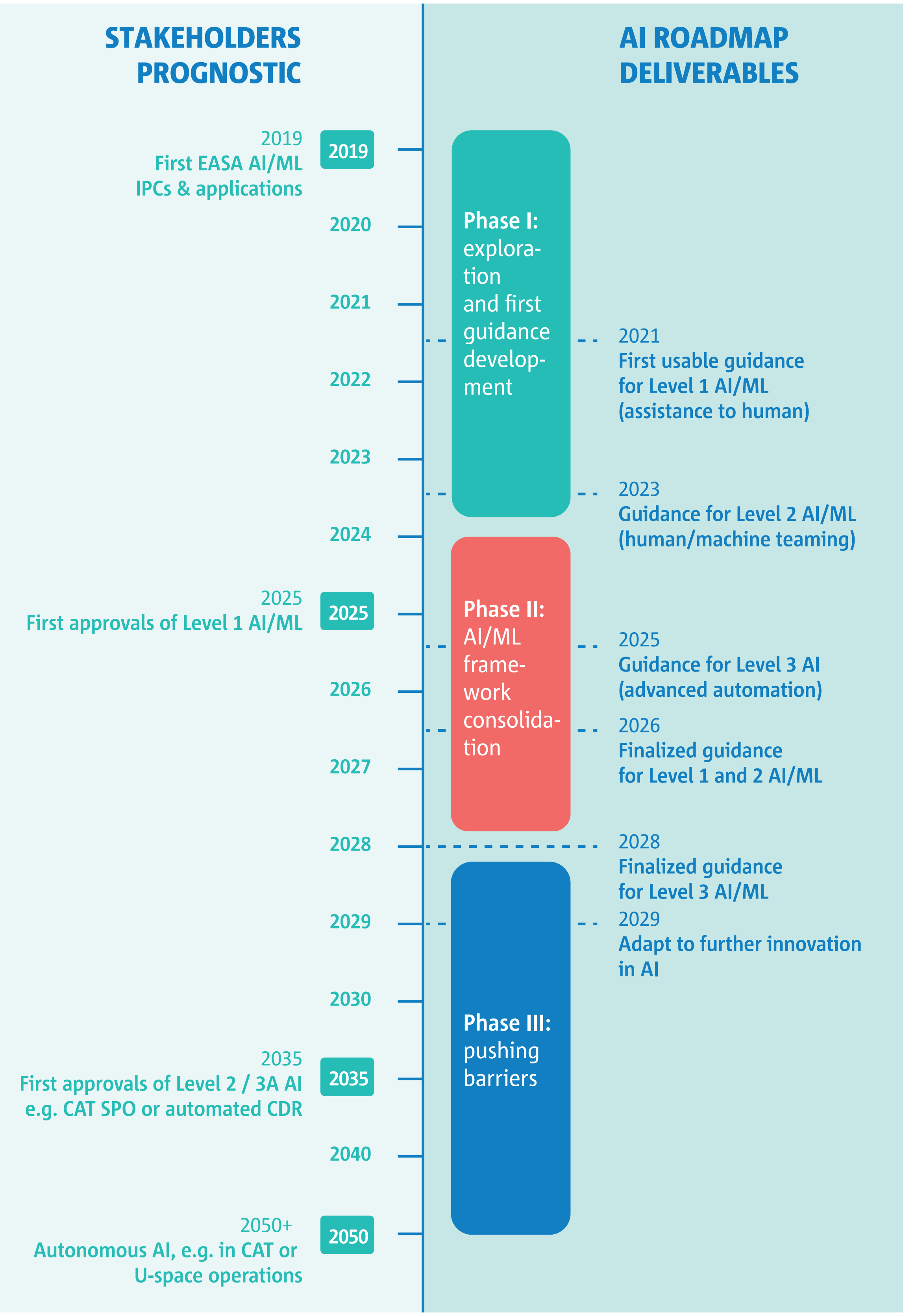


Figure 7: EASA AI Roadmap 2.0 timeline

K. Top 5 EASA AI Roadmap objectives

In line with the challenges identified above and the necessity of an AI trustworthiness concept, five top objectives have been identified:

1. Develop a human-centric AI trustworthiness framework
2. Make EASA a leading oversight authority for AI
3. Support European aviation industry leadership in AI
4. Contribute to an efficient European AI research agenda
5. Contribute actively to EU AI strategies and initiatives

Those top 5 objectives will be achieved:

A. with EASA staff ...

... by increasing internal awareness (seminars, workshops), identifying necessary skills and training needs for impacted staff and delivering training

... by gaining practical experience through involvement in industry projects and activities,

B. with EASA stakeholders...

... by developing and implementing long-term partnerships (MoC) with industry on AI and collaborating with industry on AI developments through IPCs,

... by contributing to industry standards AI development activities (working groups),

... by promoting EU policies and EASA best practices on AI,

C. with EU Commission, Member States & other institutions...

... by participating in EU Commission initiatives (e.g. AI HLEG) and ensuring that EU guidelines (e.g. on fairness, transparency, etc) are accounted for in EASA policy,

... by involving advisory bodies and innovation networks of Member States and NAAs,

D. with research institutes...

... by mapping research measures (existing/future), identifying research priorities and engaging with research organisations (scientific and technical knowledge),

... by participating in research activities.

L. EASA AI Roadmap 2.0

Top 5 EASA AI Roadmap objectives

1. Develop a human-centric AI trustworthiness framework
2. Make EASA a leading oversight authority for AI
3. Support European aviation industry leadership in AI
4. Contribute to an efficient European AI research agenda
5. Contribute actively to EU AI strategies and initiatives

EC Ethical Guidelines

- Human agency and oversight
- Technical robustness and safety
- Privacy and data governance
- Transparency
- Diversity, non-discrimination and fairness
- Societal and environmental well-being
- Accountability

EASA Trustworthy AI building blocks

- AI trustworthiness analysis
- AI assurance
- Human factors for AI
- AI safety risk mitigation

AI ROADMAP DELIVERABLES

- 2021 First usable guidance for Level 1 AI/ML (assistance to human)
- 2023 Guidance for Level 2 AI/ML (human/machine teaming)
- 2025 Guidance for Level 3 AI (advanced automation)
- 2026 Finalized guidance for Level 1 and 2 AI/ML
- 2028 Finalized guidance for Level 3 AI/ML
- 2029 Adapt to further innovation in AI

Phase I: exploration and first guidance development

Phase II: AI/ML framework consolidation

Phase III: pushing barriers

2019 2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030 2035 2040 2050

STAKEHOLDERS PROGNOSTIC

- 2019 First EASA AI/ML IPCs & applications
- 2025 First approvals of Level 1 AI/ML
- 2035 First approvals of Level 2 / 3A AI e.g. CAT SPO or automated CDR
- 2050+ Autonomous AI, e.g. in CAT or U-space operations

M. Consolidated action plan

Objectives
1. Develop a human-centric AI trustworthiness framework
2. Make EASA a leading oversight authority for AI
3. Support European aviation industry leadership in AI
4. Contribute to an efficient European AI research agenda
5. Contribute actively to EU AI strategies and initiatives

Actions...	1.	2.	3.	4.	5.	2023	2024	2025	2026	2027	2028+
A. ...with EASA staff											
1. Increase internal awareness (seminars, workshops)		X									
2. Identify the necessary skills and training needs for affected staff and deliver training		X									
3. Gain practical experience through involvement in industry projects and activities		X									
B. ...with EASA stakeholders											
1. Develop and implement long-term partnerships on AI with industry	X	X	X								
2. Collaborate with industry on AI developments through IPCs	X	X	X								
3. Contribute to industry standards development (working groups)	X	X	X								
4. Evaluate worldwide best practices on AI oversight	X	X	X								
5. Promote EU policies and EASA best practices on AI		X	X		X						
C. ...with EU Commission, Members States & other institutions											
1. Participate in EU Commission initiatives (e.g. AI Alliance)	X	X	X		X						
2. Ensure that EU AI regulation and directives are accounted for in the EASA AI policy	X	X	X		X						
3. Involve advisory bodies and innovation networks of Member States and NAAs		X	X								
4. Participate in other EU initiatives (e.g. FLY AI group)	X	X	X		X						

Objectives:
1. Develop a human-centric AI trustworthiness framework
2. Make EASA a leading oversight authority for AI
3. Support European aviation industry leadership in AI
4. Contribute to an efficient European AI research agenda
5. Contribute actively to EU AI strategies and initiatives

Actions...	1	2	3	4	5	2023	2024	2025	2026	2027	2028+
D. ...with research institutes											
1. Map research measures (existing/future) and identify research priorities			X	X							
2. Engage with research organisations (scientific and technical knowledge)			X	X							
3. Participate in research activities		X	X	X							
E. ...on deliverables											
1. Develop guidance for Level 1 & 2 AI	X	X	X								
2. Develop guidance for Level 3 AI	X	X	X								
3. Develop use cases in the affected domains to support guidance development	X	X	X								
4. Develop EASA AI demonstrators for testing the Concept Papers guidance	X	X	X								
5. Develop the final guidance based on experience feedback (mainly through rulemaking)	X	X	X								

N. Definitions

Adaptivity (of the learning process) — The ability to improve performance by learning from experience. [In the ML context,] **adaptive learning** refers to the learning capability during the operations (see also **online learning**).

Artificial intelligence (AI) — Technology that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with[8].

AI-based system — A system that is developed with one or more of the techniques and approaches listed in Annex I to the EU AI Act and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with[9].

Artificial neural network (ANN) or neural network (NN) — A computational graph which consists of connected nodes ('neurons') that define the order in which operations are performed on the input. Neurons are connected by edges which are parameterised by weights (and biases). Neurons are organised in layers, specifically an input layer, several intermediate layers, and an output layer. This document refers to a specific type of neural network that is particularly suited to process image data: convolutional neural networks (CNNs) which use parameterised convolution operations to compute their outputs.

Commonly used types of neural networks are to be highlighted:

— **Convolutional neural networks (CNNs)** — A specific type of deep neural networks that are particularly suited to process image data, based on convolution operators. [7]

— **Recurrent neural networks (RNNs)** — A type of neural network that involves directed cycles in memory.

Authority — The ability to make decisions and take actions without the need for approval from another member involved in the operations.

Automation — The use of control systems and information technologies reducing the need for human input, typically for repetitive tasks.

Autonomy — Characteristic of a system that is capable of modifying its intended domain of use or goal without external intervention, control or oversight[10].

Advanced automation — The use of a system that, under specified conditions, functions without human intervention[11].

Bias — Different definitions of bias have to be considered depending on the context:

— **Bias** (in the data) — [The common definition of data bias is that] the available data is not representative of the population or phenomenon of study.

— **Bias** (in the ML model) — An error from erroneous assumptions in the learning [process]. High bias can cause a learning algorithm to miss the relevant relations between attributes and target outputs (= underfitting).

Big Data — A recent and fast evolving technology, which allows the analysis of a big amount of data (more than terabytes), with a high velocity (high speed of data processing), from various sources (sensors, images, texts, etc.), and which might be unstructured (not standardised format).

Data-driven AI — An approach focusing on building a system that can learn a function based on having been trained on a large number of examples.

Data governance — A data management concept concerning the capability of an organisation to ensure that high data quality exists throughout the complete life cycle of the data, and data controls are implemented that support business objectives. The key focus areas of data governance include data availability, usability, consistency, integrity, and sharing. It also relates to establishing processes to ensure effective data management throughout the

8 Source: adapted from (EU Commission, 2021).
9 Source: (EU Commission, 2021)
10 Source: adapted from ISO/IEC 22989:2022(en), 3.1.7.
11 Source: adapted from ISO/IEC 22989:2022(en), 3.1.7.

enterprise, such as accountability for the adverse effects of poor data quality, and ensuring that the data which an enterprise has can be used by the entire organisation[12].

Data set[13] (in ML in general) — The sample of data used for various development phases of the model, i.e. the model training, the learning process verification, and the inference model verification.

— **Training data set** — Data that is input to an ML model in order to establish its behaviour.
— **Validation data set**— Used to tune a subset of the hyper-parameters of a model (e.g. number of hidden layers, learning rate, etc.).
— **Test data set**— Used to assess the performance of the model, independent of the training data set.

Data for safety (EASA) — Data4Safety (also known as D4S) is a data collection and analysis programme that supports the goal of ensuring the highest common level of safety and environmental protection for the European aviation system.

The programme aims at collecting and gathering all data that may support the management of safety risks at European level. More specifically, the programme facilitates better knowledge of where the risks are (safety issue identification), determine the nature of these risks (risk assessment) and verify whether the safety actions are delivering the needed level of safety (performance measurement). It aims to develop the capability to discover vulnerabilities in the system across terabytes of data [Source: EASA].

Decision — A conclusion or resolution reached after consideration[14]. A choice that is made about something after thinking about several possibilities[15].

Decision-making — The cognitive process resulting in the selection of a course of action among several possible alternative options[16]**.** Automated or automatic decision-making is the process of making a decision by automated means without any human involvement[17].

Deep learning (DL) — A specific type of machine learning based on the use of large neural networks to learn abstract representations of the input data by composing many layers.

Determinism — A system is deterministic if when given identical inputs, it produces identical outputs.

Development assurance — All those planned and systematic actions used to substantiate, to an adequate level of confidence, that errors in requirements, design, and implementation have been identified and corrected such that the system satisfies the applicable certification basis.

Domain — Operational area in which a system incorporating an ML subsystem could be implemented/used. Examples of domains considered in the scope of this guideline are ATM/ANS, air operations, flight crew training, environmental protection or aerodromes.

End user — An end user is the person that ultimately uses or is intended to ultimately use the AI-based system. This could either be a consumer or a professional within a public or private organisation. The end user stands in contrast to users who support or maintain the product[18].

Failure — An occurrence which affects the operation of a component, part or element such that it can no longer function as intended (this includes both loss of function and malfunction). Note: Errors may cause failures, but are not considered to be failures.

Fairness — Refers to ensuring equal opportunities and non-discriminatory practices applied to individuals or groups of users (or end users). (Definition based on EU guidelines on non-discrimination[19])

Inference — The process of feeding the AI model an input and computing its output. See also the related definition of **Training**.

12 Source: adapted from (EU High-Level Expert Group on AI, 2020).
13 Source: adapted from (ER-022 - EUROCAE, 2021).
14 Source: OxfordLanguages.
15 Source: adapted from the Cambridge Dictionary.
16 Source: adapted from Wikipedia.
17 Source: adapted from ico.org.uk.
18 Source: adapted from (EU High-Level Expert Group on AI, 2020).
19 Article 21 'Non-discrimination' | European Union Agency for Fundamental Rights (europa.eu)

Information security — The preservation of confidentiality, integrity, authenticity and availability of network and information systems.

Machine learning (ML) — The branch of AI concerned with the development of learning algorithms that allow computers to evolve behaviours based on observing data and making inferences on this data.

ML strategies include three methods:

— **Supervised learning** — The process of learning in which the learning algorithm processes the input data set, and a cost function measures the difference between the ML model output and the labelled data. The learning algorithm then adjusts the parameters to increase the accuracy of the ML model.

— **Unsupervised learning (or self-learning)** — The process of learning in which the learning algorithm processes the data set, and a cost function indicates whether the ML model has converged to a stable solution. The learning algorithm then adjusts the parameters to increase the accuracy of the ML model.

— **Reinforcement learning** — The process of learning in which the agent(s) is (are) rewarded positively or negatively based on the effect of the actions on the environment. The ML model parameters are updated from this trial-and-error sequence to optimise the outcome.

ML model — A parameterised function that maps inputs to outputs. The parameters are determined during the training process.

— **Trained model** — the ML model which is obtained at the end of the learning/training phase.

— **Inference model** — the ML model obtained after transformation of the trained model, so that the model is adapted to the target platform.

Natural language processing (NLP) — Refers to the branch of computer science — and more specifically, the branch of AI — concerned with giving computers the ability to understand text and spoken words in much the same way as human beings can [8].

Predictability — The degree to which a correct forecast of a system's state can be made quantitatively. Limitations on predictability could be caused by factors such as a lack of information or excessive complexity.

Robustness — The ability of a system to maintain its level of performance under all foreseeable conditions. At model level (trained or inference), the robustness objectives are further split into two groups: the ones pertaining to 'model stability' and the ones pertaining to 'robustness in adverse conditions'.

Surrogate model (or substitute model or emulation model) — is generally a mathematical model that is used to approximate the behaviour of a complex system. In the aviation industry, surrogate models are often used to represent the performance of aircraft, propulsion systems, structural dynamics, flight dynamics, and other complex systems. They can be particularly useful when it is not practical or cost-effective to use physical models or prototypes for testing or evaluation.

System — A combination of inter-related items arranged to perform a specific function(s) [ED-79A/ARP4754A]

Training — The process of optimising the parameters (weights) of an ML model given a data set and a task to achieve on that data set. For example, in supervised learning the training data consists of input (e.g. an image) / output (e.g. a class label) pairs and the ML model 'learns' the function that maps the input to the output, by optimising its internal parameters. See also the related definition of **Inference**.

User — A user is a person that supports or maintains the product, such as system administrators, database administrators, information technology experts, software professionals and computer technicians[20].

Variance — An error from sensitivity to small fluctuations in the training set. High variance can cause a learning algorithm to model the random noise in the training data, rather than the intended outputs (=overfitting).

20 Source: adapted from (EU High-Level Expert Group on AI, 2020).

0. Acronyms

AI	artificial intelligence
AMC	acceptable means of compliance
ANN	artificial neural network
ANS	air navigation services
ATCO	air traffic controller
ATM	air traffic management
CAT	commercial air transport
CNN	convolutional neural network
CDR	conflict detection and resolution
CS	certification specification
D4S	Data for Safety
DL	deep learning
DOA	design organisation approval
EASA	European Union Aviation Safety Agency
eMCO	Extended minimum crew operation
EU	European Union
EUROCAE	European Organisation for Civil Aviation Equipment
FDR	flight data recorder
HLEG	AI High-Level Expert Group
ML	machine learning
MLEAP	machine learning application approval
MOA	maintenance organisation approval
NAA	national aviation authorities
NLP	natural language processing
NN	neural network
RNN	recurrent neural network
SPO	single-pilot operation
SAE	Society of Automotive Engineering
UTM	unmanned aircraft systems traffic management

P. References

[1] EASA, 'EASA Artificial Intelligence Roadmap 1.0', European Union Aviation Safety Agency (EASA), Cologne, 2020.

[2] EASA, 'Concept Paper: First usable guidance for Level 1 machine learning applications — Issue 01', European Union Aviation Safety Agency (EASA), Cologne, 2021.

[3] EASA, 'Concept Paper: Guidance for Level 1 & 2 machine learning applications — Proposed Issue 02', European Union Aviation Safety Agency (EASA), Cologne, 2023.

[4] EU Commission, 'EU Commission — Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts, COM/2021/206 final', 2021.

[5] EU High Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI', European Commission, 2019.

[6] EU Commission, 'EU Commission — Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive)', 2022.

[7] Daedalean, 'Concepts of Design Assurance for Neural Networks (CoDANN)', EASA, Cologne, 2020.

[8] IBM Cloud Education, 'Natural Language Processing (NLP)', IBM, 2 July 2020. [Online]. Available: https://www.ibm.com/cloud/learn/natural-language-processing.

[9] EU Commission, 'Communication AI for Europe', 2018.

[10] ENISA, 'SECURING MACHINE LEARNING ALGORITHMS', (accessible at https://www.enisa.europa.eu/publications/securing-machine-learning-algorithms), December 2021.

[11] DEEL Certification Workgroup, 'White Paper - Machine Learning in Certified Systems', IRT StExupery, Toulouse, 2021.

[12] ER-022 — EUROCAE, 'Artificial Intelligence in aeronautical systems: Statement of concern', EUROCAE, 2021.

EASA
European Union Aviation Safety Agency

Postal address
Postfach 10 12 53
50452 Cologne
Germany

Visiting address
European Union Aviation Safety Agency
Konrad-Adenauer-Ufer 3, D-50668
Köln, Germany

Tel. +49 221 89990 - 000
Mail ai@easa.europa.eu
Web **www.easa.europa.eu/ai**