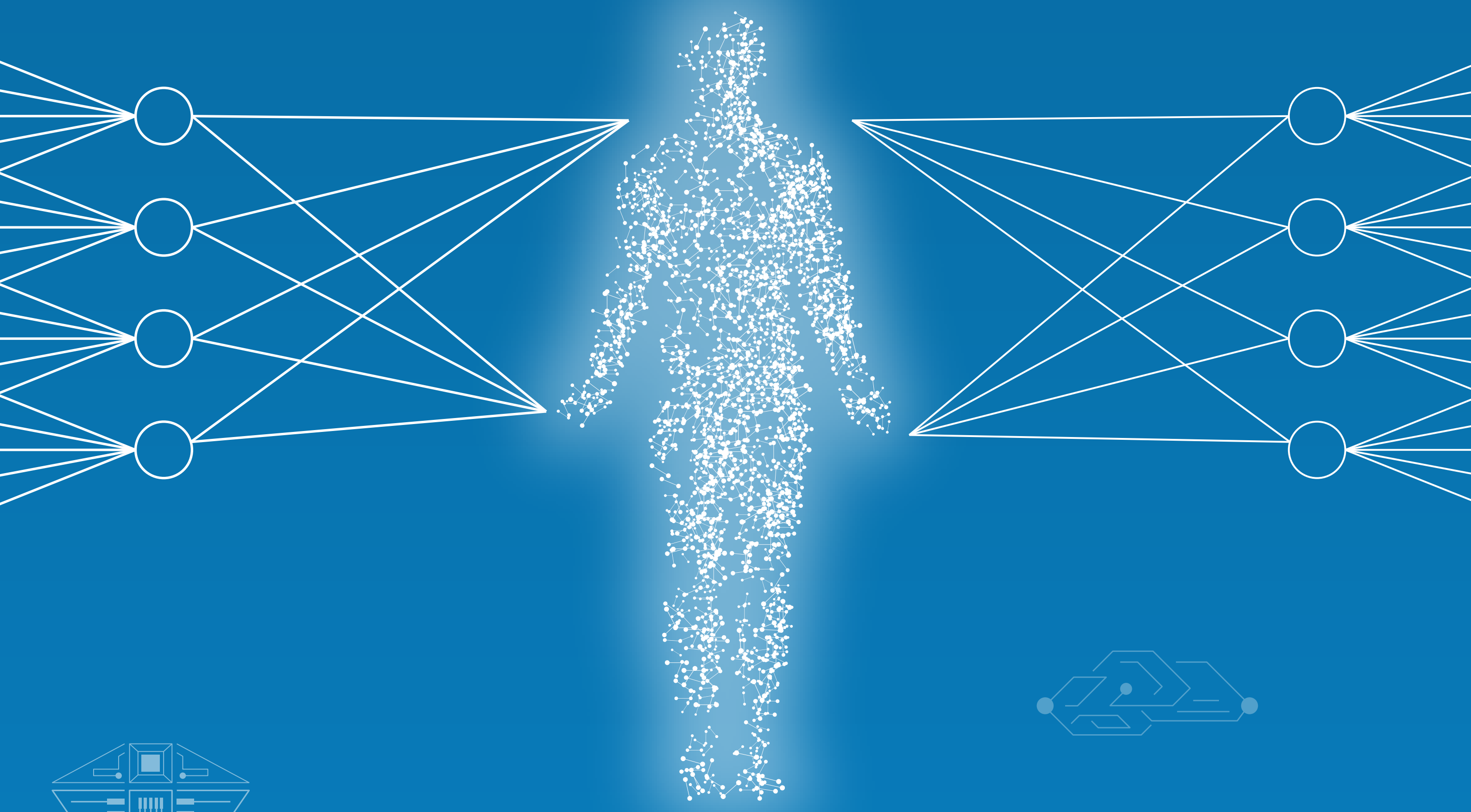
EASA
European Union Aviation Safety Agency

Artificial Intelligence Roadmap
A human-centric approach to AI in aviation



February 2020
Version 1.0

easa.europa.eu/ai

Disclaimer

The occurrence data presented is strictly for information purposes only. It is obtained from Agency databases comprised of data from ICAO, EASA Member States, Eurocontrol and the aviation industry. It reflects knowledge that was current at the time that the report 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 report should not be construed as legal advice. Acknowledgements The authors wish to acknowledge the contribution made by the Member States to thank them for their support in the conduct of this work and in the preparation of this report.

Photocredits

istock, Agustawestand

Contents

A. Foreword **2**
B. Introduction **3**
C. What is AI? **4**
D. The EU AI Strategy **5**
E. Impact of Machine Learning (ML) on aviation **7**
1. Aircraft design and operation 7
2. Aircraft production and maintenance 8
3. Air traffic management 8
4. Drones, urban air mobility and U-space 9
5. Safety risk management 9
6. Cybersecurity 10
7. Environment 11
8. EU regulations 11
F. Timeframe **12**
G. Trustworthiness of AI **14**
H. AI Trustworthiness building blocks **16**
1. AI trustworthiness analysis (including human-AI interface) 16
2. Learning assurance concept 17
3. Explainability of AI 18
4. AI safety risk mitigation 19
5. Relationship with AI trustworthiness requirements 20
I. Other challenges for EASA raised by the introduction of ML in aviation **21**
1. Staff competency 21
2. Support to the Industry 21
3. Research 22
4. Support to the EU AI Strategy and Initiatives 22
J. Top 5 EASA AI Roadmap Objectives **23**
K. EASA AI Roadmap **24**
L. Consolidated action plan **25**
M. Definitions **26**
N. References **29**
O. Internet sources **30**

A. Foreword

Aviation, and air transport in particular, has always been at the forefront of innovation. In its relatively brief history, innovations have significantly improved the passenger experience in terms of comfort, efficiency and safety.

The last two significant evolutions were the introduction of jet engines in the 1950s and fly-by-wire in the 1980s. Now, at the verge of the 2020s, one can foresee another evolution, probably unprecedented in terms of impact. With the tremendous improvements in the processing power of computers, the possibilities of Artificial intelligence (AI) will increasingly be used in aviation and make autonomous flights, preventive maintenance, ATM optimisation possible.

Among innovations, AI is probably the most disruptive. It raises a number of questions that the European Union Aviation Safety Agency (EASA), as a leading aviation safety agency, will have to answer:

- How to establish the public trust into AI-based systems?
- How to integrate the ethical dimension of AI (transparency, non-discrimination, fairness, etc.) in safety certification processes?
- How to prepare the certification of AI systems?
- What standards, protocols, methods do we need to develop to ensure that AI will further improve the current level of safety of air transport?

The purpose of this AI Roadmap is to discuss the implication of AI on the aviation sector and identify high-level objectives to be met and actions to be taken to respond to the above questions.

It develops in particular the central notion of **trustworthiness of AI** and addresses a number of challenges that the Agency will have to meet, in terms of staff competency, support to the industry, support to research, and support to the overall EU strategy and initiatives on AI.

Certainly, this work cannot be done by the Agency in isolation. EU Member States and EU industry shall be involved in the further elaboration and implementation of the roadmap. The EU Commission and other EU institutions shall support and facilitate our work.

EASA AI Roadmap 1.0 is an initial proposal, is a starting point intended to serve as a basis for discussion with the Agency stakeholders. It is intended to be a living document, which will be amended once a year and augmented, deepened, improved through discussions and exchanges of views, but also, practical work on AI development in which the Agency is already engaged.

Patrick KY
Executive Director
European Union Aviation Safety Agency

B. Introduction

Artificial intelligence (AI) is coming with a fast pace and being adopted widely, 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 already starts to be affected by this technological revolution.

As concerns the aviation sector, AI will not only affect the products and services provided by the industry; it will also trigger the rise of new business models. This will affect most of the domains under the mandate of the Agency. Its core processes (certification, rulemaking, organisation approvals, and standardisation) will be impacted. This will in turn affect 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 set up an internal task force on AI, with a view to developing a roadmap that would identify 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 courses of action that the Agency should undertake to meet those challenges.

The purpose of this Roadmap is not only to establish the Agency vision on the development of AI in the aviation domain, but also to create a basis for interaction with its stakeholders on this topic. In this perspective, this AI Roadmap 1.0 is 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.

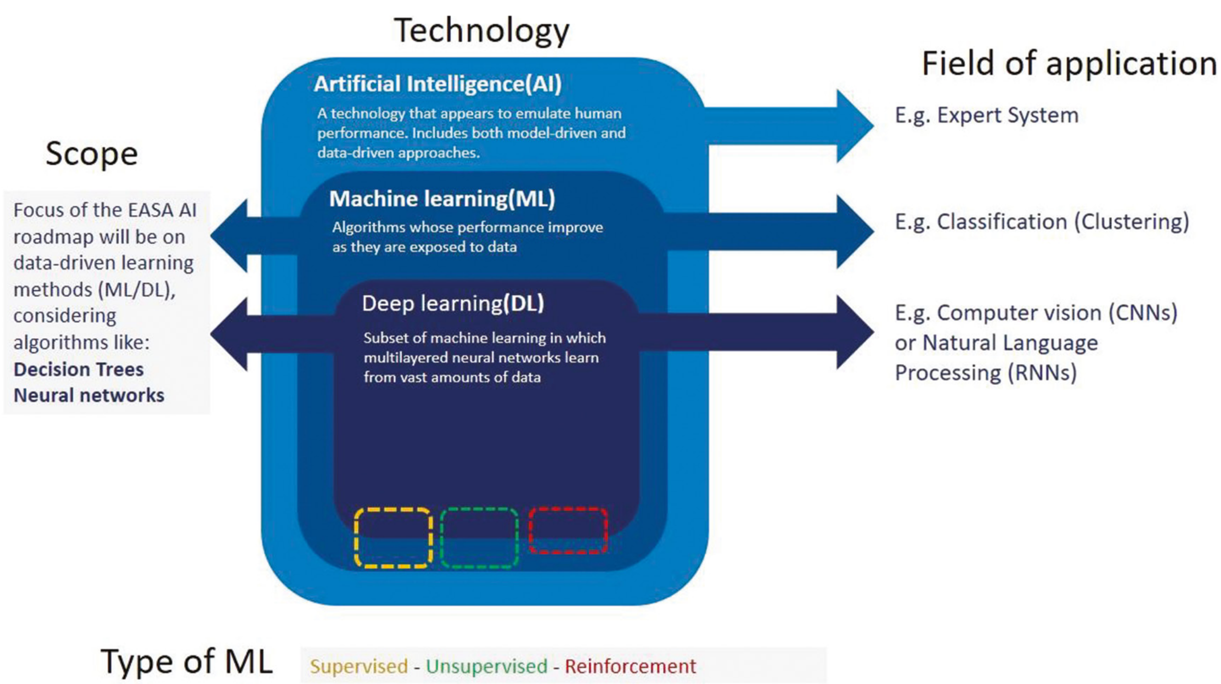
C. What is AI?

AI is a relatively old field of computer science that encompasses several techniques and covers a wide spectrum of applications. AI a broad term and its definition has evolved as technology developed. EASA therefore chose a wide-spectrum definition in this roadmap that is '**any technology that appears to emulate the performance of a human'**.

AI applications can be divided in **model-driven AI** (also known as symbolic AI) and **data-driven AI** (also known as statistical AI). Model-driven AI applications encompass for example the well-known 'expert systems' that are still programmed in a traditional way.

The current breakthrough is linked with **machine learning (ML)**, which is the use of data to train algorithms to improve their performance. **Deep learning (DL)** is a subset of ML that emerged through the use of deeper neural networks (NNs) that bring the learning capability a bit closer to the function of a human brain. It enables applications like computer vision and natural language processing (NLP) that were out of reach before the emergence of deep learning.

Figure 1. AI taxonomy in this Roadmap



Data-driven learning techniques are disruptive in essence and, by opposition to software development techniques, cannot be assessed through traditional approaches. Thus, they raise the need for developing novel approaches and trigger a specific focus in this roadmap.

For this reason, the EASA AI Roadmap will **focus on ML techniques** using, among others, learning decision trees or NN architectures. It is important to note that both data-driven and model-driven AI applications may be used in combination (also known as **hybrid AI**), which is also considered in the scope of 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. EU public and private sector will invest 20 billion EUR a year over the next decade. The European Commission will invest 1 billion EUR annually, mainly to support research, health and transport, and common data spaces, which are the AI 'fuel'. This is in addition to the 2.6 billion EUR already allocated in the frame of the Horizon 2020 programme.

Together with Member States and stakeholders, the Commission will start preparatory discussions to develop and implement a model for data sharing and making best use of common data spaces, with a focus notably on transport, healthcare and industrial manufacturing.

EU strategy for AI at a global level

International discussions on AI ethics have intensified after Japan's G7 Presidency put the topic high on the agenda in 2016. Given the international interconnections of AI development in terms of data circulation, algorithmic development and research investments, the Commission continues its efforts to bring the Union's approach to the global stage and build a consensus on a human-centric AI.

Therefore, the Commission stated priorities are to:

- explore the extent to which convergence can be achieved with third countries' draft ethics guidelines (e.g. Japan, Canada, Singapore), exploring how companies from non-EU countries and international organisations can contribute to the 'pilot phase' of the guidelines through testing and validation;

- continue to play an active role in international discussions and initiatives — in particular, to coordinate proposals for addressing the complex security challenges involved; and

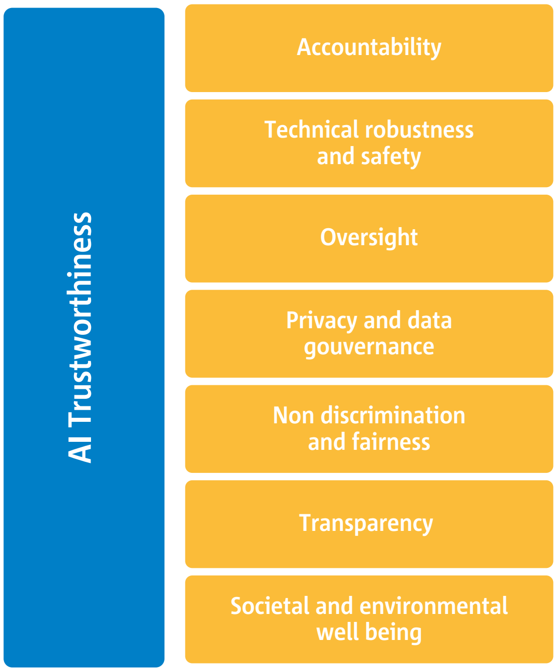
- contribute to relevant standardisation activities in international standards development.

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), 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 April 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 [1]:

Figure 2. Overview of the Ethics Guidelines from the EC High Level Expert Group



The guidelines developed by AI HLEG are non-binding and as such do not create any new legal obligations. The Commission proposes to all stakeholders, including regulators, to test the non-binding practical implementation of these ethical guidelines. The EASA strategy embraces this approach from an aviation perspective and EASA will participate in the testing and improvement of these guidelines.

Finally, the trustworthiness concept will necessarily be a key enabler of the societal acceptance of AI.

Other relevant initiatives for the development of AI in Europe

The DIGITALEUROPE[1] 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 40 national trade associations from across Europe as well as 67 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.

1 https://www.digitaleurope.org/

E. Impact of Machine Learning (ML) 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 provides a general overview of the anticipated impact on 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.

The current breakthrough of deep learning (DL) brings about a wide range of applications that could benefit aviation; in particular, computer vision and natural language processing (NLP). In aviation, these types of application could open the door to solutions such as high-resolution camera-based traffic detection or virtual assistance to the pilot.

The most discussed application of ML is autonomous flight. The drone market has paved the way and we can see now the emergence of new business models striving for the creation of air taxi systems to respond to the demand for urban air mobility. Autonomous 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 will come into play: to enable full autonomy, very powerful algorithms will be necessary to cope with the huge amount of data generated by the embedded sensors and by the machine-to-machine communications.

Beyond the 'holy grail' of autonomous flight, AI/ML is anticipated to open the way to the design of new systems that will change the relation between pilot and systems:

- reducing the use of human resources for tasks a machine can do, thus allowing them to better concentrate on high added-value tasks, in particular the safety of the flight;
- putting humans at the centre of complex decision processes, assisted by the machine; and
- addressing the impact of human performance limitations.

Although cockpit automation and AI are two different topics, 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.).

AI and ML 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.

Moreover, such techniques 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.

Finally, AI/ML can provide a solution for the modelling of physical phenomena. It could also be used for optimising qualification processes that rely on physical phenomenon demonstration (e.g. EMI, EMC, HIRF).

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 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 any more, 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 researches estimate that predictive maintenance can increase aircraft availability by up to 35 %[2].

3. Air traffic management

While today a number air traffic controllers' (ATCOs') automated assistance tools are already deployed and more advanced applications are covered within the SESAR R&D programme, AI is still not widely exploited in this programme. Nevertheless, automation is already a core focus of SESAR R&D, particularly in the automation of repetitive ATCO tasks. Providing more support here will enable pilots and controllers to focus on safety-critical tasks. Automation also proves invaluable for ensuring the seamless exchange of information and improved collaboration between all actors, including with the airborne side.

More recently, SESAR built up a portfolio of projects with specific Al applications, often using ML to process big data. These applications have been put to the test to better understand and address the underlying patterns of traffic and ATCO instructions.

Here are a few examples of some of the complex problems where Al can lend support, addressing all phases of flight from strategic and pre-tactical planning to tactical de-confliction and operations themselves. These are already being explored today [5].

- **Improving strategic planning**: The INTUIT project has developed visual analytics and ML techniques to understand trade-offs between key performance areas (safety, environment, capacity, efficiency).

2 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

- **Enhancing trajectory prediction**: The DART and COPTRA projects have developed a trajectory prediction capability based on ML to estimate aircraft performance before or during the flight. Results are based on models previously trained with recorded trajectories.

- **Enabling higher automation in ATC**: The Singapore ATM research institute has developed an AI application fed with recordings on human operator strategies for resolving conflicts. The prototype was finally capable of formulating resolution proposals assisting the ATCO.

- **Better understanding passenger behaviour**: The BigData4ATM project has investigated how different passenger-centric geo-located data can be analysed, to identify patterns in passenger behaviour, door-to-door travel times, and choices of travel mode. The aim is to enable optimised decision-making for the benefit of passengers and goods, and improved ready times provided to the network for more predictable operations.

- **Increasing the operational efficiency of air traffic control**: ATC instructions are for most of the time still given via VHF voice communication to the pilots. This means that controllers are making a lot of manual inputs to keep the system data correct. This is where automatic speech recognition can offer a viable alternative converting speech into text for input into the system. The MALORCA project (Machine Learning of Speech Recognition Models for Controller Assistance) has designed a versatile low-cost solution that adapts the speech recognition tools for use at any airport.

- **Refining time and wake separation**: When there are strong headwinds, aircraft ground speed is reduced on final approach. This results in a reduced landing rate, causing delays and even flight cancellations. SESAR's time-based separation aims at reducing the gap in landing rates in headwind conditions. Already deployed in Heathrow, the solution is currently further enhanced by ML algorithms that refine wake separation minima by combining downlinked parameters from the aircraft.

4. Drones, urban air mobility and U-space

The operations of drones, the implementation of U-space/UTM and operations in urban environment will only be possible with high levels of automation and use of disruptive technologies like AI or blockchain. Early implementation of ML solutions are a must for operations of small drones with different performance, performing very fast and unscheduled trajectories in a congested environment, such as the urban environment.

The implementation of U-space solutions to cope with large numbers of drones will not be possible using traditional approaches.

The development of detect and avoid (DAA) solutions is also expected to require the support of ML solutions, in particular for analysing images from radar or camera-based systems. AI/ML could also support contingency management, e.g. in case of C2-link loss.

Moreover, autonomous localisation/navigation (without GPS) solutions are anticipated to reap benefits from AI/ML techniques; for instance, by improving and simplifying current positioning sensors, data aggregation and the overall performance of the functions.

5. 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 seen as 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.

6. Cybersecurity

The cybersecurity domain encompasses three main actors:

- The system/organisation which has vulnerabilities that lead to the risk of being exploited causing 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, which mitigates one or more security risks.

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

- 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 behavioural basis, rather than rule-based (detection).

7. 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*.), machine learning 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.

8. EU regulations

Generally speaking, in the aircraft design domain, the current implementing rules (Part 21 and Certification Specifications) already create an open framework for the introduction of AI/ML solutions. In particular, paragraphs like CS 25.1309 could still be valid for evaluating the safety of AI/ML-based systems, provided additional means of compliance and standards are developed to answer the gap identified in the building blocks of this roadmap.

In the other domains (operations, maintenance, ATM, aerodromes), the current regulations provide an open framework for the use of AI/ML. However, it must be noted that those regulations will need to be adapted to the **specific applications** of AI/ML; for instance, reduced crew operation, predictive maintenance, etc.

It should be noted that the adaptation of regulations should be facilitated by the latest amendment of the EASA Basic Regulation (EU 2018/1139), which allows the Agency to better support the development of innovation, in particular through the use of the so-called performance-based regulations (see in particular recital 12[3] and Article 1.2[4] of said Regulation)

Considering the potential application of AI/ML solution in all the above domains, EASA will define a common policy that can be applied to any domain-related regulations (e.g. whether from an airworthiness CS or from the ATM rules), rather than issuing domain-specific guidance.

3 '(12) The measures taken in accordance with this Regulation to regulate civil aviation in the Union, and the delegated and implementing acts adopted on the basis thereof, should correspond and be proportionate to the nature and risks associated with the different types of aircraft, operations and activities they address. Such measures should also, in as far as possible, be formulated in a manner which focuses on objectives to be achieved, while allowing different means of achieving those objectives, and should also foster a systemic approach to civil aviation, taking into account interdependencies between safety and other technical domains of aviation regulation, including cybersecurity. This should contribute to a more cost-efficient achievement of required safety levels and to the stimulation of technical and operational innovation. Use should be made of recognised industry standards and practices, where it has been found that they ensure compliance with the essential requirements set out in this Regulation.'

4 '2.This Regulation further aims to: [...] (i) promote research and innovation, inter alia, in regulatory, certification and oversight processes;'

F. Timeframe

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

In the quest for autonomous flights, most of the industrial actors are envisaging first assistance scenarios to ramp up gradually to single-pilot operations through the use of virtual co-pilots, a next step being autonomous flights with human supervision and ultimately fully autonomous flights. We see, however, a push to get to the last step more quickly, especially from the drones industry; nevertheless, for CAT operations, a stepped approach is observed.

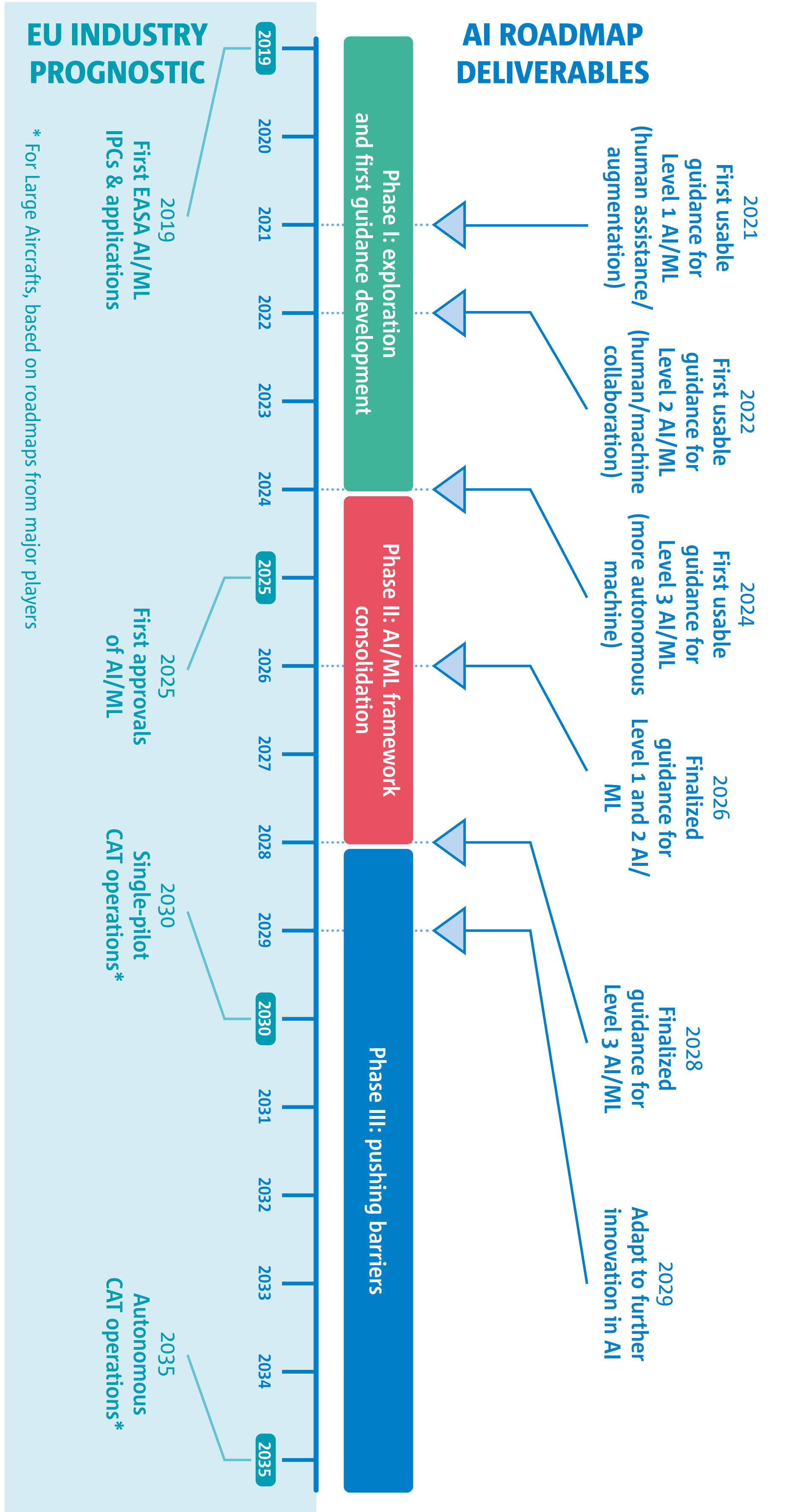
According to the industrial roadmaps known to the Agency, the first certifications of assistance to pilots are expected in 2025, with a gradual ramping up towards full autonomy in around 2035.

In the domain of commercial air transport, the timeline associated with the three steps described above could be:

- First step: crew assistance/augmentation (2022-2025)
- Second step: human/machine collaboration (2025-2030)
- Third step: autonomous commercial air transport (2035+)

In the UTM field, the first guidance should be available by 2021 to support the first applications of U-space and automated/semi-autonomous drones.

The deliverables of the AI Roadmap should be as far as possible aligned with the different industry roadmaps. In the initial phase, it will start 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/machine collaboration) in 2022 and Level 3 (autonomous machines) in 2024. A consolidation phase will follow, which foresees the finalisation of EASA AI/ML policy by 2028.


EU INDUSTRY PROGNOSTIC
AI ROADMAP DELIVERABLES
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
Phase I: exploration and first guidance development
Phase II: AI/ML framework consolidation
Phase III: pushing barriers
2021 First usable guidance for Level 1 AI/ML (human assistance/ augmentation)
2022 First usable guidance for Level 2 AI/ML (human/machine collaboration)
2024 First usable guidance for Level 3 AI/ML (more autonomous machine)
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
2019 First EASA AI/ML IPCs & applications
2025 First approvals of AI/ML
2030 Single-pilot CAT operations*
2035 Autonomous CAT operations*
* For Large Aircrafts, based on roadmaps from major players

G. Trustworthiness of AI

The power of ML lies in the capability for a system to learn from a set of data rather than requiring development and programming of each necessary decision path in a software.

This is a major opportunity for the aviation industry to shorten development cycles, but comes also with a consequent number of challenges with respect to the **trustworthiness** of ML/DL software. Here is an analysis of the ones we know of at the present time:

- **Traditional development assurance frameworks are not adapted to machine learning**
  ML puts additional emphasis on other parts of the process, namely data preparation, architecture and algorithms selection, hyper-parameters tuning, etc. There is a need for a shift in paradigm to develop specific assurance methodologies to deal with learning processes (see the associated 'learning assurance' building block).

- **Difficulties in keeping a comprehensive description of the intended function**
  It is anticipated that development assurance principles will still be used to capture the intended function at a higher level, and at low level to define the platform requirements (hardware + core software) that will be used as a resource for executing ML applications.

  Nevertheless, when it comes to learning processes, the behaviour will be by nature contained in the data set that is used to train the algorithm and will also depend on the learning process itself. It may become more challenging to maintain a traceability link with higher-level requirements and to ensure the completeness and correctness of the data set. Moreover the quality of the dataset will have a great importance as incomplete or incorrect data could influence the behaviour of the training model. The level of challenge will also depend on the ML process type: supervised learning typically implies a defined expected functional behaviour, whereas unsupervised learning or reinforcement learning may imply more unpredictable behaviour.

- **Lack of predictability and explainability of the ML application behaviour**
  ML applications are by nature probabilistic. Even if an ML model is deterministic from a mathematical perspective (e.g. fixed weights in a NN), for any new input, the output will depend on the correlation between that input with the data set that was used for the training process. This can lead to unpredictable outputs that may be difficult to explain. This is often confused with a lack of determinism of the algorithm, so it is clearer to speak about unpredictability of the ML application.
  Consequently, there is a need to increase the capability of making more understandable the conditions that led to a given output, further investigating notion of 'Explainability of AI' (see the associated 'explainability' building block).

- **Lack of guarantee of robustness and of no 'unintended function'**
  Due to this statistical nature of ML applications, they are subject to variability on 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 robustness of ML/DL applications, as well as to evaluate the completeness of the verification. The use of formal methods is anticipated by some research papers as a possible verification means and may as well be a means to compensate for the lack of coverage analyses.

- **Lack of standardised methods for evaluation of the operational performance of the ML/DL applications**
  Reference metrics on accuracy or error rate of an ML/DL application need to be investigated and established.

- **Issue of bias and variance in ML applications**
  While gaining experience, data scientists have figured out that ML solutions are subject to *bias* and *variance*, which can compromise the integrity of their outputs.

  One of the most challenging aspects when collecting, preparing or using data, is the capability to identify, detect and finally mitigate adequately any *bias* or *variance* that could have been introduced at any time during the data management and/or of the training processes.

- **Complexity of architectures and algorithms**
  - Convolutional neural networks (CNNs) will be used extensively in the field of computer vision. Even if more complex from an architecture perspective, if properly bounded, it is anticipated that they should be manageable as classical ANNs.
  - Recurrent neural networks (RNNs) raise however more complex issues (e.g. use of feedback loops) and may need specific guidance and boundaries.
  - Generative adversarial networks (GANs) could seemingly be used to enhance the training or complement the verification set for ML/DL. They are, however, the result of quite recent research and raise a lot of questions regarding the explainability of their output.

- **Adaptive learning processes**
  Real-time learning in operations is a parameter that will 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. This is at this stage considered a much more complex issue that may require to be bounded.

H. AI Trustworthiness building blocks

This document identifies four building blocks' that are considered essential in creating a framework for AI/ML trustworthiness and for enabling readiness for use of AI/ML in aviation. All four building blocks will require further investigation and research, activity that constitutes the backbone of this EASA AI Roadmap.

1. AI trustworthiness analysis (including human-AI interface)

General approach for the trustworthiness analysis building block

In order for the Agency to be in a position to evaluate the adherence of AI/ML applications to EU ethical principles, specific guidance for a comprehensive AI trustworthiness analysis need to be developed.

It should encompass the seven gears of the EU ethical guidelines (*accountability, technical robustness and safety, oversight, privacy and data governance, non-discrimination and fairness, transparency, societal and environmental well-being*), and for each of them, should provide a canvass of essential aspects to be considered by applicants when using AI/ML applications in their products or appliances design.

The thorough handling of this trustworthiness analysis and answering of the questions that it may raise regarding the ethical impact of the proposed AI/ML application, should be a prerequisite for proceeding with the evaluation of acceptability of this application in the frame of a project.

Focus on human-AI interface

To focus on one specific EU Guideline: the one dealing with 'oversight' considers different perspectives of the role of the human and of the machine. Three major scenarios are envisaged: human in the loop (HITL), human on the loop (HOTL), and human in control (HIC). The definitions of these scenarios still require further discussion as they could mean different things for different actors. However, pushing this concept to an extent adaptable to the aviation domain, we could come up with a classification of AI/ML applications in three levels, considering the degree of oversight of a human on the machine:

➨ **Figure 3.** Possible classification of AI/ML applications

Level 1 AI/ML : assistance to human	Level 2 AI/ML : human/machine collaboration	Level 3 AI/ML : more autonomous machine
• Level 1A – Routine assistance • Level 1B – Reinforced assistance	• Level 2A – Human performs a function / Machine monitors • Level 2B – Machine performs a function / Human monitors	• Machine performs functions with no human intervention in operations. Human is in the loop at design and oversight time

It appears that these scenarios can easily be mapped to the staged approach that most of the industrial stakeholders are planning for the deployment of AI/ML applications, starting with assisting functions, then making a step towards more human-machine collaboration, and at last seeking for more autonomy of the machine. This maps ideally to the increased automation scenario described in the aircraft operations domain impact analysis (in section E.2), but could seemingly also be generalised to any type of AI/ML application.

Taking the example of an autonomous flight:

- Level 1 AI/ML would concentrate on applications like assistance and augmentation of crew for tasks ranging from flight preparation to flight execution. Flight documentation preparation (MTO, performance, weight and balance, NOTAMs, aircraft status) is a routine task; however, it may lead to significant safety issues. The use of AI/ML on such tasks may generate important safety gains, as for instance, an ML application to prioritise NOTAMs for assisting the pilot in selecting the most critical ones (e.g. identification of NOTAMs related to dangerous areas). As a next step, the flight decision-making process may benefit from AI advice, with various degrees of criticality depending on the type of applications.

- Level 2 AI/ML would consist in enhanced human/machine collaboration still based on the fundamental crew role sharing flying/monitoring, with the human retaining full responsibility. In the most advanced step, the machine will perform the functions in autonomy, but still under the supervision of the human.

- Level 3 AI/ML would be the path towards the full autonomy scenario where the human is not in the operational loop; still, the human will be in the loop at the design and oversight phases. A further step would be that design and oversight phase would be to a great extent performed by the machine under the supervision of the human.

Proportionate guidance could be driven in particular by such a classification. The AI trustworthiness analysis building block should also be a tool to investigate further such a risk-based approach to AI/ML applications.

2. Learning assurance concept

In the current regulatory framework, the associated risk-based approach for systems, equipment and parts is mainly driven by a requirement-based 'development assurance' methodology during the development of their constituents. Although the system-level assurance might still require a requirement-based approach, it is admitted that the design-level layers that are relying on learning processes cannot be addressed with the existing 'development assurance' methods.

Intuitively, the assurance processes should be shifted on the training/verification data sets' correctness and completeness, on the identification and mitigation of biases, on the measurement of the accuracy and performance of an ML application, on the identification and the use of novel verification methods, etc. A brand new 'design assurance' paradigm is needed.

It is where a new concept of 'learning assurance' could provide novel means of compliance. The objective is to gain confidence at an appropriate level that an ML application supports the intended functionality, thus opening the 'AI black box' as much as practicable. This is a building block that may serve any application using ML or DL techniques, whatever the domain.

Anticipated key aspects of the learning assurance building block*:*

- **System development assurance**
  - Guidance on how to account for AI/ML in system safety assessment processes
  - Specification of the intended function when using AI/ML at design level
  - Impact of AI/ML on system architecture considerations

- **Data assurance process**
  - Definition of end-to-end process for data lifecycle management
  - Considerations on data quality

- **Training/verification data sets selection and validation**
  - Completeness/correctness of data sets v functional specifications
  - Identifying & mitigating bias in data sets
  - Consistency of distribution of test set vs training set
  - Independence between training and test data sets

- **Learning model selection and tuning**
  - Model architecture and algorithms
  - Hyperparameter tuning

- **Learning model evaluation**
  - Guidance on accuracy/error rate metrics
  - AI/ML performance and limitations and their inclusion in safety analyses
  - Guidance on bias and variance trade-off

- **Verification aspects**
  - Guidance regarding learning phase verification
  - Guidance on the inference phase verification (e.g. target testing, preservation of model properties)
  - Considerations on the use of novel verification methods (e.g. formal methods, adversarial methods like use of GANs)
  - Tool qualification considerations

- **Mixability aspects**
  - Interface with other components developed using development assurance
  - Mixability of AI/ML-based systems with other systems and with human operators

- **Adaptivity aspects**
  - More complex issue when learning continues in operations

- **Changes to learning models**
  - Changes to training data sets
  - Reuse/transfer learning/re-training of portions of ANN versus full learning process

- **Accident/incident investigation**
  - Explainability is a key concept (see next paragraph)
  - Need for enhanced data recorders

3. Explainability of AI

Explainability of AI is a concept that is resolutely human-centric. It deals with the capability to provide human understandable explanation on how an AI application is coming to its results and outputs.

An example could be a CNN that would classify pictures between wolves and dogs. There was this noticeable case where the CNN was using the snowy background as a criterion for classifying the animal as a wolf, with a relatively good accuracy. The identification of the evident bias may be possible only through an analysis of the features that are retained by deeper layers of the CNN to perform the classification, which can be a form of explainability of the way the CNN is coming to its conclusions.

The question is how to generalise, as far as practicable, such a capability. What would explainability mean when using AI/ML for applications other than computer vision, in particular when it comes down to decision-making processes? It will surely involve a great deal of human-machine interface and human factors considerations.

Several research initiatives have already started on this topic, the most relevant ones being:

- The DEEL (DEpendable Explainable Learning) project, which associates industrials and laboratories to develop robust, certifiable and dependable AI technological bricks for critical systems. DEEL is also being carried out in collaboration with Quebec partners (IVADO and CRIAQ) for an amount of €30M over 5 years. DEEL Partners: Airbus, Thales Avionics, Continental, Renault, Renault SW Lab, SAFRAN, Airbus DS, TAS, SNCF, etc.

- The DARPA Explainable AI programme aims to create a suite of ML techniques that:

  › produce more explainable models, while maintaining a high level of learning performance (prediction accuracy); and
  › enable human users to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners.

4. AI safety risk mitigation

AI safety risk mitigation is based on the anticipation that the 'AI black box' may not always be opened to a sufficient extent and that supervision of the function of the AI application may be suitable to the necessary extent.

This could be achieved by several means, among others:

- keeping a human in command (HIC) or in the loop (HITL);

- monitoring of the output of the AI/ML and passivation of the AI/ML application with recovery through a traditional backup system (e.g. safety net);

- encapsulation of ML with rule-based approaches (e.g. hybrid AI);

- monitoring of AI through an independent AI agent; and

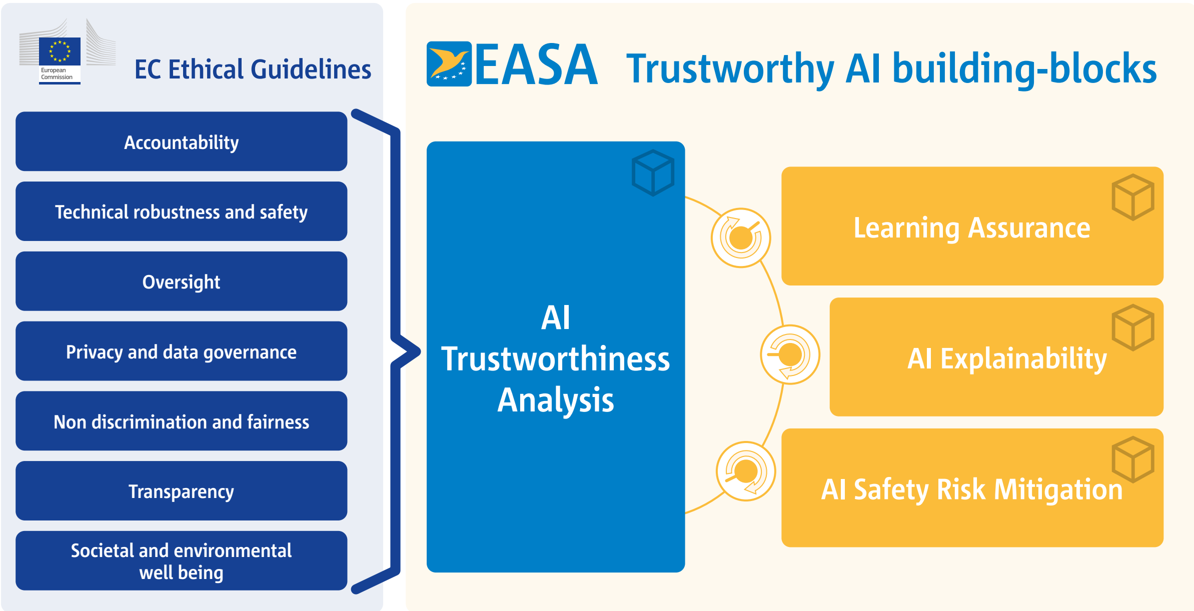
- in a wider horizon by considering the notion of 'licensing' to an AI, as anticipated in [7] and developed further in [4].

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

5. Relationship with AI trustworthiness requirements

The four building blocks above could be organised as follows:

Figure 5. Relationship between AI Roadmap building blocks and AI trustworthiness



The trustworthiness analysis would serve as an interface with the set of EU ethical guidelines and as such would create a gate to the three other technical building blocks. A certain degree of obvious traceability already exists between those three technical building blocks and some of the seven key guidelines (e.g. the 'Learning Assurance building-block' relates directly to 'Technical Robustness and Safety'). Other guidelines like 'Societal and environmental well-being' will require dedicated guidance within the trustworthiness analysis building block.

All four building blocks are anticipated to have an importance in gaining confidence in the trustworthiness of an AI/ML application. Whereas the trustworthiness analysis will be always necessary, the three other blocks could be weighted (represented on Figure 9 through potentiometers), depending on parameters like the category of AI/ML application, the criticality of the function or the relevance of the safety mitigations.

The development of the four AI trustworthiness building blocks is the main objective of the EASA AI Roadmap. To that end, EASA will work with all involved stakeholders in projects that can contribute to the enrichment of this future certification framework. At the time of publication of this document, EASA has already launched Innovation Partnership Contracts (IPCs) with key aviation industry partners and with start-ups developing AI-based solutions. EASA is also part of the EUROCAE WG-114 and SAE G-34 working groups that deal with the approval of AI and sees a great potential for progressing on standard guidance for the different building blocks of the roadmap.

I. Other challenges for EASA raised by the introduction of ML 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 the 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 or involved in the development of AI. This poses the risk of having 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 will it be necessary to train EASA personnel on AI but also to expose them to AI/ML practices as soon as possible.

A comprehensive training package should be developed to ensure peer-to-peer discussions with our industry and regulatory authorities.

Partnership agreements with the industry, as explained in paragraph 2 below should also be the vehicle for practical exposure to AI.

2. Support to the Industry

A growing number of industry projects will include AI/ML in the close 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 be able to adapt our regulatory framework to allow the early implementation of innovation;
- to adapt our processes to new technologies, including certification of AI;
- 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.

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 (MoC) on innovation are the tools that EASA has recently developed 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 particular 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, to identify possible regulatory gaps and safety challenges.

MoC on innovation are broader partnership agreements, intended to facilitate the exchange of information on innovation between industry key players and the Agency. They will 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.

MoC have already been signed, or are currently under discussion, with key players in the field of aeroplane, helicopter, engine, systems, aerodrome equipment manufacturers and maintenance, repair and overhaul (MRO) organisations.

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

3. Research

The need to support to 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 High Level Expert Group on AI indicate that Europe must aim for scientific AI leadership, by focusing the power of its multi-faceted and distributed research excellence with several worldwide renowned Centres of Excellence to establish and demonstrate its intellectual and commercial leadership in AI. 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 should also come up with a dedicated coordinated plan, following a three-step approach:

- Mapping of research initiatives
- Identification of key research gaps
- Setting-up of research priorities for the Agency in terms of AI/ML in collaboration with all research stakeholders.

Once priorities are set, EASA should follow up relevant research projects and promote or launch the missing ones.

4. Support to the EU AI Strategy and Initiatives

The development of AI in the aviation domains should not be considered in isolation from the other industrial domains. It is important that the development of AI in aviation remain within the framework of the high-level guidelines developed by the EU. The Agency should play a significant role by ensuring that those guidelines are known and applied by the EU aviation industry, but also by providing feedback to the EU Commission.

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** 1.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. [...]'

J. Top 5 EASA AI Roadmap Objectives

In order to meet the five above challenges, five top objectives have been identified:

1. **Develop a human-centric AI Trustworthiness framework**
2. **Make EASA a leading certification authority for AI**
3. **Support European Aviation leadership in AI**
4. **Contribute to an efficient European AI research agenda**
5. **Contribute actively to EU AI strategy and initiatives**

Those top 5 objectives will be achieved:

A. with EASA staff ...

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

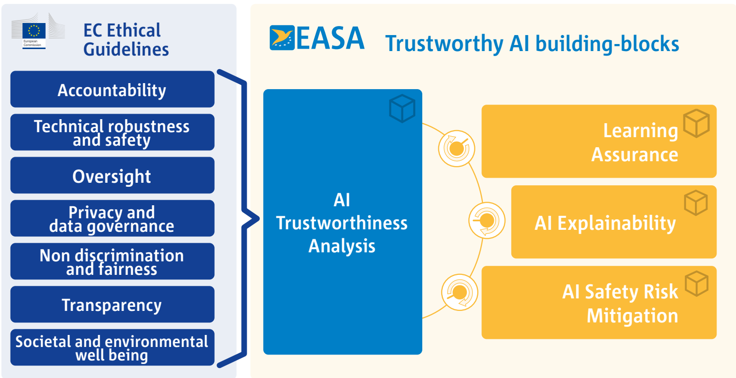
C. with EU Commission, Members States & other Institutions...

...by liaising with 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 innovation networks of NAAs,
...by participating in industry standards AI development activities (working groups),

D. with Research Institutes...

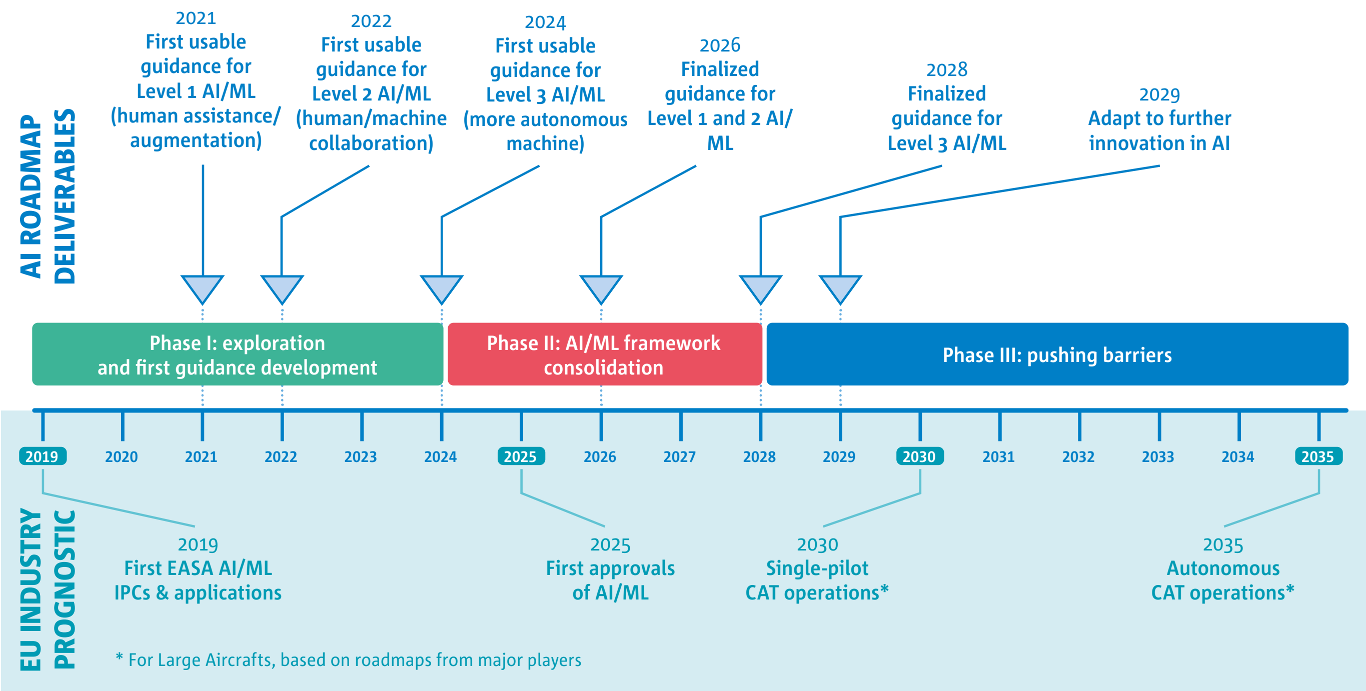
...by mapping research measures (existing/future), identifying research priorities and engaging with research organisations (scientific and technical knowledge),
...by participating in research activities.

K.
EASA AI Roadmap



Top 5 EASA AI Roadmap Objectives

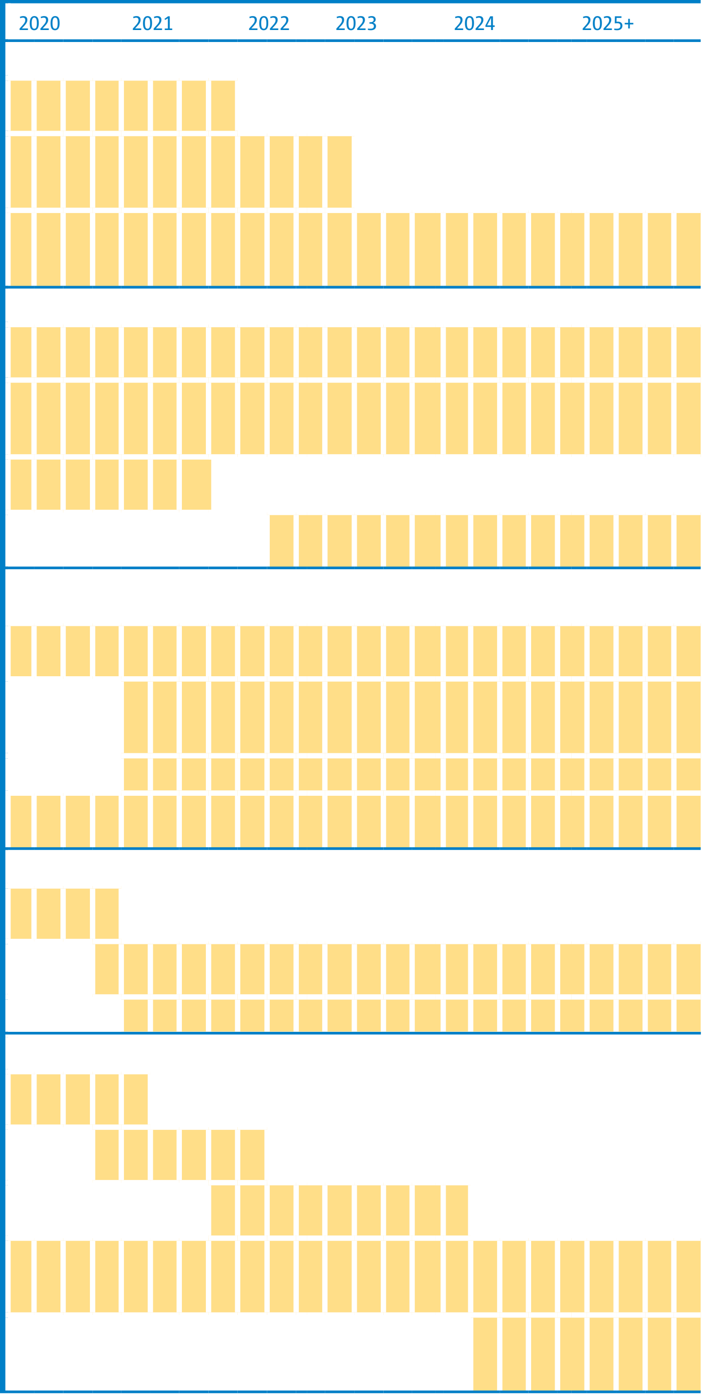
1. Develop a human-centric AI Trustworthiness framework
2. Make EASA a leading certification authority for AI
3. Support European Aviation leadership in AI
4. Contribute to an efficient European AI research agenda
5. Contribute actively to EU AI strategy and initiatives



L.
Consolidated action plan

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

Actions...	1.	2.	3.	4.	5.
A. ...with EASA staff					
1. Create 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 of industry projects and activities		x			
B. ...with EASA stakeholders					
1. Develop and implement long-term partnerships (MoC) with industry on AI	x	x	x		
2. Collaborate with industry on AI developments through Innovation Partnership Contracts (IPCs)	x	x	x		
3. Evaluate worldwide best practices on AI oversight	x	x	x		
4. Promote EU policies and EASA best practices on AI		x	x		x
C. ...with EU Commission, Members States & other Institutions					
1. Liaise with EU Commission initiatives (e.g. AI HLEG)	x	x	x		x
2. Ensure that EU guidelines (e.g. on fairness, transparency, etc.) are accounted for in the EASA policy	x	x	x		x
3. Involve innovation networks of NAAs		x	x		
4. Participate in industry standards development (working groups)	x	x	x		x
D. ...with Research Institutes					
1. Map research measures (existing/future) & 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 first usable guidance for level 1 AI/ML	x	x	x		
2. Develop first usable guidance for level 2 AI/ML	x	x	x		
3. Develop first usable guidance for level 3 AI/ML	x	x	x		
4. Capitalise on return of experience on initial guidance in projects (IPC, research, certification, etc.)	x	x	x		
5. Develop final guidance based on experience feedback (mainly through rulemaking)	x	x	x		



M. Definitions

Adaptivity - the ability to improve performance by learning from experience[7]. [In the Machine Learning context], a system is considered **adaptive** when it continues to learn in real-time operation.

Artificial intelligence (AI) - technology that appears to **emulate human performance** typically by learning, coming to its own conclusions, appearing to understand complex content, engaging in natural dialogues with people, enhancing human cognitive performance (also known as cognitive computing) or replacing people on execution of non-routine tasks. Applications include autonomous vehicles, automatic speech recognition and generation, and detection of novel concepts and abstractions (useful for detecting potential new risks and aiding humans to quickly understand very large bodies of ever-changing information)[8].

Artificial neural network (ANN) or neural network (NN) - a computational graph that 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 specialised kind of neural network for processing data that has a known grid-like topology. [...] Convolutional networks use convolution [a specialised kind of linear operation] in place of general matrix multiplication in at least one of their layers. [8, p. 326]

- **Recurrent neural networks (RNNs)** - a type of advanced artificial neural network (ANN) that involves directed cycles in memory. One aspect of recurrent neural networks is the ability to build on earlier types of networks with fixed-size input vectors and output vectors[9].

- **Generative adversarial networks (GANs)** - a type of construct in neural network technology that is composed of two neural networks: a generative network and a discriminative network. These work together to provide high-level simulation of conceptual tasks[10].

Automation - the use of control systems and information technologies reducing the need for human [input][11], typically for repetitive tasks.

Autonomy - the ability to perform tasks in complex environments without [input] by a [human][12].

Bias - an error from erroneous assumptions in the learning [process]. High bias can cause an algorithm to miss the relevant relations between attributes and target outputs (=underfitting).

7 Source: https://course.elementsofai.com/
8 Source: https://www.gartner.com/it-glossary/artificial-intelligence
9 Source: https://www.techopedia.com/definition/32834/recurrent-neural-network-rnn
10 Source: https://www.techopedia.com/definition/32515/generative-adversarial-network-gan
11 Source: adapted from http://aviationknowledge.wikidot.com/aviation:automation
12 Source: adapted from https://course.elementsofai.com/

Big Data - a new technology, which allows the analysis of 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).

DARPA (Defence Advanced Research Projects Agency) - a US Government Agency whose mission is to make pivotal investments in breakthrough technologies for national security.

Data-driven AI - the data-driven [approach] focuses on building a system that can [learn] what is the right answer based on having [trained on] a large number of [labelled] examples.[13]

Data science - a broad field that refers to the collective processes, theories, concepts, tools and technologies that enable the review, analysis and extraction of valuable knowledge and information from raw data[14].

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. This includes safety reports (or occurrences), flight data (i.e. data generated by the aircraft via the flight data recorders), surveillance data (air traffic data), weather data - but those are only a few from a much longer list.

As for the analysis, the programme's ultimate goal is to help to 'know where to look' and to 'see it coming'. In other words, it will support the performance-based environment and set up a more predictive system.

More specifically, the programme will facilitate 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].

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.

Inference - the process of feeding the machine learning model an input and computing its output. See also related definition of 'Training'.

Internet of things (IoT) - the network of physical objects that contain embedded technology to communicate and sense or interact with their internal states or the external environment[15].

Machine learning (ML) - rooted in statistics and mathematical optimisation, *machine learning* is the ability of computer systems to improve their performance by exposure to data without the need to follow explicitly programmed instructions[16]. [*Machine learning* is a branch of artificial intelligence].

13 Source: https://hackernoon.com/building-ai-software-data-driven-vs-model-driven-ai-and-why-we-need-an-ai-specific-software-640f74aaf78f

14 Source: https://www.techopedia.com/definition/30202/data-science

15 Source: https://www.gartner.com/it-glossary/internet-of-things

16 Source: https://www.enterpriseai.news/2018/01/19/ai-definitions-machine-learning-vs-deep-learning-vs-cognitive-computing-vs-robotics-vs-strong-ai/

Machine learning strategies go by three methods:

- **Supervised learning** - the process of learning a function that maps an input to an output based on example input-output training samples.

- **Unsupervised learning** - this strategy is used in cases where there is no labelled data set available to learn from. The neural network analyses the data set, and then a cost function tells the neural network how far off target it was. The neural network then adjusts to increase accuracy of the algorithm.

- **Reinforcement learning** - in this algorithm, the neural network is reinforced for positive results, and punished for a negative result, forcing the neural network to learn over time.

Machine learning model - a parameterised function that maps inputs to outputs. The parameters are determined during the training process.

Model-driven AI - model-driven AI (or symbolic AI), instead, attempts to capture knowledge and derive decisions through explicit representation and rules.[17]

Predictability/determinism - a system is predictable/deterministic if when given identical inputs, produces identical outputs.

Robustness - for an input varying in a region of the state space, the system is producing the same outputs.

Safety science - a broad field that refers to the collective processes, theories, concepts, tools and technologies that support safety management.

Training - the process of optimising the parameters (weights) of a machine learning 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 machine learning model 'learns' the function that maps the input to the output, by optimising its internal parameters. See also related definition of 'Inference'.

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

17 Source: https://hackernoon.com/building-ai-software-data-driven-vs-model-driven-ai-and-why-we-need-an-ai-specific-software-640f74aaf78f

N.
References

[1]	EU High Level Expert Group on AI, 'Ethics Guidelines for Trustworthy AI', European Commission, 2019.
[2]	Yoav Shoham, Raymond Perrault, Erik Brynjolfsson, Jack Clark, James Manyika, Juan Carlos Niebles, Terah Lyons, John, 'The AI Index 2018 Annual Report', Stanford University, Stanford, CA, 2018.
[3]	EU Commission, 'Communication AI for Europe', 2018.
[4]	ECATA Group, 'ECATA Technical Report 2019 - The exploitation of Artificial Intelligence in future Aircraft Systems', 2019.
[5]	SESAR JU, 'Automation levels of ATC Systems - SESAR Model and Roadmap', 2018.
[6]	D. Gunning, 'Explainable Artificial Intelligence (XAI) - Program Update November 2017', DARPA, 2017.
[7]	Siddhartha Bhattacharyya and Darren Coffer, 'NASA/CR-215-218702 - Certification Considerations for Adaptive Systems', NASA, 2015.
[8]	Goodfellow-et-al, Deep Learning, MIT Press, 2016.

0.
Internet sources

Digital Europe Source: https://www.digitaleurope.org/

Science|Business 'US and China lead investments in artificial intelligence start-ups'
Source: https://sciencebusiness.net/news-byte/us-and-china-lead-investments-artificial-intelligence-start-ups

Government of China 'Next Generation Artificial Intelligence Development Plan'
Source: http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm

Tim Dutton, Senior Policy Advisor at Office of the Premier of Ontario, 'An Overview of National AI Strategies'
Source: https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd

Reaktor and the University of Helsinki, 'Elements of AI'
Source: https://course.elementsofai.com/

Gartner IT Glossary, 'Artificial Intelligence (AI)'
Source: https://www.gartner.com/it-glossary/artificial-intelligence

Techopedia, 'Recurrent Neural Network (RNN)'
Source: https://www.techopedia.com/definition/32834/recurrent-neural-network-rnn

Techopedia, 'Generative Adversarial Network (GAN)'
Source: https://www.techopedia.com/definition/32515/generative-adversarial-network-gan

AviationKnowledge, 'Automation in aviation'
Source: adapted from http://aviationknowledge.wikidot.com/aviation:automation

Reaktor and the University of Helsinki, 'Elements of AI'
Source: adapted from https://course.elementsofai.com/

Techopedia, 'Data Science'
Source: https://www.techopedia.com/definition/30202/data-science

Gartner IT Glossary, 'Internet of Things (IoT)'
Source: https://www.gartner.com/it-glossary/internet-of-things

Enterprice AI, 'AI Definitions: Machine Learning vs. Deep Learning vs. Cognitive Computing vs. Robotics vs. Strong AI....'
Source: https://www.enterpriseai.news/2018/01/19/ai-definitions-machine-learning-vs-deep-learning-vs-cognitive-computing-vs-robotics-vs-strong-ai/

Hackernoon Blog, 'Data-driven and Model-driven AI'
Source: https://hackernoon.com/building-ai-software-data-driven-vs-model-driven-ai-and-why-we-need-an-ai-specific-software-640f74aaf78f

EASA
European Union Aviation Safety Agency

Stay informed:

easa.europa.eu/ai

European Union Aviation Safety Agency

Postal address
Postfach 10 12 53
50452 Cologne
Germany

Visiting address
Konrad-Adenauer-Ufer 3
50668 Cologne
Germany

Tel. +49 221 89990 - 000
Fax +49 221 89990 - 999
Web www.easa.europa.eu