Human-Centered Aerospace Systems and Sustainability Applications, Vol. 26, 2022, 35–45
https://doi.org/10.54941/ahfe1001432

Impact of Artificial Intelligence in the Certification of Human-Centered Aviation Systems

Rosa Maria Arnaldo Valdés, Victor Fernando Gómez Comendador, Maria Zamarreño Suarez, Francisco Pérez Moreno, and Raquel Delgado-Aguilera Jurado

Universidad Politécnica de Madrid, Madrid 28040, Spain

ABSTRACT

In recent years we have witnessed the emergence of applications based on artificial intelligence (AI) in the aviation industry. To address the challenges of enabling readiness for use of human-centric AI, civil aviation authorities must anticipate the unprecedented impact of AI on human-centric aerospace systems and answer a number of critical questions. The starting point for the certification of human-centric AI in aerospace systems. It develops in particular the core notion of trustworthiness of AI, and proposes a framework based on four human-centric AI trustworthiness building blocks: Trustworthiness analysis, Learning assurance, Explainability, and Safety risk mitigation. This paper discusses and revises the 4 elements of the trustworthiness of human-centric AI framework proposed by EASA, and based on this discussion anticipates the possible impacts of the introduction of human-centric AI in the Aviation Certification Regulation.

Keywords: Artificial Intelligence, Human-centric, Aviation, Certification, Trustworthiness, Learning assurance, Explainability, Safety risk mitigation

INTRODUCTION

In recent years we have witnessed the emergence of applications based on artificial intelligence in the aviation industry. This technology is said to be promoting a new era or evolution, such as the introduction of jet engines in the 1950s and fly-by-wire in the 1980s. To maintain aviation safety standards in this transition, civil aviation authorities responsible for certifying aerospace systems must anticipate the unprecedented impact of AI on human-centric aerospace systems and answer a number of critical questions:

- How to establish public trust in human- centric AI-based systems?
- How to integrate the ethical dimension of human- centric AI (transparency, non-discrimination, fairness, etc.) in safety certification processes?
- How to prepare for the certification of human- centric AI systems?
- What standards, protocols, methods need to be developed to ensure that human- centric AI further improves the current level of air transport safety?

EASA, the European Aviation Safety Agency, has recently developed a roadmap for the certification of AI applications in aviation, which analyses the involvement of human- centric AI in the aviation sector and identifies the objectives that must be met, and the actions that must be taken to respond to the previous questions.

This effort constitutes a starting point for the certification of human-centric AI in aerospace systems. It develops in particular the core notion of trustworthiness of AI in human centred systems, and proposes a framework based on four human-centric AI trustworthiness building blocks:

- Trustworthiness analysis,
- Learning assurance,
- Explainability,
- Safety risk mitigation.

This paper discusses and revises the 4 elements of the trustworthiness of human-centric AI framework proposed by EASA, and based on this discussion anticipates the possible impacts of the introduction of human-centric AI in the different Implementation Rules (IR), Certification Specifications (CS), Acceptable Means of Compliance (AMC) and guidance material (GM) in the domains covered by the EASA Basic Regulation.

TRUST: THE CORNERSTONE OF HUMAN-CENTRIC AI ACCEPTANCE

Currently, most governments are putting the focus on the main ethical concerns raised by the advent of AI in all areas of our lives and society (INTEL, 2020), (Hashmi, 2019), (Office of the Director of National Intelligence, 2020), (Australian Government, 2020), (IEEE, 2018). This ethical approach is seen as key enabler to AI gaining and strengthening citizen trust and societal acceptance, to the point that most governments and institutions postulate that AI can only be considered trustworthy if its development and use respect the ethical values widely shared by modern societies. This conviction generates the need to translate and build these ethical guidelines in the existing regulatory frameworks. The High-Level Expert Group on Artificial Intelligence (High-Level Expert Group on Artificial Intelligence, 2019) has proposed guidelines with seven key requirements for trustworthy AI that any certification framework should include. Figure 1 illustrate these seven requirements. The seven are considered of equal importance and support each other.

Although these guidelines are not binding, EASA has taken this approach from an aviation perspective to meet the challenge of certifying human-centric AI in aviation. However, the trustworthiness of IA involves a significant amount of challenges (Dario Amodei, 2016):

- **The frameworks applied in aviation up to day for SW development assurance need to be adapted to AI** (Kritzinger, 2017). AI put additional emphasis on data preparation and management, learning process and management, model training and validation, etc … Conventional development assurance principles will still be used at high level to elicit functional

Figure 1: AI HLEG seven key requirements for trustworthy AI. Adapted from (High-Level Expert Group on Artificial Intelligence, 2019).

requirements and at platform level for HW and core SW requirements. However, it will be necessary to develop specific assurance methodologies for dealing with the specificities of the learning processes.

- **Traceability of high-level requirements integrity, and accuracy of the data set** (Federal Aviation administration, s.f.). The AI learning process is based on both, the data and the learning process itself. Supervised learning generally involves the definition of expected functional behavior; while unsupervised learning or reinforcement learning may involve more unpredictable behaviour. Ensuring traceability to high-level requirements and ensuring the integrity and correctness of the dataset is key, as these could influence the behaviour of the training model.
- **Predictability and explainability** (Xu, 2019). Although the mathematical foundation of AI techniques (ML, DL etc., eg fixed weights on a NN) is deterministic, there is a high degree of unpredictability in AI applications, because the output will depend on the correlation between each new input and the dataset used in training. It is therefore necessary to make the conditions that lead to a given result more transparent and understandable. This concept is commonly referred as 'Explainability of AI'.
- **Robustness and unintended function** (Alignment, 2019). There is a need to propose new methods to verify the robustness of AI applications, as well as to guarantee that their validation is complete. It needs to be determined whether the use of formal methods could be a sufficient means of verification while compensating for the lack of coverage analysis.
- **Standardization** of methods and metrics to quantify and evaluate the operational performance of AI applications, (accuracy, error rate, etc…) (Tim G. J. Rudner, 2021).
- **Biases and variations**. (Nelson, 2019) Not surprisingly, AI applications are subject to bias and variation just as people are. In many cases, these biases pose a risk to the integrity of the decisions made and the results obtained. An important challenge in data management is the identification and mitigation of any inherent data bias or variation that may propagate .

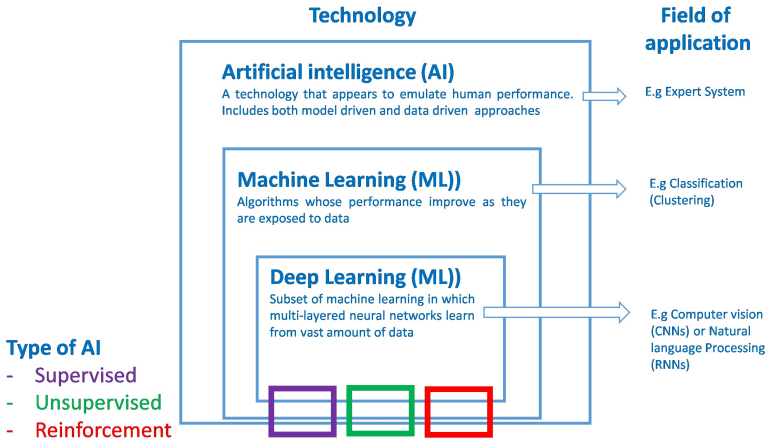

Figure 2: EASA AI Roadmap taxonomy for AI (Source: EASA AI Roadmap).

- **Complexity of architectures and algorithms**. Newer and more powerful algorithms and architectures pose high levels of complexity. Neural networks are a clear example with increasing complexity from classical ANNs, convolutional neural networks (CNN), recurrent neural networks (RNN), up to generative adversarial networks (GAN).
- **Adaptive learning processes** (Tumaini Kabudi, 2021). Real-time learning is incompatible with current certification processes, since it implies a constantly changing SW. Before even being able to plan solutions to this problem, the complexity of this problem may need to be narrowed down.

EASA FRAMEWORK FOR CERTIFICATION OF HUMAN CENTRIC AI APPLICATIONS

The EASA AI Roadmap 1.0 (EASA, 2020), acknowledging that AI is a broad term whose definition has evolved as technology has developed, considers a wide-spectrum definition of AI as "any technology that appears to emulate the performance of a human", and defines the taxonomy illustrated in figure 2. Figure 3 details the decomposition of an AI-based system. As can be seen an AI-based system is made up of several subsystems, at least one of them being an AI-based system subsystem. An AI-based subsystem incorporates at least one AI component, which is a collection of hardware and/or software elements, and at least one AI Item. The AI item is a specialized software and/or hardware element that contains at least one AI inference model. Hardware and software elements do not include any aspect related to AI/ML model inference.

The deployment of learning processes in civil aircraft certification projects is a sudden reality, to the point that EASA has already received certification requests proposing limited use of AI solutions. In the pursuit of autonomous flight, a phased approach is proposed for CAT (Commercial Aviation Transport), starting the first pilot assistance certifications by 2025, with a gradual increase towards full autonomy around 2035. The expected timeline for commercial aviation might be:

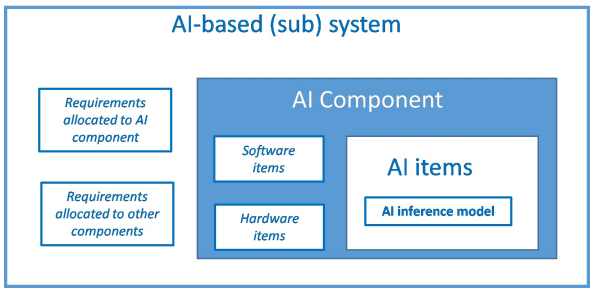

Figure 3: Decomposition of the AI-based (sub)system.

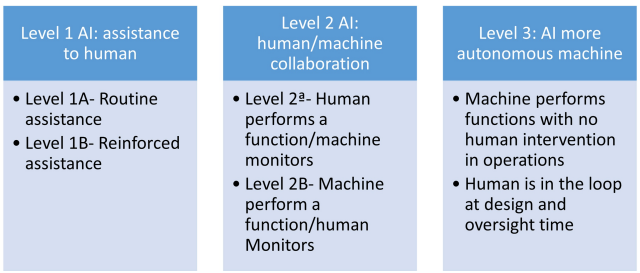

Figure 4: Classification of AI applications in three levels considering the role of the human.

- First step: AI applications for crew support, assistance and augmentation (2022–2025).
- Second step: AI applications for man/machine collaboration (2025–2030).
- Third step: AI applications for autonomous commercial air transport (2035+).

The drone industry is pushing to the last step faster. In the UTM/drone domain, the Agency has delivered first guidance material at the end of 2021 that could support the first U-space applications and Automated/semi-autonomous drone's certification.

EU Guidelines 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). Although the detailed definition of these scenarios still require further discussion, EASA has extended this concept to the aviation domain and come out with a classification of AI applications in three levels, considering the degree of oversight of a human on the machine, as indicated in figure 4.

OVERVIEW OF THE FRAMEWORK FOR TRUSTWORTHINESS OF HUMAN CENTRIC AI

To address the challenges of enabling readiness for use of human-centric AI, four main 'building blocks' have been defined in the framework for AI trustworthiness, as indicated in figure 5.

The **trustworthiness analysis** block serves as interface between the EU Ethical Guidelines (High-Level Expert Group on Artificial Intelligence, 2019)

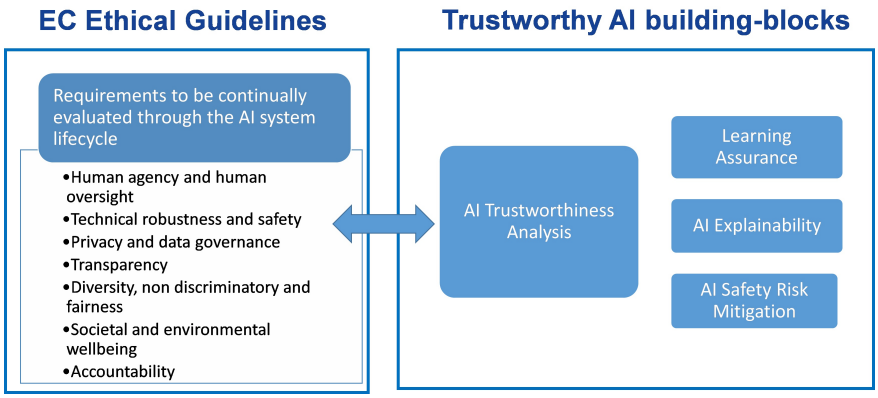

Figure 5: Four main 'building blocks' of the framework for AI trustworthiness.

and the three other technical building blocks. It provides the guidance on how to address each of the seven key trustworthiness guidelines (accountability, technical robustness and safety, oversight, privacy and data governance, non-discrimination and fairness, transparency, societal and environmental well-being) in the context of civil aviation. This block includes the safety, security and ethics assessment key in the reliability analysis concept.

The **learning assurance** block addresses the paradigm shift from programming to learning and provides guidelines for new development assurance methods adapted for the learning processes specific to AI. The existing regulatory framework quantifies and controls the risks of systems, equipment and parts by applying a "development assurance", based on requirements, during the development of its constituents. For AI applications, system-level assurance will apply, but current "development assurance" methods are not applicable to design-level layers that rely on learning processes. Intuitively, assurance processes must evolve to take into account the accuracy and completeness of training/verification data sets, the identification and mitigation of bias, the accuracy and performance of an AI application, novel verification methods, etc. This new concept of 'learning guarantee' will bring with it new means of compliance. Figure 6 syntheses the key aspects identified within the learning assurance block.

Explainability of AI is a human-centred concept implemented related with the ability to explain how an AI application achieves its results and outputs in a way that can be understand by the operator. The challenge begins with understanding the meaning of the concept of explainability when using AI particularly for decision-making processes. It involves a lot of human-machine interface/ factors considerations.

The **AI safety risk mitigation** block takes into account that it might not always be possible to open the 'AI black box' to the extent required by the certification process. For those cases, the derived safety risks need to be evaluated to identify proper mitigations. Guidelines will be provided on how to account for the inherent uncertainty of AI. Risk mitigation could be achieved by various means, including:

Figure 6: Key aspects within the learning assurance block.

- Keeping a human in command (HIC) or in the loop (HITL);
- Monitoring of AI/ML output and passivation of the AI/ML application with recovery through a traditional backup system (eg safety net);
- AI encapsulation with rule-based approaches (eg hybrid AI);
- AI monitoring through an independent AI agent;
- License to an AI.

The framework envisages that trustworthiness analysis should be performed in its full spectrum for any AI application, whereas for the other three building blocks, the depth of the requirements and guidance should be adapted on the application.

IMPACT ON THE CERTIFICATION PROCESS

The EASA Basic Regulation main objective is to establish and maintain a high uniform level of civil aviation safety in the Union. New certification guidelines will be applicable to any system developed using AI techniques or incorporating AI algorithms, as far as they will be used in safety-related applications or in applications related to environmental protection covered by the EASA Basic Regulation. The following domains are affected:

- **Initial and continuing airworthiness:** systems or equipment required for type certification or by operating rules, including those which improper functioning would might lead to failure conditions Catastrophic, Hazardous, Major or Minor;
- **Air operations**: systems, equipment or functions to support, complement, or replace pilot tasks (e.g. information acquisition, information analysis, decision making, action implementation and monitoring of outputs);

- **ATM/ANS**: equipment or procedures intended support, complement or replace end-user tasks delivering ATS or non-ATS services;
- **Maintenance**: systems supporting scheduling and performance of tasks intended to timely detect or prevent unsafe conditions; or critical maintenance tasks', which could create unsafe conditions .
- **Training**: systems used for monitoring the training efficiency or for supporting the organisational management system, both in terms of compliance and safety;
- **Aerodromes**: systems that automate key aspects of aerodrome operational services;
- **Environmental protection**: systems or equipment affecting the environmental characteristics of products.

Adaptations in the organizational rules of each domain will also be required, e.g. Holders of Design Organization Approvals (DOAs), Holders of Maintenance Organization Approvals (MOAs), Continuing Airworthiness Management Organizations (CAMOs), Air Navigation Service Providers (ANSPs), Approved Training Organizations (ATOs), operators, etc…Organizations should ensure compliance with EU regulations and should assess the impact of new AI applications on their internal processes (e.g. competency management, design methodologies, change management, supplier management, incident reporting, security aspects, cybersecurity, record keeping, etc...). Table 1 syntheses the analysis of the anticipated impact on aviation regulations and on the means of compliance to the current regulations for the various impacted domains.

CONCLUSION

Ethical criteria and requirements are at the core of the certification of AI human centred applications in aviation. Fundamental to this ethical consideration is the concept of trustworthiness that has been translated by EASA into a framework for AI trustworthiness with four main 'building blocks'. Through this paper the 4 building blocks the trustworthiness of human-centric AI framework proposed by EASA have been revised to identity the possible impacts of the introduction of human- centric AI in the different Implementation Rules (IR), Certification Specifications (CS), Acceptable Means of Compliance (AMC) and guidance material (GM) in the domains covered by the EASA Basic Regulation.

In the domain of aircraft design, the current implementation rules (Part 21 and Certification Specifications) provide an open framework for the introduction of AI human-centred applications. More specifically, paragraphs such as CS 25.1309 could still be valid for assessing the safety of AI systems, provided additional means and compliance standards are developed to address the identified gap in the core components outlined through this paper. In the other domains (operations, maintenance, ATM, aerodromes), current regulations provide an open framework for the use of AI human centred applications. However, such regulations will need to be tailored and extended to cover the specificities of new AI applications.

Table 1. Anticipated AI applications impact on aviation regulations and on the means of compliance.

Domain	Anticipated impact on current regulation	Anticipated impact on current AMC/MoC framework
Product design and certification	Current implementing rules (Part 21) and CSs offer an open framework for the introduction of AI/ML solutions. Requirements such as CS25.1301, 1302, 1309/SC-VTOL.2500, 2505, 2510 are still valid. However additional means of compliance and standards should be developed to cover the gap identified in the building blocks of the EASA AI Roadmap. For AI Level 1 applications, no impact on the EU regulatory framework is foreseen. Higher AI Levels (2 and 3) require further analysis.	Initial AMC /MoC have been identified for Level 1 AI applications. Certification Review Items (CRI) could be used to address installations issues. Current guidance (e.g. AMC25.1309 or MOC VTOL.2510) is fully applicable. The technical particularities of AI technology might require a need to adapt or introduce new AMC & GM related to the following Part 21 points: 21.A.3A 'Failures, malfunctions and defects'; 21.A.31 'Type design'; 21.A.33 'Inspections and tests' and 21.A.615 'Inspection by the Agency'; 21.A.55, 21.A.105 and 21.A.613 'Record-keeping'; 21.A.91 'Classification of changes to a type-certificate'
Air Operations domain	Current regulatory framework (Regulation (EU) No 965/2012 (Air OPS Regulation) (Part-ORO) contains safety management principles to identify and mitigate risks and manage changes in their organisation and their operations (ORO.GEN.200). It allows the introduction of AI solutions. However, new AMC and GM will need to be developed. AI Level 1 application require specific provisions in the Air OPS Regulation. AI Levels 2 and 3 might require a deeper assessment.	Current means of compliance for system, software and hardware development assurance are not sufficient and need to be complemented through the guidelines for learning assurance. The need for explainability is a new MOC. It builds however on some existing guidance; in particular, the applicable human factors guidance already used in certification could provide a sufficient layer of MOC for Level 1A AI/ML applications.
Training / FSTD	Covered by Annexes to Regulation (EU) 1178/2011 (Aircrew Regulation) Annex I (Part-FCL) and Annex II (Part-ORA) and the Air OPS Regulation, is referring to traditional methodologies. The main impact will be on: definitions; description of training programme delivery methodologies; crediting criteria; organisation requirements .	AMC shall be reviewed and updated to account for: new training needs for new categories of aircraft (e.g. VTOL or RPAS); new training devices (e.g. virtual or augmented reality). Aircrew Regulation is not intended to certify products and does not address the design process, therefore all the elements of the AI model would need an effort to be tailored to the purpose.

Table 1. Anticipated AI applications impact on aviation regulations and on the means of compliance.

Domain	Anticipated impact on current regulation	Anticipated impact on current AMC/MoC framework
Ground / ATM/ANS	Current Regulation (EU) 2017/373 with common requirements for providers of ATM/ANS, Regulation (EC) No 552/200414 for interoperability, and Regulation (EU) No 376/2014 for occurrence reporting open the path to the use of Level 1 AI. For higher AI Levels (2 and 3) might require a deeper assessment.	Initial adaptations include: ANNEX III — Part-TM/ANS.OR – AMC6 ATM/ANS.OR.C.005(a)(2), AMC1 ATM/ANS.OR.C.005(b)(1), AMC4 ATS.OR.205(a)(2) for Safety assessment and assurance of changes to the functional system; and ANNEX XIII — Part-PERS – AMC1 ATSEP.OR.210(a) Qualification training. Associated GM could be impacted as well.
Aircraft production and maintenance	Regulation (EU) No 1321/2014 wording is generic and assumes a maintenance task-based approach Level 1 systems do not contradict this philosophy. For Level 2 systems current regulation may limit the actions which can be carried out by systems. Level 3 systems are not in line with the current regulation.	In the maintenance domain, there is no MoC framework comparable to the one used in certification, and significant part of the approval is done by the competent authorities (NAAs) rendering thus the impact of AI/ML not that easy to be evaluated. Future standards on AI developed by recognised official bodies (like SAE, ISO) could be used for demonstrating compliance with requirements .
Environmental protection	Environmental protection requirements for products are laid out in the Basic Regulation Articles 9 and 55 for manned and unmanned aircraft respectively, and in its Annex III. These requirements are further detailed in Part 21 (in particular point 21.B.85) as well as in CS-34 'Aircraft engine emissions and fuel venting', CS-36 'Aircraft noise' and CS-CO2 'Aeroplane CO2 Emissions'. No impact is envisaged.	AMC and GM linked to requirements are defined in the appendices to ICAO Annex 16 and in Doc 9501 'Environmental Technical Manual'. The AI/ML guidance for Level 1 systems is anticipated to have no impact on the current MoC framework for environmental protection. The impact of Level 2 or 3 AI/ML guidance will be assessed at a later stage.
Airports	Current Basic Regulation, Regulation (EU) No 139/201415 does not represent a hinderance to the use of Level 1 AI use cases. For higher AI Levels (2 and 3), further analysis is required by industry and overseen organisations, as well as manufacturers of safety-relevant aerodrome equipment.	Most of the AMC and GM do not refer to specific technologies, they do not impede the approval of Level 1 AI applications. For higher AI Levels (2 and 3) further analysis is required. Relevant AMC and GM are: ADR.OPS.B.015 Monitoring and inspection of movement area and related facilities; ADR.OPS.B.020 Wildlife strike hazard reduction and ADR.OPS.B.075 Safeguarding of aerodromes

The required regulations adaptation of the regulations should be eased by the latest amendment to the EASA Basic Regulation (EU 2018/1139), which allows the Agency to better support the development of innovation through the use performance-based regulations. Considering the multi-domain implications of AI EASA will define a common policy that can be applied to any regulation related to the domain.

REFERENCES

Alignment, R. V. i. E. U. F. f. A. V., 2019. Requisite Variety in Ethical Utility Functions for AI Value Alignment. Computer Science. Artificial Intelligence.

Australian Government, 2020. Australia's Artificial Intelligence Ethics Framework, s.l.: s.n.

Dario Amodei, C. O. J. S. P. C. J. S. D. M., 2016. Concrete Problems in AI Safety. Computer Science. Artificial Intelligence.

EASA, 2020. EASA Artificial Intelligence Roadmap 1.0 published. A human-centric approach to AI in aviation, s.l.: s.n.

Federal Aviation administration, s.f. Software Assurance Approaches, Considerations, and Limitations: Final report, s.l.: s.n.

Hashmi, A., 2019. AI Ethics: The Next Bif Thing in Government, s.l.: s.n.

High-Level Expert Group on Artificial Intelligence, 2019. Ethics Guidelines for Trustworthy AI, s.l.: s.n.

IEEE, 2018. ETHICALLY ALIGNED DESIGN. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems, s.l.: s.n.

INTEL, 2020. ARTIFICIAL INTELLIGENCE ETHICS FRAMEWORK FOR THE INTELLIGENCE COMMUNITY, s.l.: s.n.

Kritzinger, D., 2017. Development Assurance. En: Aircraft System Safety. s.l.:s.n.

Nelson, G. S., 2019. Bias in Artificial Intelligence. North Carolina Medical Journal, 80(4), pp. 220–222.

Office of the Director of National Intelligence, 2020. Principles of Artificial Intelligence Ethics for the Intelligence Community, s.l.: s.n.

Tim G. J. Rudner, H. T., 2021. Key Concepts in AI Safety: Specification in Machine Learning, s.l.: s.n.

Tumaini Kabudi, I. P. D. H. O., 2021. AI-enabled adaptive learning systems: A systematic mapping of. Computers and Education: Artificial Intelligence, Volumen 2, p. 100017.

Xu, F. U. H. D. Y. F. W. Z. D. &. Z. J., 2019. Explainable AI: A brief survey on history, research areas, approaches and challenges. In CCF international conference on natural language processing and Chinese computing. s.l., s.n.