Flying High for AI? Perspectives on EASA's Roadmap for AI in Aviation*

Ivo Emanuilov** & Orian Dheu***

In early 2020, the European Union Aviation Safety Agency (EASA) published its long anticipated 'Roadmap for Artificial Intelligence in Aviation'. This document builds upon previous European initiatives such as the High-Level Expert Group's Ethical Guidelines on artificial intelligence ('AI'), where the concept of 'trustworthiness' is embedded as a key pillar and a pre-requisite for developing and deploying AI technologies. The roadmap assesses the associated ethical, safety and regulatory challenges that may arise from the deployment and use of AI applications in aviation. This article provides an overview of the main takeaways, strengths and weaknesses of this roadmap. It critically analyses the main challenges of AI-driven technologies throughout the entire aviation domain. The article argues the roadmap would benefit from considering new regulatory tools and processes, such as regulatory sandboxing and AI-driven certification, and contends any efforts for standardization of AI in aviation must be reconciled with existing standardization of automation and that this may not always be a straightforward process as far as interoperability is concerned. Finally, the article argues that further exploration of the identification and allocation of liability will be indispensable in fostering increased levels of trust in AI-enabled aviation.

Keywords: AI, Innovation, Regulation, Aviation, Autonomous Systems, Regulatory Sandboxing

1 INTRODUCTION

Since the dawn of aviation, technological innovation has played an important role in fostering increased levels of safety and reliability, as demonstrated by Lawrence

* This research has received funding from the SBO OmniDrone project, https://www.omnidrone720.com, the European Union's Horizon 2020 research and innovation programme under the Secure Collaborative Intelligent Industrial Automation (SeCoIIA) project, grant agreement No 871967, https://secoiia.eu, and the European Union's Horizon 2020 research and innovation – Marie Skłodowska-Curie actions program under the Safer Autonomous Systems (SAS) project, grant agreement No 812.788, https://etn-sas.eu/. The authors have contributed equally to this work. The authors would like to thank the reviewers for their insightful comments and suggestions for improvement and future research.

** A practising lawyer and doctoral researcher in intellectual property law at the University of Leuven, Belgium. In his research, he explores interdisciplinary perspectives on the overlapping protection of software by copyright, patents and trade secrets. Email: ivo.emanuilov@kuleuven.be.

*** A doctoral researcher in private and transport law at the University of Leuven, Belgium. In his research, he explores the technology induced legal and liability challenges facing the development and deployment of autonomous transportation systems. Email: orian.dheu@kuleuven.be.

Emanuilov, Ivo & Dheu, Orian. 'Flying High for AI? Perspectives on EASA's Roadmap for AI in Aviation'. *Air & Space Law* 46, no. 1 (2021): 1–28.

Sperry's autopilot invention as early as 1912. In the aftermath of World War II, technological breakthroughs enabled much more sophisticated automation tools to be developed and deployed on both military and civilian aircraft. Since then, the advent of electronics, computers and modern communications networks have pushed the boundaries even further.

Nowadays, autoland and flight control systems have become standard features within commercial aviation and serve as powerful technical tools for assisting pilot (s) in nominal and non-nominal flying conditions. Technological evolution has increased the overall safety levels and has therefore played an important role in leveraging public trust within a transport medium once thought to be an 'Icarus'-inspired fantasy.

Rapid technological developments have been on the rise with the accumulation of domain-specific big data and of artificial intelligence ('AI') gaining momentum. These advances have revealed various applications in the field of aviation. Indeed, AI is a decision-making safety and optimization tool extending far beyond the cockpit. With AI, the technological cursor is slowly pushing from automation towards what is now called 'autonomy' where increased decisional power is being delegated to computational artefacts.[1] These technologies, however, come along with several socio-technical, legal and regulatory challenges that will have to be addressed before AI can be effectively implemented within this safety-critical domain.

The European Union (EU) has been at the forefront of exploring and addressing these challenges through an important regulatory effort which is currently under way. The European Union Aviation Safety Agency (EASA) has also taken an active role in this process and has published its roadmap on AI in aviation.[2] This article explores the document's key findings and shortcomings and proposes recommendations for consideration.

The article is structured in six main parts. It sets out the policy context which is rooted in the EU's strategic vision for AI (Part 2); introduces public trust and ethics as bedrock principles for AI in aviation (Part 3); analyses sectoral applications of AI in aviation (Part 4) and the roadmap's identified challenges and trustworthiness building blocks (Part 5). Finally, the article highlights some of the roadmap's shortcomings, that is, what it does not say and what direction its 'flight plan' should take (Part 6).

---

1 Computational artefacts are, in the broadest sense, 'made things (…) process[ing] symbol structures signifying information, data or knowledge', i.e. utilitarian, human-made things that reflect their creators' goals. *See* Subrata Dasgupta, *Computer Science: A Very Short Introduction* 30–32 (1st ed., Oxford University Press 2016).

2 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap: A Human-Centric Approach to AI in Aviation* (2020), https://www.easa.europa.eu/sites/default/files/dfu/EASA-AI-Roadmap-v1.0.pdf (accessed 14 Apr. 2020).

2 IT TAKES TWO TO TANGO: EASA'S ROADMAP FOR AI IN AVIATION AND THE EU'S STRATEGIC VISION FOR AI

The roadmap is aligned with some of the EU's key positions of the 'Ethical Guidelines' for AI (2.1), while it emphasizes the need to specifically build 'trustworthiness' in AI driven aviation (2.2) which is considered to be a high-risk application (2.3).

2.1 A roadmap aligned with the 'ethics guidelines' for Artificial Intelligence

The Artificial Intelligence Roadmap for a human-centric approach to AI in aviation is the result of a process initiated by EASA. This sectoral initiative is aligned with the EU's strategic vision for AI laid down by the European Commission ('the Commission') in two related communications.[3] The roadmap also seems to be generally aligned with much of the fundamental positions expressed by the Commission in its White Paper on AI.[4] The following paragraphs will outline the main tenets of this vision.

The essence of the Commission's approach is to promote and boost AI-driven innovation,[5] tackling 'socio-economic changes'[6] and ensuring 'an appropriate ethical and legal framework'.[7] To support the implementation of its vision, the Commission established a High-Level Expert Group on AI ('AI-HLEG'), comprising fifty-two experts from academia, civil society and industry and tasked with the development of recommendations on a broad range of issues.

In April 2019, AI-HLEG published its much touted 'Ethics Guidelines on Trustworthy AI' proposing a set of non-binding recommendations regarding AI.[8] The guidelines suggest three essential requirements for 'trustworthy AI', namely

---

[3] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: 'Artificial Intelligence for Europe', SWD(2018) 137 final (2018) and Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: 'Coordinated Plan on Artificial Intelligence,' COM/2018/795 final (2018).

[4] European Commission, *White Paper on Artificial Intelligence – A European Approach to Excellence and Trust*, COM(2020) 65 final, https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf (accessed 14 Apr. 2020).

[5] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: 'Artificial Intelligence for Europe', *supra* n. 3, at 5.

[6] *Ibid*., at 11.

[7] *Ibid*., at 13.

[8] High-Level Expert Group on Artificial Intelligence, *Ethics Guidelines for Trustworthy AI* 41 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 (accessed 14 Apr. 2020).

that AI should be 'lawful', 'ethical' and 'robust'. The broad definition of AI in the guidelines encompasses multiple approaches to AI:

> *software (and possibly also hardware) systems designed by humans [who design AI systems directly, but they may also use AI techniques to optimise their design] that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.*[9]

The three wholesale conditions of lawfulness, ethics and robustness translate into seven key requirements of trustworthiness. These include oversight, technical robustness and safety, privacy and data, transparency, non-discrimination and fairness, societal and environmental well-being, and accountability.

While not (yet) formally endorsed by the Commission, the guidelines have been hailed as a 'valuable input for its policy-making'.[10] The Commission shares the view that in order to gain public trust, AI must be 'predictable, responsible, verifiable, respect fundamental rights and follow ethical rules'.[11] It is therefore not surprising that the guidelines are continuously referred to throughout the roadmap. Indeed, as an agency of the EU, EASA is bound by the principle of consistency between the policies and activities of the EU.[12] Therefore, it is not surprising the roadmap's approach is to follow closely the Ethics Guidelines. Thus, building upon these recommendations,[13] it, expectedly, positions trustworthiness centre stage in the strategic vision of AI in aviation.

While largely aligned with the Ethics Guidelines, EASA's roadmap also departs from it in some key respects. One of the examples is its broad apprehension of the term 'artificial intelligence'. Unlike the Ethics Guidelines, the roadmap defines AI much more generally as 'any technology that appears to emulate the performance of a human'.[14] This definition is certainly closer to the common

---

[9] *Ibid*., at 36. *See also* High-Level Expert Group on Artificial Intelligence, *A Definition of AI: Main Capabilities and Disciplines* 6 (2019), https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60651 (accessed 14 Apr. 2020).

[10] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Building Trust in Human-Centric Artificial Intelligence', COM(2019) 168 final 4 (2019), https://ec.europa.eu/transparency/regdoc/rep/1/2019/EN/COM-2019-168-F1-EN-MAIN-PART-1.PDF (accessed 17 Apr. 2020).

[11] Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions: 'Coordinated Plan on Artificial Intelligence', *supra* n. 3, at 7.

[12] Article 7 Treaty on the Functioning of the European Union.

[13] European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 2, 5–6, 16–20.

[14] *Ibid*., at 4. In the 'Definitions' section of the roadmap a more elaborate definition reads that AI is a 'technology that appears to emulate human performance typically by learning, coming to its own

understanding of general AI rather than that of instrumental, narrow AI, it is surprising given that AI is primarily seen as a support tool in aviation.

The definition adopted by the roadmap could be challenged on three main points. First, automation could also *appear to* emulate the performance of a human, for example, in simple and repetitive tasks. Emulation is not the key distinguishing feature of AI systems, rather it is the attainment of a human-defined goal. Furthermore, human performance can sometimes be undermined by complex internal and external factors and this is precisely the challenge AI is called on to resolve, not emulate. Second, the roadmap seems to focus predominantly on machine learning (ML), a subset category of AI, that it defines as 'the use of data to train algorithms to improve their performance'.[15] At the same time, however, references to AI/ML as interchangeable notions appear sporadically throughout the document. Finally, the extended definition provided in the 'Definitions' section of the document appears to suggest an (anthropomorphising) degree of agency inherent in the system itself. Hints to this can be found in references to the system 'coming to its own conclusions', 'understand[ing] complex content' or 'engaging in natural dialogues with people'. In our view, in using this vocabulary, EASA risks creating confusion among the aviation community as to the true purpose of AI in aviation, i.e. to support decision making.

This noticeable departure from the definition of AI adopted by AI-HLEG is undoubtedly surprising. The main concern here is that, in the long run, this poses the risk of continuing terminological fragmentation and confusion. The policy debate on AI has already been plagued by this phenomenon and the roadmap has clearly missed an opportunity to bring clarity. This is particularly true regarding the concept of 'trustworthiness'.

2.2 'Trustworthiness': Key pillar in developing AI-driven aviation

The roadmap recognizes trustworthiness as a bedrock principle and a key pillar in the development and deployment of AI technologies in aviation.[16] EASA explains the relationship between the building blocks of trustworthy AI, as identified by AI-HLEG, and their implications for aviation through a process called 'trustworthiness analysis'.

---

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)', *Ibid*., at 26.

15 *Ibid*., at 4.

16 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 5.

Trustworthiness analysis interfaces the building blocks for trustworthy AI in aviation with the principles embodied in AI-HLEG's Ethics Guidelines.[17] EASA aims to create additional, specific technical building blocks that are critical for aviation. These building blocks are learning assurance, explainability and safety risk mitigation. Trustworthiness analysis can be described as a regulatory analysis tool whose purpose is to support EASA in evaluating the extent to which AI applications in aviation embed the principles set out in AI-HLEG's Ethics Guidelines. In other words, this tool provides regulatory bodies with a specific guidance[18] accounting for the complex, safety-critical nature of aviation activities. Trustworthiness analysis can thus be seen as a preliminary step which serves as a 'go/no go' decision-making tool in respect of whether to continue or not with the development or deployment of AI technologies in aviation.

In our understanding, trustworthiness analysis therefore aims to translate the results of this analysis into actionable information for stakeholders who consider developing AI technologies,[19] e.g. by eliciting requirements to be implemented in a system's design. This way trustworthiness analysis could nurture a higher accountability culture not only in the regulators, but also in other aviation stakeholders as far as AI technologies are concerned.

2.3 Aviation AI applications as 'high-risk applications'

The degree of trustworthiness allegedly depends upon the risks of an AI application. The Commission recognized this when it announced its White Paper on AI. In striving to establish an 'ecosystem of trust' around AI, the Commission highlighted the need of a risk-based approach grounded in clear criteria differentiating between different AI applications (e.g. high-risk or low-risk).[20] It acknowledged that such an assessment should be based on both the sector and the intended use of an application.[21] Many AI applications in aviation, particularly those concerning optimization of safety-critical flight activities,[22] could be considered high-risk and

17 *Ibid*., at 20.
18 *Ibid*., at 16.
19 *Ibid*., at 16.
20 European Commission, *White Paper on Artificial Intelligence*, *supra* n. 4, at 17.
21 *Ibid*., at 17. The Commission has identified transport as meeting the first criterion and the safety risks of injury, death or significant damage as meeting the second. Therefore, AI applications in safety-critical aviation activities would almost certainly and almost always be considered high-risk applications.
22 An early example of AI in flight activities is the Runway Overrun Prevention System (ROPS) deployed in certain Airbus aircraft. *See* Airbus' ROPS certified by EASA on A330 Family, Airbus (2015), https://www.airbus.com/newsroom/press-releases/en/2015/07/airbus-runway-overrun-prevention-system-rops-certified-by-easa-on-a330-family.html (accessed 19 May 2020).

should therefore be subject to increased legal and regulatory scrutiny and mandatory requirements.[23]

In its White Paper, the Commission suggested several criteria that could shape the mandatory requirements for high-risk AI applications. These include:

- Good practices concerning the supply and subsequent use of training data, considering the objectives of safety and protection of fundamental rights.[24]
- Mitigating opaqueness of AI through measures for storing data for accountability purposes, incl. accurate records, data sets and documentation of the programming, training and testing processes, methodologies and techniques.[25]
- Ensuring transparency vis-à-vis the deployers and customers of AI applications regarding their capabilities, risks and limitations must be provided proactively and clearly in addition to established internal accountability processes.[26]
- Fostering AI's technical robustness and accuracy through requirements aimed at ensuring correct reflection of an AI application's accuracy throughout its lifecycle, reproducibility of its outcomes, mechanisms to deal with errors and inconsistencies and resilience against attacks.[27]
- Mitigating the risk of undermining human autonomy through design and operational human oversight requirements.[28]
- Requirements to implement safeguards, for example, concerning remote biometric identification for biometrics applications, which are considered high-risk *ipso facto*.

These criteria are indeed reflected also in the four high-level questions formulated by the roadmap that serve as a basis for engagement with the stakeholder community. In our opinion, however, the vagueness of some of the notions which lie at the heart of the roadmap could undermine this much-needed dialogue. The following sections provide a critique of the conceptual framework proposed by the roadmap and then move on to discuss specific issues in the sectoral applications identified in the document.

---

23 European Commission, *White Paper on Artificial Intelligence*, *supra* n. 4, at 17.
24 *Ibid*., at 18–19.
25 *Ibid*., at 19.
26 *Ibid*., at 20.
27 *Ibid*., at 20–21.
28 *Ibid*., at 21.

3 PUBLIC TRUST AND ETHICS: CORNERSTONES OF THE ROADMAP

In his foreword to the roadmap, the executive director of EASA, Patrick Ky, highlighted four high-level questions that are to serve as basis for discussion within the stakeholder community.[29] Essentially, these questions also outline the roadmap's conceptual framework, organized around the notions of public trust (3.1), ethics (3.2), certification and standardization (3.3).

3.1 Trust as a (context-specific) socio-technical enabler

The first question relates to how stakeholders could bring public trust into AI-based systems. In our view, this is a challenging undertaking, not least because of the abstract nature of 'public trust'.[30] Our criticism echoes a more general critical stance towards the approach adopted by the Commission in organizing its policy responses around ill-defined policy notions such as 'trustworthiness'.

In the risk society, trust is often seen as a '"protective cocoon" which stands guard over the self in its dealings with everyday reality'[31] and which 'enables individuals with cognitive limitations to make decisions'.[32] In the context of aviation, trust can easily be reduced to the provision of the public goods of safety and adequate risk management due by the institutions and individuals charged with it.[33] It is therefore contended here that trust is vested in the individuals and institutions tasked with the evaluation and approval of a AI applications. Trustworthiness is not a feature of AI applications but of the institutions established to manage and mitigate risks. In other words, it is the approving authorities that are in a position of trust, not the technical artefacts.[34] Furthermore, in addition to

---

29 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 2.

30 Indeed, the Ethics Guidelines for Trustworthy AI have adopted a scholarly definition of 'trust' which reads, as follows: '[t]rust is viewed as: (1) a set of specific beliefs dealing with benevolence, competence, integrity, and predictability (trusting beliefs); (2) the willingness of one party to depend on another in a risky situation (trusting intention); or (3) the combination of these elements', High-Level Expert Group on Artificial Intelligence, *supra* n. 8, at 38. AI-HLEG highlighted that 'trust can be ascribed to all people and processes involved in the AI system's life cycle', but it did not make it clear how stakeholders' trusting beliefs and intentions could translate into and induce trusting beliefs in the general public.

31 Anthony Giddens, *Modernity and Self-Identity: Self and Society in the Late Modern Age* 3 (1st ed. 1991).

32 George Leloudas, *Risk and Liability in Air Law* 52 (2013).

33 *Ibid*., at 53.

34 A painful reminder of the need of trustworthy regulatory and institutional capacity was given by the recent events in the wake of the accidents involving Boeing 737 Max 8 aircraft and the subsequent questioning of the transparency, independence and soundness of the Federal Aviation Authority's approval process. Despite the identified shortcomings of the current certification processes and the degree to which regulators rely on manufacturers to provide compliance artefacts take them on face value, this cooperation remains essential to promoting trust among the industry's stakeholders.

public trust, which seems to be equated by the roadmap with citizens' trust,[35] there are other types of trust which are equally important. For example, trust between the industry's stakeholders in the aviation value chain, when one or all of them relies on AI in their products or services, needs to be instilled as well. It is surprising that the roadmap seemingly neglects this aspect, focusing instead mostly on the citizens' trust viewpoint.

In our view, the policy goal should instead be to instil contextual public trust in the institutions, manufacturers, service providers, airspace users and individuals who design, develop, manufacture, evaluate and approve AI systems serving public interests, such as a high measure of safety and security. Obviously, different types of systems would require distinct measures on the part of the authorities and institutions so as to impart a high level of public trust. This would undoubtedly depend also on the degree of autonomy exercised by the system, so it is not practicable to define 'public trust' and 'trustworthiness' in the abstract. Any definition of 'trust' should always be tied to the context, institutional capacity, risks and level of operational autonomy of the concrete application.

3.2 'Ethical AI' as an elusive and uncertain requirement

The second question concerns the implementation of the ethical dimension of AI, understood in the roadmap as referring to transparency, non-discrimination, fairness etc. in safety certification processes. The concept of 'ethical AI' is blurry, if not obscure. The so-called 'ethical' dimension of AI is essentially a shorthand for voluntarism and industry-driven self-regulation.[36]

Relying on ethics as a 'requirement' can also be counterproductive, especially when (more) clearly defined legal notions are brought into the mix. Obviously, this could create confusion regarding the binding force of one *rule* or another. Scholars have convincingly argued that fairness, for example, is a highly context-sensitive concept.[37] Furthermore, the content of and the legal basis for such convergence within the various domains of aviation are far from obvious.

For instance, would 'fairness' in the context of processing aircraft operational data, crew behaviour data and manufacturing data always have the same normative content? Imagine a manufacturer's machine learning model which

---

[35] European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 5.

[36] *See* The Human Rights, Big Data and Technology Project – Written evidence (AIC0196) – Submission to the House of Lords Select Committee on Artificial Intelligence by the Human Rights, Big Data and Technology Project (HRBDT), 9, 12, 13 (2017), http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/artificial-intelligence-committee/artificial-intelligence/written/69717.html#_ftn11 (accessed 19 May 2020).

[37] Michael Veale & Reuben Binns, *Fairer Machine Learning in the Real World: Mitigating Discrimination Without Collecting Sensitive Data*, Big Data & Soc'y 4 (2017).

optimises the safety performance of an aircraft engine for one customer but not for another. The criterion for discrimination could be based on the profitability of the commercial data sharing agreement in place, number of engines in exploitation etc. Would this be considered a 'fairly' performing model? Similarly, imagine a model which predicts the likelihood of aircrews experiencing high levels of fatigue based on in-flight crew behavioural data. If an airline optimises the allocation of its aircrews based on these predictions, some aircrews might be assigned to non-profitable short-distance air routes. In the long run, this may change how different crews stack up against each other, potentially resulting in differential treatment.

Obviously, fairness in these two cases has a very different normative content. This content may depend on normative choices and factors such as beliefs and priorities of an organization, or willingness to accept moral responsibility. Ethical choices may equally be informed by real or perceived negative legal consequences of one preference over another. In other words, fairness cannot be reduced to a purely ethical or technical issue; it is a sociotechnical challenge with high contextual dependency. The problem of quantifying and implementing contextual fairness could become particularly sensitive, for example, in the framework of safety occurrence reporting or accident and incident investigations.

This goes to show that while it is true that reasons of policy coherence dictate that the roadmap follows the trajectory set by AI-HLEG's Ethics Guidelines, it cannot leave the contextualization of vague ethical or policy notions to mere chance. This is a particularly legitimate concern as far as clearly defined regulatory processes are concerned, such as safety certification.

3.3 Certification and standardization as key points of attention

The third point raises practical questions on how to prepare for the certification of AI systems. This point is assumedly linked also to the 'public trust block'. As will be demonstrated in the following sections, this is justifiably one of EASA's main areas of concern since it falls directly in its sphere of competence.

The roadmap does not specifically discuss the problem of certification of AI systems from the perspective of the aviation ecosystem. This is a matter which has not yet garnered the attention of policymakers, but one that is particularly important in the context of the complex supply chains in the aimed 'ecosystem of trust'.[38] The certification of AI systems cannot neglect the fact that AI is more

[38] European Commission, *White Paper on Artificial Intelligence*, *supra* n. 4, at 2, 3, 14.

often than not a systems artefact.[39] Thus, AI cannot be thought of as a single, independent product or service but should be considered in the context of the ecosystem of systems in which it operates. This article argues that the roadmap would benefit from considering a system-of-systems approach to certification of AI inspired by cybernetics.[40] This would enable a certification process which is informed by the interdependencies between the actors and artefacts which mediate their conduct in an ecosystem of which the AI application may be just one piece.

Finally, the document raises the question on what industry standards, protocols, and methods the aviation sector will need to develop in order to ensure that AI technologies will further improve the current level of safety of air transport. This last question is also open in light of the global nature of aviation and the need of uniform standards under the aegis of the International Civil Aviation Organization (ICAO) and its mandate to adopt and review Standards and Recommended Practices (SARPs).

In summary, the roadmap's conceptual framework is based on four distinct concepts of trust, ethics, certification, and standardization. These concepts underpin the four building blocks of AI trustworthiness in aviation, namely trustworthiness analysis, learning assurance, explainability and safety risk mitigation. The following section focusses on how the roadmap sees the contextual implementation of this framework. This is key to understanding the impact of AI on aviation in different sectoral applications.

4 AI AND ITS SECTORAL APPLICATIONS IN AVIATION: CRITICAL NOTES

The following paragraphs look at some of the challenges of these applications from the perspective of the roadmap's conceptual framework. More specifically, they look at the sectoral applications that the roadmap pinpoints such as aircraft design and operation (4.1), aircraft production and maintenance (4.2), air traffic management and urban air mobility (4.3) as well as safety management and cybersecurity (4.4).

4.1 Aircraft design and operation

First and foremost, the roadmap points out that AI may impact 'aircraft design and operation',[41] most notably through the enabling of autonomous flying. It is alleged

[39] *See* on the need of a 'system' view to regulation of AI in another safety-critical domain (i.e. medical devices), Sara Gerke et al., *The Need for a System View to Regulate Artificial Intelligence/Machine Learning-Based Software as Medical Device*, 3(1) npj Digital Med. 1–4 (2020).

[40] *See* for the concept of 'symmathesy' and mutual learning in living systems, Nora Bateson, *Symmathesy–A Word in Progress*, 1(1) Proc. 59th Annual Meeting of the ISSS – 2015 Berlin, Germany (2016).

[41] European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 7.

that increased levels of automation could, among others, improve the safety and efficiency of flight operations as well as meet some environmental concerns.[42] Going 'beyond the holy grail of autonomous flight',[43] different automation models co-exist where the human operator would retain more or less control over the aircraft, but could increasingly be assisted in nominal (e.g. cockpit assistance, flight profile optimization, etc.) and non-nominal flying conditions (e.g. safety-critical flight decisions in high workload situations). The roadmap mentions a 'change [in] the relation between pilots and systems'[44] where humans would be at the centre of a complex decisional process. They would be assisted by a machine, just like what is currently in place with 'fly by wire' automation. Further in its development, the roadmap briefly refers to 'the human-AI interface'[45] and identifies three different levels of automation according to the degree of human oversight in which applications could be classified: 1) level one, where AI would provide 'assistance to humans'; 2) level two, where there would be a 'human-machine collaboration'; 3) and level three, where the machine is more 'autonomous'. As indicated in the roadmap's provisional calendar,[46] the introduction of such technologies, as well as the corresponding regulatory guidance, would follow a stepped approach spanning over many years. Lower automation levels would allegedly be reached before higher automation would be implemented. The roadmap also ponders on the necessity of adopting a 'risk-based approach'[47] according to the degree of automation and human oversight.

Surprisingly, specifically in relation to levels one and two, the roadmap does not sufficiently address the 'automation paradox'[48] issue which has been known to be a contributory factor in at least one major accident (e.g. flight Air France 447).[49] This paradox refers to the possible overreliance of human operators on the technological abilities of an on-board system. It emphasizes the importance of clearly delineating functions and responsibilities in human-machine interaction. This paradox is likely to be exacerbated by the introduction of AI technologies in the cockpit, leading to 'overtrust' and 'misjudgement' of pilots over a system's capacities. This latter argument is often used in favour of pushing for full autonomy since AI is leveraged as a technological tool that would dramatically enhance safety by *in fine* eliminating the human

---

42 European Aviation Artificial Intelligence High Level Group, *The FlyAI Report: Demystifying and Accelerating AI in Aviation/ATM* 12 (2020).
43 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 7.
44 *Ibid*., at 7.
45 *Ibid*., at 16–17.
46 *Ibid*., at 24.
47 *Ibid*., at 17.
48 Robert Charette, *Automated to death*, IEEE Spectrum (2009).
49 Robert Charette, *Air France Flight 447's Final Minutes Reconstructed: Hints of the Automation Paradox Exacerbated by Inadequate Pilot Training at Work*, IEEE Spectrum (2011).

factor component (which is often found to have had a causal role in many accidents[50]) through the replacement of human pilots by a computer system. However, such an argument is farfetched (and presently, utopian) since it fails to consider the complexity of aircraft systems and accidents which involve a myriad of flight parameters (human, environmental and machine related) and it fails to account for the cases where the human's creative abilities enabled to overcome or mitigate a hazardous situation that a fully automated system may not have been able to manage.[51]

4.2 Aircraft production and maintenance

Secondly, the roadmap recognizes that 'aircraft production and maintenance'[52] could also benefit from the introduction of AI. Indeed, the growing amount of data held by producing and maintenance organizations could be optimally used, and value can be generated through the deployment of technologies such as Industrial Internet of Things (IIoT), predictive maintenance and digital twins.

The roadmap foresees the advent of digital twins in manufacturing, leveraging IIoT and predictive maintenance as anticipated opportunities for aviation.[53] In our view, EASA's policy vision would benefit from considering the impact of collaborative smart manufacturing practices on the duties of the actors involved as well as the diverse roles AI can play in aircraft manufacturing. Collaborative manufacturing refers to smart manufacturing developments in Industry 4.0, such as IIoT, cloud manufacturing, product customization and real-time asset monitoring etc., which lead to an end-to-end horizontal and vertical alignment of supply chain actors, manufacturers, and customers.

The Commission has expressed its support for the shift towards smart manufacturing, for example, by reinforcing public-private partnerships and digital industrial platforms. In a recent communication, it acknowledged the need to rethink the legal framework to accommodate smart manufacturing. The emerging regulatory issues call for further research, particularly regarding the safety and liability rules which are discussed in the following sections.[54] This is a clear gap in the roadmap which needs to be closed.

---

50 Husam Kharoufah et al., *A Review of Human Factors Causations in Commercial Air Transport Accidents and Incidents: From to 2000–2016*, 99 Progress in Aerospace Sci. 1 (2018), where they state that '[h]uman factors contribute to approximately 75% of aircraft accidents and incidents'.

51 *See* e.g. the case of United Airlines flight 232 which on 19 July 1989 suffered inflight structural damage (to its tail mounted engine) which led to the loss of many flight controls. Despite this adverse situation, thanks to airmanship and adequate crew resource management, the flight crew managed to crash land the aircraft saving two thirds of the people on-board.

52 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 8.

53 *Ibid*., at 8.

54 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Digitising European Industry:

Another gap concerns the different roles played by AI in aircraft manufacturing. AI can play at least four distinct roles in the context of aerospace manufacturing, namely monitoring, optimization, control and resilience. In our view, the roadmap should consider the different functions played by AI in aircraft production and maintenance to adjust its policy response based on the different challenges presented by each function.

Another interesting aspect that has remained outside the roadmap's scope concerns the emergence of new business models with novel operational approaches. The roadmap alludes to a 'shift in what engine manufacturers sell - not engines but flight hours'[55] – which raises the question of what the 'product' in the collaborative smart manufacturing production line would be. Manufacturers may see their functional boundaries extended, therefore blurring the traditional separation between 'product' and 'service'.[56] The traditional 'product-oriented paradigm'[57] may be challenged with manufacturers occupying increased operational-level functions[58] in cyber-physical environments.

4.3 Air traffic management and urban air mobility

Thirdly, air traffic management (ATM),[59] which is the field of aviation that deals with the safe and seamless management of air traffic, could also see growing use of AI. The roadmap portrays AI as enhancing 'data exchange between all actors', improving 'strategic planning', enhancing 'trajectory planning', increasing 'operational efficiency of Air Traffic Control' (ATC), and enabling 'higher ATM automation'. In parallel to this roadmap, the European Aviation/ATM AI High Level Group (EAAI HLG) prepared and published its FlyAI report which touches upon similar topics, therefore signalling the importance that aviation stakeholders give to AI in the ATM field.[60]

However, ATM is a particularly challenging use case of AI. A socio-technical system of systems, ATM is organized as a collaborative environment where major

---

Reaping the full benefits of a Digital Single Market,' COM/2016/0180 final (2016), https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0180 (accessed 17 Apr. 2020).

55 *Ibid*. *See also* European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 8.

56 *See* analogy with autonomous and connected cars in Orian Dheu, Charlotte Ducuing & Peggy Valcke, *The Emperor's New Clothes: A Roadmap for Conceptualizing the New Vehicle*, 75 Revue Transidit 14–16 (2020).

57 *Ibid*., at 13.

58 *See* Expert Group on Liability and New Technologies Formation, *Liability for Artificial Intelligence and Other Emerging Digital Technologies* 39, 44 (European Commission 2019), which refers to '[p]roducers, whether or not they incidentally also act as operators'.

59 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 8–9.

60 This High Level Group, which gathers EUROCONTROL, the European Commission and multiple (industry) partners, aims at advancing 'understanding among aviation/ATM actors of AI and its potential, demystifying the topic, and helping accelerate the uptake of AI in our sector', *supra* n. 42.

functions are delegated to or shared by increasingly automated systems and human operators. In the deployment of AI, incumbent actors may have to assume new responsibilities not covered by the legal frameworks currently in place.[61] This may be the case specifically for industry-led alliances which seek to engage in collaborative approaches to promote higher levels of resilience.[62] Deploying AI in operational settings can also change significantly the responsibilities of air traffic controllers.[63]

The roadmap also mentions AI technologies as key tools in fostering the development and deployment of 'drones, urban mobility and U-space'.[64] Urban air mobility, which may involve the use of unmanned aircraft, is seen as a promising medium of decongesting highly dense urban environments and facilitating point to point transportation and delivery within cities and between cities and rural environments. Because of drones' operational specificities[65] and the potentially large number of them navigating within a highly dense environment, it is argued that a change in paradigm in flight traffic management and operations will be required. Though not limited to urban environments, this is where U-Space/Unmanned air Traffic Management (UTM)[66] could come in handy. In Europe, the Single European Sky ATM Research Joint Undertaking (SESAR JU) advanced the U-Space[67] concept of operations[68] where AI technologies would prove as essential enabling tools. And fairly recently, EASA published its Opinion on a *High-level regulatory framework for the U-space,*[69] which was eventually re-drafted by the Expert Group on Drones which updated it for the Commission to consider,[70] paving the way for the future adoption of binding regulation by the EU. UTM, which is fundamentally a 'service

---

61 Ivo Emanuilov, *Shared Airspace, Shared Liability?*, SESAR Innovation Days 2018, 5 (2018), https://www.sesarju.eu/sites/default/files/documents/sid/2018/papers/SIDs_2018_paper_88.pdf (accessed 12 May 2020).

62 SESAR Joint Undertaking, *A Proposal for the Future Architecture of the European Airspace* 16 (2019), https://www.sesarju.eu/sites/default/files/documents/reports/Future%20Airspace%20Architecture%20Proposal.pdf (accessed 12 May 2020).

63 *Ibid*., at 44.

64 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 9.

65 *See* description of Mikko Huttunen, *The U-space Concept*, 1 Air & Space L. 44, 69–90 (2019).

66 *See* NASA, Concept of Operations for Unmanned Aircraft Systems (UAS) Traffic Management (UTM) V2.0 (2020), https://utm.arc.nasa.gov/docs/2020-03-FAA-NextGen-UTM_ConOps_v2.pdf (accessed 12 May 2020).

67 SESAR Joint Undertaking, *U-Space Blueprint* (2018), https://www.sesarju.eu/sites/default/files/documents/reports/U-space Blueprint brochure final.PDF (accessed 12 May 2020).

68 CORUS, *U-Space Concept of Operations* (2019), https://www.sesarju.eu/sites/default/files/documents/u-space/CORUSConOpsvol2.pdf (accessed 12 May 2020).

69 EASA, *Opinion n° 01/2020: High-Level Regulatory Framework for the U-space* (2020), https://www.easa.europa.eu/sites/default/files/dfu/Opinion No 01–2020.pdf (accessed 12 May 2020).

70 Expert Group on Drones (main group), *Draft Commission Implementing Regulation on a Regulatory Framework for the U-Space* (2020), https://ec.europa.eu/transparency/regexpert/index.cfm?do=groupDetail.groupMeetingDoc&docid=41693 (accessed 15 Sept. 2020).

oriented'[71] and partially de-centralized concept, is based on the premise that urban drone traffic management will be dealt on a local manner and that some ATM functions may partially be transferred to the operators themselves.

As mentioned in the roadmap,[72] AI will serve as a set of tools, approaches and technologies in the implementation of UTM and urban air mobility. The roadmap suggests that AI will help realize *self-separation* of unmanned aircraft within a highly dense environment through the use 'of "detect and avoid" (DAA) solutions'[73] which would require 'the support of ML solutions systems',[74] and could also 'support contingency management'.[75] Furthermore, such technologies will help make 'optimised decision making' and foster the efficient use and sharing of safety and non-safety critical data.

4.4 Safety management and cybersecurity

Finally, the roadmap mentions the potential applications of AI in different fields such as safety risk management, cybersecurity and the environment.[76] This article focusses on safety risk management and cybersecurity.[77]

In terms of safety risk management, the roadmap recognizes the potential of AI to support emerging risks detection, risk classification of occurrences, Safety Risk Portfolio design and prioritization of safety issues, understanding of safety data, identifying hidden correlations between different data silos and anomaly detection. Indeed, AI can potentially contribute to the development of a wholesale data-driven approach to safety risk management. However, the roadmap does not mention any of the major stumbling blocks before the adoption of such an approach.

The identification of safety hazards in air traffic operational data is contingent upon the exceptionally low probabilities involved and the need for rich datasets required to identify these (luckily) rare events. The authors suggest that the application of AI to safety risk management should be rolled out in stages and be

---

71 Cristina Barrado et al., *U-Space Concept of Operations: A Key Enabler for Opening Airspace to Emerging Low-Altitude Operations*, 24 Aerospace 7, 2 (2020).

72 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 9.

73 *Ibid*.

74 *Ibid*., at 9.

75 *Ibid*.

76 *Ibid*., at 9–11.

77 As recital 59 and Art. 88, para. 1 of the Basic Regulation prescribe, the Agency should take part in the cooperation concerning the area of aviation security, including cyber-security. It should contribute its expertise to the implementation, by the Commission and by Member States, of Union rules in that area. More specifically, the Commission, EASA and the Member States shall cooperate on security matters related to civil aviation, including cyber security, where interdependencies between civil aviation safety and security exist.

case-driven. Initially, the focus should be on the cases where AI can produce immediate safety gains. EASA should take a proactive role in defining these use cases and supporting the industry in implementing proof-of-concepts.

In terms of cybersecurity, the roadmap acknowledges that the effectiveness of AI comes at the price of increasing the attack surface. On the threats side, the inherent vulnerabilities of data-driven AI, such as data poisoning and adversarial attacks, represent a marked challenge. On the defender side, the roadmap recognizes the potential for AI to be leveraged in countermeasures and security controls.[78]

The authors believe the roadmap would benefit from highlighting the fundamentally different nature of traditional cyber-attacks and attacks against AI.[79] For example, in traditional cyber-attacks, the attacker exploits existing software vulnerabilities or social engineering techniques. In contrast, attacks against AI may exploit inherent and often well-known limitations of the algorithms used. Importantly, unlike traditional cyber-attacks, attacks against AI can be committed by a much broader scope of persons, who do not necessarily have to possess advanced knowledge of cybersecurity. In the context of aviation, future policy and legislative actions targeting cybersecurity of AI should account for these differences. Finally, when AI is deployed on the defender side to detect, correlate and disseminate knowledge derived from large-scale data analytics, the allocation of responsibilities in such an agile environment may prove challenging.

5 IDENTIFIED CHALLENGES AND 'TRUSTWORTHINESS BUILDING BLOCKS': WHAT THE ROADMAP SAYS

The roadmap alleges that current safety assurance frameworks may not be fully adapted to these technologies. Compliance with strict risk-based design requirements during the development of systems and equipment in aviation is a well-known and proven way to create development assurances. However, such a design-level assurance methodology is not entirely applicable to processes which depend on (continuous) learning, as is the case with some of the most prominent AI applications.

The 'lack of standardized methods for [the] evaluation of the operational performance of the ML/DL applications' as well as 'bias and variance issues' is another critical challenge.[80] An understandable explanation of this behaviour is equally crucial for humans to trust the system, particularly in an aviation context

[78] Such as malware detection. *See* on AI-enabled defences, Miles Brundage et al., *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation* 101, 59–60 (2018).
[79] Marcus Comiter, *Attacking Artificial Intelligence* 90, 1, 47–51 (2019).
[80] European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 15.

where automation is already complex enough.[81] The roadmap recognizes the 'lack of predictability and explainability of the ML application behaviour' as particularly problematic.[82] It confirms that the 'complexity of architectures and algorithms' as well as the 'adaptive learning processes' are 'incompatible with current certification processes'.[83]

The roadmap offers several potential remedies called 'trustworthiness building blocks' that could help tackle these issues.

The first remedy is 'trustworthiness analysis' that 'encompasses the seven gears of the EU ethical guidelines'[84] and that should be carried out with the focus put on 'oversight' over the human – AI interface. Trustworthiness analysis is also seen as a 'tool to investigate further […] a risk-based approach to AI/ML applications'.[85]

The second remedy concerns the novel (complementary) concept of 'learning assurance' as a means to overcome the safety assurance challenges. Learning assurance entails a shift to the training and verification of data sets' completeness and accuracy, bias mitigation, performance etc.[86] This alleged move from software engineering to data engineering requires the development of methods to check the correctness and completeness of data sets and to mitigate biases as an essential part of attaining the policy objective of public trust. Ultimately, learning assurance aims to provide stakeholders with a high degree of confidence that an AI application is doing what it is supposed to do.[87] The roadmap acknowledges there will be severe 'difficulties in keeping a comprehensive description of the intended function(s)'.[88] However, it adds, 'learning assurance' processes could supplement traditional safety assurance methodologies and formal methods[89] have also been identified as potential candidates.

The third remedy concerns explainability of AI decision-making processes. The roadmap recognizes this field as 'resolutely human-centric'.[90] However, it remains laconic in this part and limits itself to identifying existing research

---

81 *See* the recent case of the allegedly faulty Manoeuvring Characteristics Augmentation System in Boeing 737 Max 8 aircraft. *See also* the case of Air France Flight 447. *See* more in Nick Oliver, Thomas Calvard & Kristina Potočnik, *The Tragic Crash of Flight AF447 Shows the Unlikely but Catastrophic Consequences of Automation*, Harv. Bus. Rev. (2017), https://hbr.org/2017/09/the-tragic-crash-of-flight-af447-shows-the-unlikely-but-catastrophic-consequences-of-automation (accessed 16 Sept. 2017).

82 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 14, 18–19.

83 *Ibid*., at 15.

84 *Ibid*., at 16.

85 *Ibid*., at 17.

86 *Ibid*.

87 *Ibid*.

88 *Ibid*., at 14.

89 Matt Webster et al., *Formal Methods for the Certification of Autonomous Unmanned Aircraft Systems*, Computer Safety, Reliability, and Security. SAFECOMP 2011. Lecture Notes in Computer Science, vol. 6894 (Springer 2011).

90 European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 18–19.

initiatives in the field. It correctly identifies that the problem of defining what 'explainability' means and entails lies at the heart of the challenge.[91]

Finally, the roadmap recognizes AI safety risk mitigation as a way of mitigating the impact of the 'black boxes' problem which, it rightly acknowledges, may not always be sufficiently opened. Supervision of the application's behaviour is thus seen as a mitigation measure, e.g. by embedding the human actor within the decision-making loops, creating safety nets, hybridization of AI, monitoring the behaviour of the supervised AI agent by another AI agent or even licencing of AI.[92]

6 A 'FLIGHT PLAN TO BE CONTINUED': WHAT THE ROADMAP DOES NOT SAY

The roadmap is a good start and sets a heading for future discussions as it synthesizes the main issues and challenges and proposes several interesting avenues to address them. However, the document would benefit if it borrowed from the experience gained in other industries – something that seems like a missed opportunity in the current draft. For example, one way to implement trustworthiness analysis could be by creating safe spaces and processes, such as regulatory sandboxes,[93] where AI applications can be trialled, and their performance and accuracy assessed (6.1). Moreover, new regulatory tools such as data analytics and machine learning could be used within the aviation sector in order to foster a more agile and proactive management of safety risks (6.2). Considering the specificities of AI driven aviation, standardization may also have to be re-thought (6.3). Finally, though not falling directly under the remit of EASA, legal certainty in relation to liability need to be addressed as supplementary 'trust building blocks' (6.4).

6.1 Safe regulatory environment

Sandboxing is a term of art in computer security. Broadly, it refers to a security mechanism used to separate running software applications from the rest of the

---

91 *Ibid*., at 18.

92 *Ibid*., at 19. *See* for an assessment of some of these options from a legal and regulatory perspective, Ivo Emanuilov, *Autonomous Systems in Aviation: Between Product Liability and Innovation*, in *SESAR Innovation Days 2017: Selected Scientific Papers on Air Traffic Management* 98–110, 104–106 (2018).

93 Presently, there is no universally agreed legal or regulatory definition on the concept of 'regulatory sandboxing'. However, the term generally refers to the idea of setting up an environment where certain rules are alleviated thereby enabling the testing of innovative products, services and business models. *See* Federal Ministry for Economic Affairs and Energy (BMWi), *Making space for innovation – The handbook for regulatory sandboxes* 88 10 (2019). Ultimately, a regulatory sandbox is set up by a regulatory authority mostly with the objective of learning about the potential and risks of a particular technology with a view to adjusting the applicable frameworks based on first-hand information.

system. This enables control of the usage of resources and mitigates potential adverse consequences for the whole system.[94] Sandboxing is implemented, for example, when running unverified or untrusted code. In this sense, it can be described as '[a]n encapsulation mechanism that is used to impose a security policy on software components'.[95]

In a similar vein, a regulatory sandbox is created by an authority responsible for the implementation of the corresponding legal rules against which the innovation of a product, service or a business model is supposed to be tested. It could be described as a test area and process which benefits from certain regulatory leeway, but which also informs the regulator and legislator about possible future improvements in the applicable frameworks.[96]

As a regulatory tool, sandboxing involves the creation of a safe space and the development of a process whereby businesses can test new innovative products and services, business models or delivery mechanisms with mitigated risk of imposed sanctions and in close collaboration with and assistance from national regulators. In our view, the ambition of a regulatory sandbox for testing of aviation AI applications should be threefold. It should (1) enable the regulator to observe, steer the development and ensure the compliance of innovative approaches to certification of AI applications; (2) encourage innovation through competition, demonstrating a friendly regulatory view on innovation; and (3) allow for better prediction of risks, development of guidance material and building of holistic social risk management plans which can set an example for the entire aviation community.

Regulatory experience with sandboxing has so far been limited to the FinTech industry and, lately, also data protection.[97] Only recently have aviation authorities taken firm steps to establish regulatory sandboxes offering the participating entities to work with the regulator to test and trial innovative solutions in a safe environment.[98]

In our view, the roadmap would benefit from considering such novel regulatory approaches which may help bridge the knowledge and expertise gap between regulators and aviation stakeholders. In the long term, introducing a regulatory sandboxing process can inspire the development of targeted incentives

---

94 *See* on the various definitions of sandboxing, Michael Maass et al., *A Systematic Analysis of the Science of Sandboxing*, PeerJ Computer Sci. 2:e43, at 2–6 (2016), https://doi.org/10.7717/peerj-cs.43 (accessed 27 May 2020).

95 *Ibid*., at 5.

96 *Ibid*., at 7.

97 *See* e.g. the UK Information Commissioner's Office's Regulatory Sandbox, *The Guide to the Sandbox (Beta Phase)*, Information Commissioner's Office (2020),, https://ico.org.uk/for-organisations/the-guide-to-the-sandbox-beta-phase/ (accessed 19 May 2020).

98 *See* the recent example of the United Kingdom's Civil Aviation Authority, *Regulatory Challenges for Innovation in Aviation*, UK Civil Aviation Authority, https://www.caa.co.uk/Our-work/Innovation/Regulatory-challenges-for-innovation-in-aviation/ (accessed 1 May 2020).

for early movers in order to stimulate the transition to novel delivery models of AI-based products and services.[99] There is thus a need for future research into the potential models of regulatory sandboxing in aviation that would stimulate cross-border and interoperable exchange of experience and know-how.

6.2 New regulatory tools

The current safety management system paradigm is mainly a reactive one concerning the occurrence of safety events.[100] Essentially, this means safety assessments are typically updated in the wake of a major accident or incident. The agility of learning AI systems, however, calls for an equally agile and proactive management of safety risks.[101]

Aviation regulators have developed various methodologies to evaluate the risk of occurrence of safety events. One example is bowtie risk assessment models.[102] Bowtie models are a visual tool for describing risks which provides an opportunity to identify and assess the key safety barriers either in place or lacking between a safety event and an unsafe outcome. They provide a visual depiction of risk alongside a balanced risk overview for the whole aviation system between internal and external stakeholders, including third-party risks. Bowtie models are considered best practice guidance material for safety risk management at an operational and regulatory level offering an identification of critical risk controls and an assessment of their effectiveness.[103] As an aviation regulator, the UK Civil Aviation Authority has led the global use of bowtie models. A similar approach to AI applications could increase the regulatory body's awareness of the complexities and risks of the respective application, thus bridging the knowledge gap between the industry and the regulators.

The knowledge gap is particularly manifested in the process of certification of AI-driven aviation systems. The roadmap has clearly identified this issue but failed to acknowledge that AI technologies could also be an opportunity for regulatory

---

99 *See* on the need of incentives for early movers in the context of the proposed future architecture of the European airspace, SESAR Joint Undertaking, *supra* n. 62, at 14.

100 Allegedly, this has been changing with new regulatory initiatives, such as the ones concerning cybersecurity, urban air mobility etc.

101 Such frameworks are already being developed. *See* e.g. John Alexander McDermid, Yan Jia & Ibrahim Habli, *Towards a Framework for Safety Assurance of Autonomous Systems*, 2419 Artificial Intelligence Safety 2019 1–7, 3 (2019).

102 UK Civil Aviation Authority, *Introduction to Bowtie* (2020), https://www.caa.co.uk/Safety-initiatives-and-resources/Working-with-industry/Bowtie/About-Bowtie/Introduction-to-bowtie/ (accessed 23 Apr. 2020).

103 UK Civil Aviation Authority, *What Does Bowtie Show?* (2020), https://www.caa.co.uk/Safety-initiatives-and-resources/Working-with-industry/Bowtie/About-Bowtie/What-does-bowtie-show-/ (accessed 23 Apr. 2020).

bodies. Indeed, data analytics and machine learning could become powerful tools in the approval and certification of complex systems. For example, data analytics can present a dynamic view of how an automated system performs in time. Deploying automatic safety data monitoring based on historical analysis would also facilitate validation by human personnel. Furthermore, simplified data management, storage, cleaning, indexing and analysis practices would enable a better understanding of the capabilities of AI-based systems. Data analytics could therefore become an important element of future frameworks which aim for 'continued and proactive assessment in operation – in contrast to current safety management that tends only to update safety assessments in response to problems or accidents'.[104] In other words, AI has the potential not only to support the enforcement of existing rules but also to contribute to the development of new, more agile and more context-specific rules. Of course, this comes at a price and that price is the identified increased liability risk exposure stemming from the sharing with or delegation of rulemaking functions to autonomous systems.

6.3 Rethinking standardization of AI: Back to the drawing board?

The lack of standardized methods for evaluation of the operational performance of AI applications is well recognized by the roadmap.[105] However, it fails to acknowledge that the problem with standardization of AI in aviation is in fact an inherited one.

The history of automation in aviation clearly demonstrates standardization is the product of a lengthy process of consensus-building largely pushed by the industry itself.[106] Indeed, the aviation community has accumulated more than eighty years of experience in the standardization of automation.[107] In order to be able to reach a comparable point regarding standardization of operational autonomy, legacy and new systems alike would have to be interoperable to a degree enabling their safe, efficient and seamless communication.

The concern is real that the standardization of AI applications in aviation requires efforts to reconcile what has already been achieved in terms of standardization of automation with what AI has to offer in terms of opportunities. In other words, it may be that in order to come up with good standards for AI applications in aviation, existing standards of automation may have to be 'unrolled' and the entire stakeholder community may have to go back to the drawing board.

---

[104] McDermid et al., *supra* n. 101, at 3.
[105] European Union Aviation Safety Agency, *Artificial Intelligence Roadmap*, *supra* n. 2, at 15.
[106] Madeleine C. Elish & Tim Hwang, *Praise the Machine! Punish the Human! The Contradictory History of Accountability in Automated Aviation*, SSRN Electronic J. 5–6 (2015).
[107] *Ibid.*, at 6.

6.4 Liability and legal certainty as [supplementary] trust building blocks

Other issues which obviously do not fall within the remit of EASA will also have to be dealt in order to foster long-term *trust* in AI-driven aviation. Among others, (legal) concerns over accountability and liability aspects for accidents involving AI technologies may be raised.[108]

While air carriers/operators of unmanned aircraft, which are on the front line of liability exposure, will presumably continue to be subject to international/European liability rules for damages to passengers or cargo[109] and to national liability provisions for damages to third parties,[110] the emergence of a highly digitalized and automated ATM and U-Space eco-system will raise new questions. As decisional power and control gradually shift from the human operator(s) towards cyber-physical systems, questions arise concerning the adequacy and effectiveness of traditional liability regimes. Increased digitalization, collaboration and technological interdependencies accompanying the deployment of AI may blur the final allocation of legal responsibilities. Technology-induced autonomy will question the attribution of liability to human agents and/or legal entities through the blurring of intent and causation,[111] specifically in fault-based liability.

---

[108] *See* e.g. the European Union's Horizon 2020 research and innovation – Marie Skłodowska-Curie actions program under the Safer Autonomous Systems (SAS) project, grant agreement Grant Agreement n° 812.788.

[109] In the prospective case where such unmanned aircraft would carry passengers (and subject of further analysis), the EU air carrier operating though a valid operating license may continue to be subject to provisions of the 1999 Montreal Convention for the Unification of Certain Rules for International Carriage by Air as applied in EU law through Council regulation (EC) n° 2027/97 on air carrier liability in the event of accidents and amended through Regulation (EC) n° 889/2002. The air carrier is objectively and strictly liable for damages up to 100 000 SDRs, and presumed to be liable for damages exceeding 100 000 SDRs. There are some caveats though. First, both the Montreal and the Warsaw systems establish only 'certain rules' regarding second-party liability regime and therefore do not govern all aspects of liability in a uniform or exhaustive manner. The rules of Ch. III of the Montreal Convention determine the liability of the carrier and extent of compensation for damage. Many of these rules should be interpreted in light of the specifics of unmanned aircraft. Some legislators have taken a very restrictive approach by introducing an outright prohibition of carriage or persons or even cargo using unmanned aircraft, as is the case, e.g. with Art. 6 (3) of the Belgian Royal Decree on the Use of Remotely Piloted Aircraft in Belgian Airspace.

[110] The Rome Convention of 1952 deals with damages caused by foreign aircraft to third parties on the surface and establishes a strict liability regime for aircraft operators. However, few countries have ratified it. Liability for damages to third parties is therefore mostly governed by national law. In Europe, a fragmented patchwork of legislations exists with some countries abiding to a strict based liability regime whereas other countries apply fault liability. Such provisions usually target the operator but can also concern the owner or the pilot. Finally, some countries provide a capped liability while others don't. *See* Andrea Bertolini, *Artificial Intelligence and Civil Liability*, Study requested by the European Parliament JURI committee 118 (2020). *See also* Steer Davies Gleave, *Mid-Term Evaluation of Regulation 785/2004 on Insurance Requirements of Air Carriers and Aircraft Operators*, Report for the European Commission 22 (2012).

[111] Yavar Bathaee, *The Artificial Intelligence Black Box and the Failure of Intent and Causation*, 2 Harv. J. L. & Tech. 31 (2018).

In the context of ATM, where associated activities are usually considered sovereign and are thus covered by the legal regime of State liability,[112] the proliferation of data sources and actors governed by scattered (national) legal frameworks, coupled with the transition towards data-driven ATM, may also give rise to new liability challenges. These challenges include liability for reliance on infrastructure virtualization,[113] predictive machine learning models,[114] or indeed liability of standard-setters for their design choices and of regulators for certification.[115] The evolution of ATM towards an interconnected, interdependent cross-border system of systems, could challenge the current allocation mechanism based on a territorial connection of the act or omission. The internationalization of this data-driven infrastructure and the involvement of multiple actors contributing with varying degrees to a wrongdoing comes at odds with the largely 'single actor' approach to liability. This approach does not translate well to situations of multiple attribution where conduct is also mediated through network artefacts.[116] Finally, there is also the question of liability for delegation not only of decision- but also of rulemaking functions to autonomous systems, e.g. allowing a system to deviate autonomously from a rule within a predefined safety net.[117]

In the foreseen U-Space ecosystem,[118] shifts in operational and technological functions will occur. The difficulties in delineating the responsibilities[119] of multiple actors could make it more challenging to identify and prove the source of damages.

In the digitalized and automated ATM and U-Space eco-system, various parties could therefore be held liable when a damage occurs, individually or jointly, with a very diverse set of applicable liability regimes, both extra-contractual (e.g. product liability) and contractual, most of which are national specific adding to further fragmentation. The roadmap should mention these questions as they can be both critical enablers and stumbling blocks for the adoption of AI in the aviation industry.

Another contentious matter from a liability standpoint concerns the use of operational data in the manufacturing process. The integration of actors in the aviation supply chain driven by the 'servitization' of manufacturing expands not

---

112 Emanuilov, *Shared Airspace*, *supra* n. 61, at 2.
113 Francis Schubert, *The Technical Defragmentation of Air Navigation Services – The Legal Challenges of Virtualisation*, From Lowlands to High Skies: A Multilevel Jurisdictional Approach Towards Air L. 43–65 (2013).
114 Emanuilov, *Shared Airspace*, *supra* n. 61, at 6.
115 Hanna Schebesta, *Risk Regulation Through Liability Allocation: Transnational Product Liability and the Role of Certification*, 42 Air & Space L. 107–136, 133–134 (2017).
116 Emanuilov, *Shared Airspace*, *supra* n. 61, at 5–6.
117 *Ibid*., at 6–7.
118 *Ibid*., at 5–6.
119 *Ibid*.

only the geographical scope of production, but also the personal scope of the involved actors.[120] Integrating the customer in the value chain is one of the features of collaborative manufacturing. Customers can be involved through the process of product customization and cooperation, e.g. by feeding operational data directly into the manufacturing process. Importantly, the legal nature of the customer itself could vary significantly since States could also be customers when State aircraft are concerned.[121] The emerging 'oracle'-like predictive power of the manufacturer elevates its position in the aviation value chain to that of an almost omnipresent entity with significant authority. The possibility for the manufacturer to dynamically reconfigure and predict its product or service's behaviour increases its liability risk exposure, on the one hand, and intertwines it with that of the carrier, on the other.

In any case, if assuring safety is an essential building block in the *ex-ante trustworthiness* paradigm, providing *ex post* compensation through effective liability mechanisms is another pivotal block. Establishing a clear understanding of the liability mechanisms at play would improve legal certainty for the various stakeholders. As was suggested in an EU commissioned report,[122] a '*one-size-fits-all*' liability regime may not be relevant across all sectors. The Commission seems to agree on this point as it mentions in its report accompanying the White Paper on AI that a 'targeted, risk-based approach, i.e. taking into account that different AI applications pose different risks'[123] may be necessary. Specificities even within a domain such as aviation may warrant an even more granular approach.

Finally, the way AI technologies are portrayed in the policy and legal debate, particularly in media outlets, plays a key role in shaping the public opinion. Arguably, social perceptions of risk and the role of mass media in the process of

---

120 Furthermore, industrial AI refers not only to the horizontal integration of actors, but also of objects and systems. *See* for a cybernetics approach to industrial AI, Jay Lee, *Why Do We Need Industrial AI?*, in *Industrial AI: Applications with Sustainable Performance* 5–32, 16, 19 (Jay Lee ed. 2020), https://doi.org/10.1007/978-981-15-2144-7_2 (accessed 22 Apr. 2020).

121 Typically, these refer to aircraft used for military, customs and police services. The criterion is derived from Art. 3(b) *Convention on International Civil Aviation* (1944). In EU law, recital 10 of the Basic Regulation allows Member States to apply, instead of their national law, this regulation to aircraft carrying out military, customs, police, search and rescue, firefighting, border control and coastguard or similar activities and services undertaken in the public interest. *See* Recital 10 Regulation (EU) 2018/1139 of the European Parliament and of the Council of 4 July 2018 on common rules in the field of civil aviation and establishing a European Union Aviation Safety Agency, and amending Regulations (EC) No 2111/2005, (EC) No 1008/2008, (EU) No 996/2010, (EU) No 376/2014 and Directives 2014/30/EU and 2014/53/EU of the European Parliament and of the Council, and repealing Regulations (EC) No 552/2004 and (EC) No 216/2008 of the European Parliament and of the Council and Council Regulation (EEC) No 3922/91 (Text with EEA relevance.), 212 OJ L (2018), http://data.europa.eu/eli/reg/2018/1139/oj/eng (accessed 27 Nov. 2018).

122 Expert Group on Liability and New Technologies Formation, *supra* n. 58, at 36.

123 European Commission, *Report on the Safety and Liability Implications of Artificial Intelligence, the Internet of Things and Robotics, Report to the European Parliament, the Council and the European Economic and Social Committee*, COM(2020) 64 final, at 17 (2020).

risk construction could have significant impact on the liability exposure of carriers.[124] In times of uncertainty 'mass media find ample ground to influence the social construction of reality by infusing doubt and scepticism over risk management choices'.[125] This is of course valid not only of liability exposure of carriers, but equally of other actors involved in the process of manufacturing and using AI. Indeed, aviation is an 'industry that is very image-conscious and risk management-intensive, as well as (…) historically linked to high levels of customer care'.[126] The introduction of AI technologies capable of undermining trust makes it even more important to engage in meaningful trust-building exercises on concerns of 'operational reliability, transparency, consumer services, media relations, and environmental protection'.[127] It is therefore essential not only for regulators but also for the aviation community at large to consider the social perceptions of the risks of AI in a broader and holistic social risk management plan. Such a plan should be the product of coordination and mutual understanding of all stakeholders[128] in the 'ecosystem of trust'.

7 CONCLUSION

The Roadmap for AI in Aviation marks a first tentative step towards exploring the needed innovations and adaptations to the legal and regulatory frameworks in order to accommodate AI-driven aviation. Such a roadmap is a dynamic process whose output is expected to evolve overtime following consultations and discussions with industry stakeholders. The heterogeneous nature of potential applications and the intrinsic features of these technologies make it a challenging endeavour for regulatory authorities to approve and monitor AI-enabled products and/or services using the traditional tools in their toolbox.

Building upon the EU's strategic vision for AI, and more specifically on the AI-HLEG's concept of 'trustworthy AI', the roadmap provides some interesting routes for dealing with these issues. Among others, it heavily relies on an 'ethically driven' approach to AI in aviation which is seen as a key requirement in building 'public trust'. However, the blurry and subjective notion of 'ethical' AI alongside the uncertain (and context-specific) apprehension of 'public trust', raises many questions regarding its concrete implementation.

The roadmap highlighted as most pressing the regulatory hurdles concerning the certification and approval of AI-driven products and services. If this roadmap

---

[124] Leloudas, *supra* n. 32, at 56–57.
[125] *Ibid*., at 239.
[126] *Ibid*., at 239.
[127] *Ibid*., at 240.
[128] *Ibid*., at 239.

represents a good start, its assessment of the challenges and potential remedies could use further elaboration. Among the possible routes, there is the avenue of regulatory sandboxing as a powerful regulatory tool striking the balance between 'innovation for innovation's sake' and 'innovation which leads to actual safety gains'. AI technologies could also prove useful not only as a matter of regulatory attention, but as potential regulatory tools themselves. The aviation community may have to go back the drawing board and assess how AI standardization should be pursued in light of the already largely standardized automation technologies in place.

Finally, other fields which are not part of EASA's remit should also be assessed and considered in order to foster *trust* in AI, e.g. concerning *ex post* allocation of liability. Inducing *public trust* in AI-driven aviation not only entails safety-incentivizing mechanisms and processes, but also an effective and balanced allocation of responsibilities and collaborative trust-building exercises. This can only be achieved by engaging not only traditional aviation stakeholders, but also the society at large and especially the media, as constructors of public opinion.