UK
Civil Aviation
Authority

CAA's Strategy for AI

Building Trust in AI

5 Principles for AI and Autonomy

CAP2970
February 2024

The CAA's Strategy for AI

Introducing the CAA's Strategy for Artificial Intelligence (AI). With a dual focus on enabling the deployment of AI within the aviation sector and utilising it as a regulatory tool, this initiative aims to elevate safety measures, enhance operational efficiency, and foster innovation.

The aviation industry continues to embrace the transformative power of AI. It already enhances safety and efficiency through predictive maintenance, aiding air traffic management, and refining pilot training with advanced insights and simulations. But the future of AI will usher in a new era in aviation.

But what does it mean for the CAA? How will it affect the way we work, and what we regulate? These are the questions the CAA's Innovation Hub hopes to tackle with a new **CAA Strategy for AI, to be published in Summer 2024.**

This document is the second of 3 tools to support the strategy.

1. **Addressing the terminology of AI** is important to create common language so that we can have a level and transparent conversation with innovators.
2. **Providing a set of principles** that will help to steer how we regulate AI while enabling AI innovation to flourish.
3. **Horizon scanning** the future of AI, to keep us abreast of the technological developments

Artificial Intelligence and increasing degrees of autonomy have the potential to impact every part of the sector and across the CAA itself. These effects can be described in 3 broad categories for the CAA.

What we regulate

We are already seeing applications of AI in some of the proposals that reach our Innovation Advisory Services team in the CAA, and even within applications received by our regulatory approval teams.

How we regulate

The power of AI to rapidly process and analyse large volumes of data presents us with an opportunity we should not ignore. We are just scratching the surface of the potential to improve how we carry out our regulatory duties.

How we operate

As with any other organisation, the power that AI brings to help colleagues on a day-to-day basis is transformative. Whether it's helping to draft a new CAA publication, create a financial report, or produce meeting notes, AI tools will soon become a natural and essential part of our working lives.

The CAA's forthcoming strategy will explore the use and regulation of modern AI and high degrees of autonomy.

“Trust is the currency of a safe tomorrow

Stephen Covey

Trust is fundamental to success

Over many decades, the aviation system has developed a reputation for being one of the most trusted modes of travel. Introducing AI and autonomy must not degrade that trust.

If artificial intelligence and autonomy are to be accepted within the aviation ecosystem, or indeed within the CAA as a business or regulatory tool, the public, consumers, colleagues, and customers need to be able to trust it.

For software that is developed and used in the aviation sector, we must be able to assure the safety and security of it to the extent that the hazards and risks are deeply understood and appropriately mitigated. When it comes to introducing modern AI, regulations and standards will be paramount to achieving this.

But to be able to facilitate regulatory and policy development in the CAA, we are **introducing five AI Principles** described within this document. A principles-based approach enables us to assess a wide range of potential AI applications in a consistent manner, while also **allowing us to retain flexibility** as the technology develops. This marries with the existing risk-based approach to aviation regulation that is shared internationally. Most importantly, it contributes to **building public trust and acceptance of AI** in aviation.

> “When a passenger steps on an airliner, he or she demonstrates an inherent trust that the individuals and organizations that comprise the aviation system have done their jobs properly and to the best of their abilities.”
>
> Dr. Hassan Shahidi,
> President and CEO of the
> Flight Safety Foundation

Image: Microsoft Stock

Building trust with AI principles

These principles are aligned to the UK Government's Pro-Innovation Approach to Regulating AI. The CAA will follow and support the Government's approach to continuously reviewing and learning how these principles support or hinder AI regulation in aviation and will be providing feedback to help this.

Safety, security, and robustness

No harm to people, things, or environment.

Example: Medical diagnosis tool – The AI analysing scans should not lead to wrong treatments (safety), be protected from data breaches revealing patient information (security) and remain reliable even with imperfect scans (robustness).

Fairness & bias

No unfair treatment based on who you are, and free from bias.

Example: Facial recognition technology – The AI shouldn't misidentify people based on skin colour or other personal characteristics (fairness). It should be accurate and unbiased (fairness).

Accountability and governance

Someone responsible for AI's actions.

Example: Delivery drone crash – If a drone delivering your package crashes, you should know who to hold accountable (accountability). The drone company should have clear oversight over its AI systems (governance).

Transparency and explainability

Understand how AI works and why it decides.

Example: Loan application algorithm – You should understand why your loan was denied, not just get a generic "rejected" message (transparency). The reason could be explained as "insufficient income" or "negative credit history" (explainability).

Contestability and redress

Challenge unfair AI decisions and get help if harmed in some way.

Example: Automated parking ticket – You should be able to appeal a parking ticket issued by an AI system if you believe it was wrongly issued (contestability). You should have the opportunity to explain your case and potentially get the ticket overturned (redress).

The following pages provide a deeper description of each principle, notes on the CAA's interpretation in an aviation context, and an illustrative application of each principle against 4 examples.

Safety, Security & Robustness

DESCRIPTION OF THE PRINCIPLE

Applications of AI should function appropriately in a secure, safe, and robust way in normal and foreseeable use, and in cases of misuse or other adverse conditions. Risks should be identified, assessed, and carefully managed, with an ability to analyse the system's lifecycle in response to an inquiry.

CAA NOTES

Aviation already has a strong safety-first culture. Safety Management Systems are a systematic and proactive approach to managing safety risks. Introducing AI to an aviation system will bring about new risks that will need to be captured by the SMS.

The ability to adequately describe the safety, security, and robustness at all stages of the lifecycle is linked closely with the transparency and explainability of the system.

Safety, security, and robustness require assurance – clearly defined target levels of safety, security, and robustness, and methods to demonstrate that the system always maintains those levels.

1 | Detect & Avoid for RPAS

The system achieves target levels of safety with regards to loss of separation and collision risk.

The security of the system is such that it cannot be modified (purposefully or mistakenly) or tampered with.

The system is robust to uncertainties and changes in its operating environment, such us weather conditions, visibility levels, airspace traffic density, and behaviour.

2 | Automated ATM

The system achieves target levels of safety with regards to loss of separation and collision risk and prioritises these objectives over throughput and airspace utilisation objectives.

Security hazards such as airspace infringements by non-cooperative airspace users, and traffic not complying with advisories, are identified and the system can demonstrably mitigate against them.

The system is robust to uncertainties in its operating environment, such us weather conditions, visibility levels, airspace traffic density and behaviour.

3 | Qualification of MORs

The system accurately identifies and categorises safety related issues raised in the MORs, ensuring that concerns are never overlooked.

The system demonstrates robustness to differences in reporting and writing styles from various organisations and individuals, ensuring that safety related information is captured accurately regardless of who is reporting it.

The CAA could validate outputs against human knowledge, implement measures to guarantee the security of sensitive information contained in MORs, and ensure that the automated system doesn't compromise the confidentiality or integrity of the reports.

4 | Licencing Theory Questions

The CAA must ensure that the system generates accurate, dependable, and contextually appropriate questions that sufficiently cover the examinable content and evaluate depth of knowledge requirements. It could be possible to verify the system's reliability in producing questions that align with established aviation safety standards.

The use of AI shall not increase the risk that examination questions become known to candidates in advance of their examination.

The system could be secure from cyber-attacks that might attempt to exploit it to influence or disclose questions prior to exams.

Transparency & Explainability

DESCRIPTION OF THE PRINCIPLE

Organisations and individuals developing and deploying AI should clearly communicate when, how, and why it is used, and explain the system's decision-making process in an appropriate level of detail and timeliness that matches the risks posed by it. It should also be transparent to a human such that those decisions and outcomes can be traced and explained.

CAA NOTES

The key point to note here is the proportionality to risk. The degree of transparency and explainability is dependent on the complexity of the software. For example, machine learning techniques can develop software that is incomprehensible to an experienced software engineer.

It may be proposed that if an aviation system which has a high degree of risk associated to it is not sufficiently explainable or transparent, it will not be approved for use. The level of acceptance will adapt with technology and skills maturity, as they develop to an extent where complex machine-learnt systems can be explained through novel means.

1 | Detect & Avoid for RPAS

The system's manufacturer can communicate the functioning of the system in all operating modes.

When, how, and why a manoeuvre is made can be explained in real-time and retrospectively to the remote pilot and operator.

System detections and manoeuvring decisions are recorded and accessible for later interrogation.

2 | Automated ATM

The Air Navigation Service Provider (ANSP) provides comprehensive and accessible documentation explaining where and how the system operates independent of or alongside human controllers, including the level of autonomy, hazards and risks associated, and implications on human factors.

The system functions in an explainable fashion, allowing a human to effectively supervise and manage by exception.

All observations, calculations, instructions, and responses are recorded for later interrogation.

3 | Qualification of MORs

Establish transparency in the automated system's processes for reviewing and categorising MORs and provide insights into the algorithms and methodologies used to extract safety intelligence, facilitating understanding for stakeholders.

The CAA could ensure that the system's outputs and the logic applied to reach these are interpretable and explainable, enabling stakeholders to comprehend how safety intelligence is derived from MORs and fostering trust in the system's assessments.

4 | Licencing Theory Questions

The CAA could develop clear documentation and procedures outlining the algorithmic processes used in question generation, including the sources of information, such as the published syllabus and learning objectives, or databases accessed by the system.

The Approved Training Organisations are clear about the AI/Human intervention process involved in generating suitable question content.

Procedures could exist to ensure that questions are aligned to the syllabi, depth of knowledge requirements.

Fairness and Bias

DESCRIPTION OF THE PRINCIPLE

AI should be created, deployed, and maintained in a way which complies with applicable regulations and laws, and must not discriminate against individuals or organisations, or somehow create unfair commercial outcomes.

CAA NOTES

Creation, deployment, and maintenance is intended to describe all possible stages of an AI system's lifecycle.

Applicable regulations and laws are dependent on the context. These may range from technical regulations such as Air Traffic Management / Air Navigation Services (ATM/ANS) Regulations, through to the General Data Protection Regulation (GDPR).

It is intended that a system is updated throughout its life to reflect the requirements of any applicable laws and regulations. This is like any other type of system that must remain compliant with evolving laws and regulations.

ILLUSTRATIVE APPLICATIONS OF THIS PRINCIPLE

1 | Detect & Avoid for RPAS

The design and operation of the system comply with the applicable RPAS regulations and product standards and are consistent with the UK Rules of the Air as appropriate.

2 | Automated ATM

The system complies with the Air Traffic Management (ATM) / ANSP regulations, as well as operating appropriately within the context of the UK Rules of the Air.

It operates in such a way as to not result in unfair monopolisation of the airspace by a single airline operator.

3 | Qualification of MORs

The CAA could develop standardised criteria within the system for categorising MORs, ensuring that all reports are assessed fairly and objectively. Sources of bias could be identified, and biases avoided in the categorisation process, treating all incidents impartially.

Regular audits of the system's categorisation process could identify and rectify any biases or inconsistencies that may inadvertently affect the fairness of incident assessments. The system would likely need to comply with GDPR.

4 | Licencing Theory Questions

Ensure that the automated system does not exhibit biases or favour specific topics or styles of questions. The CAA could validate that questions cover a comprehensive range of knowledge areas relevant to pilot licensing without favouring certain aviation specialties or topics within leaning objectives.

In line with existing quality standards, there could be a regular review of question sets to identify any biases in the generated questions and adjust the algorithms or selection criteria to maintain fairness and diversity in the exam content.

The CAA provides a mechanism for post-exam feedback and review for continuous improvement.

Accountability & Governance

DESCRIPTION OF THE PRINCIPLE

Organisations should ensure the proper functioning of the AI system throughout its lifecycle and that it is created, operated, and maintained in accordance with applicable regulatory frameworks. This should be clearly demonstrated through their actions and decision-making process.

CAA NOTES

Where “fairness” is focused on the system, “accountability and governance” are aimed at the organisations involved. As such, there are organisational factors that affect the “proper functioning of the AI system” – roles, procedures, oversight, committees, and many more.

In aviation, the “operator” is a term defined in law and determines the legal responsibilities of an individual or organisation with regards to governance, safety reporting, training, and much more. Similar terms are defined for other stakeholders in the system. The application of an AI system into any of these roles should not predispose the application of existing legal responsibilities.

1 | Detect & Avoid for RPAS

Depending on the system’s assigned level of autonomy, accountability and responsibility are clearly established to provide clarity in the case of a collision or loss of separation.

The system manufacturer and RPAS operator have robust maintenance routines in place that ensure the software is updated regularly. They have procedures in place for issue identification and rectification.

2 | Automated ATM

The ANSP’s safety management system includes thorough procedures for routine maintenance of the system, identification, and rectification of issues, supported by clear organisational governance and procedures.

Depending on the system's assigned level of autonomy, and at every operating mode of the system, clear procedures exist to allocate accountability in the case of a hazardous incident to either the ATM system itself, the human ATM operator using the system, or the flight crew of the aircraft.

3 | Qualification of MORs

Clear accountability for the system’s performance and the decisions made in extracting safety intelligence from MORs could be established. Governance protocols outlining responsibilities and processes for quality assurance and oversight of the system could be developed.

The CAA could also implement mechanisms for tracking and documenting the system's decisions and actions taken based on the extracted safety intelligence, ensuring that there are procedures in place for addressing discrepancies or errors in the categorisation process.

4 | Licencing Theory Questions

Clear governance over the automated question generation system could be established, assigning responsibility for overseeing the algorithm's performance and the quality of questions produced. The CAA could implement procedures for quality assurance, validation, and periodic review of generated questions.

The system could align to existing protocols for addressing discrepancies or errors in generated questions, ensuring accountability in rectifying any issues and maintaining the integrity of the licensing exams.

Contestability & Redress

DESCRIPTION OF THE PRINCIPLE

Individuals and organisations should have clear routes to dispute harmful outcomes or decisions generated by AI. Appropriate application of this principle will be dependent on the context.

CAA NOTES

The focus here is on the ability to contest an outcome, as opposed to a specific functional output. For example, a system which adjusts aircraft control surfaces every nanosecond essentially creates 1,000,000,000 functional outputs every second. It would be impractical to enable contestability for each output. However, if the outcome of the system was to pitch the aircraft up and turn left 35 degrees to avoid a collision with another aircraft, this outcome will likely need to be contestable, particularly in the case of an incident.

There is a close relationship with many of the other principles, particularly transparency and explainability which enables contestability.

ILLUSTRATIVE APPLICATIONS OF THIS PRINCIPLE

1 | Detect & Avoid for RPAS

Operators, other airspace users, authorities, and the public can enquire about and contest the decisions made by the system.

Dispute rectification is prioritised by the OEM above non-safety activities.

2 | Automated ATM

Instructions issued by the system are accessible by users. The ANSP provides a user-friendly mechanism to raise concerns or disputes.

Disputes are prioritised above non-safety activities.

3 | Qualification of MORs

The CAA could offer a structured mechanism for stakeholders to challenge or provide additional information related to the system's categorisation of MORs. The CAA could prioritise the review and resolution of disputes or concerns raised about the extracted safety intelligence.

A transparent process for handling disputes or appeals could be developed, including a reassessment of incidents if contested, to maintain the accuracy and reliability of safety intelligence

4 | Licencing Theory Questions

The CAA could offer a mechanism for candidates or stakeholders to challenge or report issues related to the generated questions, enabling them to contest inaccuracies, biases, or other concerns. The CAA could prioritize the resolution of disputed questions, especially those impacting exam outcomes.

Clear processes could be in place to handle disputes or concerns raised by candidates, including a review and re-evaluation of contested questions to maintain the credibility of the exam

Taking a pro-innovation approach

Where did these principles come from?

First created by the Organisation for Economic Co-operation and Development (OECD), these 5 AI principles were developed to be flexible and practical enough to be adapted to any sector or application. Today, the Department for Science, Innovation & Technology (DSIT) has proposed to embody 5 similar principles within a UK framework for the pro-innovation regulation of AI. It is therefore necessary to apply and test these in the aviation sector.

With guidance and expertise of colleagues in DSIT, the Department for Transport (DfT), the CAA, other UK regulators, and other National Aviation Authorities, we have adapted the principles to reflect the responsibilities of the UK Civil Aviation Authority and the broad sector we regulate.

These principles are intended to protect people and enable innovation by providing guidance for anyone in the aviation ecosystem who is creating, deploying, or maintaining AI systems. Furthermore, they should provide a common framework for conversations with the regulator, as well as a means for the various capability teams within the CAA to develop their own training, assessment, and oversight frameworks. They are therefore designed to be flexible enough to apply to a broad range of AI applications, but also specific enough to ensure adequate consistency and assurance.

While the text primarily describes “AI”, the intent is to also reflect high levels of autonomy where in most cases (but not all) it is expected that autonomy of this type is enabled by AI technologies and methods.

What are their purposes?

To be applicable across our entire regulatory and business remits, wherever AI is introduced.

To provide guidance to innovators to help prepare for engaging with regulation.

To enable development of new policies, regulation, acceptable means of compliance, and standards

To support training and skills development within the CAA

What's Next?

The CAA's Strategy for AI will provide a 'north star' to guide how the CAA approaches the regulation of AI and autonomy, while also giving innovators guidance on how to prepare for engaging with the CAA.

During the early part of 2024, we will develop the 3 tools further (terminology, principles, technology outlook) while aiming to publish a strategy document in the Summer.

In parallel with this work, we will develop an initial portfolio of activity and deliverables across two parts:

- Part A: Regulating AI
- Part B: Using AI

During this time, and beyond the publication of the strategy document itself, the CAA is open for engagement and discussion, and ready to listen.

Visit the CAA Innovation website

for latest updates, guidance and challenges

caa.co.uk/innovation

Tell us what you think

We are keen to hear your views on the content of this publication. Please get in touch via the email address below.

To submit feedback please contact StrategyforAI@caa.co.uk

Additional information

The DSIT's pro-innovation white paper proposes 5 principles based on the OECD's 2018 analysis. It should be noted that we expect the OECD to refresh their analysis and guidance in 2024.

The DSIT principles have been adjusted slightly to accommodate the breadth of aviation applications. Our own analysis and testing of these principles will be shared with DSIT as part of their pro-innovation approach to the regulation of AI.

We expect that each area of the CAA will use these principles, at the appropriate time according to demand a resource availability, to develop new or amended policy and regulations. We do not expect to see an overarching regulatory power or requirement for aviation; however, we remain open to feedback and to learn from experience within the industry, academia, and government.

Cover image: Adobe Firefly