Fabrice Drogoul

Philippe Palanque

COLLABORATION WITH AUTOMATION

OVER THE LOOP, INSIDE THE LOOP OR OUTSIDE THE LOOP?

With the current hype for artificial intelligence, aviation and other sectors are looking at how new technologies can be exploited to benefit performance. But AI brings – and expands – problems already known about from research in automation. Fabrice Drogoul and Philippe Palanque explain why caution is required.

KEY POINTS

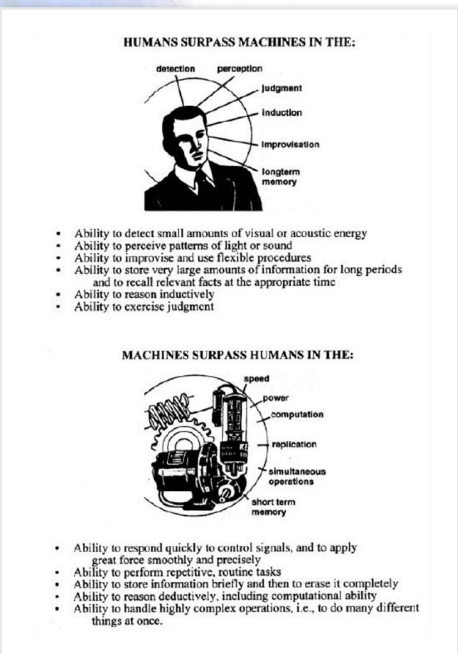
- **AI in air traffic management:** New AI tools are being proposed for air traffic management. These technologies can either replace or assist human tasks, but they bring fresh challenges, especially in roles where safety is critical. In safety-critical contexts, very high reliability is required. This is not yet achieved by AI.
- **Human-AI teamwork:** Roles can vary in human-automation setups. In some cases, the human is fully in charge with automation as a tool. In others, control is split or even reversed, where the system takes the lead. It's crucial to get the balance right to keep humans "in the loop" where it counts.
- **Complexity and risk:** Complexity brings increasing risk of system failure, and so aviation regulators are cautious, advising against quickly rolling out new AI technologies until they've proven reliable.
- **Regulatory limits:** Only certain types of automation where humans can stay in control will be approved in the near future. While full AI autonomy is a possibility, deployment, robustness, and safety issues will prevent early use of AI technologies in safety-critical contexts.

INTRODUCTION

The recent advances in artificial intelligence (AI) technologies have been perceived as a game changer in air traffic management as in many other areas. Statistical AI technologies that are used for tasks such as pattern recognition or item classification are based on machine learning (ML) technologies. Symbolic AI technologies that are used for tasks such as diagnosis or decision-making are based on rule-based technologies. These technologies are very different, but they share the fact that they are designed to replace or augment human-performed activities with computer-performed activities. Whether or not AI technologies are embedded, this process has been known for nearly a century as 'automation', and many limitations, pitfalls and drawbacks have been studied (see Drogoul & Palanque 2019). These are usually exacerbated when AI technologies come into play.

A concrete example of ML-based computer vision technologies is the one being deployed for detecting undesired objects on airports` runways (see Noroozi et al., 2023). In that contribution, the authors propose a stepwise processing of computer images for foreign object detection (FOD). Based on a widely used training dataset called YOLOv4, the highest accuracy is about 93.81%. This could be seen as a good level of precision for the AI technology. However, this means that the system is wrong once every 16 FOD. In numbers, the system produced 134 false positive (a non-existing object was detected) and 215 false negatives (a non-detected object was actually present on the runway) and 1566 true positives. With respect to the expected reliability level of safety-critical systems, this reliability level would be considered poor.

The following generation of work considered automation with electronics and algorithms. Already in 1985, Chambers and Nagel were worried about automation drawbacks "As more and more automation is incorporated in aircraft, the essential question becomes one of autonomy: Should the automated system serve as the human pilot's assistant, or vice versa?" (p. 1187).



HUMANS SURPASS MACHINES IN THE:

- Ability to detect small amounts of visual or acoustic energy
- Ability to perceive patterns of light or sound
- Ability to improvise and use flexible procedures
- Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
- Ability to reason inductively
- Ability to exercise judgment

MACHINES SURPASS HUMANS IN THE:

- Ability to respond quickly to control signals, and to apply great force smoothly and precisely
- Ability to perform repetitive, routine tasks
- Ability to store information briefly and then to erase it completely
- Ability to reason deductively, including computational ability
- Ability to handle highly complex operations, i.e., to do many different things at once.

Figure 1: A reproduction of Figure 1 from (Fitts, 1951) showing the reference to mechanical automation

Advances in AI technologies are expanding the potential for automation, enabling computers to take over tasks that are difficult to describe with algorithms, such as generating images from textual prompts. A recurring key element in the process of automatisation is that humans remain involved and must collaborate with the automation to carry out their tasks.

DIVERSE VIEWS ON HUMAN-AUTOMATION COLLABORATION

The ways in which humans can collaborate with automation is presented using the GUSPATO model in Figure 1. GUSPATO is the acronym composed with the first letter of the seven types of collaboration where control, authority and responsibility (according to the terminology of the RCRAFT framework for automation of Bouzekri et al., 2021) migrate between the technical system embedding automation and the human.

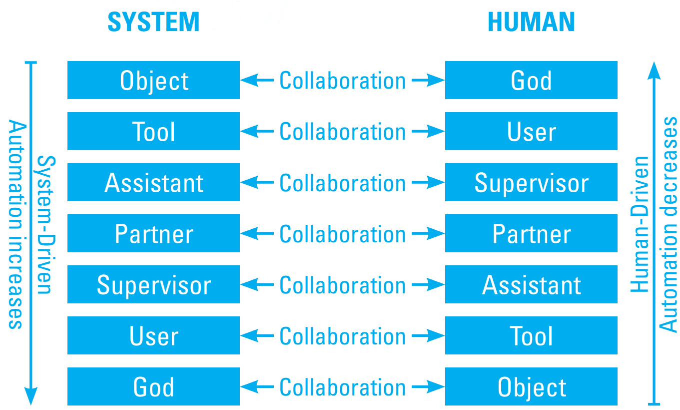


Figure 2: GUSPATO: A seven-level classification of human-system collaboration

Each line of the figure corresponds to one type of collaboration with automation. The first line describes a collaboration where the human is seen as a 'god' and creates the system and its outcome. In that case the system can be seen as an object belonging to the creator.

The second line represents the classical use of computers where the system is seen as a tool used by the user or operator. The tool may embed some automation but the control, the authority and the responsibility remain with the user/operator. The human is, here, inside the interaction loop perceiving information provided by the system, cognitively processing it and triggering system functions when appropriate.

Line three shows an unbalanced sharing of control, authority and responsibility between the system (as assistant) and the human (as supervisor). Automation is more complex, and more complex tasks are performed by the system, following a delegation of tasks by the human. The human still holds control, authority and responsibility but positioned over the interaction loop (monitoring the partly autonomous behaviour of the system).

Line four in the middle of the figure corresponds to a symmetric relationship for control, authority and responsibility between the system and the human. In this type of collaboration, both entities can delegate tasks to the other entity and monitor their performance. Authority and responsibility are shared, and the human can be considered as outside of the interaction loop when the systems perform tasks autonomously.

Line five corresponds to reversal of the collaboration presented in line three but now the human is an assistant to the system. In that context the system might require the human to perform tasks and will monitor the performance of the human. Such reversal of roles in the collaboration is similar for the last two lines of the figure. The last line corresponds, for instance, to generative AI where objects are created by the system and the human is an object amongst many others.

ROBUST AUTOMATION IN SAFETY-CRITICAL CONTEXTS

A key issue in the use of GUSPATO model is that the lower lines require more complex algorithms and, in some cases, might require the exploitation of AI technologies. While this might be acceptable for entertainment or mass-market systems, complexity in computer systems is a precursor for failures (at least in the area of software where "Complexity metrics are better predictors than simple size metrics of fault and failure-prone modules", according to Fenton and Olhsson, 2000).

When new technologies for producing computing systems appear (a new programming language, for instance) significant effort is required to harden the technology making it suitable for deployment in critical contexts. This is why it is wiser and safer to keep older technology in use and to refrain from being an early adopter in order to avoid disillusion, as is the case with the fantasy of fully autonomous driving (Cusumano, 2020).

Level 1 AI: assistance to human

- **Level 1A:** Human augmentation
- **Level 1B:** Human cognitive assistance in decision-making and action selection

Level 1 AI: human-AI teaming

- **Level 2A:** Human and AI-based system cooperation
- **Level 2B:** Human and AI-based system collaboration

Level 3 AI: advanced automation

- **Level 3A:** The AI-Based system performs decisions and actions that are overridable by the human
- **Level 3B:** The AI-based system performs non-overridable decisions and actions (e.g. to support safety upon loss of human oversight)

Figure 3: The levels of automation based on Human-AI interaction in the EASA AI Roadmap

"In the aviation domain, regulators such as EASA are defining safeguards to prevent the exploitation of AI technologies before there has been a demonstration that they have been made robust enough."

In the aviation domain, regulators such as EASA are defining safeguards to prevent the exploitation of AI technologies before there has been a demonstration that they have been made robust enough to meet the development assurance and safety levels identified for the target system. The EASA roadmap to AI (in its current version 2.0, EASA, 2023) provides a clear path toward adoption of AI in the long term, demonstrating that only the first three lines of GUSPATO model will be available in the coming years. For Human-AI teaming, EASA has now provided a clear distinction between a) cooperation where AI and humans are working together but with distinct goals and collaboration, and b) where AI and humans are working together towards the same goal (see Figure 2). The advanced automation level foreseen in the long term encompasses some full automation that can be overridden by the operator (human still in control). Only as a safety tool (when incapacitated) can AI perform in the most advanced automation level actions and take decisions that cannot be overridden (see Figure 3). This corresponds to the fact that AI trustworthiness requirements, as identified in EASA (2020), are not met yet.

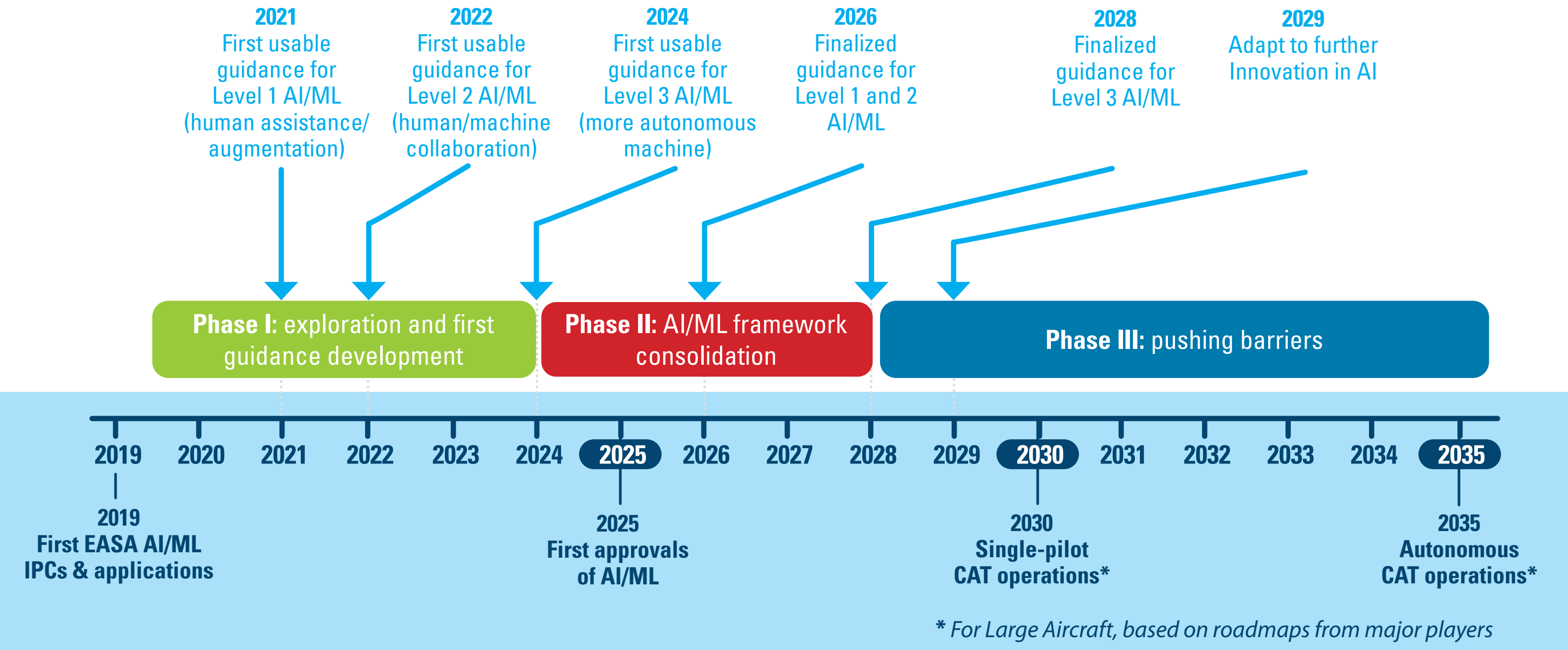


Figure 4: EASA AI Roadmap target dates for deliverables (adapted from EASA [2020], p. 13)

This paper presented the issues related to artificial intelligence technologies, and in particular machine learning, in the context of automation. The trustworthiness of these technologies is far behind the currently deployed safety-critical technologies. Based on output from the EASA AI task force, we can see that certification authorities take trustworthiness issues very seriously, and understand that deployment of AI technologies is not there yet. We are thus far away from a real collaboration between such technologies and operators, at least in such critical contexts. A lot of research work remains to be done in order to go from design options to implementation, certification and deployment of such systems.

REFERENCES

Bouzekri, E., Martinie, C., Palanque, P., Atwood, K., & Gris, C. (2021). Should I add recommendations to my warning system? The RCRAFT framework can answer this and other questions about supporting the assessment of automation designs. In Ardito, C., et al. (Eds.), Human-Computer Interaction – INTERACT 2021. INTERACT 2021. Lecture Notes in Computer Science, vol 12935. Springer. https://doi.org/10.1007/978-3-030-85610-6_24

Chambers, A., & Nagel, D. (1985). Pilots of the future: human or computer? Communications of the ACM, 28(11) (Nov. 1985), 1187–1199. https://dl.acm.org/doi/pdf/10.1145/4547.4551

Cusumano, M. (2020). Self-driving vehicle technology: progress and promises. Communication of the ACM, 63 (10), 20-22. https://dl.acm.org/doi/10.1145/3417074

Drogoul, F., & Palanque, P. (2019). How to make automation a good solution to the current problems in ATM? 2019 Recommendations education and performance in aviation: realizing and sustaining benefits. Hermes, Air Transport Organisation.

EASA (2020, February 7) EASA AI roadmap 1.0, https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-10

EASA (2023, May 10). Artificial intelligence roadmap 2.0: A human-centric approach to AI in aviation. https://www.easa.europa.eu/en/document-library/general-publications/easa-artificial-intelligence-roadmap-20

Fenton N. E., & Ohlsson N. (2000). Quantitative analysis of faults and failures in a complex software system. IEEE Transactions on Software Engineering, 26, (8), 797-814. https://ieeexplore.ieee.org/document/879815

Fitts, P. (1951). Human engineering for an effective air navigation and traffic control system. National Research Council, Washington, DC. https://apps.dtic.mil/sti/citations/ADB815893

Noroozim M., & Shah, A. (2023) Towards optimal foreign object debris detection in an airport environment. Expert Systems with Applications, 213, Part A. https://doi.org/10.1016/j.eswa.2022.118829

FABRICE DROGOUL is Human Factors and Safety Expert at EUROCONTROL, course manager of the Design and Assessment of Systems Using Human Centered Approaches and Fatigue Management course. Fabrice and Philippe started their collaboration in 2015 on automation within the Higher Automation Levels in Aviation network of excellence within SESAR WPE.
fabrice.drogoul@eurocontrol.int

PHILIPPE PALANQUE is professor in Computer Science at the University Toulouse III – Paul Sabatier in Toulouse, France where he leads the department on Reliability of Systems and Software at the IRIT research institute. The main driver of Philippe's research over the last 30 years has been to address in an even way Usability, Safety and Dependability in order to build trustable safety critical interactive systems. philippe.palanque@irit.fr