Considerations of Artificial Intelligence Safety Engineering for Unmanned Aircraft

Sebastian Schirmer ✉, Christoph Torens, Florian Nikodem, and Johann Dauer

German Aerospace Center (DLR), Institute of Flight Systems
Lilienthalplatz 7, 38108 Braunschweig, Germany
http://www.dlr.de/ft/en
{sebastian.schirmer,christoph.torens,florian.nikodem,johann.dauer}@dlr.de

Abstract. Unmanned aircraft systems promise to be useful for a multitude of applications such as cargo transport and disaster recovery. The research on increased autonomous decision-making capabilities is therefore rapidly growing and advancing. However, the safe use, certification, and airspace integration for unmanned aircraft in a broad fashion is still unclear. Standards for development and verification of manned aircraft are either only partially applicable or resulting safety and verification efforts are unrealistic in practice due to the higher level of autonomy required by unmanned aircraft. Machine learning techniques are hard to interpret for a human and their outcome is strongly dependent on the training data. This work presents the current certification practices in unmanned aviation in the context of autonomy and artificial intelligence. Specifically, the recently introduced categories of unmanned aircraft systems and the specific operation risk assessment are described, which provide means for flight permission not solely focusing on the aircraft but also incorporating the target operation. Exemplary, we show how the specific operation risk assessment might be used as an enabler for hard-to-certify techniques by taking the operation into account during system design.

Keywords: Aerospace · Certification · AI-based System · Unmanned Aircraft Systems · Verification and Validation

1 Current State in Unmanned Aviation

In aerospace, safety considerations are one of the main concerns and cost drivers. Only after certification of the aircraft, a participation in civil aviation is allowed. In order to direct companies in the process of achieving certification and to support certification authorities, several guidelines were presented.

The SAE[1] Aerospace Recommended Practices (ARP) and the RTCA[2] DO standards are prominent acceptable means of compliance. As shown in Figure 1,

[1] Society of Automotive Engineers is a global associations, developing standards for aerospace, automotive, and others.

[2] Radio Technical Commission for Aeronautics is a private, not-for-profit association, developing technical guidance.

several guidelines exist for development processes and safety assessments, both for the development phase and the operational phase. These standards impose high requirements on development and verification, e.g. requirements on code coverage increase with the criticality of a software item. The so-called MC/DC metric, which is required for the most critical assurance level, is a significant driver for costs of verification alone [5, 6]. The overall objective is to demonstrate that the system under development is working correctly or, where this is not possible, show the high quality of the development process.

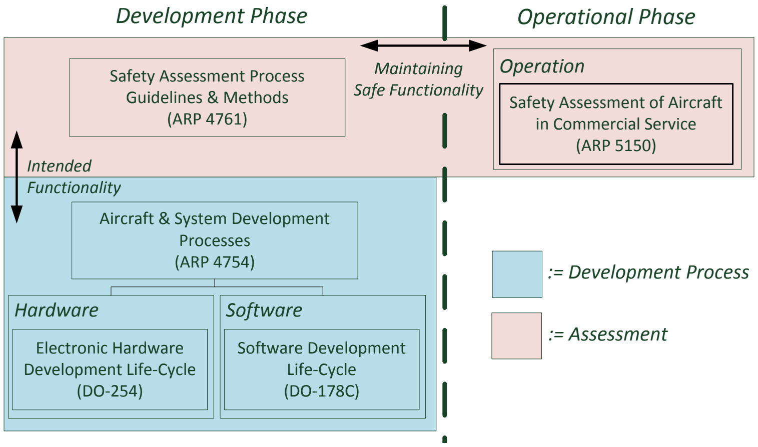
Development Phase
Operational Phase
Maintaining Safe Functionality
Safety Assessment Process Guidelines & Methods (ARP 4761)
Operation
Safety Assessment of Aircraft in Commercial Service (ARP 5150)
Intended Functionality
Aircraft & System Development Processes (ARP 4754)
Hardware
Electronic Hardware Development Life-Cycle (DO-254)
Software
Software Development Life-Cycle (DO-178C)
🟦 := Development Process
🟥 := Assessment

Fig. 1. Overview of aircraft development guidelines - How they complement each other.

Unmanned Aircraft Systems (UAS) are aircraft without a human pilot on board but, instead, with an operator on the ground having some form of control over it – rendering the UAS into a complex distributed system. Amazon[3], DHL[4], as well as other companies investigate to deliver goods with the overall objective to facilitate unmanned last-mile delivery.

An autonomous flight capable of contingency measurements in case of off-nominal behavior is highly desirable for these operations. At best fully autonomous, without an operator on-ground who is unfavorable for long-distance deliveries. Although the aircraft is unmanned, it can still do harm on-ground or in-air. We mentioned that certification of manned aircraft is cost intensive and hard to achieve. For presented business cases to work, the main concern is not anymore the performance of the autonomy functionalities – vision-based autonomous flying system do exist and perform well – but instead the certification of these functionalities [8]. A loss of reputation for these business cases could be fatal. For instance, machine learning techniques like Convolutional Neural Networks allow to infer weights, given a loss function for the domain and enough

[3] Amazon Prime Air, www.amazon.com/primeair
[4] DHL Parcelcopter, http://www.dpdhl.com/parcelcopter

training data. But, how can they be certified? As of yet, such nets are not capable of outputting in which situations they are working. Additionally, attacks on machine learning with adversarial examples were presented which cause the model to make mistakes by only changing a few pixels [11].

However, the achievements in machine learning or artificial intelligence (AI) in general are undeniable and research on e.g. Bayes Deep Learning exist which combine Bayesian approaches with deep learning to reason about the model confidence [4]. Nonetheless, to our knowledge, these confidence values are similarly dependent on the training data and, therefore, no general statement in which situations the trained nets work is possible. Therefore, a different question might be: How can we embed these techniques within the overall system design to enable certification?

In industry, a common architecture to facilitate hard-to-certify components is to switch to a more conservative backup component in case of an hazardous situation. For such an architecture, two main aspects are required. First, an alternative conservative action has to exist. Second, the hazardous situation needs to be detectable. Exemplary, in manned aviation the Brake-to-Vacate (BTV) system by Airbus is a “convenience” function on top of the safety-critical Runway Overrun Prevention System (ROPS). The objective of BTV is to optimize the path from the runway to the taxiway. The optimized path reduces the wear on brakes and tires as well as the time the aircraft spends on the runway and, therefore, increases the throughput of the airport. ROPS is a simpler function than BTV which is directly integrated within BTV. It either warns the pilots if the runway is insufficiently long before touchdown/deceleration or applies maximum braking when deceleration already started.

In late 2015, the European Aviation Safety Agency (EASA) introduced the so called specific category which allows to combine the reasoning of aircraft system and target operation. In the next section, we present the new categories, an upcoming guideline for the risk assessment, and the DLR project ALAADy. Then, we discuss how these innovations can help to apply AI techniques. Specifically, we argue how the system architecture and regulatory frameworks can interlock to enable the use of hard-to-certify components. Finally, we give a future perspective.

2 Trends in Aerospace

EASA recently introduced three categories of UAS operation that use separate sets of regulation, based on the intrinsic risks involved [2]. The three categories are referred to as open, specific, and certified. The open category is reserved for low risk operation of unmanned aircraft below 0.25 kg. This category requires no or minimal regulation. The certified category is used for operations that are of an equivalent level of risk comparable to manned aviation, therefore the same level of rigor for development and verification is applied. The new specific category allows a step-wise adaptation of regulation and certification requirements between the open and certified categories. For flight permission, the specific category relies on a risk assessment to determine the required level of certification requirements. As risk assessment, EASA recommends the so-called specific operation risk assessment (SORA) that is currently being developed by JARUS [7]. In contrast to the certified category, the specific category is not targeted solely on the UAS, but towards the operation of a specific UAS in its entirety including the mission, the environment, operation conditions, rigor during development as well as operator, and pilot qualification. SORA is currently under development and a first deployment by authorities is likely within years.

DLR (German Aerospace Center) is currently supporting the SORA development by applying it to a cargo application within the project ALAADy (Automated Low Altitude Air Delivery) [1]. The mission consists in cargo delivery on a range of around 600 km flying over sparsely populated areas. The unmanned aircraft are intended to fly in very low level to circumvent most of the air traffic. These efforts help to further research interdependencies of the SORA process with UAS design and development, specifically DLR investigates the advantages of the new specific category concept to maintain a safe operation. A controlled termination of the UAS in safe areas is used to minimize the risk to third parties on-ground or in the air and, thus, ultimately makes the operation safe. This is necessary, since a backup pilot cannot be aboard the target aircraft which is in contrast to autonomous cars. There, a backup driver supervises the system and takes control in case of emergencies. In aviation, the use of Optional Piloted Vehicle (OPV) during development is possible but the required components and pilot might be too heavy for the real system [3].

One hope utilizing the specific category is to develop cheaper UAS with new autonomous capabilities. DLR works on a system architecture that is leveraging the containment of the risk of the operation. The goal is to use technologies that are not usable in a traditional certification process for aerospace due to the high requirements on verification objectives imposed by certification standards.

3 AI Applicability

The major innovation with the specific category and SORA is that it considers not only the UAS but also factors the target operation in. Explicitly, manned certification always assumes harm on people in case of an operation being out of control. Therefore, a high confidence in the system is mandatory. SORA relieves this burden by incorporating the risk to harm people on-ground or in-air. So called harm and threat barriers need to be in place to achieve equally high confidence with respect to the operation. This is a huge change because it allows to fit the system design to the target operation. Already limitations to basic operation parameters like daytime or flight altitude have significant impact on the system design and its certification effort. An extreme example would be the case of operations only involving flying in low-altitude in the desert at daytime: for this no highly reliable ice protection system is required. Similar scenarios can be easily found for many possible UAS operations in Europe. Additionally, it allows to monitor the operation parameters instead of an explicit functionality.

For instance, surveying crops by unmanned aircraft where limiting the airspace to a pre-defined area, e.g. the field, reduces the risk of harming people on the ground or in the air. There are two main capabilities: the monitoring (MC) and the flight capability (FC). MC monitors the UAS and has the authority to activate countermeasures or the flight termination whenever required to avoid a breaking out of the pre-defined area. Regarding SORA, a certification of MC plays the major role for flight certifiability since it prevents harm on-ground or in-air, i.e. acts as harm barrier. Complementary, FC controls and manages the UAS in an efficient and economic way such that all crops are sufficiently viewed. Possibly, reacting upon environmental conditions and based on vision. Roughly spoken, FC tries to achieve the highest performance whereas MC takes care that no environmental/operation conditions are violated. There are different ways of implementing such capabilities within a system each of them having different advantages and disadvantages considering certification effort, cost, and performance. In Figure 2, possible system designs are depicted and now discussed:

(a) Both capabilities are embedded in one software component running on a dedicated hardware. Interaction-wise, this setup offers the closest interaction between MC and FC due to no resource separation. However due to certification efforts in conjunction with cost, applicable hardware is limited, e.g. multi/many-core processors and multiprocessing. Software-wise, the performance might suffer due to the hardware restrictions and certification is also difficult due to no clear separation between MC and FC.

(b) Advantages and restrictions on hardware similarly remain. Limited computing power prevent the usage of advanced algorithms. However, the clear separation between MC and FC in software improves the certification efforts software-wise. Still, there is a dependency between MC and FC. However, the certification efforts for FC are reduced. Showing the absence of effects on the hardware, i.e. the unobstrusiveness regarding MC, might suffice.

(c) MC and FC are running on dedicated hardware. This system design offers the possibility to encapsulate the MC and FC into a certified component and an uncertified component, respectively. It represents a sweet spot between flight performance and certification efforts. Certification efforts should be focused on MC where a trade-off in complexity is apparent. As a first step, simple and conservative buffers within the pre-defined area facilitate geo-fencing, i.e. preventing a break out. The certification of this approach is easier but reduces the surveillance performance, i.e. the buffers prevent the access to the complete area. The complexity can be increased from worst-case buffers to more advanced techniques which incorporate the continuously changing state of the UAS [9, 10]. Considering FC, dedicated, possibly uncertified, hardware can be used. Further, algorithms can be applied which could not be used so far. For instance, some of the most popular machine learning techniques, like deep neural networks, show promising results but are hardly human interpretable and therefore hard-to-certify.

Note that the discussed certification efforts can be addressed by the operation itself. By incorporating a sufficiently large safety buffer around the operation

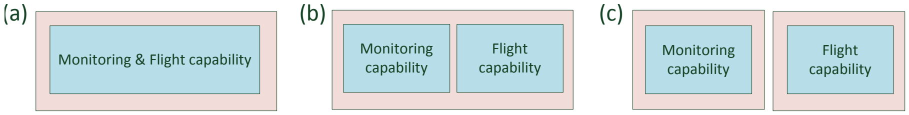

Fig. 2. Possible system design. (a) Shared hardware and combined capability (b) Shared hardware but separated capabilities (c) Dedicated hardware for each capability

which exceeds the physical limits of the aircraft, e.g. the lift-to-drag ratio combined with fuel consumption, even MC turns out to be redundant.

Crop surveillance is just one example which indicates that SORA enables the usage of previously inconceivable techniques by allowing to fit the system design of the UAS to its specific intended operation. In general and especially for more complex use cases, finding such a sweet spot between aircraft certification and operational limitations is challenging but in any case worth looking at. Once sufficient harm and threat barriers are established, it allows the usage of state of the art algorithms for high performance within the respective operation. An obvious use of AI for UAS is the detection of obstacles as well as conflicting traffic using computer vision. This use case is similar to the task that is currently performed by machine learning algorithms for autonomous cars. Furthermore, AI could be used to categorize the real time UAS performance as normal or abnormal. Machine learning algorithms could learn from simulations, test runs, and actual test flights during development as well as all operational in-service flights of all UAS of a certain type or fleet. The certified/uncertified point of view might be not applicable or too rigor in some cases. External safety frameworks enhancing AI, e.g. runtime safety monitoring, or research advances towards certification of AI in general, might soften the rigorous separation.

4 Future Perspective

Although the results of AI, in particular deep learning techniques, are very promising, the use of such techniques in safety critical areas is problematic. The aerospace domain imposes high requirements on the development and verification of software systems. The certification of AI is currently only possible with a thoroughly documented service history that can establish the necessary trust or a switching architecture relying on detecting misbehavior. It is unclear how existing coverage metrics, such as MC/DC, could be applied to neural networks to assure functional safety. More research efforts regarding the verification and validation aspects of these new AI techniques are therefore necessary.

However, the specific category approach introduced by EASA in combination with the safe monitoring of the operation could be an enabling technology for AI applications, even in a safety critical context. The specific category offers a pragmatic way to gather service history experience with uncertified components, exemplary depicted in Section 3. It is unlikely that a certification of these components is possible with traditional means of compliance in the near future. Also, in applications like visual obstacle detection, it is hard to identify a misbehavior, e.g. a human was falsely associated or not detected at all, therefore a traditional safety switch to a conservative alternative cannot be used. The specific category can uniquely support AI applications by monitoring the operation instead of the correct system-level functionality. The safety monitor ensures that there is no increased risk when using AI because a mitigation action, e.g. flight termination, can be triggered as soon as the operation is detected to be out of control. Similar to autonomous cars, where hundreds of test vehicles with human safety drivers are currently performing thousands of service miles, the service history for UAS could be supported by the specific category use cases. Of course, open questions remain, such as the comparability of the specific category use case, how much service hours are sufficient for safety, and how to ensure proper requirement coverage. Future work will also be necessary on ensuring the safety of the operation and possible mitigation strategies. Also, how AI techniques can be adjusted to a specific operation, e.g. training for operation specific inputs.

References

1. Dauer, J.C., Lorenz, S., Dittrich, J.S.: Automated low altitude air delivery. Deutscher Luftund Raumfahrtkongress DGLR (13-15 Sep 2016)
2. European Aviation Safety Agency (EASA): Introduction of a regulatory framework for the operation of drones. Advance Notice of Proposed Amendment 2017-05
3. Friehmelt, H. (ed.): Integrated UAV Technologies Demonstration in Controlled Airspace Using ATTAS, AIAA Atmospheric Flight Mechanics Conference and Exhibit (2003)
4. Guyon, I., von Luxburg, U., Bengio, S., Wallach, H.M., Fergus, R., Vishwanathan, S.V.N., Garnett, R. (eds.): Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems, USA (2017)
5. Hayhurst, K.J., Dorsey, C.A., Knight, J.C., Leveson, N.G., McCormick, G.F. (eds.): Streamlining Software Aspects of Certification: Report on the SSAC Survey (1999)
6. Hayhurst, K.J., Veerhusen, D.S. (eds.): A practical approach to modified condition/decision coverage (2001)
7. Joint Authorities for Rulemaking of Unmanned Systems: JARUS Guidelines on Specific Operations Risk Assessment (SORA). Draft for public consultation (2016)
8. Rein, W.: Autonomous drones: Set to fly, but may not comply; 5 major obstacles for unmanned aircraft systems (2018), https://www.wileyrein.com/newsroom-articles-Autonomous-Drones-Make-It-Easier-to-Fly-But-Harder-to-Comply.html; https://www.wileyrein.com/newsroom-articles-5-Major-Obstacles-For-Unmanned-Aircraft-Systems.html
9. Schirmer, S., Torens, C., Adolf, F.M. (eds.): Formal Monitoring of Risk-based Geofences, AIAA Information Systems-AIAA Infotech at Aerospace (2018)
10. Torens, C., Nikodem, F., Dittrich, J.S., Dauer, J.C. (eds.): Onboard Functional Requirements for Specific Category UAS and Safe Operation Monitoring, 6th CEAS air and space conference (2017)
11. Yuan, X., He, P., Zhu, Q., Bhat, R.R., Li, X.: Adversarial examples: Attacks and defenses for deep learning. CoRR (2017)