Thinking the Certification Process of Embedded ML-Based Aeronautical Components Using AIDGE, a French Open and Sovereign AI Platform

Filipo Studzinski Perotto[1], Anthony Fernandes Pires[1], Jean-Loup Farges[1], Youcef Bouchebaba[1], Mohammed Belcaid[2], Eric Bonnafous[2], Claire Pagetti[1], Frédéric Boniol[1], Xavier Pucel[1], Adrien Chan-Hon-tong[1], Stéphane Herbin[1], Mario Cassaro[1] and Sofiane Kraiem[1]

[1]*ONERA, The French Aerospace Lab, Toulouse, France*

[2]*CS-GROUP, Toulouse, France*

Keywords: Embedded Critical Systems, Machine Learning Certification, Aeronautical Software Components.

Abstract: AIDGE is a novel software development platform for embedded *Artificial Intelligence* (AI). It is designed to import or even learn *Deep Neural Networks* (DNN) and generate optimized code for target hardware architectures, in a completely open, transparent, and traceable manner. The purpose is to avoid dependence on opaque and non-sovereign tools or elements, ensuring competitive performance and favoring the certification of embedded *Machine Learning* (ML) components. In this paper, we present the preliminary analysis on the potential benefits of using this platform in light of the rising aeronautical certification standards concerning the use of ML into critical aeronautical systems, pointing possible steps toward certification, based on the artifacts that can be automatically generated by AIDGE.

1 INTRODUCTION

The aerospace industry is showing increasing interest in the possibility of embedding systems that integrate software functions obtained by machine learning (ML) methods, especially deep neural networks (DNN). However, aeronautical standards impose development constraints on software components, and these systems must be certified beforehand. The severity of the certification rules varies depending on the level of automation of each function and on its criticality: the more critical and/or autonomous a system is, the higher the level of assurance required.

The use of ML in the development of systems whose malfunction would contribute to a major failure with potential threat to user safety presents complex challenges in terms of certification, in particular due to the non-deterministic and data-intensive nature of the learning process, and due to the difficulty in explaining the inferences obtained from the resulting models. Another issue is the paradigm shift underlying machine learning, which is data-oriented, completely different from classical software development, thus requiring specific standards and adapted tools.

This work is supported by a French government grant managed by the National Research Agency under the France 2030 program with the reference ANR-23-DEGR-0001 – DeepGreen Project.

2 AI-DEDICATED PLATFORMS

Artificial Intelligence (AI) is at the heart of an important international economic competition. The stakes are high: AI is engendering a new industrial revolution and will be crucial for mastering many technological innovations. The USA is currently the world leader in AI thanks to the economic power of GAFAM (their giant IT companies), associated to prestigious research centers, supported by public incentive and promotion.

Today, the landscape of Deep Learning frameworks in the world is dominated by American products, notably TensorFlow, released by Google in 2015, and PyTorch, released by Facebook in 2018. The creation of AI technological components is then strongly dependent on the major orientations of these platforms. This software dependence also creates hardware dependence, since some processors are better supported, integrated and optimized, then benefiting American component suppliers, such as Nvidia, which is the dominant GPU-based processors manufacturer. Some authors highlight the risk of digital colonialism (Arora et al., 2023). For this reason, building a complete ecosystem to support the entire value chain around AI, from the algorithm to the component, is a strategic choice aimed at reducing French and European dependence on American enterprises.

Perotto, F., Pires, A., Farges, J., Bouchebaba, Y., Belcaid, M., Bonnafous, E., Pagetti, C., Boniol, F., Pucel, X., Chan-Hon-tong, A., Herbin, S., Cassaro, M. and Kraiem, S.
Thinking the Certification Process of Embedded ML-Based Aeronautical Components Using AIDGE, a French Open and Sovereign AI Platform.
DOI: 10.5220/0012965100004562

In *Proceedings of the 2nd International Conference on Cognitive Aircraft Systems (ICCAS 2024)*, pages 64-71
ISBN: 978-989-758-724-5

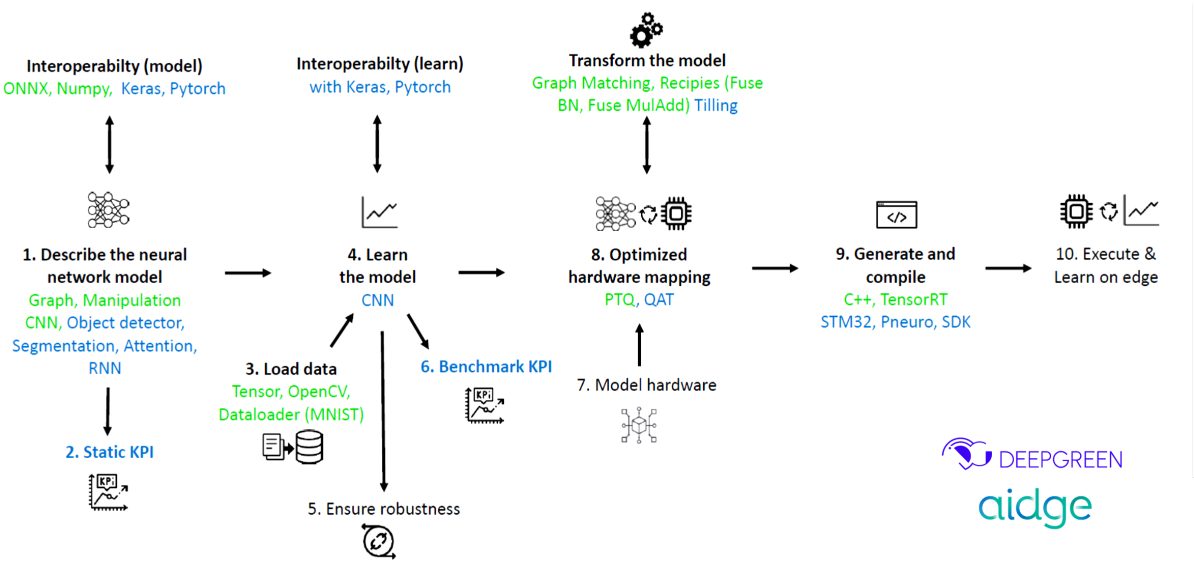

Figure 1: AIDGE workflow.

The mainstream ML research is focused on the cloud, where giant neural networks are trained through massive data processing and executed on powerful servers. However, effective industrial solutions require frugality to deploy standalone embedded AI systems, conceived for and transposed into small components. In addition, ML-based components need to be explainable and trustworthy to be used in critical systems and infrastructures.

The DeepGreen project[1] contributes to the community effort in searching for certifiable neural network based components. DeepGreen is funded by the French National Agency for Research (ANR), and is led by a consortium composed of around twenty major French research and industry actors in AI, embedded technology, and microelectronics. The project started last year and is scheduled to run until 2027.

The goal of DeepGreen is to develop an open-source software platform for embedded AI called AIDGE[2], with a first complete version expected to be available this year. The purpose of the project is to establish a European AI platform, open, sustainable, and sovereign, aimed at facilitating the deployment of deep neural networks on limited physical devices by taking into account the embedding needs and constraints that link models and hardware targets. The capacities of that platform (learning a DNN, importing a DNN, verifying the model, optimizing the model, generating source code, etc.), are illustrated in the Figure 1.

In addition, since AIDGE aims to be a certification-friendly platform, it should be able to automatically generate evidences supporting arguments for certification, and guide users on the necessary activities that must be fulfilled during the development cycle to meet the selected certification objectives.

[1]https://deepgreen.ai/en

[2]https://projects.eclipse.org/projects/technology.aidge

3 AERONAUTICAL CERTIFICATION

Critical systems, which can put lives at risk, are generally subject to a certification process. A competent authority evaluates the design conditions and the conformity of the product with regulatory requirements, before authorizing its exploitation. For a system to be certified, a set of documents constituting a structured argumentation must be presented in order to demonstrate that all the recommendations and standards have been considered and respected. Thus, certification activities aim to provide justifications explaining why the development of a specific product is trustworthy and safe, meeting the requirements of the pertinent standards. Such comprehensive documentation should contain not only the results, but also the input data, assumptions made, techniques applied, rationales for design decisions, etc.

In the aeronautics field, committees of experts draft standards in order to respond to regulatory laws. When these documents are recognized and validated as *Acceptable Means of Compliance* (AMC), the certification of a system can be achieved by demonstrating compliance with them, which are then formalized as *Aerospace Recommended Practices* (ARP). In Eu-

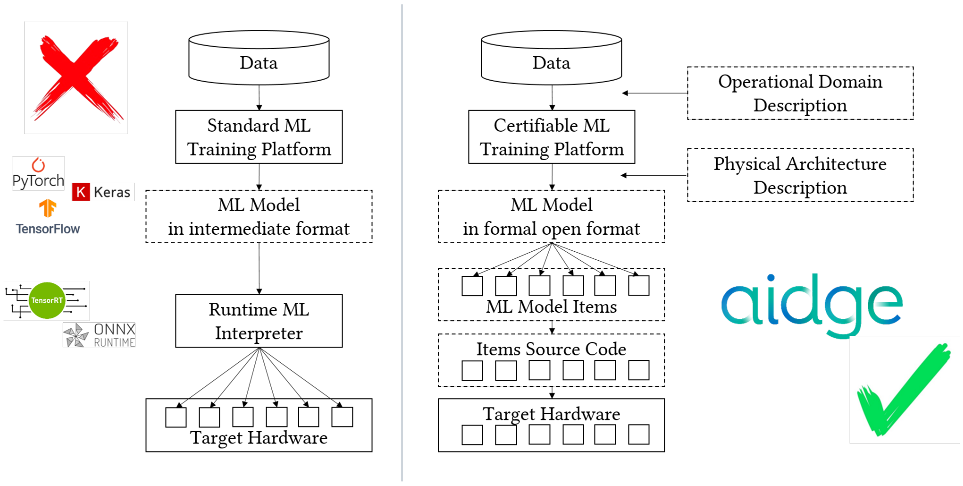

Figure 2: Difference between a standard ML platform and a certifiable one, in terms of hardware deployment (inspired from (Gauffriau and Pagetti, 2023)).

rope, the certification authority is EASA[3]. Although the certification process is well covered by current aeronautical practices in a majority of common areas and use cases, the approach to compliance is evolving for certain new and disruptive technologies. This is the case for solutions that use data-based machine learning for training artificial neural networks, since the development cycle of this type of software module is significantly different from the classic software engineering cycle, based on human programming activities, expert knowledge and well-defined mathematical models.

The standard and widely used AI-dedicated platforms, such as TensorFlow and PyTorch, even if available in open-source, do not meet certain basic certification requirements. The most restrictive issue is that deployment on target hardware is carried out by additional modules, either proprietary, or opaque in their operation (Figure 2). The AIDGE platform is intended to be a tool adapted to the construction of embedded and certifiable AI-based solutions. AIDGE, which is also open-source, adheres to principles such as transparency, traceability, and determinism in operations throughout the entire chain (from learning, to various types of optimization, and configuration for hardware deployment), enabling complete reproducibility. This is a minimum starting point for certifying ML-created software functions embedded in critical aeronautical systems, which requires a demonstration of reliability not only of the final product, but also of all stages of production. Concerning ML, this involves in particular:

(a) demonstrating the quality of the data sources, describing all pre- and post-processing operations;

(b) justifying the suitability for the operational domain to which the system must respond;

(c) demonstrating the robustness of a model learned automatically in the form of a neural network, explaining the choice of architecture and hyperparameters, being able to deterministically reproduce the entire learning cycle;

(d) demonstrating that any optimization applied to the model in order to facilitate its embedding does not degrade the required quality of the inferences, being able to reproduce any optimization operation;

(e) carrying out and documenting verification and validation operations, ensuring their reproducibility; and

(f) mastering code generation, compilation, and deployment on target hardware, being able to go back from the deployed system to the generated source code, and from that to the model (graph and parameters of the neural network), ensuring that the embedded component complies with all operational and performance constraints.

For a critical system or component to be certified, it must actually conform to the relevant standards (be compliant), and that conformity must be demonstrated (show compliance). The entire development process must verify certain qualities, and the actions put in place to ensure compliance must be clearly indicated. Security risks must be identified and actions to mitigate them must be indicated. Any decision regarding the choice of methods or alternatives during the development phase must be enumerated and justified. In other words, it is necessary to ensure the quality of the software developed, and the quality of the development process itself. In aeronautics, these

[3]European Union Aviation Safety Agency

Table 1: The 5 different development assurance levels (DAL), depending on the criticality of the system or function, following DO-178C (EUROCAE, 2011).

DAL	Failure Effect	Risk	Req. Tracing
A	**Catastrophic**: continuing the flight, takeoff or landing safely is impossible, crippling overload of work for the crew.	– Several fatalities, maybe airplane crash.	– traces to executable.
B	**Hazardous**: alarming reduction in safety margins or operational capacities, harmful overload of work for the crew.	– Serious injuries, maybe fatalities.	– traces to source code.
C	**Major**: significant reduction in safety margins or operational capacities, considerable increase in crew load.	– Discomfort to the occupants, maybe injuries.	– traces to source code.
D	**Minor**: small reduction in safety margins, light increase in crew load.	– Some inconvenience to passengers.	– verification by tests is enough.
E	**Insignificant**: no effect on aircraft operational capability or crew workload.	– No particular risks.	– no particular requirements.

Table 2: Six different levels of AI autonomy in aeronautics (EASA, 2023a).

Level	Description
1A	- Augmented perception assistance: AI can help human by analyzing or interpreting input signals.
1B	- Decision-making assistance: AI can suggest actions or evaluate preferable choices.
2A	- Human-AI Cooperation: AI performs specific tasks when asked for, under human supervision and authority.
2B	- Human-AI Collaboration: they communicate and work together with shared initiative under human authority.
3A	- Delegated AI Autonomy: AI in charge of whole task under human monitoring who can take control anytime.
3B	- Full AI Autonomy: AI in charge of whole task, without human intervention or supervision.

requirements are defined in the form of *Development Assurance Levels* (DAL), defined in DO-178C (EUROCAE, 2011) and ARP4754A (SAE, 2011), ranging a system from level A (the most critical) to E (the least critical), depending on the severity of consequences that a malfunction may cause (Table 1).

4 CONSTITUTION OF ML STANDARDS

The software-oriented standard DO-178C (EUROCAE, 2011) and the hardware-oriented standard DO-254 (EUROCAE, 2005) effectively regulate the development of on-board computer systems for avionics. However, DO-178C was not designed for software that comprises functions implemented by artificial neural networks trained using machine learning methods on a database. In fact, today, there are no standards dedicated to the use of machine learning in the context of onboard avionics systems, but several recent initiatives aim to establish them. Among those initiatives, there is the work carried out by EASA, which recently proposed a roadmap for AI as well as a classification of AI applications based on the level of autonomy of the machine in relation to human action (EASA, 2023a) (Table 2).

The joint initiative between EUROCAE[4] and SAE[5], called WG-114/G-34 working group, aims to define a standard dedicated to the development and certification of critical avionics products using ML. The result of this work, ARP6983 (SAE, 2025), still in progress, is expected to be published in the next year. This document must address five challenges (Gabreau et al., 2021):

1. Due to the “black box” nature of the data-driven training mechanism underlying ML, it is difficult to verify the link between high-level requirements and the learned model, as well as to validate data-based requirements. The standard must therefore indicate ways to open this black box in order to clearly specify, validate and verify ML requirements.
2. Since the ML model is essentially produced from data using automatic statistical methods, it is necessary to ensure their representativeness and their completeness in relation to the needs of the system, which can prove complex (bias during the collection data, lack of relevance to the problem, etc.), in coherence with the defined operational design domain (ODD).
3. The standard must guide the user in evaluating the model and learning robustness, assessing both how dependent the model is on the way it was trained and how stable and predictable the inference is in relation to variations in the input data.

[4]European Organization for Civil Aviation Equipment

[5]Society of Automotive Engineers

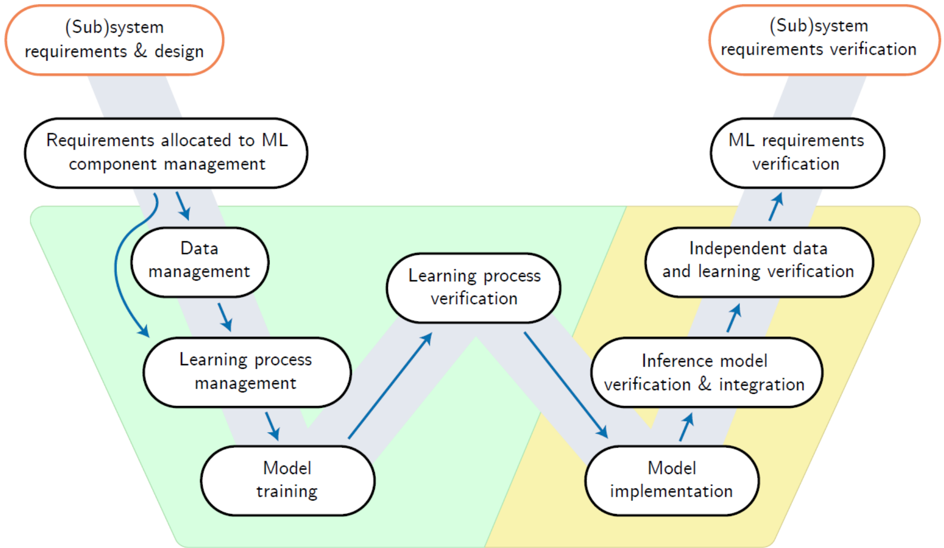

Figure 3: The learning phase (in green) and the inference phase (in yellow) in the W development process of an assured ML solution (EASA, 2020; EASA, 2021b).

4. The standard must assess the explainability and interpretability of the model, possibly identifying causality between the input dataset and machine learning outputs.
5. Risk mitigation must be assessed by defining the safe operating framework of the ML product in its ODD, as well as secure fallback procedures in case of deviation.

In fact, in order to analyze what kind of evidences (for feeding certification arguments) could be automatically, or semi-automatically generated by a certification-friendly AI-dedicated platform, like AIDGE, the first step is to enumerate a list of certification objectives, then to develop them in the form of argument patterns, identifying the kind of artifacts that can support compliance with each objective. Beyond the ARP6983, not yet published, other documents try to propose necessary certification objectives. The most complete are the *EASA Concept Papers – Guidance for Machine Learning Applications* (EASA, 2021a; EASA, 2023b) which propose a first set of objectives and means of compliance for the certification of critical ML-based systems.

The *DEEL White Paper – Machine Learning in Certified Systems* (Delseny et al., 2021) identifies properties that ML-based systems should present and which can have a positive impact on certification, like: auditability, data quality, explainability, maintainability, resilience, robustness, verification, and clear specification. The *Concepts of Design Assurance for Neural Networks* reports (EASA, 2020; EASA, 2021b), concerning the use of ML in critical avionics, identified a W-shaped development cycle for machine learning applications (Figure 3).

Other documents making the exercise of enumerating objectives or challenges, or suggesting practices for certifying ML components are (AVSI, 2020; Hawkins et al., 2021; LNE, 2021; Ashmore et al., 2021; Dmitriev et al., 2022; Gabreau et al., 2022; Gauffriau and Pagetti, 2023; MLEAP, 2024; SAE, 2021). A comparative study of the different existing reports on ML from different fields (aeronautics, automotive, etc.), notably carried out in the context of the certification objectives covered in each document and suggested means of compliance, is presented in (Kaakai et al., 2022).

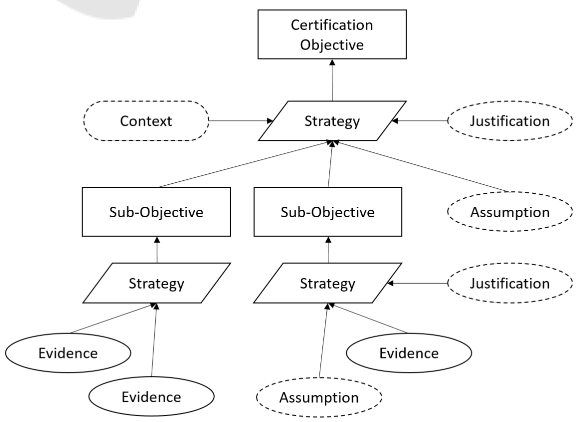

Figure 4: GSN-style graph to represent an assurance case for a given certification objective. Evidences are artifacts that, in some cases, can be generated with the help of the AI platform.

Once the complete set of certification objectives established, the next step is to analyze them in order to identify possible means of compliance. To do so, a common approach is to break down each certification objective in the form of argumentation patterns, also called "assurance case patterns" (Hawkins and Kelly, 2009; Hawkins and Kelly, 2010; Delmas et al., 2020), where an argument constitutes a structured reasoning based on assertions (goals) and facts (elements of justification). Such approach has been applied in aeronautical certification (Hawkins et al., 2013; Boniol et al., 2019; Boniol et al., 2020; Polacsek et al., 2018), and in particular for ML in (Gabreau et al., 2022; Hawkins et al., 2021; Damour et al., 2021; Grancey et al., 2022). An example of assurance case pattern using the GSN notation (Kelly and Weaver, 2004) is shown in the Figure 4.

5 PRELIMINARY RESULTS, DISCUSSION, CONCLUSION, AND PERSPECTIVES

The work within the DeepGreen project is still in its first steps. Nevertheless, some general characteristics have been identified to ensure that the proposed AI-dedicated platform (AIDGE) will respect the certification objectives and will be able to generate useful artifacts to fill the assurance cases:

1. **Determinism:** all the operations executed with the help of the platform (dataset split, learning, optimization, quantization, translation into intermediate model, decomposition into items, translation into source code, compilation, etc.) must be deterministic, i.e. all the necessary parameters and random seeds must be controllable and explicit, ensuring that repeating the operations will always lead to exactly the same results.
2. **Traceability:** all the subsequent models, generated by chained transformations, must have their parts identified in order to trace from where they come. For example, each neural network layer must be commented, annotated, and identified identically from the ML model to the generated source code.
3. **Reproducibility:** the complete process leading from the data to the implemented model must be reported, identifying the order, inputs and parameters used on each operation, allowing to reproduce all the activities of the development cycle obtaining the same results.
4. **Formalization:** to each manipulated model, object, and operation, a complete, explicit, non-ambiguous formal description language (syntax and semantics) and mathematical structure must be associated.

To be able to generate the artifacts necessary to certification, the entire process must be encapsulated in the form of a project, including the data used for training and verification, an explicit ODD description, and the description of the physical architecture into which the executable code will run (Figure 5). All those elements must be stored together, in a structured manner, to allow verification and reproducibility.

Some works in the literature propose formal structures for a rigorous ODD characterization in ML-based components (Kaakai et al., 2023; Adedjouma et al., 2024). Such kind of ODD representation can be used within AIDGE, allowing the user to enter ODD parameters, like value intervals, edge and corner cases, then analyze the learning, validation, and test datasets on the ODD, and produce artifacts to demonstrate their adequacy.

Similarly, some works in the literature propose formal structures to represent the important aspects of the hardware architecture, relative to the supported operations, numerical precision, number of cores, size of different levels of cache memories, bus speed, etc. In (Binder et al., 2022) a procedure for describing abstract processor models is presented, enabling the evaluation of predictable execution time and security assessments. Similar representation could be used within AIDGE to represent the target hardware architecture, with the possibility of assessing its compatibility with generated code, and even estimating worst case execution times (WCET).

In practice, building a high-quality ML system for a specific task relies on data-scientist expertise, demanding several iterations of trial and error experimentation to correctly fine-tuning the learning process and the model architecture. Automated Machine Learning (AutoML) is an active research topic in the field, proposing a set of techniques for automating the ML development pipeline, which includes data preparation, feature engineering, hyperparameter optimization, and neural architecture search (He et al., 2021). Several of those activities are similar to the ones demanded by certification. One of the possible ways to improve a certification-friendly AI-dedicate platform, like AIDGE, is by adapting certain of those AutoML techniques to automatically produce verification reports concerning, for example, the choice of neural architecture, by systematically testing performance and robustness of different configurations.

Finally, a certification-friendly AI-dedicated platform should allow to control and adapt the activi-

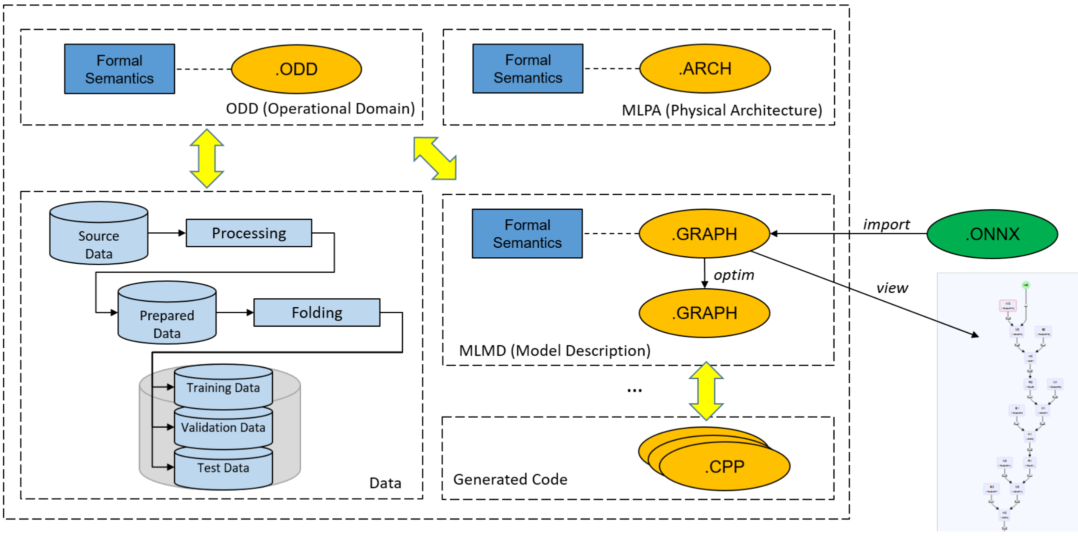

Figure 5: Draft of what a *project* `.aidge` could be, with the representation of the imported or learned neural network, in an open, formal and non-ambiguous format, associated with other elements necessary for verification, validation and certification of the network, such as the ODD, and the description of the target hardware architecture.

ties within the development cycle by maintaining the compliance state with the defined objectives, indicating to the user the remaining needed activities to generate necessary artifacts, in the same way as suggested by (Idmessaoud et al., 2024).

REFERENCES

Adedjouma, M., Botella, B., Ibanez-Guzman, J., Mantissa, K., Proum, C.-M., and Smaoui, A. (2024). Defining operational design domain for autonomous systems: A domain-agnostic and risk-based approach. In *SOSE 2024 - 19th Annual System of Systems Engineering Conference*, Tacoma, United States.

Arora, A., Barrett, M., Lee, E., Oborn, E., and Prince, K. (2023). Risk and the future of AI: algorithmic bias, data colonialism, and marginalization. *Information and Organization*, 33(3):100478.

Ashmore, R., Calinescu, R., and Paterson, C. (2021). Assuring the Machine Learning Lifecycle: desiderata, methods, and challenges. *ACM Comput. Surv.*, 54(5).

AVSI (2020). AFE 87: Machine Learning. Technical report, Aerospace Vehicle System Institute (AVSI).

Binder, B., Bensaid, S. A., Tollec, S., Thabet, F., Asavoae, M., and Jan, M. (2022). Formal Processor Modeling for Analyzing Safety and Security Properties. In *11th European Congress on Embedded Real Time Software and Systems, ERTS*.

Boniol, F., Bouchebaba, Y., Brunel, J., Delmas, K., Loquen, T., Gonzalez, A. M., Pagetti, C., Polacsek, T., and Sensfelder, N. (2020). PHYLOG certification methodology: a sane way to embed multi-core processors. In *10th European Congress on Embedded Real Time Software and Systems, ERTS*.

Boniol, F., Bouchebaba, Y., Brunel, J., Delmas, K., Pagetti, C., Polacsek, T., and Sensfelder, N. (2019). A service-based modelling approach to ease the certification of multi-core COTS processors. In *SAE AEROTECH Europe*.

Damour, M., De Grancey, F., Gabreau, C., Gauffriau, A., Ginestet, J.-B., Hervieu, A., Huraux, T., Pagetti, C., Ponsolle, L., and Clavière, A. (2021). Towards Certification of a Reduced Footprint ACAS-Xu System: a Hybrid ML-based Solution. In *Computer Safety, Reliability, and Security (40th SAFECOMP)*, pages 34–48. Springer.

Delmas, K., Pagetti, C., and Polacsek, T. (2020). Patterns for Certification Standards. In *32nd International Conference on Advanced Information Systems Engineering, CAiSE*, LNCS 12127, pages 417 – 432, Grenoble, France.

Delseny, H., Gabreau, C., Gauffriau, A., Beaudouin, B., Ponsolle, L., Alecu, L., Bonnin, H., Beltran, B., Duchel, D., Ginestet, J.-B., Hervieu, A., Martinez, G., Pasquet, S., Delmas, K., Pagetti, C., Gabriel, J.-M., Chapdelaine, C., Picard, S., Damour, M., Cappi, C., Gardès, L., Grancey, F. D., Jenn, E., Lefevre, B., Flandin, G., Gerchinovitz, S., Mamalet, F., and Albore, A. (2021). White Paper: Machine Learning in Certified Systems. DEEL Certification Workgroup / IRT Saint Exupéry / ANITI.

Dmitriev, K., Schumann, J., and Holzapfel, F. (2022). Towards Design Assurance Level C for Machine-Learning Airborne Applications. In *41st IEEE/AIAA Digital Avionics Systems Conference, DASC*, pages 1–6.

EASA (2020). Concepts of Design Assurance for Neural Networks (CoDANN) I. EASA and Daedalean.

EASA (2021a). Concept Paper: First Usable Guidance for Level 1 Machine Learning Applications, n.1. European Aviation Safety Agency (EASA), Cologne.

EASA (2021b). Concepts of Design Assurance for Neural Networks (CoDANN) II. EASA and Daedalean.

EASA (2023a). Artificial Intelligence Roadmap 2.0 – Human-centric approach to AI in aviation. European Aviation Safety Agency (EASA), Cologne.

EASA (2023b). Concept Paper: First Usable Guidance for Level 1&2 Machine Learning Applications, n.2. European Aviation Safety Agency (EASA), Cologne.

EUROCAE (2005). DO-254 / ED-80 Design Assurance Guidance for Airborne Electronic Hardware. RTCA, Inc / EUROCAE.

EUROCAE (2011). DO-178 / ED-12C – Software Considerations in Airborne Systems and Equipment Certification. RTCA, Inc / EUROCAE.

Gabreau, C., Gauffriau, A., Grancey, F. D., Ginestet, J.-B., and Pagetti, C. (2022). Toward the certification of safety-related systems using ML techniques: the ACAS-Xu experience. In *11th European Congress on Embedded Real Time Software and Systems, ERTS*.

Gabreau, C., Pesquet-Popescu, B., Kaakai, F., and Lefevre, B. (2021). AI for Future Skies: On-going standardisation activities to build the next certification/approval framework for airborne and ground aeronautical products. In *proceedings of the Workshop on Artificial Intelligence Safety (AISafety)*.

Gauffriau, A. and Pagetti, C. (2023). Formal description of ml models for unambiguous implementation.

Grancey, F. D., Ducoffe, M., Gabreau, C., Gauffriau, A., Ginestet, J.-B., Hervieu, A., Huraux, T., Pagetti, C., Clavière, A., and Damour, M. (2022). Optimizing the design of a safe ML-based system - the ACAS Xu experience. In *11th European Congress on Embedded Real Time Software and Systems, ERTS*.

Hawkins, R., Habli, I., Kelly, T., and McDermid, J. (2013). Assurance cases and prescriptive software safety certification: a comparative study. *Safety Science*, 59:55–71.

Hawkins, R. and Kelly, T. (2009). A systematic approach for developing software safety arguments. In *Proceedings of the 27th International Systems Safety Conference*.

Hawkins, R. and Kelly, T. (2010). A structured approach to selecting and justifying software safety evidence. In *Proceedings of the 5th IET System Safety Conference*.

Hawkins, R., Paterson, C., Picardi, C., Jia, Y., Calinescu, R., and Habli, I. (2021). Guidance on the Assurance of Machine Learning in Autonomous Systems (AMLAS).

He, X., Zhao, K., and Chu, X. (2021). Automl: A survey of the state-of-the-art. *Knowledge-Based Systems*, 212:106622.

Idmessaoud, Y., Farges, J.-L., Jenn, E., Mussot, V., Fernandes Pires, A., Chenevier, F., and Conejo Laguna, R. (2024). Uncertainty in Assurance Case Template for Machine Learning. In *Embedded Real Time Systems (ERTS)*, Toulouse, France.

Kaakai, F., Adibhatla, S. S., Pai, G., and Escorihuela, E. (2023). Data-centric operational design domain characterization for machine learning-based aeronautical products. In Guiochet, J., Tonetta, S., and Bitsch, F., editors, *Computer Safety, Reliability, and Security*, pages 227–242, Cham. Springer Nature Switzerland.

Kaakai, F., Dmitriev, K., Adibhatla, S., Baskaya, E., Bezzecchi, E., Bharadwaj, R., Brown, B., Gentile, G., Gingins, C., Grihon, S., and Travers, C. (2022). Toward a machine learning development lifecycle for product certification and approval in aviation. *SAE International Journal of Aerospace*, 15(2):127–143.

Kelly, T. and Weaver, R. (2004). The goal structuring notation – a safety argument notation. In *Proceedings of the dependable systems and networks 2004 workshop on assurance cases*, volume 6. Citeseer Princeton, NJ.

LNE (2021). Certification Standard of Processes for AI. Technical report, Laboratoire National de métrologie et d'essais (LNE).

MLEAP (2024). EASA Research -– Machine Learning Application Approval (MLEAP) – final report. European Union Aviation Safety Agency (EASA).

Polacsek, T., Sharma, S., Cuiller, C., and Tuloup, V. (2018). The need of diagrams based on toulmin schema application: an aeronautical case study. *EURO Journal on Decision Processes*, 6(3):257–282.

SAE (2011). ARP4754A / ED-79A – Guidelines for development of civil aircraft and systems-enhancements, novelties and key topics. SAE / EUROCAE.

SAE (2021). EUROCAE WG114 / SAE G34 Artificial Intelligence in Aviation – AIR6988 / ER-022 – Artificial Intelligence in Aeronautical Systems: Statement of Concerns. SAE / EUROCAE.

SAE (2025). ARP6983 / ED-324 – Process Standard for Development and Certification/Approval of Aeronautical Safety-Related Products Implementing AI (to appear). SAE / EUROCAE.