Proceeding Paper

Evaluation of an AI-aided Document Framework for Certification of Novel Aircraft Propulsion Systems †

Sebastian Stoppa [1,*], Durga Sri Sharan Katabathula [2], and Robin Frank [3]

1 sebastian.stoppa@dlr.de, 4
2 durga.katabathula@dlr.de, 4
3 robin.frank@dlr.de, 4
4 German Aerospace Center (DLR), Institute of Electrified Aero Engines, 03046 Cottbus, Germany;
* Correspondence: sebastian.stoppa@dlr.de; Tel.: +49-355-28888-066
† Presented at the 15th EASN International Conference, in Madrid, Spain, from 14-17 of October 2025.

Abstract: In contrast to the development of conventional civil aircraft propulsion systems, novel propulsion technology or non-standard energy sources disclose a lack of flexibility in the current aviation certification framework. Additionally, aircraft certification in 2025 relies heavily on manual effort, causing major difficulties in maintaining the traceability of data across vast sets of regulation documents. One promising solution to overcome such challenges offers the field of Artificial Intelligence (AI), particularly Large Language Models (LLMs). This paper introduces an AI-aided framework designed to streamline the certifiability of novel aircraft propulsion systems. To meet the AI trustworthiness demands set by the European Union Aviation Safety Agency (EASA), the framework proposes a concept for achieving data and model transparency in Machine Learning (ML) applications. To address the demand for data transparency, aviation regulatory context data will be unified and stored in a Unified Regulations Database (URD). This unified data is classified and enriched with related information for ML purposes. The URD enables the creation of modern, transparent AI features for civil aircraft certification. This AI-aided framework will enable certification measures for the development and allows for certifiability checks for novel aircraft technologies. Both, the aviation industry and regulatory authorities may equally benefit from the existence of the URD as starting point for certification AI features.

Keywords: Aircraft Certification, Aircraft Development, Artificial Intelligence (AI), Large Language Model (LLM), Transparency, AI Trustworthiness, Aviation

Published:

Citation: Stoppa, S.; Katabathula, DSS.; Frank, R. Evaluation of an AI-aided Document Framework for Certification of Novel Aircraft Propulsion Systems. *Eng. Proc.* **2025**, *1*, 0. https://doi.org/

1. Introduction

1.1. European Regulatory Framework

Aircraft certification is a fundamental necessity for ensuring a high and uniform level of safety within civil aviation. The *European Commission (EC)*, as executive body of the European Union (EU), establishes the regulatory framework for the European Aerospace, which defines certification as the primary method for verifying compliance (BR-2018-1139 [1]). Compliance means consideration of comprehensive documents from *European Union Aviation Safety Agency (EASA)*, *European Organization for Civil Aviation Equipment (EUROCAE)*, *SAE International*, DIN EN, IEEE, among others [2–8]. Table 1 shows a list of common institutions for aviation. Next to the institutions their typical established document types are mentioned. This list makes no claim to completeness - it provides insights about the complexity of the regulatory framework for aviation. Aside to these standard documents the aviation industry has developed and uses their own standards e.g. standard practices for testing of airborne equipment.

Institution	Typical Document Types
European Commission (EC)	Basic Regulations (BR), Legislation Information, Harmonization Rule Proposals, Annex Part 21
European Aviation Safety Agency (EASA)	Certification Specifications (CS), Acceptable Means of Compliances (AMC), Special Conditions (SC), Part 21, Guidelines, Easy Access Rules, Roadmaps
European Organisation for Civil Aviation Equipment (EUROCAE)	EUROCAE Documents (ED), Technical Reviews (ER), Standard Procedures, Guidelines and Guidance Material for Development, Safety, Processes, Testing, Operation, ...
Federal Aviation Administration (FAA)	Federal Aviation Regulations (FAR), Advisory Circular (AC), Airworthiness Directives (AD), Temporary Flight Restrictions (TFR), Notices to Airmen (NOTAM)
Institute of Electrical and Electronics Engineers (IEEE)	IEEE Technology Standards (e.g. 802.11, 802.3), Methodologies, System Specifications and Requirements, Best Practices
International Civil Aviation Organization (ICAO)	Standards and Recommended Practices (SARP), Procedures for Air Navigation Services (PANS), Manuals (Doc), Annexes, Safety Reports, Service Descriptions
International Organization for Standardization (ISO)	International Standards (ISO), European Standards (EN), German Standards (DIN), International Electrotechnical Commission (IEC) Standards, Processes, Guidelines
Radio Technical Commission for Aeronautics (RTCA)	Standards & Means of Compliances (DO), Safety & Performance Requirements (SPR), Interoperability Requirements (IRR), Minimum Aviation Performance Standards (MASPS & MOPS), Operational Services and Environment Definition (OSED)
SAE International	Aerospace Information Report (AIR), Aerospace Recommended Practice (ARP), AESQ Reference Manuals (RM), Technical Standards Committee (TSC) Docs, Compliance Test Specification

Table 1. List of common aviation standardization institutions as representative part of the aviation regulatory framework including their established document types

These regulatory documents provide standardized guidelines, procedures and methodologies e.g. for aircraft development, component & equipment design, installation, reliability, testing, maintenance concepts, crew training and level based safety assessments. Once a development has reached a sufficient level of maturity (Technology Readiness Level (TRL)) the Original Equipment Manufacturer (OEM) submits the project to the National Aviation Authority (NAA) and the applicable rules for certifying are set out in a binding document, the *Certification Basis (CB)* [9].

1.2. Certification Integrated Development Approach

A recommended approach is to consider certification holistically right from the beginning of any serious aircraft development. Siemens depicted this as controversy to traditional development in a software promotional video and outlines the importance of this approach [10].

Often promised solutions, provided by global players such as IBM and Siemens, are mainly collaborative life-cycle management (CLM) tools. These tools employ a systems engineering approach and enable an end-to-end traceability from initial specifications to the verified, tested implementation. But the relationship between the aviation certification framework and the complex system development is still missing; it is not covered by use of CLM tools entirely. Certification conformity for Aviation regulation still has so far been left to experts and innovative concepts currently lead to high development cost and unforeseeable impact on certify ability.

1.3. Aviation Industry Demands

During the *EASA-FAA International Aviation Safety Conference 2025* several discussions about certify ability of hybrid electric propulsion arose. Matthieu Pettes-Duler, Head of

Powertrain at Beyond Aero, described many discussions with EASA regarding Beyond Aero's new concept development of a novel hybrid-electric jet. He describes a ...

> *"need for flexibility in the certification process, recognizing the unique challenges posed by emerging technologies"* [1]

Another speaker, Antonio Tripoli from ATR, was talking about the non-existing certification base for the implementation of a hybrid electric propulsion system in large aeroplane (CS-25) [11]:

> *"(...) if the new rules are not built with sufficient flexibility on one side and sufficient clarity on the other side, we risk to not work, we risk to work with no efficiency, and so we can also shift our planning for development, increase the cost, and ultimately lead to the cancellation of the stock of the product."* [2]

The aeronautical industry is calling for reforms to its regulatory framework to better accommodate novel technologies. There is a clear need for *improved regulatory guidance*, for higher *transparency* in the regulatory framework and for an increased *flexibility* in certification processes.

2. Artificial Intelligence (AI) in Aviation

2.1. AI integration guidelines for Aviation

Among others, EASA and EUROCAE are actively working to address the challenges and opportunities of integrating AI into aeronautical systems. They mainly focus on developing AI integration guidelines for operational, in-service **AI trustworthiness** applications. The *EASA AI Roadmap 2.0* [12, p.15ff] follows a "human-centric" approach, emphasizing that AI should augment, not replace, human expertise and decision-making, see figure 1.

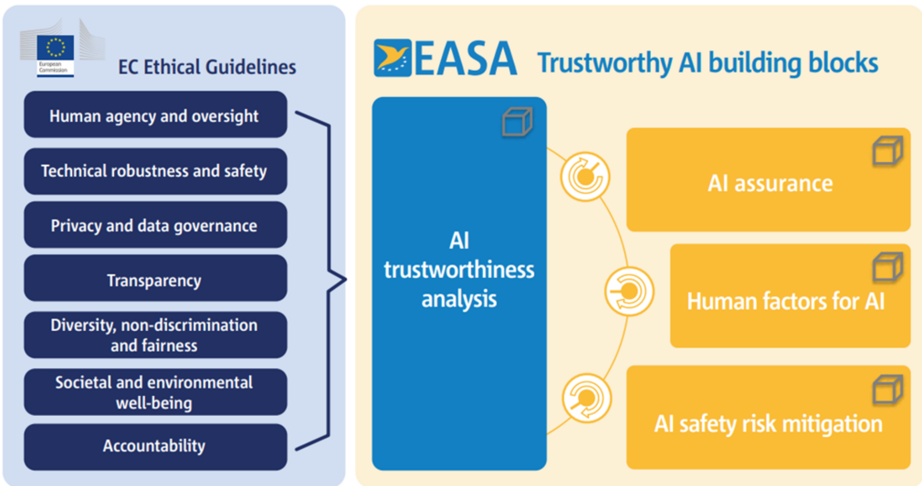

Figure 1. EC Ethical Guidelines for achieving AI trustworthiness in aviation [12, p.15ff]

Guidelines for using AI in certification-compliant development are not mentioned yet, whereas the implementation of AI features aligned with the aviation regulatory framework allows for a certification compliant development. This fundamental framework is accessible for aviation industry. Its published documents from institutions and agencies of the aeronautical sector are well known and accepted in aviation.

[1] Pettes-Duler, BEYOND AERO Head of Powertrain, Panel Discussion "Integration of New Technologies", 2025 EASA-FAA International Aviation Safety Conference

[2] Tripoli, ATR Preliminary Design and R&T Expert, Workshop Discussion "Electric/hybrid propulsion", 2025, EASA-FAA International Aviation Safety Conference

Its high accuracy provides a ground truth because of reviewed, proved, accepted and published documents within the aviation regulatory landscape from aviation experts. This fundamental ground truth can be used for training of AI models. Its accessibility ensures the transparency of the data - one of the key demands by EASA for trustworthy AI.

2.2. AI as Game Changer for Aviation

Table 2 outlines potential use cases for LLM-, Regression- or Classification Applications in aviation, offering benefits to both the industry and regulatory certification authorities (see note information *A* and *B*).

Model	Description
Certify Ability Checker [A,B]	regulatory tracing based on provided system description
Certification Gap Analyzer [B]	model check on missing or incomplete dev. data
Requirements Analyzer [A,B]	requirements quality check model
System Safety Analyzer [B]	partial safety analysis checks at system level
Dev Monitoring Models [B]	simultaneously textual and numerical data analysis
Referencing Models [A,B]	semantically similar endpoint linking across documents
Class Transposer [A,B]	transposition of e.g. one spec into multiple requirements
Certification Chat Bots (hosted) [A,B]	policy-based, provision of exact regulatory knowledge

[A] useful for Certification Authority [B] useful for OEM

Table 2. Examples of potential AI-Aided tools for aircraft certification and development

Such powerful models can provide early awareness of certification demands, thereby reducing late-stage development cycles and the need for extensive discussions between certification agencies and OEMs. While several approaches can be used to create these models, only a few are relevant to aviation.

3. Transparency first

To meet EASA's demands for AI trustworthiness and transparency, the following three major steps must be addressed individually: Data Engineering, Machine Learning Method Selection and Model Verification, see figure 2. AI-tools, built up on this particular transparency concept can provide clarity regarding their accuracy and response behavior.

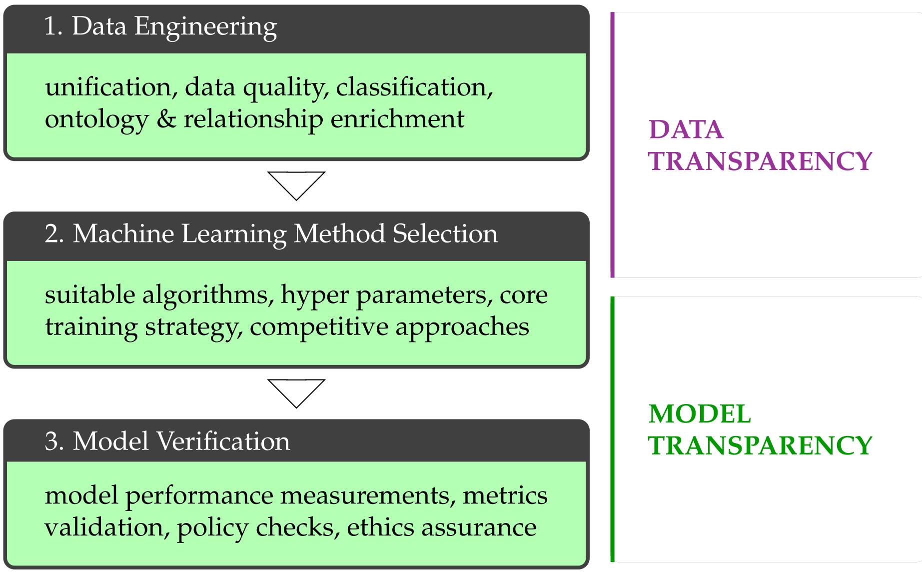

Figure 2. Basic Concept of Data- and Model-Transparency to achieve AI trustworthiness

The demand for transparency of AI models is one of the seven ethical guidelines for trustworthy AI set by the European Commission and published by EASA in the AI Roadmap 2.0 [12]. A precise understanding of, and reproducible probabilistic behaviour based on, the model's outputs are key to ensuring trust in AI solutions. However, their complex and opaque architectures can make it challenging to discern how they arrive at a particular output or exhibit uncertain behaviour [13, sec. 2.2].

Currently, only a handful of well-known aviation LLMs are available online to both industry and regulatory institutions: *Aviation-GPT*, *Aviator-GPT*, *Aviation Regulatory Advisor*, and *US General Aviation Expert*, among others. These models are website-embedded chatbots that have been fine-tuned from GPT on a wide range of aviation resources—including regulatory specifications, maintenance manuals, airplane flying handbooks (AFH), aeronautical information manuals (AIM), aviation weather publications, flight protocols, and more. The fine-tuned LLMs perform well and are often accurate on aeronautical queries. However, unlike cutting-edge GPT models, there are no dedicated, open-source aviation evaluation metrics. Consequently, their performance is assessed via general-purpose benchmarks such as Massive Multitask Language Understanding (MMLU), Graduate-Level Google-Proof Q&A (GPQA), and Humanity's Last Exam (HLE) [14,15], among others.

Until there are no effective evaluation metrics and rankings for aviation LLMs existent, these models are unlikely to be seriously integrated into aircraft development, certification, or aircraft equipment.

3.1. Unified Regulations Database (URD)

To address the data transparency in the Data Engineering process (cf. figure 2) a **Unified Regulations Database (URD)** is being created. Within this database all relevant document content from the Aviation field is captured in a single, consistent data structure underlying a simplified referencing and filing system. This approach ensures that information from every document is weighted equally. It enables a seamless data processing with further usage. The URD will provide transparency of the aviation regulatory framework for both, aviation industry and standardization authorities equally and in a unified format. One of the challenges of data *unification* is to overcome the diverse styles and complex document structures of the aviation regulatory document framework. All documents have varying metadata structures and content bodies, based on their institution's terminology and the document's character (presentation printout, regulation, specification, glossary, report, ...).

Therefore, the **File Content Transformer (FCT)** is used for correct data extraction and creation of the initial URD (I) based on the unified extracted data points. Unlike standard text-embedding approaches that rely on fixed chunk sizes, the FCT dynamically extracts each entry in a unique manner. An entry's size can vary and might include a sentence, a text block, a figure, or a table. It processes each certification file and parses its information into a JSON database according to the specification, see Table 3.

URD target data structure	more precise content
Section content	Text (sentence, text block) ∣ Image ∣ Table
Section ID	Uniquely generated ID
Chapter Info	Chapter No. ∣ Chapter Name
References	Ref. content to page1 ∣ Ref. content to file2, page2
Document Info	Date ∣ Name ∣ Path
Page Info	Number ∣ Footer ∣ Header
Human Feedback Loop	HFID ∣ HFIN (Dataset ∣ Issue Type ∣ Feedback)

Table 3. Example of a URD target data structure ∣ A Major Specification Request for the FCT for correct content transformation

The development of the FCT has progressed through several key stages. The selection of `pymupdf` as primary PDF parser was made after early attempts at heading detection using font heuristics and regex failed due to varied document formatting. `DocLing`, an open-source framework for document parsing and information extraction, is implemented next to layout parser and screenshot featuring models. The FCT's internal logic captures hyperlinks and textual references correctly - existing information from the documents does not get lost through the parsing process. Table 3 shows one example of typical attributes, to be transformed into the URD data structure format. For considering the Human Feedback, the *Human Verification Identifier (HFID)* and the *Human Feedback Information (HFIN)* attributes have been foreseen in the URD data structure as well. As database format a dedicated JSON structure has been finalized to store and refer extracted elements like images, tables and metadata.

The readily available, documented information, such as the data in the URD, can be used to extract essential content, that is being searched for. By contrast, information that is not explicitly recorded but exists as ontological context must be connected to the search result along with its relation description.

To bridge this gap, the **Data Enrichment Model (DEM)** augments URD entries with related topics and implicit keywords, functioning similarly to the assignment of tags and labels. But unlike simple tagging, the DEM provides structured transparency by differentiating between database topics and database keywords. While explicit and implicit keywords may be combined, topics must be consistently identified and harmonized across the entire database. A **Classifier Model (CM)** is responsible for categorizing all database entries. It assigns a tag from a predefined list to any entry that wasn't already classified during the initial data capturing process. Table 4 shows the initial list of classifiers. Applying defined classes to each data point within the database enables precise data extraction for various purposes, such as input filtering for specific purposes or subsequent training- and verification use cases.

abbreviation	**item type**	**data origin**[1]
SPC	Specification	p, c
MOC	Means of Compliance Information	c
GUI	Guideline / Guidance Information	c, g
GLT	Glossary Term	p, c, g
LL	Lessons Learned Information	p, g
REQ	Requirement	p, c, g

[1] p=project | c=certification | g=general document

Table 4. Exemplary Types classification for Aviation Regulatory Document's content

After the enriching, labeling and classification of data it is necessary to make the data base valid for approved usage. But human verification of such a vast amount of information, like the aviation regulatory framework, requires an extremely high effort. To compensate for this effort, certain strategies will be developed and evaluated for the Human Feedback Loop (HFL) of the URD. Within the URD, two additional attributes will be used to capture the human feedback: the *Human Verification Identifier (HFID)* and the *Human Feedback Information (HFIN)*. For each database entry, the HFID indicates its verification status and the necessity of using the captured feedback information. The HFIN attribute captures the human feedback content including the name of involved person. Both attributes come with a dynamic nature and will be frequently updated at certain verification steps of the URD data management, increasing the traceability and verification state of the database.

3.2. Explicit and Implicit Knowledge, Ontologies

In sophistical terms, direct accessible data is known as **explicit knowledge** - such as data source, meta and content information. Adjacent data and indirect information, belonging to the data, is known as **implicit knowledge**. This knowledge enables the data enrichment process, e.g. through classification and labeling with tags. Labeling involves adding relevant keywords and topics to specific data entries, while classifying assigns elements from predefined lists (supervised approach). Both processes *turn implicit information into direct accessible explicit data* and increase the transparency of the data set. A common term this within this context is *ontology*. Ontologies encompass the entirety of word families with context based relationships. Ontology visualizers allow for data relationships representation and information cluster printings in multi-dimensional diagrams: *ontology graphs*. Such ontology representations might be helpful for elaborating the data source of a ML model, that has been created on top of the URD. The development steps for the creation of the URD from the Aviation Regulatory Content as well as the allocation of explicit and implicit knowledge within the context, are depicted in figure 3.

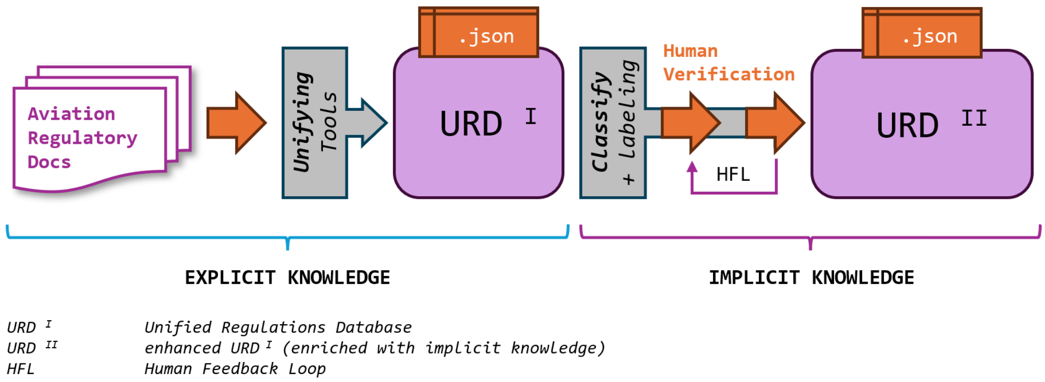

Figure 3. Schematic of the simplified URD creation process

4. Conclusion, Further Development Plans

Transparent AI features can support the aviation industry and the regulatory authorities equally at the complex development of future aircraft systems such as hybrid-electric propulsion technologies. To that end, the Unified Regulations Database (URD) is built as a single source of truth derived from the aviation regulatory framework. The URD is created within the Data Engineering step - one of the three identified stages for archiving EASA's AI-trustworthiness: Data Engineering, Machine Learning Utilization and Model Verification.

The first URD entries from EASA, EUROCAE and SAE documents have been created by the FCT tool. Work continues to ingest additional regulatory documents content automatically. Parallel prototyping is testing random, point-wise human verification of URD entries. Specifications for the Classification Model (CM) and Data Enrichment Model (DEM) are under construction. A key element is the Human Feedback Loop (HFL), enabled by new Human-Machine Interface (HMI) tools and graphical user interfaces. Intuitive interfaces accelerate human review, help converge on URD ground truth, and reinforce data and model transparency. Together, this yields a ML-ready framework that satisfies EASA's demand for AI trustworthiness.

Author Contributions: Conceptualization, S.S.; Methodology, S.S., D.K. and R.F.; Software, D.K.; Validation, R.F.; Writing—original draft and editing, S.S.; Writing—reviews, R.F. and D.K.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

LLM	Large Language Model
AI	Artificial Intelligence
EC	European Commission
EASA	European Union Aviation Safety Agency
EUROCAE	European Organization for Civil Aviation Equipment
TRL	Technology Readiness Level
OEM	Original Equipment Manufacturer
NAA	National Aviation Authority
CB	Certification Basis
CLM	Collaborative Life-cycle Management
GUI	Graphical User Interface
HFL	Human Feedback Loop
HMI	Human-Machine Interface
MMLU	Massive Multitask Language Understanding
GPQA	Graduate-Level Google-Proof Q&A
HLE	Humanity's Last Exam
RLHF	Reinforcement Learning from Human Feedback
URD	Unified Regulations Database
ML	Machine Learning
FCT	File Content Transformer
CM	Classifier Model
SPC	Specification Item
MOC	Means of Compliance Information
GUI	Guideline / Guidance Information
GLT	Glossary Term
LL	Lessons Learned Information
REQ	Requirement
DEM	Data Enrichment Model
HFID	Human Verification Identifier
HFIN	Human Feedback Information
RAG	Retrieval-Augmented Generation

References

1. EUROPEAN PARLIAMENT AND OF THE COUNCIL. BASIC REGULATION (EU) 2018-1139 - on common rules in the field of civil aviation. *Official Journal of the European Union* **4 July 2018**.
2. EASA. Certification Specifications and Acceptable Means of Compliance for Normal Category Aeroplanes (CS-23), 2023.
3. EASA. Certification Specifications and Acceptable Means of Compliance for Large Aeroplanes (CS-25), 2023.
4. EASA. Certification Specifications and Acceptable Means of Compliance for Engines (CS-E), 2023.
5. EASA. European Technical Standard Orders (CS-ETSO), 2022.
6. EASA. Electric / Hybrid Propulsion System (SC E-19), 2021.
7. SAE International. Guidelines for Conducting the Safety Assessment Processon Civil Aircraft, Systems, and Equipment, 2023.
8. SAE International. Guidelines for Development of Civil Aircraft and Systems, 2023.
9. Hinsch, M. *Industrielles Luftfahrtmanagement*; Springer Berlin Heidelberg: Berlin, Heidelberg, 2022. https://doi.org/10.1007/978-3-662-66452-0.
10. Siemens. "Shift Aircraft Certification left", 2024.
11. Tripoli, A. (ATR). Day 2 Presentation - Integrating hybrid propulsion system on large transport airplane. *EASA-FAA International Aviation Safety Conference* **2025**.
12. EASA. EASA Artificial Intelligence Roadmap 2.0, 2023.
13. Liao, Q.V.; Vaughan, J.W. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap.
14. Artificial Analysis, I. Artificial Analysis — artificialanalysis.ai. https://artificialanalysis.ai/evaluations. [Accessed 22-09-2025].
15. Long Phan, Alice Gatti, Z.H.N.L. Humanity's Last Exam.