A Specialized LLM Framework for Aviation Regulation and Compliance

Bianca Ebrasu
Faculty of Engineering in Foreign Languages
University Politehnica of Bucharest
Bucharest, Romania
bianca.ebrasu@gmail.com

Avram Brîndușa
Faculty of Engineering in Foreign Languages
University Politehnica of Bucharest
Bucharest, Romania
trufan_brindusa@yahoo.com

***Abstract*— Advances in Artificial Intelligence are transforming the management of complex regulatory frameworks across industries. In aviation, where safety and compliance are vital, timely and reliable access to regulatory information is essential. This study presents the development of a specialized Large Language Model system meant to improve compliance management activities and decision-making. It combines fine-tuning with Retrieval-Augmented Generation, with Regulation (EU) 2017/373 serving as the primary reference corpus. The system is designed to interpret and analyze aviation regulations, process the specialized language of the field, retrieve relevant information, and provide support in line with applicable regulatory requirements.**

***Keywords*— large language models, Retrieval-Augmented Generation, aviation regulation, compliance, artificial intelligence, natural language processing**

I. Introduction

In recent years, the progress in Artificial Intelligence (AI) technology saw a transformation across numerous domains, from healthcare and education to finance and urban planning, advancement which redefines the boundaries of what machines are capable of. But this revolution is not only a modern phenomenon and concept, but is deeply engraved in the history of the human thought itself, dating back to Ancient Greece where philosophers such as Aristotle began to ponder the potential for machines to mimic human behavior. The focus of the study is the exploration of the use of Artificial Intelligence in the form of a large language model applied to one of the most vital and important sectors today, aviation. In this domain, regulatory management is essential, as it serves as the solid foundation of operational safety and overall compliance. The management of regulations impacts a broad range of stakeholders, from airlines, airport operators to air navigation service providers and meteorological services, therefore the effectiveness of understanding and implementing of aviation regulations is not only a matter of legal compliance, but also of safety. Along with the evolution of technology and air traffic, the need for a system that would accommodate the continuous changes in compliance and oversight in aviation domains becomes apparent. By aligning regulatory practices with technological advancements, especially in Artificial Intelligence, a new bridge is forming which closes the gap between aviation law and its practicability and management. As the aviation sector strives for higher safety standards, efficiency, and customer satisfaction, AI offers indispensable potential to achieve these objectives.

While regulatory text mining and retrieval tasks have been explored in the legal domain, their application in aviation compliance is still nascent. Regulatory oversight systems such as ICAO's Global Aviation Safety Plan (GASP, Doc 10004) [1] emphasize proactive safety management, but tools for parsing the legal language of regulations remain underdeveloped. Creating a foundation for an effective Safety Management System involves establishing a framework where stakeholders have access to the regulations that they must follow. It is essential that these regulations remain uniform across Europe, meaning that high-level rules should be consistently defined for all Member States. LLMs have shown success in customer service, legal reasoning, and medical domains, but aviation-specific compliance frameworks require tailored adaptations, including mitigation of hallucinations and continuous document updating.

The regulatory framework and safety standards in aviation have developed over many years and are consistently revised and enhanced to continually improve safety performance and address the challenges brought on by new air navigation concepts and the necessity to sustain the long-term viability of civil aviation. Therefore, keeping up to date with the regulatory information is crucial for anyone working in such domains. Analyzing the way in which the information is transmitted in the context of maintaining the safety levels will determine the effectiveness of the process. Ensuring the best possible means of accessing and understanding the information is a factor in the compliance with the standards imposed by the regulations. Studies show that problems with the implementation of laws and regulations are often attributed to the complex policy-making structure of the EU and the vague and poorly drafted policies [2].

The use of a generic large language model for regulations is not efficient, due to the many amendments and changes these regulations undergo constantly: unlike other sectors, to be updated with regulations in the aviation field is very difficult due to its dynamic nature, being a sector with frequent regulation changes. As the models have no knowledge of the latest amendments and changes, the risks of generating incorrect responses increases, which makes them too unreliable to use in such industry.

Here is where the usage of a specialized large language model comes into play. A domain-specific language model is more advantageous than the general language model when working in specific, niche areas, such as aviation.

Based on these findings and the need to accommodate these needs, as well as the continuous technological advancements in aviation, the developed model needs to be able to:

• interpret and analyze aviation regulations;

• understand the complex and specific language used in aviation;

• retrieve precise and relevant information;

• be able to offer support based on the requirements set out in regulations.

This work investigates the development of a specialized LLM for aviation regulation management, using Regulation (EU) 2017/373 [3] as its primary reference and object of study. It proposes an integrated architecture that combines fine-tuning with a Retrieval-Augmented Generation framework, thereby blending generative reasoning with context-driven information retrieval. The resulting system is designed to provide aviation professionals with precise, verifiable, and up-to-date regulatory support, helping to reduce the time and effort required to maintain compliance in this domain.

II. METHODS

The training and implementation of a specialized language model for regulatory management and compliance in the aviation industry is a complex process, and several logistical and technical considerations were identified.

1. Resource Requirements – Large language models are known for being computationally demanding, requiring heavy resources to be able to run. The infrastructure that is needed exceeds the hardware usually used for the normal, "personal use". When talking about the components, the more powerful the Graphical Processing Unit (GPU) is, the faster the training process is. Central Processing Unit (CPU) also plays a major role in loading the data, preprocessing and tasks related to model configuration. RAM and Storage are also essential for training parameters and managing the large amounts of data. During training, there are often required tens or hundreds of gigabytes of RAM, making the hardware infrastructure a costly but crucial aspect of training a large language model.

2. Training Costs - The models have billions of parameters and are being trained on terabytes of data. For example, the training of GPT-4 is estimated to cost between 60 and 100 million dollars, while the training of the next generation model, GPT-5, is closing to 1.2 billion dollars [4]. In addition to this, to be able to preserve the model's relevance and accuracy when dealing with regulations, having a continuous process of maintenance is necessary, which involves additional costs for training.

3. Data Collection – The training data is a crucial aspect, as the quality and the size of the dataset impact the performance and reliability of the model. The amount and structure of the training data depends on the tasks of the model and its role. Gathering the data is a time-consuming process, as the data needs to be preprocessed and cleaned, and the quality and correctness need to be assured.

Considering the constraints and challenges, the optimal solution for the environment was Google Colab, a notebook that uses Google's cloud infrastructure. This decision was taken due to Google Colab providing access to powerful resources and computing power and offering alternatives in choosing the most fitting CPUs and GPUs. Therefore, the model was trained and deployed on Google Colab Pro, using the following specifications:

• Python 3

• GPU: L4

• RAM GPU: 25 GB

• RAM CPU: 51 GB

• Storage: 201.3 GB

Since the resources are limited, the chosen model is a quantized instruct version of Mistral 7B, a model with 7.3 billion of parameters. Because of the constraint in cost and resources, this was considered the best solution based on its performance compared to other models with more parameters: it outperforms Llama 2 13B, with 13 billion parameters, and Llama 2 24B, with 34 billion parameters [5].

In this sense, the system uses RAG, as well as being fine-tuned to meet the specifications.

A. Fine-tuning

Fine-tuning is the process of continuing the training of a pre-trained model (base model) using a new set of data, smaller but more specific to the task. This is done by adjusting key parameters and updating the weights of the model in order to optimize its performance for specific tasks.

Although fine-tuning is more costly and resource demanding, there are some benefits added to it:

• Domain adaptation;

• Boost of performance;

• Stylistic alignment of the responses;

The choice to use fine-tuning was made taking into consideration the uses-cases of the model, the need of adaptability in different scenarios and situations, and the improved efficiency in domain specific tasks.

At a high level, fine-tuning the model involved several steps:

1. Preparation of training data

2. Training the model on the training data by adjusting parameters

3. Evaluate results and go back to previous step if needed

The fine-tuning process utilized Low-Rank Adaptation (LoRA) to adjust parameters with computational efficiency, reflecting aviation-specific terminology and compliance queries. Training data were prepared as natural language question–answer pairs linked to precise excerpts of the regulatory texts, preserving contextual fidelity and interpretability.

B. *Retrieval Augmented Generation*

Retrieval Augmented Generation works by taking the user's query and creates a response by looking through the new data source. It will then use its training knowledge combined with the information from the document and give a response based on both, compared to only the training knowledge if RAG is not implemented.

The choice to implement Retrieval Augmented Generation for the model of this study was made considering the following points:

• One of the model's objectives is the retrieval of correct information from aviation regulations;

• For the feasibility of the project, because of the high cost to maintain a language model up to date with the new regulations and amendments by new trainings;

• Exploring RAG in the context of aviation regulations and law;

• Possibility to use sensitive/private data;

• Easy evaluation of the model by cross-checking with the original document.

For the implementation of the Retrieval-Augmented Generation (RAG) model presented in this paper, a specific choice was made to use as the primary source for the retrieval component the Commission Implementing Regulation (EU) 2017/373 of 1 March 2017, regarding requirements for providers of air traffic management/air navigation services and other air traffic management network functions and their oversight. This regulation is stored and managed in PDF format, allowing for direct and efficient retrieval by the system.

III. RESULTS

Evaluation of the system included two representative use cases: regulation retrieval and compliance audit support. Performance was measured on criteria of accuracy, relevance, and contextual appropriateness. In a test set of 15 unseen regulation queries, the model achieved an average accuracy of approximately 79%, demonstrating a high capacity to retrieve and interpret domain-specific information. In an audit-assistance scenario, the model displayed an average performance and behavior. While some responses were accurate and reliable, with the model understanding basic concepts, it also displayed difficulty in a deeper understanding of specific topics.

The results demonstrate the synergy between generative modeling and retrieval augmentation: fine-tuning improved coherence and domain-specific alignment, while the RAG framework reduced hallucinations and supported consistency with up-to-date knowledge. Response evaluations show a strong performance in handling domain-specific language, references, and cross-linked regulatory clauses, providing a solid foundation for domain-critical use.

IV. DISCUSSION

While the overall performance is strong, there are areas where the model scored lower: these areas provide opportunities for further improvement, such as responses requiring deeper understanding or interpretation beyond direct regulation citation. Another important finding is that the model sometimes struggles with complex or too vague queries.

As with any tool, there is room for improvement. Therefore, future strategic improvements have been identified:

- Focused Training: diverse examples in training, especially those that reflect the complexities of real-world scenarios, would be beneficial.
- Real-Time Updates on Amendments of the Regulatory Framework.
- Coverage over the entire Regulatory Framework.

By focusing on the identified areas for improvement and continually adapting to the evolving landscape of aviation standards, the model can become even more integral to the industry, improving safety, compliance, and operational efficiency. This analysis reflects not only the model's current value but also opens a path forward for its future development, as well as the integration of large language models in aviation.

V. CONCLUSIONS

Because of the complexity of the aviation industry, when discussing aviation, the discussion moves to the foundation of it: safety. At this point, the advancements of large language models make this technology a powerful tool, but further work and research is needed in this direction.

The implementation of a specialized LLM enhanced with RAG for aviation regulation management marks a significant advance in AI-assisted compliance. The model demonstrates the ability to deliver explainable, domain-faithful, and contextually relevant answers.

The findings of this paper show promising results in the training of a large language model for the aviation industry, but, at the same time, highlights the challenges and considerations of the technical side when developing it and calling for a balanced approach to its development and implementation.

The story of Artificial Intelligence is only at its beginning, and its potential is waiting to be fully realized and harnessed. As we are standing now on the brink of what may be the most transformative technological era in human history, AI continues to unfold new possibilities across every sector of society.

VI. References

[1] INTERNATIONAL CIVIL AVIATION ORGANIZATION (ICAO), "Doc 10004, Global Aviation Safety Plan," 2020-2022.

[2] M. H. S. L. O. T. Gerda Falkner, "Non-Compliance with EU Directives in the Member States: Opposition through the Backdoor?," 2004.

[3] E. Commission, *Commission Implementing Regulation (EU) 2017/373,* 2017.

[4] D. Meyer, "The cost of training AI could soon become too much to bear," [Online]. Available: https://finance.yahoo.com/news/cost-training-ai-could-soon-101348308.html?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAECC9CAC-35c8HQBg_RSdc72ALk8scI6z_Ttxqh-I6UpOasyQjt2CCoIerayYwBxDZ_tQnMmcPsKjeyJC-4jSp94MZE-prT3NCcz&guccounter=2.

[5] A. S. A. M. C. B. D. S. C. D. d. l. C. F. B. G. L. G. L. L. S. L. R. L. M.-A. L. P. S. T. L. S. T. Albert Q. Jiang, 2023.