# Automating Regulatory Alignment: A RAG-Based Architecture for Navigating ESRS E1 Climate Disclosures

Yajas Kandel

February 12, 2026

## 1 Project Topic and Significance

The implementation of the Corporate Sustainability Reporting Directive (CSRD) and its associated European Sustainability Reporting Standards (ESRS) represents a major shift from voluntary sustainability disclosure to mandatory, audit-grade reporting across the European Union. Organizations now face substantial challenges in translating high level regulatory language into operational data collection workflows. Prior research in regulatory technology (RegTech) shows that the "conceptual ambiguity" and "inherent complexity" of emerging regulatory frameworks create significant barriers to compliance, particularly for small-to-medium enterprises (SMEs) [1]. These barriers often lead to inconsistent reporting practices and an increased dependence on external consultants.

Within the field of Computer Science, Retrieval-Augmented Generation (RAG) has emerged as a promising solution for navigating complex regulatory environments. Unlike traditional Large Language Models (LLMs), which rely solely on static training data, RAG systems dynamically retrieve authoritative, up-to-date documents during inference. This capability is especially important for sustainability reporting, where accuracy, verifiability, and source traceability are essential to avoid risks such as greenwashing. Research in Natural Language Processing demonstrates that RAG architectures significantly reduce hallucinations by grounding model outputs in retrieved evidence [2]. Additional studies show that retrieval based constraints improve factual consistency and reliability in high stakes domains such as law, policy, and scientific documentation [3].

The primary objective of this project is to develop a specialized AI tool designed to support sustainability teams in aligning their internal data with the ESRS E1 (Climate Change) standard. The AI will function as a "Knowledge Assistant" that provides evidence-backed answers to reporting queries by retrieving relevant passages from the ESRS E1 and the relevant European Financial Reporting Advisory Group (EFRAG) guidelines, generating grounded explanations through a RAG pipeline. The system will help organizations interpret disclosure requirements, identify necessary data inputs, and reduce reliance on external consultants by offering transparent, traceable, and regulation-aligned guidance.
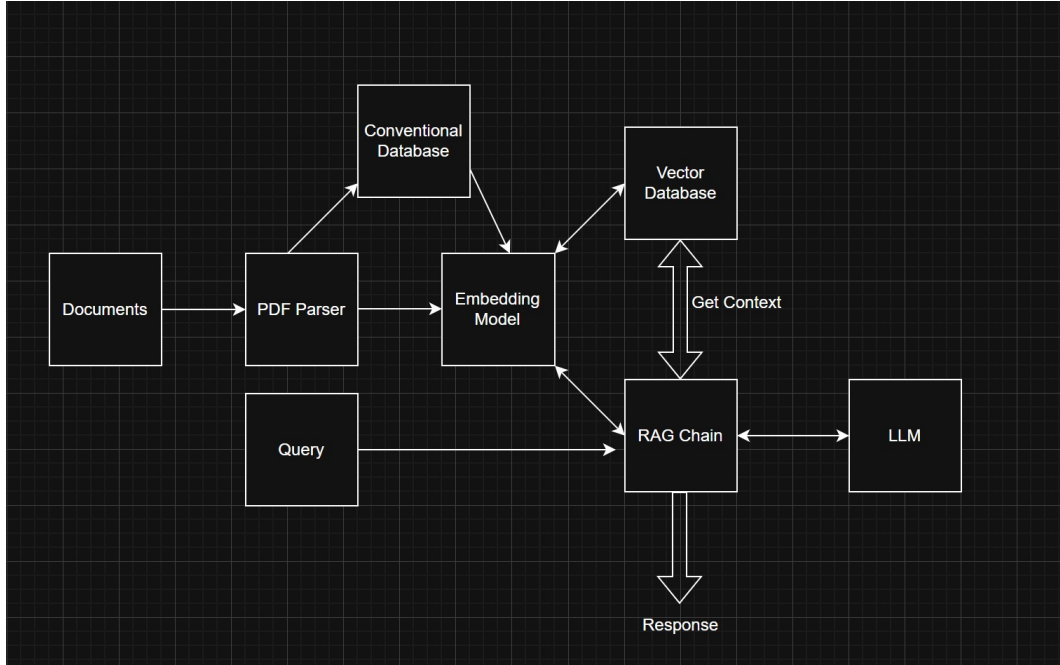
Figure 1: RAG Pipeline Diagram

# Appendix

Planned Features for the AI assistant include:

- **Data Preparation**: Manually extract and clean raw text from official ESRS E1 and EFRAG documentation. This involves converting PDF content into a structured text format while ensuring that headers, footers, and page numbers are accounted for as metadata to prevent data noise during AI processing.

- **Document Chopper**: Implement a function to partition extracted text into small, manageable segments to ensure efficient AI processing.

- **Vector Database**: Establish a Vector Database to store text segments and their associated metadata (e.g. page number, header), allowing the system to retrieve information with verifiable source references.

- **Search Tool**: Create a retrieval mechanism that identifies the most relevant text segments from the database in response to a user query.

- **Strict Instructions**: Author a "System Prompt" that mandates the AI generate responses using only the provided segments and disclose when information is unavailable.

- **RAG Chain Integration**: Orchestrate a Retrieval-Augmented Generation pipeline to programmatically link the retrieval mechanism with the language model, ensuring that generated responses are grounded in the retrieved context.

- **Basic Chat Box**: Build a web-based user interface featuring a text input field for natural language inquiries.

- **Answer Display**: Design a clear output area to present the AI-generated responses to the user.

- **Source "Receipts"**: Implement a citation feature that displays the specific page or section of the regulation used to generate each answer.

- **Clear Chat Button**: Add a session reset function to allow users to clear history and begin new inquiries.

- **PDF Viewer (Stretch Goal)**: Integrate an on-screen document viewer that displays the actual PDF page alongside the AI's response.

- **Loading Bar (Stretch Goal)**: Add a visual progress indicator while the system performs retrieval and generation tasks.

- **Extend to other ESRS Standards (Stretch Goal)**: Expand the system's capabilities to include additional ESRS standards beyond E1, such as E2 (Pollution) and S1 (Own Workforce).

- **Automated PDF Parsing (Stretch Goal)**: Develop a script to extract raw text from PDF files while maintaining the structural layout, reducing the need for manual data preparation in future iterations.

# References

[1] ARNER, D., BARBERIS, J., AND BUCKLEY, R. Fintech, regtech, and the reconceptualization of financial regulation. *Northwestern Journal of International Law and Business 37* (01 2017), 373–415.

[2] LEWIS, P., PEREZ, E., PIKTUS, A., PETRONI, F., KARPUKHIN, V., GOYAL, N., KÜTTLER, H., LEWIS, M., TAU YIH, W., ROCKTÄSCHEL, T., RIEDEL, S., AND KIELA, D. Retrieval-augmented generation for knowledge-intensive nlp tasks, 2021.

[3] SHUSTER, K., POFF, S., CHEN, M., KIELA, D., AND WESTON, J. Retrieval augmentation reduces hallucination in conversation, 2021.