FEDERAL UNIVERSITY OF JUIZ DE FORA
EXACT SCIENCES INSTITUTE
POSGRADUATE PROGRAM IN COMPUTER SCIENCE

Matheus Brugger Jenevain

Contextualized Data Sharing in the Energy Industry:
A Retrieval-Augmented Ontology Solution

Juiz de Fora
2025

Matheus Brugger Jenevain

Contextualized Data Sharing in the Energy Industry:

A Retrieval-Augmented Ontology Solution

Dissertation presented to the Posgraduate Program in Computer Science of the Federal University of Juiz de Fora as a partial requirement for obtaining the title of Master in Computer Science.

Advisor: Prof. Dr Mario A. R. Dantas

Coadvisor: Prof. Dra Milena F. Pinto

Juiz de Fora

2025

Cataloging record prepared using the UFJF CDC Latex Model with data provided by the author

Jenevain, Matheus Brugger.

Contextualized Data Sharing in the Energy Industry : A RetrievalAugmented Ontology Solution / Matheus Brugger Jenevain. – 2025. 61 f. : il.

Advisor: Mario A. R. Dantas Coadvisor:

Milena F. Pinto

Dissertation (Master) – Federal University of Juiz de Fora, Exact Sciences Institute. Posgraduate Program in Computer Science, 2025.

1. Large Language Models. 2. Context-Aware. 3. Ontology. I. Dantas, Mario Antônio Ribeiro, orient. II. Pinto, Milena Faria. III. Title.

Matheus Brugger Jenevain

Contextualized Data Sharing in the Energy Industry:
A Retrieval-Augmented Ontology Solution

Dissertation presented to the Posgraduate Program in Computer Science of the Federal University of Juiz de Fora as a partial requirement for obtaining the title of Master in Computer Science.

Approved on June, 27th 2025

EXAMINING BOARD

Prof. Dr. Mario Antonio Ribeiro Dantas - Orientador
Universidade Federal de Juiz de Fora

Profª. Dra. Milena Faria Pinto - Coordenadora
Centro Federal de Educação Tecnológica Celso Suckow da Fonseca

Prof. Dr. Victor Ströele de Andrade Menezes
Universidade Federal de Juiz de Fora

Prof. Dr. Nelson Francisco Favilla Ebecken
Universidade Federal do Rio de Janeiro

Documento assinado eletronicamente por **Victor Stroele de Andrade Menezes**, **Professor(a)**, em 27/06/2025, às 16:07, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Milena Faria Pinto**, **Usuário Externo**, em 14/07/2025, às 15:34, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Nelson Francisco Favilla Ebecken**, **Usuário Externo**, em 16/07/2025, às 17:52, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Mario Antonio Ribeiro Dantas**, **Professor(a)**, em 22/07/2025, às 09:10, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **2423463** e o código CRC

I would like to dedicate this work to everyone who supported me along the way, especially my advisors, my wife Laís, my family, and my friends.

ACKNOWLEDGMENTS

I would like to thank my advisors, Professor Mario and Professor Milena, for their guidance, support, and patience throughout my research. I would also like to express my gratitude to INESC P&D Brasil for the invaluable opportunities to contribute to impactful projects and for the supportive environment that greatly enriched my professional and academic development.

"The mystery of human existence lies not in just staying alive, but in finding something to live for."
(Fyodor Dostoievsky, 1879)

RESUMO

A crescente globalização dos mercados de energia, o surgimento de novas fontes de energia e a demanda por troca fluida de dados entre as partes interessadas ressaltam a necessidade crucial de comunicação eficaz e interoperabilidade no setor energético. No entanto, os protocolos existentes frequentemente priorizam especificações técnicas em detrimento da compreensão contextual, resultando em interpretações equivocadas e falhas na comunicação. Este trabalho propõe uma abordagem inovadora que combina ontologias com Retrieval Augmented Generation (RAG) e Large Language Models (LLMs) para garantir a interpretação precisa dos dados compartilhados, independentemente dos diversos contextos dos atores envolvidos. Ao aproveitar os pontos fortes das ontologias para a representação do conhecimento e do RAG para a compreensão contextual, a estrutura proposta visa aprimorar a interoperabilidade e facilitar a colaboração eficaz no setor de energia, contribuindo para uma transição energética mais suave e eficiente. Os resultados indicam que a solução proposta é eficaz na atribuição de significados contextualmente relevantes aos termos, melhorando o compartilhamento de dados e reduzindo a probabilidade de mal-entendidos decorrentes de diferentes contextos e perspectivas entre emissores e receptores.

Palavras-chave: context-aware; large language models; ontologies; renewable energies; retrieval augmented generation.

ABSTRACT

The increasing globalization of energy markets, the advent of new energy sources, and the growing demand for seamless data exchange among stakeholders emphasize the vital need for effective communication and interoperability within the energy industry. However, existing protocols often prioritize technical specifications over contextual understanding, leading to misinterpretations and communication breakdowns. This work proposes a novel approach that combines ontologies with RAG and LLMs to ensure accurate interpretation of shared data, irrespective of the diverse contexts of the actors involved. By leveraging the strengths of ontologies for knowledge representation and RAG for contextual understanding, the proposed framework aims to enhance interoperability and facilitate effective collaboration within the energy sector, ultimately contributing to a smoother and more efficient energy transition. The findings suggest that the proposed solution effectively assign contextually relevant meanings to terms, thereby improving data sharing and reducing the likelihood of misunderstandings arising from differing contexts and perspectives between senders and receivers.

Keywords: context-aware; large language models; ontologies; renewable energies; retrieval augmented generation.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

AEMO	Australian Energy Market Operator
AINA	Advanced Information Networking and Applications
AI	Artificial Intelligence
ANEEL	Agência Nacional de Energia Elétrica
ANN	Approximate Nearest Neighbor
BLEU	Bilingual Evaluation Understudy
CIEEMAT	Congresso Ibero-Americano de Empreendedorismo, Energia, Ambiente e Tecnologia
CIM	Common Interface Model
DPR	Dense Passage Retrieval
EEA	European Environment Agency
ENTSO-e	European Network of Transmission System Operators for Electricity
FAIR	Findable, Accessible, Interoperable, and Reusable
FDO	FAIR Digital Objects
ICT	Information Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IRENA	International Renewable Energy Agency
LLM	Large Language Models
NLP	Natural Language Processing
NLTK	Natural Language ToolKit
OEO	Open Energy Ontology
RAG	Retrieval Augmented Generation
ROUGE	Recall-Oriented Understudy for Gisting Evaluation
SG	Smart Grids
TF-IDF	Term-Frequency Inverse Document Frequency
UN	United Nations

SUMMARY

1 INTRODUCTION

Due to the growing necessity for reliable and zero-carbon energy supplies, the current electrical grids require upgrades to become more intelligent, flexible, and capable of autonomous operation, monitoring, and self-recovery. SG emerged as a primary focus in modernizing power systems. However, due to its heterogeneity regarding components, purposes, and protocols (Ho et al., 2012), appropriate semantic interoperability needs to ensure seamless information exchange between components (Salameh et al., 2019). It is also important to highlight that, with the increased penetration of renewable energies on the grid, new technologies and standards had to be developed to maximize the returns while enabling better communication between all the players in the energy sector (Anisie et al., 2023).

As concerns about environmental issues, like gas emissions, so does the necessity for cleaner, sustainable energy (Pata et al., 2023). Brazil, for example, registered its highest clean energy production, up until that point, in 2023 (Ministério de Minas e Energia, 2023). Also, the REGACE Project (RegAce Project, 2023), funded by the European Commission's Horizon Europe Coordination and Support Actions, among others, attempts to do the same by using agrovoltaics technology using $CO2$ to enhance greenhouse yields and improve electricity production. Also, COP 27 (Cop 27: Energy security and prosperity for the future - wind and solar, 2022) focused entirely on renewable energies, their importance, and how production on a larger scale could be achieved.

The 2015 IEEE Smart Grid Report (IEEE Smart Grid Annual Report, 2015) highlighted the importance of operability within the energy sector, leading to projects like ENTSO-e's (ENTSO-E, 2024) Tdx Assist (TDX Assist, 2024), which focused on interoperability technologies and protocols. Also, as seen with TDX-Assist (Suljanović et al., 2019), there is an emphasis on the interoperability aspect and on the importance of exchanging data between operators, market agents, and distributed energy resources. Therefore, it is plausible to assume that advances in interoperability and the increase in the production and use of clean energies go hand in hand.

According to (Ministério de Minas e Energia, 2010), interoperability is the possibility of two distinct systems exchanging data with each other securely and transparently. For this, creating protocols is necessary to enable both sides to communicate, no matter how different their standards may be. NIST, the body responsible for defining standards in the United States, the ENTSO-e above, and IRENA play a role in this.

This emphasis on seamless operation underscored the complexities inherent in energy data exchange, complexities influenced by factors like the nature of the data itself and the specific geographic context (Radi et al, 2019). For instance, ENTSO-e promotes the CIM (IEEE 61970-30) ENTSO-E Common Information Model, 2024), while AEMO uses AseXml tailored for Australia (Australian Energy Market Operator, 2024) (ASEXML Schemas – AEMO, 2024). IRENA's 2023 World Energy Transitions Outlook further addressed these challenges (International Renewable Energy Agency, 2024) (World Energy Transitions Outlook, 2023).

Another factor in increasing interest in interoperability is globalization and tech data exchange. These phenomena amplify challenges related to information exchange, regardless of the source (Liu et al., 2023). Increased communication necessitates understanding to avoid miscommunications, especially given the various protocols and data formats in use. Green hydrogen exemplifies this Alamwith new projects/policies aiding its production (Raimondi; Spazzafumo, 2023) (Zainal et al., 2024). As renewables grow, so does the need for cooperation and modernizing processes (Alam; Murad, 2020).

For this reason, interoperability can be quite challenging, particularly where common understanding between senders and receivers of any data is concerned. One such example can be seen with ENTSO-e and the countries under its umbrella, each with its own protocols, regulations, and standards. Green hydrogen, again, can be prone to this since each country may interpret its meanings, concepts, and uses in different manners.

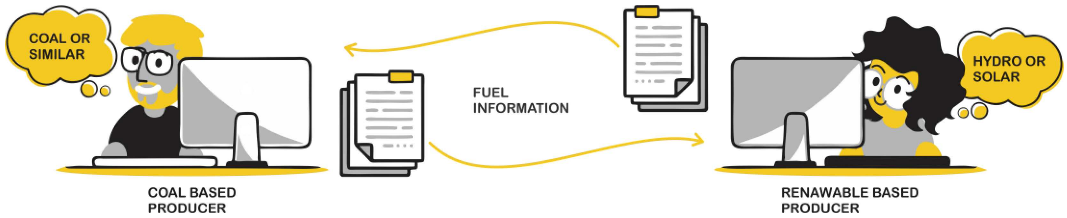
Such miscommunication arises from words having different meanings, altering data interpretation based on the context that each actor belongs to. Santos et al. (Santos, 2012) defines "context" as factors shaping reality and influencing the perception and interaction of those under its influence. For example, actor A's context *storage* could mean the capacity to store and retrieve electricity. However, actor B knows it as the amount of data stored in a hard drive or the cloud. Any data exchange involving this concept could be severely misinterpreted in this scenario (Jenevain et al., 2024).

Addressing interoperability benefits diverse settings like Brazil, where the large country size is home to considerable differences in culture and reality. This work aims to mitigate challenges by conveying intended meaning accurately, ensuring effective communication.

Researchers such as Booshehri et al. (2021), Kemp et al. (2023) and Costa et al. (2019) studied interoperability and its challenges from various angles, with some creating ontologies for structured data exchange, while others examined context and the definition and perception of context on processes. However, ontologies alone are not initially built to convey context-specific meaning. This can be achieved, either manually or automatically, by using other techniques or solutions. In linguistics, for instance, RAG (Retrieval-Augmented Generation) can be used for such a task (Zhao et.al, 2024).

Figure 1 shows two actors attempting to communicate but understanding a term in different ways due to their interpretations and contexts. In this scenario, a user initiates an inquiry with a query such as "What is...". The RAG solution, leveraging the pre-trained knowledge and parameters of an LLM, integrates this with information extracted from external sources such as documents or databases.

Figure 1 - Communication in Different Contexts



Source: Created by the author (2025)

This synthesized information is then processed to generate a response to the user's question. Finally, the robot on the right delivers the answer, stating "The answer is...". This process ensures that the response leverages both the LLM's vast knowledge and the up-to-date information retrieved via RAG.

Furthermore, authors like Li et al. (2022) categorize RAG's information sources into three types: the training corpus, accessed on-demand; external data, retrieved from datasets; and unsupervised data, proposed for scenarios such as machine translation.

Lewis et al. (2020) conducted experiments showcasing RAG's effectiveness in tasks like question answering, Jeopardy, and fact-checking. Their results demonstrate RAG's capability to understand input and generate contextually relevant answers, mirroring human comprehension.

This emphasis on context aligns with the solution proposed in this work by addressing the question from two fronts. Initially, it establishes the fundamental structure of an ontology designed to capture and manage the context of any actor involved in data exchange, be it the sender or receiver. Secondly, it focuses on assigning meanings to key terms, leveraging LLMs and RAG to identify the most suitable match within documents and texts relevant to the energy sector.

1.1 MOTIVATION

This research is driven by the necessity to facilitate seamless communication between actors in the energy sector, ensuring accurate interpretation. Commonly used protocols, as discussed in section 1, often prioritize data exchange and structure (Wang et al., 2020), rather than nuanced context and interpretation. This creates challenges in a globalized energy landscape where diverse countries engage in trade involving both renewable and non-renewable sources.

This diversity is not limited to international interactions; it also exists within countries. For example, Australia’s energy market trades domestically produced energy from various sources (Australian Energy Regulator, 2024). Similarly, Brazil, with its vast territory and regional variations, faces challenges due to differing interpretations of energy terminology. Policies focused on energy transition demand nuanced understanding and communication, which may not be fully supported by protocols solely focused on terms and regulations (Werner; Lazaro, 2023).

These complexities highlight the need for innovative solutions that address both structured data sharing (maintaining the regulatory compliance and trust of existing protocols) and mutual understanding of information’s meaning beyond mere terminology. This dual need was the primary catalyst for this research and the development of the proposed solution.

After participating in projects in the energy field, particularly concerning renewable energy and where I conducted in-depth research on the interoperability efforts of various TSOs and DSOs, this reality became apparent. A common challenge emerged in countries operating within markets spanning large geographical distances. These countries faced difficulties in understanding and cooperation when contexts differed significantly. Additionally, problems arose when commonly used protocols handled information without considering specific contextual details.

1.2 RESEARCH EVOLUTION

Since its inception, driven by needs observed in the energy sector, the proposed solution and related research have evolved through four key phases, each published in recognized events. These phases, in chronological order, are:

- First Stage: The initial focus was on developing an ontology to capture context, enabling the collection of data necessary to contextualize terms and data exchanged between actors. This was presented at AINA 2024.
- Second Stage: Building on the ontology, this phase integrated Python libraries and logic, utilizing dictionaries and NLTK's WordNet to infer the most appropriate meaning of a term within its context. It was presented at CIEEMAT 2024.
- Third Stage: The meaning assignment process was refined using RAG techniques and trained models, incorporating documents as additional context when needed. This allowed for a broader understanding and suggestion of meanings beyond WordNet's capabilities.
- Fourth Stage: This final stage presented the complete solution, including the knowledge required to build it and the insights it generated. It also showcased the evolution and maturation of the idea. It was submitted as the final requirement for a master in computer science at the Federal University of Juiz de Fora.

1.3 PROBLEM STATEMENT

The increasing globalization of energy markets, coupled with the rise of diverse energy sources and the growing need for interoperability, demands seamless communication and data exchange among energy stakeholders. However, current protocols prioritize technical aspects, often overlooking crucial nuances of context and interpretation (Jenevain et al., 2024).

This can lead to miscommunication and misunderstandings, particularly in cross-cultural or cross-regulatory interactions, hindering effective collaboration and progress toward a sustainable energy future.

This work addresses this challenge by proposing an ontology based on the structure proposed in Jenevain et al. (2024), which captures contextual nuances and assigns the most appropriate meanings to terms requiring clarification during information exchange. By integrating computer science and linguistics technologies and techniques, the proposed solution aims to enhance the concept of ontologies for the renewable energy field, facilitating clearer and more effective communication in this complex and evolving landscape.

1.4 RESEARCH QUESTION

The central research question guiding this work is: How can data shared between actors in the energy sector, regardless of their diverse contexts and backgrounds, be understood precisely as intended by the sender, preventing any miscommunications or loss of meaning?

To explore this further, the research question can be broken down into the following sub-questions:

- How can ontologies be effectively utilized to capture and represent the diverse contexts of various actors within the energy sector?
- In what ways can LLMs and RAG be leveraged to assign contextually appropriate meanings to technical terms, enhancing understanding and minimizing ambiguity?
- How can the proposed framework be seamlessly integrated with existing energy data exchange protocols, ensuring compatibility and smooth adoption?
- What methods can be employed to thoroughly evaluate the effectiveness of this framework in real-world scenarios involving diverse stakeholders and energy sources?

Additionally, a key objective is to assess the feasibility of integrating LLMs and RAG with ontologies in tools that have traditionally prioritized regulations and standardization over conveying meaning and analyzing context, potentially bridging the gap between technical compliance and effective communication.

1.5 RELATED PUBLICATIONS

The research conducted during the master's studies resulted in two publications, each documenting the state of both the solution and the methodology at that time. A careful analysis of both publications clearly demonstrates the evolution from a simpler ontology and dictionary-matching approach to the more robust and novel RAG approach presented in the final proposed solution. They are as follows:

Enabling Intelligent Data Exchange in the Brazilian Energy Sector: A Context-Aware Ontological Approach [AINA 2024] by Matheus B. Jenevain, Milena F. Pinto, Mario A. R. Dantas, Regina M. M. B. Villela, Jose M. N. David, and Victor S. A. Menezes: The growing need for interoperability in the energy sector, coupled with an increasing number of participants, presents a challenge: diverse actors with unique perspectives must exchange data and knowledge across significant contextual differences. This can lead to misunderstandings and difficulties, as information exchanged is susceptible to misinterpretation based on the sender's and receiver's individual contexts. To address this, this work proposed an extension to the OEO that focuses on context. It investigates how an actor's understanding is shaped by their context, explores methods for inferring this context, and proposes strategies to enhance interoperability. The results demonstrate the potential of this approach to improve communication and data exchange within the energy sector.

- A Context-Aware Approach to Data Exchange in the Energy Sector [CIEEEMAT 2024] by Matheus B. Jenevain, Mario A. R. Dantas, Laís R. Berno, and Milena F. Pinto: The paper introduced a novel approach to enhance data exchange within the energy sector. Specifically, it proposes a context-aware method that leverages NLP techniques and ontologies, such as the OEO, to address the challenges arising from differing interpretations of data. This is particularly crucial for actors like ENTSO-E and its member countries, who often operate with varying perspectives. By capturing, representing, and assigning meanings based on the specific realities of the energy sector, this method aims to minimize miscommunications and complications during data exchange. The proposed solution is built upon previous research exploring the interplay between ontology development and contextual perception.

- Using UAVs and Retrieval Augmented Generation for Situational Awareness in Rescue Operations [AINA 2025] by Matheus B. Jenevain, Mario A. R. Dantas, Laís R. Berno, and Milena F. Pinto: The increasing frequency and severity of natural disasters in Brazil, particularly in 2024, highlight the urgent need for innovative solutions in Search and Rescue (SAR) operations. This work presents a novel framework integrating Retrieval-Augmented Generation (RAG) techniques with Unmanned Aerial Vehicles (UAVs) to enhance real-time data processing, usability, and operator decision-making. By incorporating advanced technologies such as FrameNet Brasil, Robot Operating System 2 (ROS2), and Large Language Models (LLMs), the system transforms UAV-captured data into actionable insights accessible through natural language interfaces. Testing demonstrates its ability to improve situational awareness, identify critical points of interest, and streamline mission execution. This modular and scalable approach lays the groundwork for future advancements in SAR technologies and their application in disaster-prone regions.

- Context-Aware Data Exchange in the Energy Sector: A Retrieval-Augmented Ontology Approach [IEEE ISCC 2025 - To be Presented] by Matheus B. Jenevain, Mario A. R. Dantas, Laís R. Berno, and Milena F. Pinto: The increasing globalization of energy markets, the emergence of new energy sources, and the growing need for seamless data exchange among stakeholders highlight the importance of effective communication and interoperability in the energy sector. However, current protocols often prioritize technical aspects over contextual understanding, leading to misinterpretations and communication breakdowns. This paper proposes a novel solution that combines ontologies with Retrieval Augmented Generation (RAG) and Large Language Models (LLMs) to ensure accurate interpretation of shared data, regardless of the actors' diverse contexts. By leveraging the strengths of ontologies for knowledge representation and RAG for contextual understanding, the proposed framework aims to enhance interoperability and facilitate effective collaboration in the energy sector, ultimately contributing to a smoother and more efficient energy transition. The results indicate that the proposed solution effectively assigns contextually relevant meanings to terms, as evidenced by the Cosine Similarity and ROUGE Recall scores exceeding 0.5 for most generated answers, enhancing data sharing and mitigating potential misunderstandings originating from differing contexts and perspectives among senders and receivers.

- Augmented Energy Ontology [Intelectual Property] by Matheus B. Jenevain, Mario A. R. Dantas, Laís R. Berno, and Milena F. Pinto: Python solution to inject context defined meaning into ontologies using retrieval-augmented generation, aiming to improve intercommunication between actors in the energy sector.

1.6 STRUCTURE

This work is structured as follows:

- Material and Methods: This section provides the background knowledge and context necessary to understand and develop the proposed solution.

- Other applications of RAG in the energy Sector: This section explores additional potential use cases for RAG within the energy sector that may not have been covered in depth in this work.
- Study Case: This section details the proposed solution's development, implementation, and evaluation within a specific test scenario.
- Conclusion: This section summarizes the findings, discusses potential future research directions, and highlights the key takeaways from the proposed solution and its results.

The overarching goal of this work is to provide a comprehensive exploration of all the steps involved in addressing the identified problem—from recognizing the need for a solution and conceptualizing it to its practical implementation, evaluation, and potential future implications.

2 MATERIAL AND METHODS

This section establishes the foundational knowledge necessary to comprehend the proposed solution’s key aspects. It aims to elucidate concepts like ontologies, RAG, and the current renewable energy landscape in Brazil, fostering a deeper understanding of their relationship and mutual benefits. By the end, the fundamental concepts essential for understanding the proposed solution’s inner workings and components will be clear.

2.1 BACKGROUND

2.1.1 Ontology

The energy sector is experiencing a dramatic shift worldwide, fueled by the rise of renewable energy sources, the decentralization of power production, and a growing focus on environmental responsibility. This change has led to a significant increase in the number and variety of participants in the energy industry, encompassing traditional utilities, renewable energy producers, grid operators, and consumers.

With the energy ecosystem becoming increasingly intricate and interconnected, the necessity for smooth communication and data exchange becomes more urgent, as do the potential risks of miscommunication and misunderstanding. Existing data protocols effectively represent and transmit information within a defined structure, but their emphasis on standardization can sometimes prioritize technical accuracy over clarity, meaning, and understanding. This can create confusion, particularly when stakeholders come from diverse cultural, linguistic, or regulatory backgrounds.

Motiwalla et al. (Motiwalla; Thompson, 2009) underscored that even with shared data and terminology, challenges in information exchange can arise even within well-established ICT systems. In order to tackle issues of information sharing and standardization, a range of solutions have been explored. Alongside protocols like CIM, ontologies, as originally defined by Gruber (2009), have been employed to effectively represent concepts and their interconnections across diverse domains, thus streamlining data organization and comprehension. In particular, the practicality and sophistication of ontologies in storing, organizing, and manipulating information have made them particularly valuable within computer science, especially in areas related to knowledge representation and reasoning.

Chandrasekaran et al. (1999) further supported the fundamental definition of ontologies and highlighted their importance in object-oriented software design, emphasizing a growing convergence between these two fields. They also stressed the role of vocabularies and taxonomies as integral parts of ontologies, drawing upon the in-depth information presented in (Guarino et al., 2009). Projects such as those referenced in Section 1 leveraged protocols that could also have been developed as ontologies, especially for the purpose of collecting, structuring, and disseminating knowledge and information.

Within the domain of smart grids and energy systems, ontologies offer a structured approach to the representation of knowledge by defining relationships between concepts and entities within a specific domain. This organized representation facilitates the seamless exchange of information between various entities in the energy sector, ensuring a shared understanding of data relationships and the overall structure of information.

The field of energy presents numerous possibilities for semantic applications, emphasizing the critical role of conceptual modeling in diverse energy-related initiatives. Domain ontologies are particularly valuable in this modeling stage, aiding the representation of concepts and their connections within the energy sector and its subdomains. These ontologies are instrumental in addressing interoperability issues that often arise between different energy applications.

Despite their significance, the development of ontologies specifically tailored for the energy sector remains somewhat limited, as noted by Küçük and Küçük (2018). While there are notable examples, such as ontologies focused on electrical power quality parameters and events (Küçük et al., 2010), wind energy (Küçük; Arslan, 2014), and a broader electrical energy ontology (Küçük, 2011), these primarily address various aspects of electrical energy. Additionally, specialized ontologies have been developed for energy efficiency in smart grids (López et al., 2015) and ontology matching within the smart grid domain (Santodomingo, 2014). However, these existing ontologies are not specifically designed to capture and convey meaning, ensuring shared understanding between parties, nor do they offer tools to address the nuances of meaning and context.These instances underscore the importance of utilizing ontological structures to enhance the effectiveness of semantic applications and address challenges related to interoperability within the dynamic energy sector.

As smart grids continue to generate vast amounts of data, the necessity for FAIR data management solutions becomes increasingly apparent, particularly in scientific and operational contexts. FDOs offer access to metadata, and ontologies are essential in explicitly defining both metadata and application data objects across diverse domains. Schweikert et al. (2023) put forward the idea of integrating FDOs and ontologies as metadata models to enable data access for a wide range of energy stakeholders.

One such solution, designed to address the escalating need for interoperability and data exchange in the energy field, is the OEO, introduced in (Booshehri et al, 2021). OEO's core purpose lies in providing clear and precise definitions of data tailored specifically for the analysis of energy systems. However, as highlighted by Küçük and Küçük (2018) , the development of ontologies specific to the energy sector, with a focus on capturing and conveying meaning, remains an area with room for growth.

As stated, the OEO is an open-source ontology focused on representing the multifaceted reality of the energy sector, along with its particularities and challenges. It aims to be an ontology that is both easy to use and extend, whether by modifying its structure with tools like Protegé (Stanford Center for Biomedical Informatics Research, 2024) or by creating extensions and plugins.

The first paper resulting from the research and development of the proposed solution (Jenevain et al., 2024), presented at AINA 2024, proposed an ontology to extend the OEO. This extension aimed to enable the OEO to capture, understand, and process a wealth of data related to contexts, using the work of Dos Santos et al. (2012) as the basis for the construction of all relations, properties, and objects.

The use of OEO was more prominent in the early stages of the research, particularly concerning the production of the first paper. However, this doesn't imply that this artifact should be disregarded in the future. The possibility of an extension to the ontology or other such applications could be of great interest for future endeavors.

For instance, one notable domain ontology focuses on electrical power quality parameters and events, as detailed in Kuccuk et al. (2019). Another example, documented in Küçük and Arslan (2010), involves a domain ontology for wind energy developed through a semi-automated procedure. Expanding to a higher level, an ontology covering the broader domain of electrical energy is presented in Küçük (2015).

The presence of ontologies specifically designed to tackle distinct challenges within the energy sector, including areas like electrical power quality, wind energy, and energy efficiency in smart grids, underscores the real-world importance and usefulness of utilizing ontological structures in this field.

2.1.2 RAG Principles

This section presents the fundamental concepts of how RAG operates through each phase, from retrieval to generation. Additionally, it illustrates how RAG can be advantageous in mitigating miscommunication issues between actors.

The application of AI to create or enhance data, particularly text, has seen a remarkable expansion in recent times. The latest developments in AI have rendered tasks such as information retrieval and response generation, alongside a range NLP tasks like text generation or classification, more accessible and streamlined than ever before. Prominent instances of such tools encompass ChatGPT, LLAMA Google Gemini for text manipulation and generation, DALL-E for image creation, and Sora for video (Zhao et al., 2024), among others.

Chatbots serve as a prime example of the practical implementation of RAGs. Prior to their advent, solutions depended on searching within predefined datasets to retrieve information aligned with user inquiries. While this approach was effective for straightforward questions, any departure from established patterns or gaps in the database's knowledge often resulted in unsatisfactory responses. Furthermore, the capacity for improvisation or handling novel situations was constrained (Martineau, 2024).

The approach of returning queries or answering questions, as adopted by certain models, has found innovative applications across various domains. Financial records are being queried to derive insights, and medical data is being leveraged to assist healthcare professionals, both demonstrating impactful use cases for this technology (Merritt, 2024).

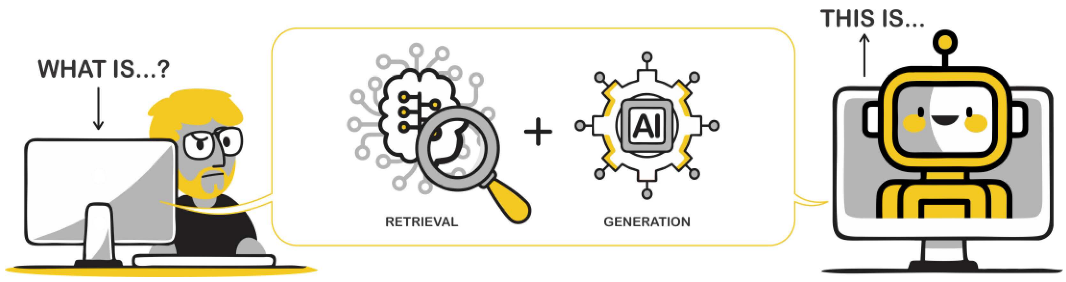
Organizations such as IBM are actively engaged in the development and enhancement of this technology, particularly for applications like customer support and other areas focused on responding to user questions (Martineau, 2024).

Large Language Models (LLMs) are trained on extensive datasets, empowering them to comprehend and produce human-like text based on their training data and adjust to various topics. Their vast number of parameters, coupled with extensive fine-tuning and adjustments, bolsters their capacity to address a broad spectrum of Natural Language Processing (NLP) requirements, including intricate tasks that involve text generation as noted by Zhao et al. (2023) and (Martineau, 2024).

Nevertheless, despite being trained on enormous volumes of data, LLMs can become outdated as new information emerges. In such instances, Retrieval Augmented Generation (RAG) presents an advanced solution to ensure applications like chatbots consistently access adequate and current data for processing and content generation with text and images, as seen in Zhao et al. (2023), International Business Machines Corporation's article on LLMs (2024) and Martineau (2024)

Figure 2 provides a visual representation of how a Retrieval Augmented Generation (RAG) system functions. When a user provides input, the RAG system combines the pre-existing knowledge embedded within a language model (for example, LLaMA, GPT, or HuggingFace) with external data sourced from repositories like text files or databases to generate a response.

Figure 2 - RAG solution combining model and external knowledge



Source: Created by the author (2025)

In this scenario, the user on the left initiates the interaction by posing a question, such as "What is...". The RAG system then leverages the pre-existing knowledge of an LLM, along with its extensive training data and parameters, in conjunction with information extracted from external sources such as documents or databases.

This combined knowledge is then processed to formulate a response to the user's query. Finally, the robot on the right presents the answer, stating "The answer is...". This mechanism ensures that the generated answer is not solely based on the vast knowledge base of the pretrained LLM but also incorporates the latest information obtained from external sources via RAG.

Moreover, researchers such as Li et al. (2022) categorize the data sources utilized in RAG into three distinct types: the training corpus, accessed when necessary; external data, retrieved from datasets and unsupervised data, proposed for situations not covered by the first two, such as machine translation.

Lewis et al. (2020) conducted experiments with RAG models on tasks like question answering, Jeopardy, and fact-checking. Their findings demonstrated RAG's capability to comprehend input and produce contextually relevant answers, mirroring human understanding. This emphasis on context in generating responses is crucial to the solution proposed in this work.

2.1.3 Related works

This section presents the most impactful papers considered during the research and implementation of the proposed solution, among all referenced works. Additionally, it highlights the articles generated at each step of the research, from its inception to its conclusion.

Tables 1 and 2 provide a chronological summary of key publications concerning ontologies and their implementation in the energy sector. It also highlights studies on Retrieval Augmented Generation (RAG) and Large Language Models (LLMs) that elucidate their internal mechanisms and their potential to produce contextually appropriate text. This overview places special emphasis on research that integrates context and perception into ontologies and RAG for generating context-specific responses, which directly relates to the central theme of this work.

Table 1 – Key Contributions in Ontologies for the Energy Sector

Year	**Researcher(s)**	**Major Contributions**
1993	Gruber	Introduced the concept of ontologies and highlighted their increasing significance in knowledge representation and processing.
1999	Chandrasekaran et al.	Offered a comprehensive perspective on the role of ontologies within objectoriented software design.
2014	Santodomingo et al.	Described an ontology matching system for the smart grid domain, aiming to address interoperability challenges.
2015	Lopez et al.	Proposed an ontology specifically designed to address energy efficiency within smart grid neighborhoods.
2024	Jenevain et al.	Proposes an ontology adressing the use of context in data exchanges on the energy sector.

Source: Created by the author (2025)

Table 2 – Key Contributions in RAG, Surveys, and LLMs

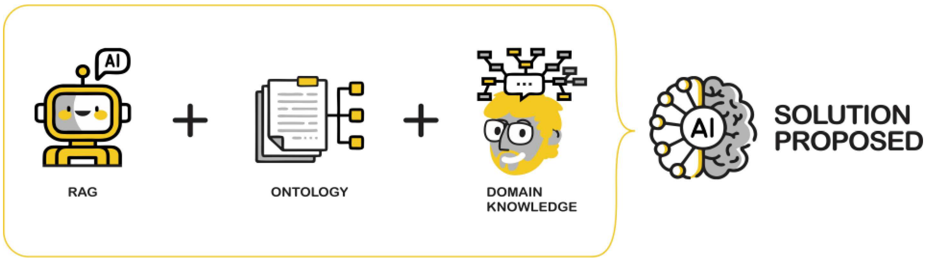
Year	Researcher(s)	Major Contributions
2017	Vaswani et al.	Presented the foundational ideas and concepts behind self-attention techniques for LLMs.
2020	Lewis et al.	Introduced the fundamental principles and concepts of RAG.
2022	Li et al.	Provided a survey of RAG techniques applicable to NLP and related fields, with a specific focus on methodologies and applications in machine translation.
2022	Tunstall et al.	Introduced Hugging Face, the Transformers library, and real-world applications of RAG with LLMs.
2023	Touvron et al.	Introduced Llama 2, a collection of pretrained and fine-tuned LLMs with varying sizes, ranging from 7 billion to 70 billion parameters.
2023	Zhao et al.	Provided an overview of LLM functionalities, their inner workings and relationship with RAG for enhanced text generation capabilities.
2024	Jenevain et al.	Proposes the combined use of NLP and ontologies to embed meaning and context into data exchanged within the energy sector.

Source: Created by the author (2025)

Importantly, as discussed in Section 3, this work introduces a groundbreaking approach within the energy sector by merging ontologies and RAG techniques to improve data interoperability, ensuring meaningful communication and mitigating the risk of misunderstandings. Figure 3 presents a visual representation of how ontologies, LLMs, and RAG intersect with the established knowledge base within the energy sector.

This convergence establishes the fundamental knowledge base upon which the proposed solution is built. The diagram visually reinforces the concept that the solution merges ontologies, LLMs, and RAG with the existing knowledge within the energy sector, resulting in the foundation of knowledge of what is being proposed.

Figure 3 - Integration of ontologies, LLMs, and RAG in the energy sector



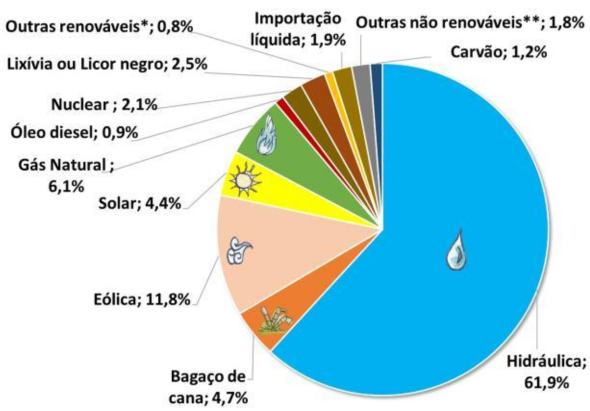
Source: Created by the author (2025)

2.1.4 Brazilian Energy Scenario

In 2024, more than 80 percent of Brazil's energy comes from renewable sources, with hydroelectric power being the dominant contributor (Agência Nacional de Energia Elétrica, 2024). However, government policies are directed towards further expanding the proportion of renewable energy in the nation's energy mix, moving away from non-renewable sources such as coal and fossil fuels (Souza et al., 2023).

Figure 4, developed by Brazil's Energy Research Company (2024), provides a visual representation of the distribution of energy generated by source in 2022, highlighting the ongoing expansion in energy production.

Figure 4 - Distribution of energy generated by source in Brazil (2022)



Source: Adapted from Brazil's Energy Research Company (2024)

ANEEL anticipates a continued increase in the proportion of renewable energy sources in the years ahead. To facilitate this, they are actively considering or already implementing various initiatives, such as subsidies, credit lines for new installations, direct sales, and compensation mechanisms for small-scale producers.

Brazil has also made significant progress in the areas of energy markets and auctions. However, challenges such as a lack of skilled labor, technological obstacles, and the need for further incentives can hinder more rapid growth in the renewable energy sector (Souza et al., 2023).

The transition to renewable energy is strongly dependent on smart grids and the seamless exchange of data. For instance, Distributed Energy Resources (DERs), like homes and small businesses, leverage interoperability to share data with the grid (Jenevain et al., 2024). Given Brazil's vast size and diverse economic and social landscapes (Silva; Bezerra, 2023), it is logical to conclude that the challenges related to miscommunication and misunderstandings, as previously discussed, are also applicable to the Brazilian context.

Heterogeneous systems, in particular, are heavily reliant on interoperability (Teixeira et al. 2018), while also being susceptible to miscommunications and misunderstandings due to the different contexts in which each participant operates. This challenge also affects the shift from non-renewable to renewable sources, as effective communication between diverse energy producers and grid operators is crucial.

2.2 RAG

This section describes the essential workings of the RAG framework, concentrating on the most frequently employed techniques identified in relevant research, as well as those crucial for constructing the proposed solution. To enhance clarity, the discussion is split into two parts, elaborating on the retrieval and generation phases.

Fundamentally, the RAG workflow comprises two separate stages: retrieval and generation. During the retrieval stage, the model actively seeks answers that align with the user's query and specific criteria, utilizing diverse sources like indexed documents or the web. The results are then forwarded to the LLM, which is responsible for the generation stage. By harnessing both its pre-existing knowledge and the newly retrieved data, the LLM constructs a comprehensive and customized response specifically addressing the user's unique question as noted by Zhao et. al (2024), Martineau (2024) and Wang et al. (2024).

The following subsections will delve deeper into each part of this process, highlighting how it can be used to augment ontologies and deepen understanding of interoperability within the energy sector.

2.2.1 Retrieval

The retriever's primary function is to locate and retrieve data pertinent to the user's query. It operates by systematically comparing keys and values, pinpointing the top-k most similar keys using a predefined similarity function, and subsequently extracting the corresponding values.

These models can be classified as dense or sparse, along with other categories, depending on their specific internal workings. Zhao et al. (2018) provides a concise overview of these two retrieval methods:

- Sparse Retrieval: A word-based retrieval approach, primarily utilized for textual data, that leverages techniques like term-frequency inverse document frequency (TF-IDF), query likelihood, and BM25 for efficient search within document collections.
- Dense Retrieval: Involves embedding queries and external knowledge into vector spaces, facilitating faster searches across various data types (text, code, audio, image, video) using ANN indexing and methods like DPR.

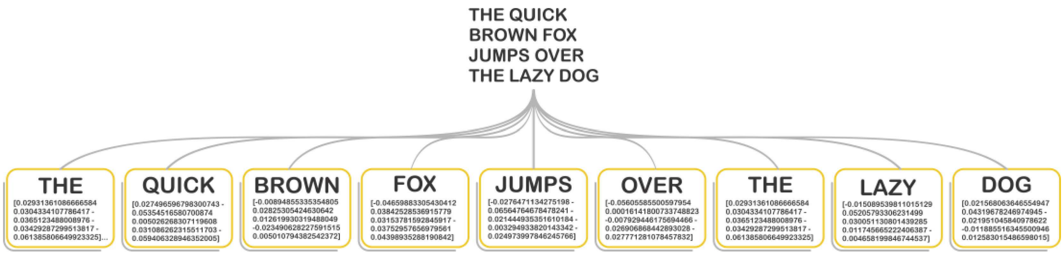
The retrieval model employs two key techniques to handle queries and text for comparison and retrieval: chunking and embedding as shown in Zhao et al. (2018), Merrit (2024), Wang et al. (2024). Essentially, these techniques can be described as follows:

Chunking is a preprocessing technique where large texts are divided into smaller, more manageable units, or chunks. These chunks can be sentences, paragraphs or custom-defined sections based on the specific use case. The primary purpose of chunking is to make large documents easier to process by machine learning models like LLMs, which often have input size constraints. Additionally, chunking improves contextual relevance by allowing models to focus on smaller, potentially more relevant portions of text when responding to queries.

Embedding, in contrast, is a transformation process where text chunks are converted into numerical representations known as embeddings. These embeddings capture the semantic meaning and relationships between words and phrases within the text. It serves two main goals: facilitating efficient semantic search by enabling similarity comparisons between embedded queries and documents; and allowing embedded chunks to be used as input to various machine learning models for tasks like classification, clustering, or question answering.

Figure 5 provides a visual demonstration of the chunking and embedding process applied to an illustrative phrase, showcasing how it would be prepared for input to the retriever and subsequently to the LLM. This demonstration utilizes LangChain (Langchain, 2024), a powerful framework for RAG implementation, and Hugging Face (Hugging Face, 2024), a prominent platform for LLMs. The corresponding code snippet can be found in the appendix.

Figure 5 - Sentence broken into numerical representations



Source: Created by the author (2025)

In this procedure, the sentence "The quick brown fox jumps over the lazy dog" is segmented into nine distinct chunks. Utilizing the capabilities of LangChain and Hugging Face, each of these chunks is then embedded and converted into a vector representation. These vectors will be subsequently employed to determine the most semantically meaningful and coherent combinations, ultimately culminating in the generation of a response to the user's initial inquiry.

This process can be metaphorically described as translating human language into a format that the machine can comprehend. It utilizes mathematical methods such as statistics and probability to infer meanings and establish relationships between topics.

Furthermore, the retrieval step can be invaluable in keeping models up to date. The ability to incorporate external data, not previously learned by the model, allows it to work with current information, both general and specific, without the immediate need for further training or fine-tuning.

2.2.2 Augmentation

The information gathered during the retrieval phase is then combined with the original query, often through the utilization of attention mechanisms. This results in a more comprehensive and context-aware input for the language model. The model, such as GPT-3 (Brown et al., 2020), then generates a response based not only on the initial query but also on the retrieved context, ensuring the output is both informative and factually accurate. Advanced models like GPT-2 and GPT-3 (Redford et al., 2019) demonstrate the potential of language models to produce high-quality, human-like text.

The generation process itself comprises multiple steps. The retrieved context, consisting of relevant passages or documents, is first transformed into a format that the language model can process, often utilizing techniques like tokenization and conversion into numerical representations. This encoding frequently employs models such as BERT (Devlin, 2018). The encoded context is then merged with the original user query, sometimes through attention mechanisms (Vaswani et al., 2017), enabling the model to dynamically focus on the most relevant parts of the context.

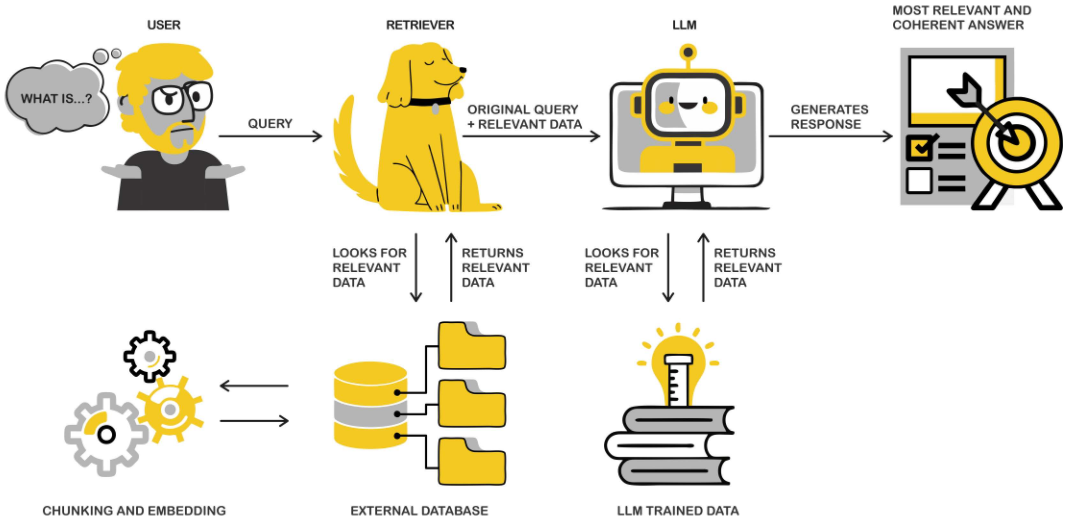
Given this fused input, the language model generates a response by iteratively predicting the subsequent word or token based on the preceding context, with the chosen decoding strategy influencing the quality of the response. Optionally, filtering or re-ranking techniques can be employed to further refine the generated responses, selecting the most relevant and coherent one.

This generation process can be fine-tuned to optimize both the retriever and generator for specific solution requirements. For instance, fine-tuning can be used to ensure that the generator effectively leverages the retrieved data or to ensure that the retriever selects only pertinent information. While this process can be challenging, the resulting improvements in performance and reliability typically justify the effort (Wang et al., 2018).

Fine-tuning, for instance, could be used to adapt existing models, trained on more general datasets, to be better suited for tasks within the energy sector, as is the case with the proposed solution. This added layer of specialization has the potential to yield even better results and assign meanings that are even more accurate to the context of both actors involved in the communication.

Figure 6 expands on the concept illustrated in Figure 2 by providing a more elaborate visual representation of this process. In this figure, the augmentation is carried out by the LLM, symbolized as a robot, upon receiving the enhanced query along with the documents furnished by the retriever, symbolized as a dog. The LLM then constructs a precise and pertinent answer by leveraging both its pre-existing knowledge base and the information acquired from the retriever.

Figure 6 - Text generation following retrieval workflow



Source: Created by the author (2025)

Figure 6 emphasizes the collaborative nature of the retriever and the generator in comprehending and providing the most complete and coherent answer. The retriever acts as a gatherer of pertinent information from external sources, while the generator utilizes this retrieved knowledge, alongside its own training data, to craft a response. This process also integrates self-attention mechanisms (Vaswani et al., 2017), wherein the model evaluates the importance of each word in connection to all other words in the input. This enables the model to weigh the significance of different words and capture intricate dependencies between them, resulting in a more nuanced grasp of the input and the generation of a more refined response. Regarding generation methods, Zhao et al. (2024) summarized the following:

- **Query-based RAG**: A prevalent technique in Retrieval-Augmented Generation where the user's query and retrieved information are merged to form the input for the text generation model. This approach has proven successful in various domains such as text generation, code generation, question answering, and image generation. By combining the query and retrieved context, the language model gains a better understanding of the user's intent and can produce more relevant and informative responses. This method is often favored for its adaptability and ease of implementation, particularly when utilizing pre-trained language models. However, careful prompt design is crucial to guide the model in effectively utilizing the retrieved data to generate accurate and coherent outputs.

- **Latent representation-based RAG**: A framework where retrieved information is incorporated into generative models as abstract representations, enhancing the model's comprehension and elevating the quality of generated content. This approach has found applications in various fields, including text generation, code generation, question answering, image generation, 3D modeling, audio processing, and video captioning.

- **Logit-based RAG**: A technique where the generation model integrates retrieved information during the decoding process by combining probabilities from both the language model and the retrieval process. This method has been applied in various fields like text, code, and image generation. For example, in text generation, it leverages similar prefixes from local databases to enhance performance, especially for rare words and phrases. In code generation, similar concepts are employed to improve output control and enhance code summarization quality. Image captioning benefits from this approach to generate more accurate descriptions. Overall, logit-based RAG utilizes historical data to infer current states and is particularly suitable for sequence generation tasks.

- **Speculative RAG**: A method within Retrieval-Augmented Generation that prioritizes the use of retrieved information over generating new text, aiming to conserve resources and enhance response speed. This is accomplished by substituting text generation with the retrieval of existing content whenever possible. For instance, instead of generating entire responses from scratch, the system might retrieve relevant phrases or sentences from a knowledge base and incorporate them into the final output. This approach is particularly advantageous when dealing with sequential data and can leverage pre-trained models as components, offering flexibility in how retrieved content is utilized.

Each method possesses unique characteristics, making them better suited to specific situations. For instance, Speculative RAG, within the context of the proposed solution, could prove valuable in scenarios where performance is paramount or computational resources are limited. Additionally, the emphasis on utilizing relevant sentences and phrases from the retriever could be beneficial when working with dictionaries or other data that would be advantageous to append directly to the ontology, without further augmentation or generation.

It is worth noting that the field of AI generation is expansive and multifaceted, reaching far beyond the boundaries of this conversation. Research publications like those by Lewis et al.

(2020) and Vaswani et al. (2017), among others, delve more deeply into the complexities of the technology underpinning each phase of RAG, from the initial user query to the final response generation. These studies offer valuable perspectives and insights into the subtleties and challenges involved in crafting effective and efficient retrieval-augmented generation systems.

2.3 RAG IN THE ENERGY SECTOR

While RAG has gained widespread adoption in well-known applications such as chatbots (like ChatGPT and Google Bard) and AI copilots (Torrent et al., 2023), its potential within the energy sector remains largely untapped. A significant portion of machine learning projects in this domain focus on predicting and generating numerical values, as opposed to text. Nevertheless, this doesn't imply that RAG is irrelevant to the energy industry.

Researchers such as Fortuna et al. (2024), for example, suggest utilizing RAG to enhance household energy consumption analysis. They propose leveraging data retrieved from official documents and other relevant sources, such as external databases.

However, this shouldn't be interpreted as a limitation of RAG's applicability or potential to advance the energy field. Quite the contrary, as its capabilities in text augmentation and generation, coupled with image and video processing, open a vast array of potential applications and solutions. This section will delve deeper into some of these possibilities, of which the solution proposed in this work is just one instance.

2.3.1 Multimodal RAG

Multimodal Retrieval Augmented Generation (MM-RAG) represents a variant of RAG where the data supplied to the model during retrieval or used for training the generation model encompasses text along with other modalities such as images or videos a (Chen et al., 2022).

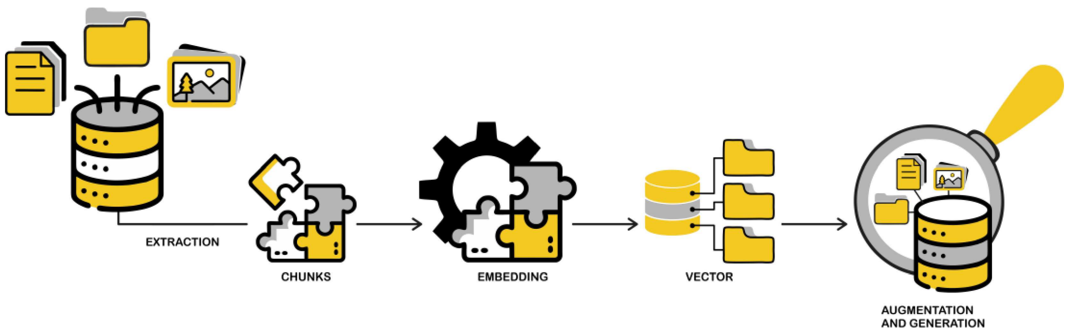
Chen et al. (2022) put forth a RAG-based solution in which a model received input comprising both text and images and could accurately answer questions about the content of those images. For example, the model successfully determined whether trees in a waterpark were taller than the waterslides. This was accomplished by combining multimodal transformers for retrieval, along with question answering and RAG-based augmentation techniques (Chen et al., 2022).

One illustrative example is the work by Zhou et al. (2024), proposing an innovative solution called Img2Loc. This approach harnesses RAG and a multimodal LLM to enhance image geolocation. By framing the problem as a text generation task and utilizing RAG techniques, their solution achieved superior performance using prompts and improved image generation.

Another potential application lies in the field of image captioning, specifically where generating accurate and informative captions is paramount. Sarto et al. (2022), for instance, suggest employing MM-RAG in conjunction with k-nearest neighbors to predict the most suitable caption for a given image, leveraging external knowledge and previously known data.

Figure 7 provides a visual representation of the process, emphasizing the stages of data input, chunking, and embedding. The sequence begins with the combination of images, text, and other data to generate chunks. This aggregated data is then transformed into vectors via embedding and subsequently passed on to the generator for further processing.

Figure 7 - MM-RAG workflow



Source: Created by the author (2025)

The rightmost image visually represents the initial stage of the augmentation process, where the LLM receives data passed from the retrieval process to generate responses. With the incorporation of image data, it can perform functions similar to those proposed by the authors mentioned in this section, such as image captioning, enhancing geolocation accuracy, and answering questions based on image content.

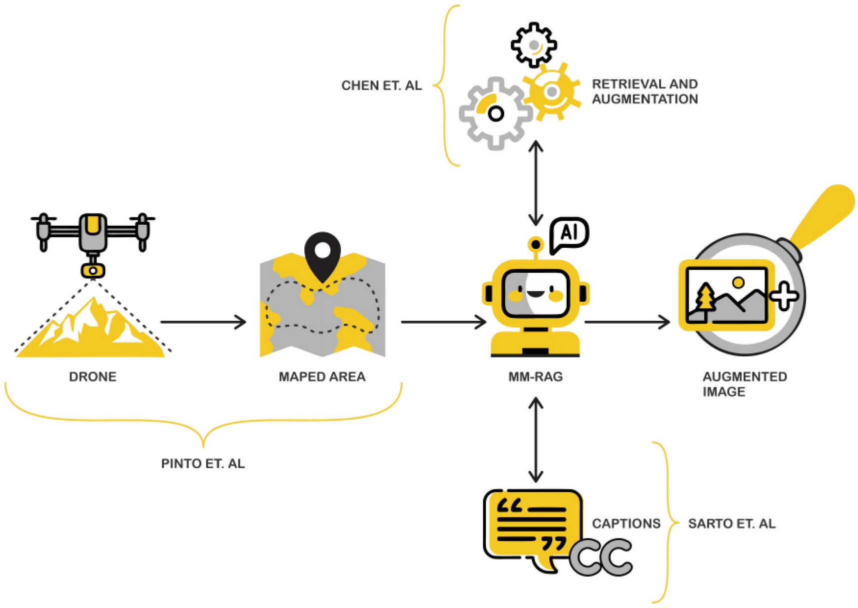
As previously noted, MM-RAG can be a potent tool for image manipulation, especially in tasks like question answering and captioning. Such solutions could prove immensely beneficial to the energy sector in scenarios where valuable data can be extracted from images obtained from diverse sources.

A case in point is the work of Pinto et al. (2024), who proposed a solution aimed at mapping unexplored terrain using drone-captured images. The resulting maps could then be used to improve the efficiency of unmanned aerial vehicles (UAVs) and to identify and map boundaries. One way to incorporate MM-RAG into this process would be at the data extraction phase, assisting with the mapping process and enhancing the results. By integrating the question-answering logic proposed by Chen et al. (2022) at the conclusion of the process, a level of understanding and manipulation not achievable with the original solution alone could be attained.

In a similar vein, captions generated by RAG could be employed to clarify and enhance images or other data obtained in this process. Beyond just providing additional training data, the strategic use of clear and concise captions could greatly aid human comprehension and analysis for those who require it.

Figure 8 provides a visual representation of the scenario, where images captured by a drone are enriched using an MM-RAG solution. The outcome is a collection of images that contain richer, more easily understandable information.

Figure 8 - MM-RAG enhancing UAV image data



Source: Created by the author (2025)

Likewise, additional image-based solutions, whether derived from drones or other sources, can be developed to streamline and enhance existing practices within the energy sector. By harnessing the capabilities of MM-RAG, novel opportunities emerge to propel advancements in both fields.

The utilization of MM-RAG as described in this section presents a promising avenue for the proposed solution. A model specifically tailored to the intricacies of the energy sector, coupled with drone imagery, could generate valuable knowledge and insights. This combination has the potential to enhance image analysis and provide deeper understanding in various energy-related applications.

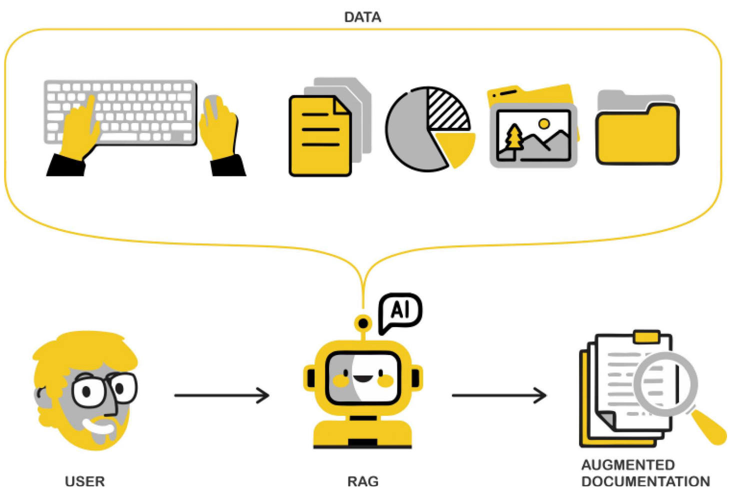
2.3.2 Automated Reporting and Documentation

Clear and effective documentation is crucial in any organization, regardless of the field, especially where sensitive data and complex processes are part of daily operations. In fact, documentation that is confusing or poorly written can lead to issues and inefficiencies, potentially causing costly errors and contributing to suboptimal performance (Atlassian, 2024).

RAG's ability to generate insightful data from various sources, both internal and external to the LLM, can be invaluable in situations where concise and precise reports or documentation are required, regardless of the specific domain. As shown in Figures 6 and 7, the retrieval process can provide the RAG application with diverse data types for various processes, such as answering questions based on the information available.

Therefore, innovative applications of RAG capabilities, like MM-RAG, have the potential to lead to significant improvements in documentation generation and overall comprehension of presented data. Figure 9 visually depicts this process, where diverse data is fed to the retriever, which, in conjunction with the LLM, can respond to questions based on the documentation provided. In this scenario, a user can leverage RAG capabilities to query all input documentation and receive answers based on the content. Additionally, RAG can generate new text and content based on the combined knowledge derived from the retrieval process and the LLM's pre-trained data.

Figure 9 - Usage of RAG to augment documentation



Source: Created by the author (2025)

Similar applications have been explored in other fields. For instance, Yepes et al. (2024) proposed using RAG, focusing on chunking techniques, to enhance financial reports. Unlu et al. (2024) suggested that RAG could improve the performance of clinical trial screening by identifying and reporting on inclusion and exclusion criteria.

Furthermore, Xu et al. (2024) proposed leveraging RAG to streamline customer support processes by thoroughly analyzing past ticket data, integrating it with knowledge graphs, and thereby assisting technical support teams.

3 MAIN CONTRIBUTION

The energy sector, with its intricate nature and complex network of stakeholders, is open to a broad range of interpretations and perspectives. This diversity is notably pronounced in situations like ENTSO-E and its affiliates, where data, despite having consistent definitions, can be understood in contrasting ways depending on the specific context and characteristics of each party involved. Strategies aimed at facilitating clear communication, customized to an actor's unique circumstances, hold great promise in mitigating the issues that arise from misunderstandings and misinterpretations.

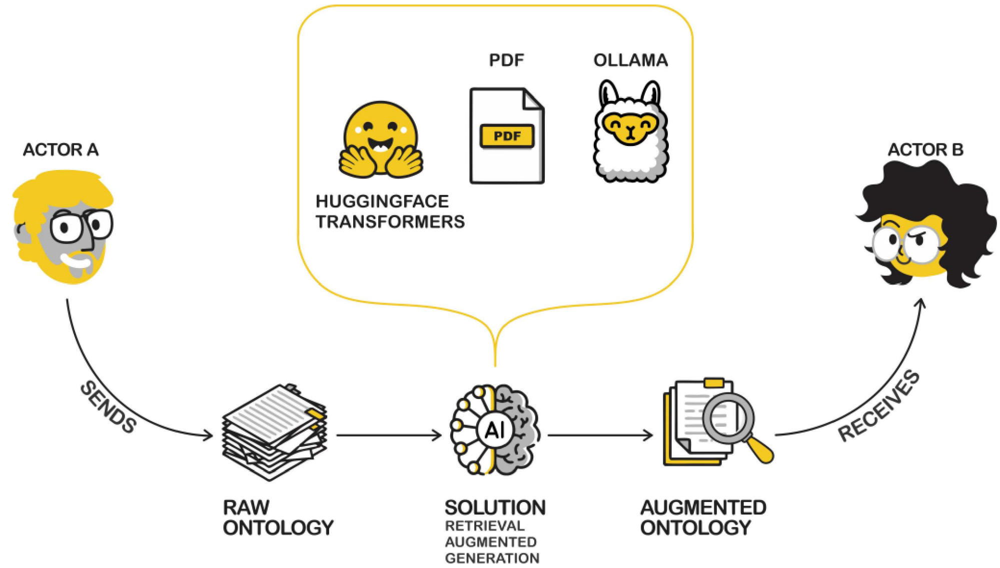
Motivated by this reality, the principal contribution of this research lies in minimizing the probability of miscommunications and their resulting complexities. The overarching objective of this study is to establish the groundwork for a comprehensive solution capable of effectively addressing the effects and challenges stemming from contextual disparities among players in the energy sector.

The key achievements of this work encompass the following:

- Presenting a fundamental solution that harnesses the capabilities of OwlReady2, LLMs, RAG, and Python to skillfully manipulate ontologies.
- Devising a methodology to attribute context-dependent meanings to terms by capitalizing on insights extracted from ontology manipulation.
- Laying the initial steps toward a solution designed to manage data interchange between diverse actors, thereby reducing the risks of miscommunication.
- Pioneering the integration of ontology, RAG, and LLM concepts and techniques to bridge the divide between structured data and contextual comprehension.

Figure 10 offers a visual illustration of this concept, depicting two actors exchanging data with the proposed solution acting as a middleware, dynamically adapting the meanings of terms based on the sender's context. Here, two actors are shown operating within different contexts. They successfully communicate with each other, achieving mutual understanding through the meanings assigned by the proposed solution.

Figure 10 - Communication with contextual alignment



Source: Created by the author (2025)

Within this framework, the ontology, RAG, and the LLM collaborate to ensure that any data conveyed by the actors on the left is interpreted by the receiver following the sender's intentions. This mechanism is reciprocal, guaranteeing accurate understanding when the data is returned to the original sender, who then assumes the role of receiver.

4 STUDY CASE

4.1 SCENARIO

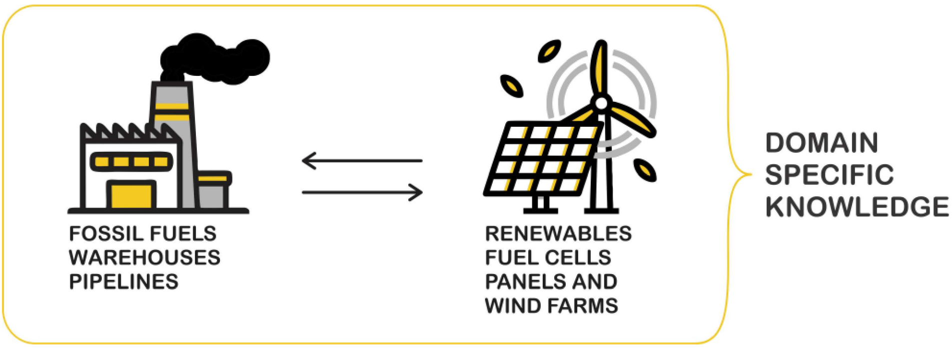
It is worth highlighting the escalating need to transition from non-renewable energy sources, like fossil fuels such as oil and gas, towards renewable alternatives. Data from the European Environment Agency (EEA) (2024) reveals that many European Union countries have set targets to increase their renewable energy production, even though production levels differ among nations. Other organizations, like the United Nations (UN) (2024), also stress the significance of such a substantial change in energy sources and production for the future.

In Germany, there's an ongoing effort to shift the country's energy mix away from fossil fuels and towards more sustainable renewable sources as seen in initiatives such as those shown in "A future based on renewable energy" (2023), "Without renewables, there can be no future" (2023) and "Energy Transition – A Project for Generations" (2024) focused on the reality and context of that country. This transition is partly driven by the limited domestic availability of fossil fuels. To accomplish this, efforts are focused not only on modernizing the grid and associated infrastructure but also on fostering collaboration between fossil-based and renewable energy producers. This collaboration is critical to accelerate the process and guarantee a seamless transition without outages or disruptions. Moreover, integrating small-scale producers into the grid via smart meters and data exchange will be essential for the success of this energy transformation.

EEA data further underscores a rising need for continuous information exchange between various players in the energy sector, including large and small-scale producers, as well as entities across different countries. Such a scenario, involving data exchange between actors in diverse contexts and with varying scales and sources of energy production, provides an ideal use case for solutions like the one proposed in this work. This solution aims to enhance understanding between actors and mitigate miscommunications and problems stemming from differences in context and realities.

Therefore, this case study will be grounded in a scenario akin to the one outlined for Germany, where fossil fuel and renewable energy producers need to exchange data. Figure 11 visually represents this by showcasing an example of each actor, emphasizing three key points of differentiation between them. In this illustration, the actors on each side need to exchange data, despite operating in significantly different domains with distinct contexts. However, employing the innovative solution proposed in this work can help alleviate the challenges stemming from these disparities.

Figure 11 - Example of energy sector actors



Source: Created by the author (2025)

The subsequent subsections will provide a more in-depth exploration of how the proposed solution can be utilized within such a scenario, detailing its setup process and the necessary tools for its intended functionality.

4.2 ONTOLOGY DESCRIPTION

This segment will outline the data leveraged by the proposed solution to address the challenges encountered in scenarios where actors operating in vastly different contexts need to exchange data effectively.

The ontology put forth in this study seeks to capture the relationships between actors, their operational context, and how this context influences their understanding of shared data. Figure 12 provides a visual representation of this concept, positioning the actor at the core of their context, influenced by various factors such as prior knowledge, culture, language, and more.

Figure 12 – An actor and his context

Source: Created by the author (2025)

As Santos et al (2012) suggested, context can be defined as a collection of relevant conditions and influences that shape a distinct and complete understanding of a particular situation. The ontology was constructed using Stanford University's Protégé software (Musen et al., 2015) a tool designed for building ontologies that provides features such as inferencing and relationship modeling. The ontology was deliberately designed to be broadly applicable, facilitating easier integration with other solutions as add-ons or plugins, like the OEO. However, future adaptations can be made to align it with specific standards if required.

The tables that follow offer a comprehensive explanation of the most crucial parts of the ontology structure, including attributes, their relationships, and pertinent aspects of the proposed work. The complete ontology can be accessed on GitHub (Energy Context Ontology, 2024). Table 3 presents the entities associated with Actors, and how they are represented and understood within both the proposed solution and the ontology. It's worth noting that the proposed ontology is a product of the research's overall evolution, building upon the work presented in Jenevain et al. (2024), with modifications made to better integrate and accommodate the use of RAG.

Table 3 – Entities related to the Actor

Name	**Represents**
actorRole	The role of the actor in the data exchange.
receiver	Actor receiving the exchanged data.
sender	An actor sending the exchanged data.
distributedEnergyResource	A type representing Distributed energy resources, such as houses with solar panels.
organization	Represents an Organization.
energySector	Represents the energy sector.
industrySector	Represents the industry sector.

Source: Created by the author (2025)

The purpose of these entities is to express the various types of actors involved in the data exchange process, either as a sender or a receiver. They capture context by defining the necessary information to infer the context of the actor, such as its role or the industry in which it operates.

Table 4 presents the entities related to terms and contexts, and how they are defined and understood within both the proposed solution and the ontology. These entities demonstrate the connection between contexts and terms, underscoring how the meaning of a term can be shaped by the context of both the sender and the receiver of the information.

Table 4 – Entities related to Terms and Contexts

Name	Represents
context	Represents a context in which an actor is inserted, or a term should be understood
term	The term being exchanged
language	The language in which the term should be understood
meaning	The default meaning of a term
byContext	The meaning of a term determined by the sender’s context
termContent	How the term should be written.

Source: Created by the author (2025)

Table 5 outlines the object properties, showcasing the relationships and interactions between the entities previously introduced. Similar to entities, information will be provided on the most central object properties necessary to understand how actors, terms, and contexts relate to each other. The full ontology, which can be found on GitHub, provides a complete list of these properties.

Table 5 – Object properties linking actors, contexts, and terms

Name	Represents
actorHasContext	Links an actor to its corresponding contex
actorHasRole	The term being exchanged.
contextDefinedbyActorType	Indicates that the context of a term or actor is defined by its type.
contextDefinedbyActorRole	Indicates that the context of a term or actor is defined by its role.
termHasMeaning byContext	Indicates that the meaning of a term is defined by a particular context.
termHasMeaningbyLiteralDefinition	Indicates that the meaning of a term is defined by its default definition.

termSentBy	Indicates the actor assigned as the sender of a term.

Source: Created by the author (2025)

As for inferences, the ontology works with two, which can be defined through the following *subproperties of chain*:

- **exchangedDataHasSender o actorHasContext**:
  results in *exchangedDataHasContext*, defining that the context of a given term(data) is dependent on the context of its sender.
- **exchangedDataHasTerm o exchangedDataHasContext**:
  results in *termHasMeaningByContext*, defining that the meaning of a given term is defined by the context of the term being exchanged and the context of the data, as defined by the previous *subproperties of chain*.

These properties illustrate the interconnectedness between the fundamental elements of the proposed solution—actors, contexts, and terms—and how they mutually define each other. In this specific case, an actor's context, particularly their role as a sender or receiver, directly impacts the interpretation of a term. As noted, the subtleties of reality, culture, and context can lead to misunderstandings and variations in meaning, even for the most commonly used terms.

4.3 HARDWARE AND SOFTWARE SETUP

The proposed solution was implemented on a standard personal computer, utilizing a more lightweight model and a single document to enhance generation through retrieval. However, this does not suggest that the reduced specifications made it less effective. In fact, it indicates that even less resource-intensive machines can successfully run the proposed solution, or similar approaches, and yield satisfactory outcomes.

The computer used for the development and testing of the proposed solution was equipped with an NVIDIA GeForce RTX 3050 with 6 GB GDDR6, a 1 TB M.2 PCIe NVMe Solid State Drive, a 13th Gen Intel Core I5 processor, and 24GB of RAM.

The proposed solution was implemented in Python using Visual Studio Code, and it leveraged the following libraries: os, pandas, numpy, owlready2, sklearn, langchain community, langchain text splitters, langchain core, rouge, nltk, and sentence transformers. These libraries support the entire RAG process, from reading documents used in retrieval, as well as chunking and embedding, (langchain) to manipulating the ontology (owlready2) and managing all data-related processes (pandas). All libraries are readily available and can be installed using commands like 'pip' within Visual Studio Code or environments like Anaconda.

All models needed for the proposed application require the installation of Ollama, which provides LLMs locally. The installation process is straightforward and can be found on the downloads section of the official website.

The readme file on the GitHub page for the proposed solution provides all the necessary installation and pull instructions needed for the process to function correctly. This includes information on langchain, its dependencies related to PDF reading, and the retrieval and generation processes.

The 'nomic-embed-text' model was selected for the embedding process due to its efficiency and capacity to manage both extensive and concise contexts while accommodating substantial text lengths. It's worth highlighting that its role is exclusively limited to embedding. The LLM employed for the generation process is Mistral 7b , chosen for its robustness, comprehensiveness, and its aptitude for question-answering and text generation tasks.

The PDF file utilized in the retrieval process for testing purposes is the glossary (Appendix B) available on the United States Environmental Protection Agency's "Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans" page. This document was selected because of its paired words and definitions, manageable size, and direct connection to the energy sector, with a particular focus on renewables. While Mistral and nomic-embed-text are both capable of handling files of all sizes, the one provided with the proposed solution is sufficient for testing and attaining satisfactory results.

As demonstrated in the code structure available on the proposed solution's GitHub repository, the file is accessible to the solution for chunking, embedding, and incorporation into the context provided to the LLM. The resulting knowledge is then utilized to deduce the meaning of terms specified by the ontology.

The prompt used is created by joining the terms with the context inferred by the ontology. In this case, it is passed by the code to the LLM in the format: *'What is {term} in the context of {context}?'*. The answer template is framed as *"You are an AI assistant. Using only your internal knowledge, choose the best answer, and only that, to the following question:{question}"*

5 EVALUATION RESULTS

Although LLM evaluation largely relies on vectors and statistics, it's predominantly conducted by comparing sets of questions and answers - both the expected responses and those generated by the model (Kamalloo et al., 2023). This can be done in multiple ways, such as counting the number of matching words or phrases (n-grams) between the expected answer and the model's output. One such evaluation model, known as ROUGE, operates in this manner. BLEU is another similar example.

An alternative approach to evaluating LLMs and their answers is to utilize another LLM. Langchain , for example, suggests using a variant of GPT to generate answers and comparing them against those of the first LLM (in this case, the proposed solution) to assess their similarity. This comparison generates a mean score that provides insights into the model's effectiveness. This strategy can also be used to provide data for ROUGE and BLEU metrics, whether evaluating n-grams or longer sentences. This can be achieved by running both LLMs against a collection of questions and answers derived from the same dataset as noted by Rasool et al. (2023) and Alinejad et al. (2024).

Kamaloo et al. (2023) proposes a third approach that involves using human evaluators to review and assess all answers. This entails preparing a set of questions based on a dataset and specifying the answers that the human reviewers expect to see. While this process might be less cost-efficient and more time-consuming, results indicate that human judgment can boost model performance by at least 28%.

5.1 QUESTIONS AND ANSWERS

The expected responses will be generated based on the PDF file referenced in section 4.3 and the general guidelines provided by Torrent et al. (2024), with the aim of producing the most specific answers to the questions. In this scenario, five questions will relate to the document, while another five will pertain to terms utilized within the proposed ontology and its data.

Table 6 presents the collection of questions to be processed by the model. The generated answers will then be compared against the gold standard established by humans, using ROUGE to evaluate the likelihood of the answers being contextually accurate. Some questions will be based on the document's content, while others will focus on the terms exchanged within the proposed ontology and its associated data.

Table 6 – Questions to be answered by the model

Question Number	**Query**
1	What is storage in the context of solar energy?
2	What is carbon in the context of renewable energies?
3	What is metering points in the context of renewable energies?
4	What is the relationship between Megawatt (MW) and Megawatt-hour (MWh)?
5	Who prepared the Annual Energy Outlook?
6	What is the Clean Air Act?
7	What is an Electric Generating Unit?
8	How Electricity Dispatch Models optimize the dispatch of a system?
9	Who maintains the Emissions & Generation Resource Integrated Database?
10	Give me examples of Energy Efficiency
11	What are ISOs?
12	What are Power Pools?

Source: Created by the author (2025)

Table 7 presents a sample pair of answers: the answer produced by the model and the expected or gold standard response. A careful examination reveals that while the answers may vary in length, the context and pertinent information remain consistent. This discrepancy in length can contribute to variations in the scores, highlighting the importance of combining human evaluation with automated metrics. Additional questions and answers can be found on the proposed solution's GitHub repository.

Table 7 – Example of generated and reference answers

Generated Answer	**Reference Answer**
In the context of solar energy, storage refers to the technology or system used to store excess solar power generated by photovoltaic (PV) panels for later use. This stored energy can be utilized during periods when solar generation is not available, such as nighttime or during cloudy weather conditions. Storage systems can	Storage in the context of solar energy refers to technologies or systems that can store excess electricity generated by solar panels for later use, especially

include batteries, capacitors, or other energy storage devices that capture and save the energy produced by solar panels, allowing it to be used when needed, thereby increasing the overall efficiency of the solar power system.	during periods of low or no sunlight.

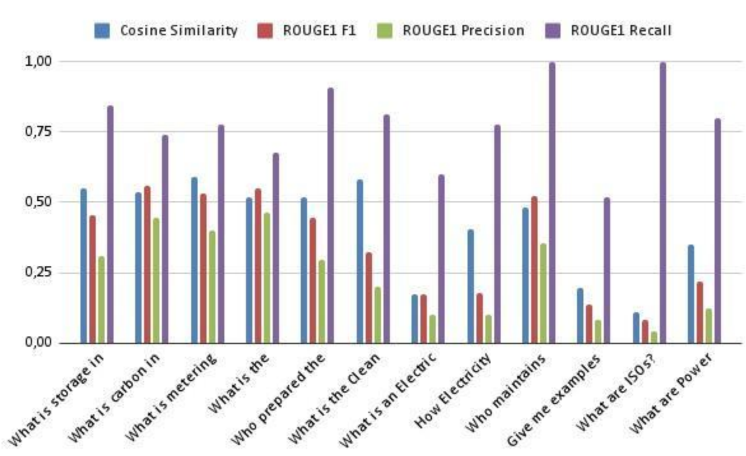
Source: Created by the author (2025)

5.2 RESULTS

The results underscore a key challenge in LLM evaluation and scoring: the output produced by the models can vary depending on factors such as how the prompt is structured and the specific model chosen. Figures 5.0.0.2, 5.0.0.2, and 5.0.0.2 illustrate the results for ROUGE-1, ROUGE-2, and ROUGE-L, respectively, all using cosine similarity, precision, recall, and F1 scores for the OLLAMA model Mistral Orca. Additional results and tests, conducted with the base Mistral model and Mini Orca, can be found on the proposed solution's GitHub repository.

The results, especially those from ROUGE-1, suggest that the model demonstrates proficiency in generating contextually appropriate responses to the terms and questions posed, aligning well with the sender's intended meaning. As shown in Table 7, the structure of the prompt can influence the generated answer, impacting the matches used for scoring. Cosine similarity indicates a strong contextual relationship between the generated and reference answers.

Figure R1 - ROUGE-1 results



Source: Created by the author (2025)

5.3 ABLATION STUDY

The ablation study for the proposed solution will focus on evaluating the model's performance when it is given terms from the ontology without additional context or supporting information that could help clarify the role or nature of the actor involved. The study will also emphasize terms that are prone to misinterpretation, particularly those that are not exclusive to, or commonly used within, the energy sector.

The first step will involve modifying the prompt given to the model to prevent it from using either the contextual information provided by the ontology or the content retrieved from the PDF via RAG. In practice, this removes the influence of the proposed ontology, forcing the model to operate in a simplified manner, relying solely on the isolated terms and selecting the most probable or readily available meanings, without any additional contextual cues. Both are created based on the prompt defined in Section 4.3. Table 7 shows each one:

Table 8 – Questions to be answered with and without context

Question	**Ontology Answer**	**Simple Answer**
What is storage?	Storage in the context of energy and sustainability refers to technologies that can capture and store excess energy generated by renewable sources for later use, when demand is high or suTepply is low. Examples include batteries, pumped hydro storage, thermal storage, and flywheels. The purpose of energy storage is to ensure a consistent and reliable power supply, improve grid stability, and facilitate the integration of variable renewable energy sources into the power grid.	Storage, in the context of computers and information technology, refers to a place or device where digital data is saved, accessed, managed, and protected for future use. This can include hard drives, solid-state drives (SSDs), external drives, cloud storage services, and even tapes. The purpose of storage is to provide a persistent location for data so that it can be retrieved when needed.
What is carbon?	Carbon is a chemical element with symbol C and atomic number 6. It	Carbon is a chemical element with the symbol C and

	belongs to Group 14 (IVa) of the periodic table and is non-metallic and tetravalent making four electrons available to form covalent chemical bonds. Carbon exists in three allotropes, including diamond, graphite, and fullerene. In the context of renewable energy and sustainability, carbon often refers to carbon dioxide (CO2), a greenhouse gas that contributes significantly to global warming when emitted into the atmosphere from burning fossil fuels or deforestation. Reducing carbon emissions is crucial for mitigating climate change and transitioning towards more sustainable energy systems.	atomic number 6. It is non-metal and belongs to group 14 (IV A) of the periodic table. Carbon is unique among the elements in that it can form four covalent bonds, making it tetrahedral in structure. This property, combined with its relative abundance and versatility, allows carbon to be found in a vast array of organic and inorganic compounds.
What are metering points?	Metering points, in the context of energy and power grids, refer to specific locations within a grid where electricity consumption or production is measured. These measurement points can be found at various levels within the electrical infrastructure, such as customer premises (for measuring household or business usage), distribution networks (to monitor energy flow and loss), and generation sites (to assess output from renewable energy sources).	Metering points, in the context of energy and utility systems, refer to specific locations where measurements are taken for electricity, water, or gas consumption. These points allow for the monitoring and billing of usage, ensuring that consumers pay for the resources they have actually consumed, and can take various forms depending on

		the type of resource being measured.
Give me examples of Energy Efficiency	Energy efficiency refers to using less energy to perform the same task or achieving the same results with less energy consumption. Here are some examples of energy efficiency in various sectors: 1. Buildings: Insulating homes and buildings, using energy-efficient appliances, LED lighting, and smart thermostats can significantly reduce energy consumption. 2. Transportation: Carpooling, public transportation, electric or hybrid vehicles, bicycling, and walking are all ways to improve energy efficiency in transportation. 3. Industry: Implementing energy-efficient machinery, optimizing manufacturing processes, recycling waste materials, and using renewable energy sources can help industries reduce their energy consumption. 4. Power Generation: Combined Heat and Power (CHP) systems, which generate electricity and useful heat from a single fuel source, are more energy-efficient than separate power and heating systems.	1. Insulation: Proper insulation in buildings can significantly reduce energy consumption for heating and cooling, as it keeps the indoor temperature consistent and minimizes the need for HVAC systems to work harder. 2. LED Lighting: Compared to traditional incandescent bulbs, LED lights use less energy and last longer. They are more energy-efficient due to their lower wattage and higher lumen output per watt. 3. Solar Panels: Harnessing solar power through photovoltaic cells converts sunlight directly into electricity. This renewable energy source can help reduce the reliance on fossil fuels and lower overall energy costs. 4. Energy Star Appliances: Energy Star-certified appliances are designed to be more efficient than standard models. They consume less electricity, resulting in

	5. Agriculture: Precision farming techniques, such as soil mapping, variable rate technology, and drip irrigation, can minimize energy waste in agriculture. 6. Consumer Behavior: Turning off lights and electronics when not in use, using power strips to eliminate standby power consumption, and washing clothes in cold water can also contribute to energy efficiency.	reduced energy bills and a smaller carbon footprint. 5. Smart Thermostats: These devices learn user habits and adjust temperatures accordingly, minimizing wasted energy when the home is unoccupied. Some smart thermostats can even optimize temperatures to save energy while maintaining comfort levels.
What are ISOs?	In the context of energy, Independent System Operators (ISOs) or Regional Transmission Organizations (RTOs) are non-profit organizations that oversee the operation of regional power grids in North America. They ensure electricity reliability, manage the flow of electricity between utilities, and operate wholesale electricity markets to optimize efficiency and promote competition among different energy suppliers. ISOs are crucial for integrating renewable energy sources into the grid and balancing supply and demand across their jurisdictions.	In the context of technology, particularly in computing and finance, ISO refers to different things: 1. In computing, an ISO file is an archive file format that contains data as if it were burnt onto an optical disc (like a CD or DVD). It's often used for distributing software, operating systems, and drivers. 2. In finance, an Initial Public Offering (IPO) or stock offering, is a type of public offering where shares of a company are sold to institutional investors and usually also retail (individual) investors. This

		takes a privately held company and makes its stock available to the public on a securities exchange for the first time. 3. ISO can also refer to International Organization for Standardization, which is an independent, non-governmental international organization that develops and publishes internationally recognized standards for a wide variety of industries all over the world.

Source: Created by the author (2025)

The results show significant variation, particularly for terms that are used across multiple domains or that are acronyms. For example, the term *storage*, when provided with context, correctly refers to methods of storing energy produced from various sources. Without context, however, the model interprets it as data storage. Similarly, *ISO* in the energy sector refers to *Independent System Operators*, but without context, the model tends to associate it with image file formats or similar unrelated meanings.

Table 8 presents the *golden answers* for each of the sample questions, representing the expected responses used to compute the ROUGE scores and cosine similarity metrics. Table 9 displays the actual results. For reference, columns labeled *no* correspond to answers generated without context, while columns labeled *with* refer to answers generated with contextual information.

Table 9 – Golden Answers for the Ablation Study

Question	**Golden Answer**
What is storage?	Storage refers to technologies that retain energy for later use, such as batteries or pumped hydro systems.
What is carbon?	Carbon is a chemical element that forms the basis of all life and is a key component in greenhouse gases like CO2.

What is metering points?	Metering points are locations where energy consumption or production is measured for billing and monitoring purposes.
Give me examples of Energy Efficiency	Examples include LED lighting, high-efficiency HVAC systems, insulation, and energy-efficient appliances
What are ISOs?	ISOs (Independent System Operators) manage the transmission grid and ensure the reliable delivery of electricity in a region.

Created by the author (2025)

Table 10 – Ablation Study Results

Question	**ROUGE-1 (with)**	**ROUGE-1 (no)**	**ROUGE-2 (with)**	**ROUGE-2 (no)**	**ROUGE-L (with)**	**ROUGE-L (no)**	**Cosine (with)**	**Cosine (no)**
What is storage?	0.2476	0.1714	0.0777	0.0290	0.2286	0.1429	0.2844	0.2120
What is carbon?	0.2712	0.1360	0.0690	0.0403	0.2203	0.1120	0.4412	0.4200
What is metering points?	0.2913	0.1899	0.1136	0.0769	0.2330	0.1519	0.3977	0.4252
Give me examples of Energy Efficiency	0.1047	0.0635	0.0208	0.0128	0.0733	0.0508	0.2791	0.2920
What are ISOs?	0.2482	0.1242	0.0408	0.0126	0.1606	0.0508	0.4150	0.1720

Created by the author (2025)

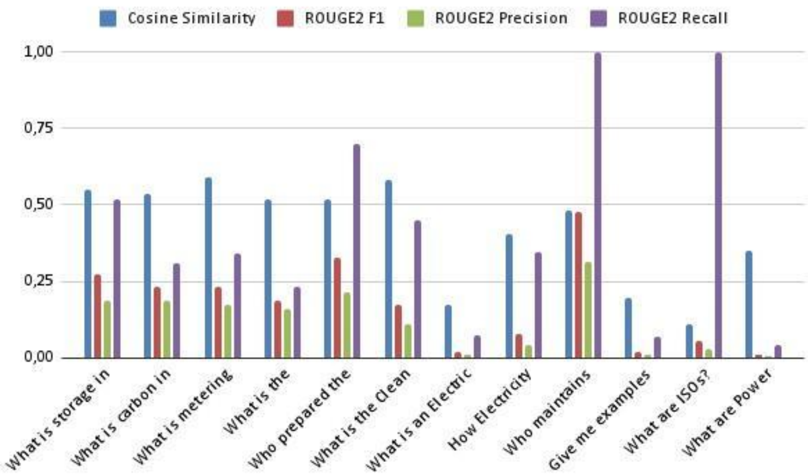
Among all the context-free responses, only one matched its contextual counterpart (the question related to metering points), and none were more closely aligned with the golden answers. This finding is consistent with the qualitative evaluation, which confirms that context significantly influences how the model interprets domain-specific terms and determines their meaning.

These results highlight that the proposed ontology functions as more than just a structure for transmitting data about the exchanged terms; it is an essential component that enables the model to make accurate assumptions about the data and its intended meaning from the sender's perspective. Without the ontology, even when using advanced models, the results are noticeably less reliable and fall short of the desired interpretative accuracy.

5.4 DISCUSSION

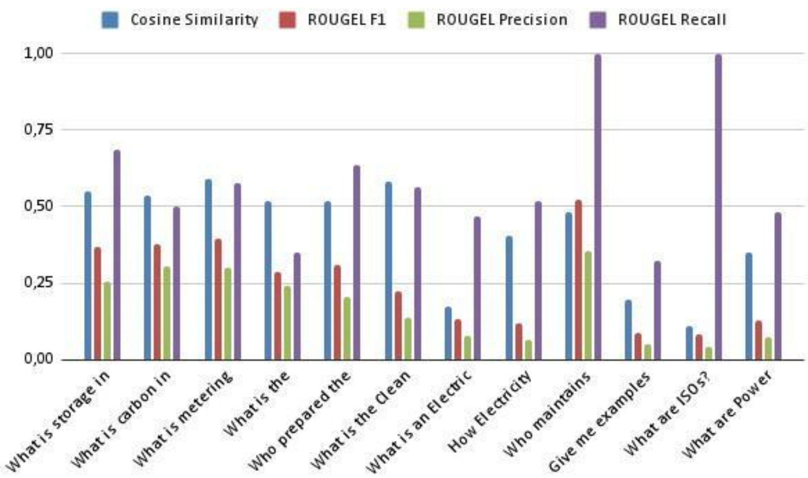
In this regard, the desired outcomes were achieved by augmenting the ontology with Ollama models, automatically adding all the meanings to terms without requiring input from the actors themselves. This can also prove invaluable when combined with other solutions commonly used in the field, like OEO.

Figure R2 - ROUGE-2 results



Source: Created by the author (2025)

Figure RL - ROUGE-L results



Source: Created by the author (2025)

Additionally, the possibility of incorporating more documents into the model's context through the retrieval process, along with the natural evolution of LLM models, paints a promising future for the proposed solution. This is especially true considering that the more content is added, the more precise the meanings become.

In terms of documents for the RAG process, new document-gathering efforts can be undertaken to expand the scope of the context generated. This ensures the LLMs have more data to work with, compensating for any lack of specific training they may have or the risk of them becoming outdated or obsolete.

Nevertheless, the results also indicate that adjustments can be made to align the generated answers more closely with the gold standards. It's crucial to remember that the model used in the experiment and the proposed solution has not been fine-tuned for this specific context, nor has it been trained to generate answers in the most efficient or desirable way for this particular task.

As noted by Siriwardhana et al.(2023) , fine-tuning the model using relevant information, prompts, and domain-specific training can be fundamental to achieving optimal results. This holds true for both the retrieval and generation phases of the process.

Finally, it is worth noting that, despite room for growth and improvement, the results suggest that utilizing RAG within the energy field offers benefits. The complexities inherent to the interoperability process, particularly between differing contexts, can potentially be mitigated through its innovative and sophisticated application.

6 CONCLUSIONS AND FUTURE WORK

This research addressed the problem of miscommunications in the energy sector, particularly between stakeholders operating in diverse contexts. It put forth the use of cutting-edge NLP techniques and text generation to embed meaning into any exchanged information, enabled by the utilization of ontologies.

Even though the proposed solution was applied within a hypothetical scenario, results demonstrated that incorporating RAG into the process successfully added meanings to terms within the ontology. This actively worked towards minimizing misunderstandings that might arise if a receiver interpreted data differently than intended by the sender.

In summary, the findings of this work indicate that utilizing RAG in the energy sector, particularly where interoperability and other forms of data exchange are involved, holds significant promise and potential. The ongoing globalization and increasing cooperation between actors from diverse backgrounds create ample opportunities for such solutions to evolve and flourish. This potential is further accentuated by the fact that the increased usage of these models and solutions generates data that can be fed back into them, resulting in even more capable and comprehensive models able to imbue ontologies with accurate and desirable meanings.

The identified limitations, such as the lack of models specifically trained for the energy sector, a wealth of documents to utilize with the RETRIEVAL part of RAG, or the need for refinements to the evaluation process (perhaps incorporating more human input), can be addressed in future iterations of the solution. This could even involve leveraging knowledge generated by the proposed solution itself.

Python code was developed with the intention of leveraging the language's power to seamlessly integrate RAG, ontologies, and other cutting-edge libraries to achieve the desired outcomes. However, given the scope of the proposed solution, this integration was implemented on a smaller, more experimental scale. Consequently, the proposed solution operated within a hypothetical scenario, although grounded in real-world observations. While this presents certain limitations, it doesn't mean that the proposed solution wouldn't be capable of tackling real-world situations and challenges.

In future work, these limitations can be addressed by applying the achievements of the proposed solution on a larger scale, involving real-world situations like the one mentioned in section 1.0.3, where actors exchange data in diverse contexts. This could be accomplished through experimentation within academia or by forging partnerships with power suppliers or government agencies. Regarding the models themselves, fine-tuning one of Ollama's models or developing an entirely new one could significantly enhance the process of assigning meaning.

While the model used in the proposed solution has the ability to answer based on context, with the assistance of the retrieval process, it's not specifically designed for the needs of the energy sector. Lastly, a wider range of documents could be fed into the retrieval part of RAG, ensuring that the model operates with the most current information possible. This goes hand-in-hand with utilizing more powerful hardware than what was used in the proposed solution.

To conclude, the results of this research point to the utilization of RAG in the energy sector, particularly in areas involving interoperability and various forms of data exchange, as being of considerable interest and possessing significant potential. The ongoing expansion of globalization and the growing collaboration between actors from diverse backgrounds create ample opportunities for such solutions to evolve and thrive. This is further underscored by the fact that the increased usage of these models and solutions generates data that can be fed back into them, leading to even more powerful and comprehensive models capable of enriching ontologies with accurate and meaningful definitions.

REFERENCES

1 Quang-Dung Ho, Tho Le-Ngoc, MS Obaidat, A Anpalagan, and I Woungang. Smart grid communications networks: Wireless technologies protocols issues and standards. *Handbook of green information and communication systems*, pages 115–146, 2012.

2 Khouloud Salameh, Richard Chbeir, and Haritza Camblong. Ssg: An ontology-based information model for smart grids. *Transactions on Large-Scale Data-and Knowledge-Centered Systems XL*, pages 94–124, 2019.

3 Arina Anisie, Juan Pablo Jimenez Navarro, Tomislav Antic, Francesco Pasimeni, and Herib Blanco. *Innovation landscape for smart electrification: Decarbonising end-use sectors with renewable power*. International Renewable Energy Agency, June 2023.

4 Ugur Korkut Pata, Abdullah Emre Caglar, Mustafa Tevfik Kartal, and Serpil Kılıç Depren. Evaluation of the role of clean energy technologies, human capital, urbanization, and income on the environmental quality in the United States. *Journal of Cleaner Production*, 402:136802, 2023.

5 Ministério de Minas e Energia. Brasil registra maior produção de energia limpa dos Últimos 12 anos, 2023. Accessed on: 19 de maio de 2025.

6 RegAce Project. Technology overview, 2023. Accessed on: 19 de maio de 2025.

7 Global Wind Energy Council (GWEC). Cop 27: Energy security and prosperity for the future - wind and solar, 2023. Accessed on: 19 de maio de 2025.

8 2015 ieee smart grid annual report. https://smartgrid.ieee.org/images/files/pdf/2015_ieee_\protect\@normalcr\relaxsmart_grid_annual_report.pdf, 2015.

9 ENTSO-E (European Network of Transmission System Operators for Electricity). https://www.entsoe.eu/. [Online; accessed 6th Feb. 2024].

10 Tdx Assist. http://www.tdx-assist.eu/. [Online; accessed 6th Feb. 2024].

11 Nermin Suljanović, Andrej Souvent, Gareth Taylor, Mohammed Radi, Jérôme Cantenot, Eric Lambert, and Hugo Morais. Design of interoperable communication architecture for tso-dso data exchange. In *2019 IEEE Milan PowerTech*, pages 1–6. IEEE, 2019.

12 MME-Grupo de Trabalho de Redes. Elétricas inteligentes ministério de minas e energia. *Relatório GT Smart Grid Portaria*, 440, 2010.

13 Mohammed Radi, Gareth Taylor, Mathias Uslar, Julia Köhlke, and Nermin Suljanovic. Energy data flow in smart grids-a conceptual model for addressing various use cases. 2019.

14 ENTSO-E Common Information Model. https://www.entsoe.eu/digital/common-information-model/. [Online; accessed 6th Feb. 2024].

15 Australian Energy Market Operator (AEMO). https://aemo.com.au/. [Online; accessed 6th Feb. 2024].

16 ASEXML Schemas - Australian Energy Market Operator (AEMO). https://aemo.com.au/energy-systems/market-it-systems/asexml-standards/asexml-schemas. [Online; accessed 6th Feb. 2024].

17 International Renewable Energy Agency (IRENA). https://www.irena.org/. [Online; accessed 6th Feb. 2024].

18 World Energy Transitions Outlook 2023. https://www.irena.org/Publications/2023/Jun/World-Energy-Transitions-Outlook-2023. [Online; accessed 6th Feb. 2024].

19 Fengqin Liu, Jae-yeon Sim, Huaping Sun, Bless Kofi Edziah, Philip Kofi Adom, and Shunfeng Song. Assessing the role of economic globalization on energy efficiency: Evidence from a global perspective. China Economic Review, 77:101897, February 2023.

20 Giulio Raimondi and Giuseppe Spazzafumo. Exploring renewable energy communities integration through a hydrogen power-to-power system in italy. Renewable Energy, 206:710–721, 2023.

21 Bidattul Syirat Zainal, Pin Jern Ker, Hassan Mohamed, Hwai Chyuan Ong, I. M. R. Fattah, S. M. Ashrafur Rahman, Long D. Nghiem, and T. M. Indra Mahlia. Recent advancement and assessment of green hydrogen production technologies. *Renewable Sustainable Energy Rev.*, 189:113941, January 2024.

22 Md. Mahmudul Alam and Md. Wahid Murad. The impacts of economic growth, trade openness and technological progress on renewable energy use in organization for economic co-operation and development countries. Renewable Energy, 145:382–390, January 2020.

23 Vaninha Vieira dos Santos, Patrícia Tedesco, and Ana Carolina Salgado. Percepção e contexto. *Sistemas Colaborativos*, 2012.

24 Matheus B. Jenevain, Milena F. Pinto, Mario A. R. Dantas, Regina M. M. B. Villela, Jose M. N. David, and Victor S. A. Menezes. Enabling Intelligent Data Exchange in the Brazilian Energy Sector: A Context-Aware Ontological Approach. In Advanced Information Networking and Applications, pages 372–381. Springer, Cham, Switzerland, April 2024.

25 Meisam Booshehri, Lukas Emele, Simon Flügel, Hannah Förster, Johannes Frey, Ulrich Frey, Martin Glauer, Janna Hastings, Christian Hofmann, Carsten Hoyer-Klick, et al. Introducing the open energy ontology: Enhancing data interpretation and interfacing in energy systems analysis. Energy and AI, 5:100074, 2021.

26 Charles Kemp, Piotr Mirowski, Zachary Simpson, Filipe Ribeiro, Victor Shi, Jongha Park, Tianlin Ma, Yixin Zhang, Mengdi Wang, Michael Andrew Wilson, et al. Context discovery: An interdisciplinary review of theory and methods. arXiv preprint arXiv:2104.09803, 2023.

27 Gabriella C.B. Costa, Cláudia M.L. Werner, Regina Braga, Humberto L.O. Dalpra, Marco Antônio P. Araújo, and Victor Ströele. Deriving strategic information for software development processes using provenance data and ontology techniques. International Journal of Business Process Integration and Management, 9(3):170, 2019.

28 Penghao Zhao, Hailin Zhang, Qinhan Yu, Zhengren Wang, Yunteng Geng, Fangcheng Fu, Ling Yang, Wentao Zhang, Jie Jiang, and Bin Cui. Retrieval-Augmented Generation for AI-Generated Content: A Survey. arXiv, February 2024.

29 Huayang Li, Yixuan Su, Deng Cai, Yan Wang, and Lemao Liu. A Survey on Retrieval-Augmented Text Generation. arXiv, February 2024.

30 Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. Retrieval-augmented generation for knowledge-intensive nlp tasks. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors,

Advances in Neural Information Processing Systems, volume 33, pages 9459–9474. Curran Associates, Inc., 2020.

31 Shouxiang Wang, Zihan Meng, and Shuangchen Yuan. IEC 61970 standard based common information model extension of electricity-gas-heat integrated energy system. *Int. J. Electr. Power Energy Syst.*, 118:105846, June 2020.

32 Australian Energy Regulator. State of the energy market 2023, October 2023. [Online; accessed 10. Jul. 2024].

33 Deborah Werner and Lira Luz Benites Lazaro. The policy dimension of energy transition: The Brazilian case in promoting renewable energies (2000–2022). Energy Policy, 175:113480, April 2023.

34 Luvai Motiwalla. V dan jeff thompson.(2009). enterprise systems for management.

35 Thomas R Gruber. A translation approach to portable ontology specifications. *Knowledge acquisition*, 5(2):199–220, 1993.

36 Balakrishnan Chandrasekaran, John R Josephson, and V Richard Benjamins. What are ontologies, and why do we need them? *IEEE Intelligent Systems and their applications*, 14(1):20–26, 1999.

37 Nicola Guarino, Daniel Oberle, and Steffen Staab. What is an ontology? Handbook on ontologies, pages 1–17, 2009.

38 Dilek Küçük and Doğan Küçük. Ontowind: An improved and extended wind energy ontology. *arXiv preprint arXiv:1803.02808*, 2018.

39 Dilek Küçük, Özgül Salor, Tolga Inan, Işık Çadırcı, and Muammer Ermiş. Pqont: A domain ontology for electrical power quality. *Advanced Engineering Informatics*, 24(1):84–95, 2010.

40 Dilek Küçük and Yusuf Arslan. Semi-automatic construction of a domain ontology for wind energy using wikipedia articles. Renewable Energy, 62:484–489, 2014.

41 Dilek Küçük. A high-level electrical energy ontology with weighted attributes. *Advanced Engineering Informatics*, 29(3):513–522, 2015.

42 Gregorio López, Víctor Custodio, José Ignacio Moreno, Marek Sikora, P Moura, and N Fernández. Modeling smart grid neighborhoods with the enersip ontology. *Computers in industry*, 70:168–182, 2015.

43 Rafael Santodomingo, Sebastian Rohjans, Mathias Uslar, JA Rodríguez-Mondéjar, and Miguel Angel Sanz-Bobi. Ontology matching system for future energy smart grids. *Engineering Applications of Artificial Intelligence*, 32:242–257, 2014.

44 Jan Schweikert, Karl-Uwe Stucky, Wolfgang Süß, and Veit Hagenmeyer. A photovoltaic system model integrating fair digital objects and ontologies. Energies, 16(3):1444, 2023.

45 Stanford Center for Biomedical Informatics Research. protégé, March 2024. [Online; accessed 25. Mar. 2024].

46 Kim Martineau. What is retrieval-augmented generation? IBM Research, May 2024.

47 Rick Merritt. What Is Retrieval-Augmented Generation aka RAG? NVIDIA Blog, June 2024.

48 What Are Large Language Models (LLMs)? | IBM, July 2024. [Online; accessed 13. Jul. 2024].

49 Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo

Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. A Survey of Large Language Models. arXiv, March 2023.

50 Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need. arXiv, June 2017.

51 Lewis Tunstall, Leandro Von Werra, and Thomas Wolf. Natural language processing with transformers, revised edition. O'Reilly Media, Sebastopol, CA, June 2022.

52 Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurelien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. Llama 2: Open Foundation and Fine-Tuned Chat Models. arXiv, July 2023.

53 Matriz elétrica brasileira alcança 200 GW, July 2024. [Online; accessed 15. Jul. 2024].

54 LA PICIRELLI DE SOUZA, L. et al. Life cycle assessment of prospective scenarios maximizing renewable resources in the Brazilian electricity matrix. **Renewable Energy Focus**, v. 44, p. 1–18, 1 mar. 2023.

55 MATRIZ ENERGÉTICA, July 2024. [Online; accessed 15. Jul. 2024].

56 Ariana Cericatto da Silva and Francisco Dietima da Silva Bezerra. DIFERENÇAS REGIONAIS NO BRASIL E A NECESSIDADE DO "PENSAR TERRITORIALMENTE". *SLAEDR*, 3(1), June 2023.

57 Brígida Teixeira, Tiago Pinto, Francisco Silva, Gabriel Santos, Isabel Praça, and Zita Vale. Multi-Agent Decision Support Tool to Enable Interoperability among Heterogeneous Energy Systems. *Appl. Sci.*, 8(3):328, February 2018.

58 Xiaohua Wang, Zhenghua Wang, Xuan Gao, Feiran Zhang, Yixin Wu, Zhibo Xu, Tianyuan Shi, Zhengyuan Wang, Shizheng Li, Qi Qian, Ruicheng Yin, Changze Lv, Xiaoqing Zheng, and Xuanjing Huang. Searching for Best Practices in Retrieval-Augmented Generation. arXiv, July 2024.

59 Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, Qianyu Guo, Meng Wang, and Haofen Wang. Retrieval-Augmented Generation for Large Language Models: A Survey. arXiv, December 2023.

60 LangChain, July 2024. [Online; accessed 17. Jul. 2024].

61 Hugging Face – The AI community building the future., July 2024. [Online; accessed 17. Jul. 2024].

62 Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini

Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language Models are Few-Shot Learners. arXiv, May 2020.

63 Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.

64 Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv, October 2018.

65 Tiago Timponi Torrent, Thomas Hoffmann, Arthur Lorenzi Almeida, and Mark Turner. *Copilots for Linguists: AI, Constructions, and Frames*. Cambridge University Press, Cambridge, England, UK, December 2023.

66 Carolina Fortuna, Vid Hanžel, and Blaž Bertalanič. Natural Language Interaction with a Household Electricity Knowledge-based Digital Twin. arXiv, June 2024.

67 Wenhu Chen, Hexiang Hu, Xi Chen, Pat Verga, and William W. Cohen. MuRAG: Multimodal Retrieval-Augmented Generator for Open Question Answering over Images and Text. arXiv, October 2022.

68 Wenhu Chen, Hexiang Hu, Chitwan Saharia, and William W. Cohen. Re-Imagen: Retrieval-Augmented Text-to-Image Generator. arXiv, September 2022.

69 Zhongliang Zhou, Jielu Zhang, Zihan Guan, Mengxuan Hu, Ni Lao, Lan Mu, Sheng Li, and Gengchen Mai. Img2Loc: Revisiting Image Geolocalization using Multi-modality Foundation Models and Image-based Retrieval-Augmented Generation. In ACM Conferences, pages 2749–2754. Association for Computing Machinery, New York, NY, USA, July 2024.

70 Sara Sarto, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Retrieval-Augmented Transformer for Image Captioning. In ACM Other conferences, pages 1–7. Association for Computing Machinery, New York, NY, USA, September 2022.

71 Fabrício O. Coelho, Milena Faria Pinto, Iago Z. Biundini, Gabriel G. R. Castro, Fabio A. A. Andrade, and Andre L. M. Marcato. Autonomous UAV Exploration and Mapping in Uncharted Terrain Through Boundary-Driven Strategy. IEEE Access, 12:92464–92483, July 2024.

72 Atlassian. Importance of Documentation | The Workstream, July 2024. [Online; accessed 25. Jul. 2024].

73 Antonio Jimeno Yepes, Yao You, Jan Milczek, Sebastian Laverde, and Renyu Li. Financial Report Chunking for Effective Retrieval Augmented Generation. arXiv, February 2024.

74 Ozan Unlu, Jiyeon Shin, Charlotte J. Mailly, Michael F. Oates, Michela R. Tucci, Matthew Varugheese, Kavishwar Wagholikar, Fei Wang, Benjamin M. Scirica, Alexander J. Blood, and Samuel J. Aronson. Retrieval Augmented Generation Enabled Generative Pre-Trained Transformer 4 (GPT-4) Performance for Clinical Trial Screening. medRxiv, 2024.

75 Zhentao Xu, Mark Jerome Cruz, Matthew Guevara, Tie Wang, Manasi Deshpande, Xiaofeng Wang, and Zheng Li. Retrieval-Augmented Generation with Knowledge Graphs for Customer Service Question Answering. In ACM Conferences, pages 2905–

2909. Association for Computing Machinery, New York, NY, USA, July 2024. 76 A future based on renewable energy, August 2023. [Online; accessed 29. Jul. 2024].

77 'Without renewables, there can be no future': 5 ways to power the energy transition, September 2023. [Online; accessed 29. Jul. 2024].

78 Energy Transition – A Project for Generations, July 2024. [Online; accessed 29. Jul. 2024].

79 Renewable Energy Sources Act (EEG) - Latest Version (EEG 2022). https://climate-laws.org/document/renewable-energy-sources-act-eeg-lat est-version-eeg-2022_1b40. [Online; accessed 6th Feb. 2024].

80 Renewable Energy - Federal Ministry of Education and Research (BMWK). https://www.bmwk.de/Redaktion/EN/Dossier/renewable-energy.html. [Online; accessed 6th Feb. 2024].

81 Mark A. Musen and The Protégé Team. The Protégé Project: A Look Back and a Look Forward. AI matters, 1(4):4, June 2015.

82 EnergyContext, August 2024. [Online; accessed 13. Aug. 2024].

83 Ollama, August 2024. [Online; accessed 14. Aug. 2024].

84 nomic-embed-text Ollama page, August 2024. [Online; accessed 19. Aug. 2024].

85 mistral 7b Ollama page, August 2024. [Online; accessed 19. Aug. 2024].

86 Oar. Energy Efficiency/Renewable Energy Roadmap Manual. *US EPA*, March 2024.

87 Ehsan Kamalloo, Nouha Dziri, Charles Clarke, and Davood Rafiei. Evaluating open-domain question answering in the era of large language models. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 5591–5606, Toronto, Canada, July 2023. Association for Computational Linguistics.

88 ROUGE - a Hugging Face Space by evaluate-metric, August 2024. [Online; accessed 19. Aug. 2024].

89 BLEU - a Hugging Face Space by evaluate-metric, August 2024. [Online; accessed 19. Aug. 2024].

90 Lance Martin. Auto-Eval of Question-Answering Tasks — blog.langchain.dev. https://blog.langchain.dev/auto-eval-of-question-answering-tasks/, 2023. [Accessed 19-08-2024].

91 Zafaryab Rasool, Stefanus Kurniawan, Sherwin Balugo, Scott Barnett, Rajesh Vasa, Courtney Chesser, Benjamin M. Hampstead, Sylvie Belleville, Kon Mouzakis, and Alex Bahar-Fuchs. Evaluating LLMs on Document-Based QA: Exact Answer Selection and Numerical Extraction using Cogtale dataset. arXiv, November 2023.

92 Ashkan Alinejad, Krtin Kumar, and Ali Vahdat. Evaluating the Retrieval Component in LLM-Based Question Answering Systems. arXiv, June 2024.

93 Copilots for Linguists, August 2024. [Online; accessed 19. Aug. 2024].

94 Shamane Siriwardhana, Rivindu Weerasekera, Elliott Wen, Tharindu Kaluarachchi, Rajib Rana, and Suranga Nanayakkara. Improving the Domain Adaptation of Retrieval Augmented Generation (RAG) Models for Open Domain Question Answering. *Transactions of the Association for Computational Linguistics*, 11:1–17, December 2023.

Appendix

APPENDIX A - Published Works

This appendix presents the papers accepted and presented during the course of this master's degree, along with their abstracts.

- **Enabling Intelligent Data Exchange in the Brazilian Energy Sector:** A Context-Aware Ontological Approach (AINA 24): The ever-expanding requirement for interoperability between systems, coupled with the growing number of participants in the energy sector, creates a scenario where diverse actors, each with unique realities and specificities, must exchange data and knowledge. This exchange often occurs among significant differences in context. These disparities can lead to many difficulties and misunderstandings, as the information exchanged is susceptible to misinterpretation based on the sender's and receiver's contexts. Therefore, this work addresses this issue by proposing an extension to the Open Energy Ontology (OEO) that focuses on context. It investigates how an actor's understanding is shaped by their context, the methods for inferring this context, and strategies to enhance interoperability. The results obtained demonstrate the potential of the approach proposed by this work.

- **A Context-Aware Approach to Data Exchange in the Energy Sector (CIEEMAT):** This paper presents a novel approach to enhance data exchange within the energy sector by developing a context-aware method utilizing Natural Language Processing (NLP) techniques and ontologies such as Open Energy Ontology (OEO). The proposed solution addresses the challenges arising from differing contexts and perceptions, particularly among actors like ENTSO-E and its member countries. By capturing, representing, and assigning meanings based on the specific realities of the energy sector, this work seeks to minimize miscommunications and complications that may arise during data exchange. The approach is built upon previous research exploring ontology development and context perception.

- **Using UAVs and Retrieval Augmented Generation for Situational Awareness in Rescue Operations (AINA 25):** The increasing frequency and severity of natural disasters in Brazil, particularly in 2024, highlight the urgent need for innovative solutions in Search and Rescue (SAR) operations. This work presents a novel framework integrating Retrieval-Augmented Generation (RAG) techniques with Unmanned Aerial Vehicles (UAVs) to enhance real-time data processing, usability, and operator decision-making. By incorporating advanced technologies such as FrameNet Brasil, Robot Operating System 2 (ROS2), and Large Language Models (LLMs), the system transforms UAV-captured data into actionable insights accessible through natural language interfaces. Testing demonstrates its ability to improve situational awareness, identify critical points of interest, and streamline mission execution. This modular and scalable approach lays the groundwork for future advancements in SAR technologies and their application in disaster-prone regions.

- **Context-Aware Data Exchange in the Energy Sector: A Retrieval-Augmented Ontology Approach (ISCC 25):** The increasing globalization of energy markets, the emergence of new energy sources, and the growing need for seamless data exchange among stakeholders highlight the importance of effective communication and interoperability in the energy sector. However, current protocols often prioritize technical aspects over contextual understanding, leading to misinterpretations and communication breakdowns. This paper proposes a novel solution that combines ontologies with Retrieval Augmented Generation (RAG) and Large Language Models (LLMs) to ensure accurate interpretation of shared data, regardless of the actors' diverse contexts. By leveraging the strengths of ontologies for knowledge representation and RAG for contextual understanding, the proposed framework aims to enhance interoperability and facilitate effective collaboration in the energy sector, ultimately contributing to a smoother and more efficient energy transition. The results indicate that the proposed solution effectively assigns contextually relevant meanings to terms, as evidenced by the Cosine Similarity and ROUGE Recall scores exceeding 0.5 for most generated answers, enhancing data sharing and mitigating potential misunderstandings originating from differing contexts and perspectives among senders and receivers.

- **Augmented Energy Ontology (IP):** Intellectual work registered under process number BR512024002512-0 at the Brazilian National Institute of Industrial Property.

APPENDIX B – Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans (RAG PDF).

EPA United States Environmental Protection Agency

Roadmap for Incorporating Energy Efficiency/Renewable Energy Policies and Programs into State and Tribal Implementation Plans

Appendix A: Glossary

EPA-456/D-12-001b
July 2012

Roadmap for Incorporating Energy Efficiency/Renewable Energy
Policies and Programs into State and Tribal Implementation Plans

Appendix A: Glossary

By:
U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Outreach and Information Division
Research Triangle Park, North Carolina

U.S. Environmental Protection Agency
Office of Air Quality Planning and Standards
Outreach and Information Division
Research Triangle Park, North Carolina

ACKNOWLEDGMENTS

We would like to acknowledge substantial contributions from members of an inter-office EPA team that included the Office of Atmospheric Programs, the Office of Policy Analysis and Review, the Office of General Counsel and Regions 1 and 6. This document also reflects comments received from a number of stakeholders, including state and local air quality agencies.

Allowance: An allowance is an authorization to emit a specific amount of a pollutant under a cap and trade program. For example, under the U.S. Sulfur Dioxide (SO_2) Allowance Trading Program, one allowance is the authorization to emit 1 ton of SO_2. Allowances are used for compliance and can be traded among sources participating in the cap and trade program.

Annual Energy Outlook: Prepared by the U.S. Department of Energy's Energy Information Administration (EIA), the Annual Energy Outlook presents long-term projections of energy supply, demand, and prices through 2035, based on results from EIA's National Energy Modeling System.

Baseline Projections: Baseline projections, including energy supply and demand and related emissions, are intended to describe future year conditions and typically assume continuation of current trends and no changes in laws and regulations. Baseline projections are sometimes referred to as "business-as-usual" projections. A baseline projection can be used for comparison with one or more alternative policy scenarios to assess the impacts of various policies.

Combined Heat and Power: Also known as cogeneration, combined heat and power (CHP) is an efficient, clean, and reliable approach to generating power and thermal energy from a single fuel source. Since less fuel is burned to produce each unit of energy output, CHP reduces air pollution and greenhouse gas emissions. As a result, the emission benefits from CHP systems can be recognized for State Implementation Plan/Tribal Implementation Plan credit. Typical CHP configurations include gas turbines or engines with heat recovery units or steam boilers with a steam turbine.

Clean Air Act: The Clean Air Act (CAA) is the law that defines the U.S. Environmental Protection Agency's responsibilities for protecting and improving the nation's air quality and the stratospheric ozone layer. The last major change in the law occurred when Congress enacted the CAA Amendments of 1990. Legislation passed since then has made several minor changes.

Criteria Air Pollutant: The Clean Air Act requires the U.S. Environmental Protection Agency to set National Ambient Air Quality Standards for six common air pollutants. These common air pollutants (also known as "criteria pollutants") are found all over the country. They are particle pollution (often referred to as particulate matter), ground-level ozone, carbon monoxide, sulfur oxides, nitrogen oxides, and lead.

Demand: The time rate of energy flow. Demand usually refers to electric power measured in kilowatt but can also refer to natural gas, usually as British thermal unit/hour (Btu/hr), kiloBtu/hr, or therms/day.

Discount Rate: A measure of the time value of money.

Electric Generating Unit: This is an entity that supplies electricity to the electricity system relying on a variety of fuels.

Electricity Dispatch Models: Electricity dispatch models (also commonly referred to as "production cost" models) simulate the dynamic operation of the electric system, generally on a least-cost system dispatch. In general, these models optimize the dispatch of the system based on the variable costs of each resource and any operational constraints that have been entered into the model. These models are helpful in assessing which existing plants are displaced. These models are also used in short-term planning and regulatory support.

Emissions & Generation Resource Integrated Database: The Emissions & Generation Resource Integrated Database is a comprehensive inventory of environmental attributes of electric power plants, providing air emissions data for the electric power sector. The U.S. Environmental Protection Agency maintains the database.[1]

Energy Efficiency: Energy efficiency is achieving the same or better level of service or performance with lower energy consumption. Examples include high-efficiency appliances; efficient lighting; high-efficiency heating, ventilating and air conditioning systems or control modifications; efficient building design; advanced electric motor drives; combined heat and power; and heat recovery systems.

Energy Efficiency Policy: Energy efficiency (EE) policy means an enacted law and/or regulation by a state, locality, or public utility commission order which requires applicable entities to adopt energy efficient technologies and/or practices, or to undertake activities to further such adoption in the marketplace. It can include: (1) policies that establish minimum efficiency requirements for new homes and buildings (building energy codes) or appliances (appliance standards); (2) policies that establish requirements on utilities (or other program administrators) to deliver a specified amount of energy savings by developing EE programs to increase market adoption of EE technologies and practices (EE resource standards); and (3) policies that commit to specified funding levels dedicated to implementing EE programs (e.g., EE rebates or combined heat and power capital cost incentives from public benefits funds). State and local governments both have authority over EE policies. EE policies are generally enforced over a multi-year period or until changed or updated by revised legislation or regulation (e.g., adopting a revised building energy code). These programs can be funded through ratepayer surcharges, federal funds, proceeds from pollution auctions such as the Regional Greenhouse Gas Initiative, or any combination of the above.

Electric-sector EE and renewable energy policies, programs and projects that will result in quantifiable reductions in emissions at existing fossil fuel-fired electric generating units and that improve air quality in a nonattainment area can be accounted for in State and Tribal Implementation Plans.

Energy Efficiency Program: Energy efficiency (EE) program means a program designed to increase adoption of energy efficient technologies and practices in particular end-use sectors (or specific market segments within a sector) through education and outreach, financial incentives, and/or technical assistance. An individual EE program can be run by a utility, state or local government, and/or third parties. In most cases, EE program administrators (i.e., utilities, state agencies, or 3rd parties) develop and implement EE programs to meet adopted EE policy objectives. State Public Utilities

[1] For more information, go to: www.epa.gov/egrid.

Commissions oversee and approve the EE programs funded with rate-payer resources. EE programs typically operate over a one to three year period.

Electric-sector EE and renewable energy policies, programs and projects that will result in quantifiable reductions in emissions at existing fossil fuel-fired electric generating units and that improve air quality in a nonattainment area can be accounted for in State and Tribal Implementation Plans.

Energy Efficiency Project or Measure: Installation of equipment, installation of subsystems or systems, or modification of equipment, subsystems, systems, or operations on the customer side of the meter, in order to improve energy efficiency (EE). These projects or measures can be taken in conjunction with or independent of an EE policy or program.

Energy Efficiency Resource Standard: An Energy Efficiency Resource Standard (also known as Energy Efficiency Portfolio Standards) consists of electric and gas savings targets for utilities, often with flexibility to achieve them through a market-based trading system or a buyout-option to purchase credits at a default price.

Energy Model: This refers to simulation models that utilize computer modeling software to analyze the electric power system Energy models are capable of simulating electric system dispatch, forecasting future capacity and technology changes, and/or representing behavior of system-wide energy markets. Examples of energy models include: Energy 2020, Integrated Planning Model, MARKAL, National Energy Modeling System, Strategist by Ventyx.

Evaluation, Measurement and Verification: The performance of studies and activities aimed at determining the effects of a program; any of a wide range of assessment activities associated with understanding or documenting program performance, assessing program or program-related markets and market operations; any of a wide range of evaluative efforts including assessing policy or program-induced changes in energy efficiency markets, levels of demand or energy savings, and program cost-effectiveness.

Federal Enforceability: This refers to what occurs in the State Implementation Plan (SIP) planning process when the U.S. Environmental Protection Agency (EPA) approves a SIP control strategy submitted to it for review and the SIP becomes federally enforceable. A federally enforceable SIP provides EPA with authority to ensure the SIP is implemented. Once energy efficiency/renewable energy policies and programs become federally enforceable, EPA has the authority under the Clean Air Act (CAA) to apply CAA-mandated penalties against the noncompliant party.

Future Attainment Year Baseline: A specific year in the future for which a state, tribal or local agency must show attainment of the National Ambient Air Quality Standard. The baseline forecast of emissions in a future attainment year refers to the emissions that will result if no future policies or programs are adopted and implemented. A baseline forecast of future emissions is made when an area prepares a State or Tribal Implementation Plan (SIP/TIP). Future year emission projections provide a basis for considering control strategies for (SIPs/TIPs), conducting attainment analyses, and tracking progress towards meeting air quality standards.

Heating, Ventilating, and Air Conditioning: Heating, ventilating, and air conditioning refers to technology to provide for indoor environmental comfort.

Independent System Operator: Independent system operators (ISOs) serve as grid operators, coordinating the power grid to ensure reliable delivery. ISOs also match generation to load instantaneously to keep electricity supply and demand balanced and administer forward capacity markets where utilities can use energy efficiency as a resource to meet demand.

Integrated Planning Model: The EPA uses the Integrated Planning Model (IPM) to analyze the impact of environmental policies on the electric power sector in the 48 contiguous states and the District of Columbia. This model simultaneously models electric power, fuel, and environmental markets associated with electric production. It is a capacity expansion and system dispatch model. Dispatch is based on seasonal, segmented load duration curves, as defined by the user. IPM can be used to model the impacts of clean energy resources on the electric sector in the short and long term.

Kilowatt-Hour: A kilowatt-hour (KWh) is a unit of work or energy, equivalent to 1 kilowatt (1,000 watts) of power expended for 1 hour. One kWh is equivalent to 3,412 British thermal units.

Load Shapes: Representations such as graphs, tables, and databases that describe energy consumption rates as a function of another variable such as time or outdoor air temperature.

Marginal Emission Rates: The emissions associated with the marginal generating unit in each hour of the day.

Measurement and verification: Data collection, monitoring, and analysis associated with the calculation of gross energy and demand savings from individual sites or projects. Measurement and verification can be a subset of program impact evaluation.

Megawatt: A megawatt is one million watts of electricity.

Megawatt-hour: A megawatt-hour is one thousand kilowatt-hours or 1 million watt-hours.

National Ambient Air Quality Standards: The Clean Air Act (CAA), which was last amended in 1990, requires the U.S. Environmental Protection Agency to set National Ambient Air Quality Standards (NAAQS) for pollutants considered harmful to public health and the environment. The CAA established two types of NAAQS. Primary standards set limits to protect public health, including the health of "sensitive" populations such as asthmatics, children, and the elderly. Secondary standards set limits to protect public welfare, including protection against decreased visibility, damage to animals, crops, vegetation, and buildings.

Nitrogen Oxides: Nitrogen oxides can refer to a binary compound of oxygen and nitrogen, or a mixture of such compounds.

North American Electric Reliability Corporation: The North American Electric Reliability Corporation (NERC) is the electric reliability organization certified by the Federal Energy Regulatory Commission to establish and enforce reliability standards for the bulk-power system. NERC ensures the reliability of the North American bulk power system.

"On the books" Energy Efficiency/ Renewable Energy Policies: Energy efficiency/renewable energy policies that have been adopted by a legislative or regulatory body.

"On the way" Energy Efficiency/ Renewable Energy Policies: Energy efficiency/renewable energy policies that are planned for adoption by a legislative or regulatory body prior to the submittal of the State or Tribal Implementation Plans in question to the U.S. Environmental Protection Agency.

Peak Demand: The maximum level of metered demand during a specified period, such as a billing month or a peak demand period.

Portfolio: A portfolio is either (1) a collection of similar programs addressing the same market, technology, or mechanisms or (2) the set of all programs conducted by one organization.

Power Pool: A power pool is an association of two or more interconnected electric systems that agree to coordinate operations and planning for improved reliability and efficiencies.

Program: A group of projects, with similar characteristics and installed in similar applications.

Public Utilities Commission or Public Service Commission: A public utilities commission or public service commission is a governing body that regulates the rates and services of a public utility. In some cases, government bodies with the title "Public Service Commission" may be civil service oversight bodies, rather than utility regulators.

Renewable Energy: Renewable energy resources are naturally replenishing but flow-limited. They are virtually inexhaustible in duration but limited in the amount of energy that is available per unit of time. Renewable energy resources include biomass, hydro, geothermal, solar, wind, ocean thermal, wave action, and tidal action.

Renewable Energy Policy: Regulations, statutes, or state public utility commission orders that require parties to acquire renewable energy or to commit to funding levels for programs aimed at acquiring renewable energy.

Electric-sector energy efficiency and renewable energy policies, programs and projects that will result in quantifiable reductions in emissions at existing fossil fuel-fired electric generating units and that improve air quality in a nonattainment area can be accounted for in State or Tribal Implementation Plans.

Renewable Energy Program: Renewable energy program means a program designed to increase the production and use of renewable energy sources through resource development and procurement, education and outreach, financial incentives, and/or technical assistance.

Electric-sector energy efficiency and renewable energy policies, programs and projects that will result in quantifiable reductions in emissions at existing fossil fuel-fired electric generating units and that improve air quality in a nonattainment area can be accounted for in State or Tribal Implementation Plans.

Renewable Portfolio Standard: A renewable portfolio standards are a requirement on retail electric suppliers to supply a minimum percentage or amount of their retail load with eligible sources of renewable energy (e.g., solar, wind, biomass and geothermal).

Retail Electricity Supplier: A person or entity that sells electrical energy to end-use customers, including but not limited to electric utility distribution companies supplying basic service or any successor service to end-use customers.

State Implementation Plans: A State Implementation Plan (SIP) is a plan developed by a state detailing how that state will comply with the requirements of the federal Clean Air Act, administered by the U.S. Environmental Protection Agency. The SIP consists of narrative, rules, technical documentation, and agreements that an individual state will use to meet the National Ambient Air Quality Standards.

State Implementation Plan/Tribal Implementation Plan Credit: Credit for State and Tribal Implementation Plans means emission reductions, achieved by using technologies or strategies, used by a state or tribe for the purpose of meeting emission reduction requirements in its reasonable further progress, attainment or maintenance (control) strategy.

Tribal Implementation Plans: Although not required to do so, a tribe with "treatment as state" eligibility may develop its own air quality control plan, called a Tribal Implementation Plan, for approval by the U.S. Environmental Protection Agency (EPA). A TIP enacted by a tribal government and approved by the EPA is legally binding under both tribal and federal law and may be enforced by the tribe, EPA and the public.

Voluntary Energy Efficiency/Renewable Energy Programs: These are energy efficiency/renewable energy programs are not directly enforceable against a source or party administering the program. Examples could include municipal government energy conservation plans or public awareness campaigns.

Emerging/Voluntary Measures Policy: In September 2004, the U.S. Environmental Protection Agency issued guidance entitled: "Incorporating Emerging and Voluntary Measures in a State Implementation Plan (SIP)." The guidance provides a policy for areas to try new types of pollutant reduction strategies such as energy efficiency/renewable energy programs to attain or maintain the National Ambient Air Quality Standards and meet Clean Air Act requirements.

Watt: The watt is a standard unit of electrical power equivalent to one ampere flowing across a potential difference of one volt. A watt is equal to 1/746 horse power.

Weight-of-Evidence: The augmentation of a State Implementation Plan/Tribal Implementation Plan modeled attainment test with supplemental analyses, which may yield a conclusion different from that indicated by the modeled attainment test results alone.

United States Environmental Protection Agency	Office of Air Quality Planning and Standards Outreach and Information Division Research Triangle Park, NC	Publication No. EPA-456/D-12-001b July 2012