

## Article

# Domain-Specific Large Language Model for Renewable Energy and Hydrogen Deployment Strategies

Hossam A. Gabber \*  and Omar S. Hemied 

Faculty of Engineering and Applied Science, Ontario Tech University (UOIT), Oshawa, ON L1G 0C5, Canada

\* Correspondence: [hossam.gaber@ontariotechu.ca](mailto:hossam.gaber@ontariotechu.ca); Tel.: +1-(647)-975-8574

**Abstract:** Recent advances in large language models (LLMs) have shown promise in specialized fields, yet their effectiveness is often constrained by limited domain expertise. We present a renewable and hydrogen energy-focused LLM developed by fine-tuning LLaMA 3.1 8B on a curated renewable energy corpus (RE-LLaMA). Through continued pretraining on domain-specific data, we enhanced the model's capabilities in renewable energy contexts. Extensive evaluation using zero-shot and few-shot prompting demonstrated that our fine-tuned model significantly outperformed the base model across renewable and hydrogen energy tasks. This work establishes the viability of specialized, smaller-scale LLMs and provides a framework for developing domain-specific models that can support advanced research and decision-making in the renewable energy sector. Our approach represents a significant step forward in applying LLMs to the renewable and hydrogen energy sector, offering potential applications in advanced research and decision-making processes.

**Keywords:** large language model; zero shot; few shots; renewable energy; artificial intelligence; hydrogen deployment; energy deployment strategies



**Citation:** Gabber, H.A.; Hemied, O.S. Domain-Specific Large Language Model for Renewable Energy and Hydrogen Deployment Strategies. *Energies* **2024**, *17*, 6063. <https://doi.org/10.3390/en17236063>

Academic Editor: Abdul-Ghani Olabi

Received: 28 October 2024

Revised: 20 November 2024

Accepted: 27 November 2024

Published: 2 December 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

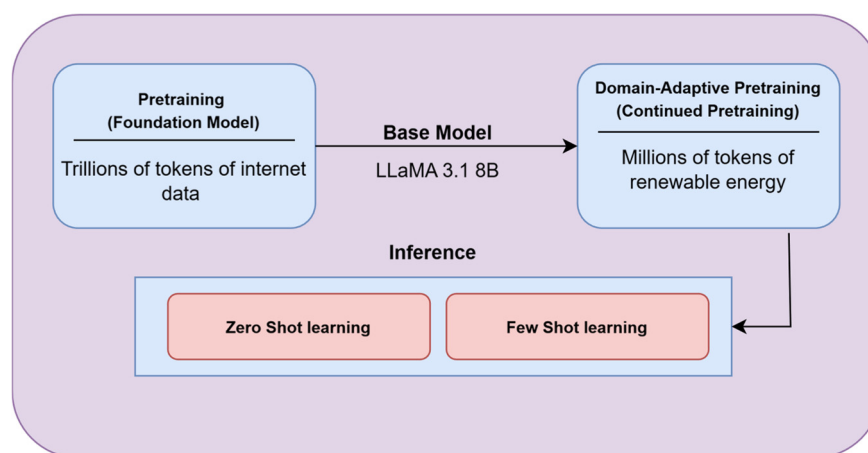
## 1. Introduction

Large language models (LLMs) have become powerful tools in the rapidly advancing field of artificial intelligence. ChatGPT, built on the Generative Pre-training Transformer (GPT) architecture, exemplifies these models' capabilities in various linguistic tasks, from article composition to code generation. The release of ChatGPT 4.0 has further demonstrated the potential of LLMs when applied to industry-specific problems [1]. While foundational language models excel in general tasks, they often fall short when dealing with specialized domains due to their lack of domain-specific knowledge [2]. Addressing this limitation is essential to improve the performance and applicability of LLMs in industry contexts. Large language models represent a significant advancement in AI applications, with the potential to transform numerous sectors. In the renewable energy field, they could revolutionize clean energy generation and decarbonization systems [2]. However, integrating domain-specific knowledge into LLMs remains a significant challenge. This paper aims to address this gap by investigating methods to fine-tune LLMs with domain-specific expertise for renewable energy applications. Our goal is to enhance LLM performance and unlock new possibilities for personalized, effective, and scalable renewable energy solutions. Our research begins with an extensive review of existing literature, focusing on methodologies for training LLMs with domain-specific knowledge and identifying potential benchmarks to evaluate their performance in the renewable energy context.

The applications of large language models to smart renewable and hydrogen energy systems represent a significant advancement in addressing the complexities of these critical domains. The inherent complexity and interconnected nature of renewable and hydrogen energy systems demand sophisticated tools capable of processing and analyzing vast amounts of multifaceted data. LLMs excel in this capacity by understanding and processing natural language queries, enabling technical information to become more accessible

and actionable for stakeholders with varying levels of expertise [3,4]. Additionally, the rapid evolution of renewable and hydrogen energy technologies creates a constant need for updated knowledge and analysis. LLMs can be fine-tuned with the latest research and technological developments, ensuring their outputs remain current and relevant [5]. Moreover, LLMs uniquely handle the multidisciplinary nature of renewable and hydrogen energy systems by integrating technical, economic, environmental, and policy-related knowledge into a cohesive framework. This capability allows for a more comprehensive and nuanced analysis of complex energy systems; LLMs bridge the communication gap between technical experts, policymakers, and other stakeholders. This holistic approach not only enhances informed decision-making but also promotes the broader adoption of renewable and hydrogen energy solutions, underscoring the transformative potential of LLMs in advancing sustainable energy systems.

In this study, we introduce RE-LLaMA, a specialized large language model for renewable energy, developed by continued pretraining of the LLaMA 3.1 8B parameter model, as shown in Figure 1. Our research explores effective methods for adapting general-purpose LLMs to specific domains, focusing on renewable energy. We develop benchmarks to measure performance in renewable energy contexts and evaluate RE-LLaMA using zero-shot and few-shot learning approaches. By comparing RE-LLaMA to the base LLaMA 3.1 8B model, we demonstrate its enhanced capabilities in renewable energy applications. Our model offers a wide range of uses in the sector, including assisting researchers with literature reviews and data analysis, supporting policymakers in understanding regulations, aiding project managers in planning and risk assessment, providing technical support, creating educational content, analyzing market trends, and facilitating public communication about renewable energy. Additionally, it can contribute to design innovation, environmental impact assessments, grid management, and financial analysis of energy projects. As the first open-source LLM specialized in renewable energy, RE-LLaMA aims to accelerate research, improve decision-making, and enhance understanding in this crucial field, potentially playing a significant role in advancing sustainable energy solutions.



**Figure 1.** RE-LLaMA training flow.

The dataset utilized for training RE-LLaMA comprises approximately 1450 papers carefully selected from the renewable energy domain. This corpus encompasses a diverse range of topics within the field, including microgrid optimization, renewable energy optimal sizing, communication systems, and design principles. To ensure comprehensive coverage of current trends, we also integrated data related to hydrogen technologies, broader renewable energy applications, and decarbonization strategies. The selection process prioritized papers that represent the cutting-edge developments and key challenges in the renewable energy sector. Following the data source phase, we conducted an intensive cleaning and processing procedure to enhance the quality and consistency of the dataset. This intensive preparation involved removing irrelevant content, standardizing formats, and ensuring

the integrity of technical information. The cleaning process was crucial in creating a high-quality training corpus that accurately reflects the current state of knowledge in renewable energy. A detailed description of the data preparation methodology, including specific cleaning techniques and processing algorithms employed, will be presented in subsequent sections of this paper. This carefully curated and processed dataset forms the foundation for training RE-LLaMA, enabling it to capture the nuances and specialized knowledge within the renewable energy domain.

Our results demonstrate the effectiveness of domain-adaptive pretraining for enhancing performance in renewable energy-specific tasks. The key research contributions of this study are as follows:

- Development and implementation of RE-LLaMA through domain-specific pretraining of LLaMA 3.1 8B, establishing a methodology for specialized language modeling in renewable and hydrogen energy applications.
- Construction and validation of a comprehensive training corpus comprising 1450 academic papers in renewable and hydrogen energy, coupled with systematic data processing workflow for domain-specific language model development.
- Systematic evaluation demonstrating quantitative improvements over the base LLaMA 3.1 8B model through assessment of zero-shot and few-shot learning capabilities in domain-specific tasks.

The remainder of this paper is structured as follows: Section 2 provides an overview of current research on domain-specific large language models and their applications across various fields, contextualizing the development of RE-LLaMA. Section 3 delves into the technical aspects of RE-LLaMA, detailing the dataset compilation, fine-tuning methodology, and adaptation process for renewable energy tasks. Section 4 presents a comparative analysis of RE-LLaMA performance against the base LLaMA 3.1 8B model, highlighting its enhanced capabilities in renewable energy-specific tasks. Section 5 discusses the limitations. Finally, Section 6 concludes with a summary of key findings and explores the future potential of RE-LLaMA in advancing renewable energy applications.

## 2. Large Language Model Analysis

The development of domain-specific large language models has become an increasingly important area of research in natural language processing. This section explores the current landscape of domain-specific LLMs, focusing on their evolution, various approaches to their creation, and the benefits of continued pretraining. We examine both proprietary and open-source models, highlighting their applications across different domains, with particular attention to medical and scientific fields. Our review sets the stage for introducing RE-LLaMA, the first LLM specialized for the renewable energy domain, and contextualizes its significance within the broader field of domain-specific language models.

The integration of large language models with renewable energy technologies presents a promising frontier in addressing climate change and ensuring energy security. Renewable sources like solar, wind, hydrogen fuel cells, electrolyzers and hydroelectric power emit minimal greenhouse gases compared to fossil fuels [6,7], offering superior sustainability and security [8]. As the International Energy Agency reports a decline in fossil fuel demand for power generation since 2019 [9], research in renewable energy technologies continues to expand, focusing on improving efficiency and replacing traditional resources with sustainable alternatives [10]. Artificial intelligence (AI) plays a crucial role in this transition, demonstrating superiority in controllability, data handling, cyberattack prevention, and smart grid implementation [11]. The strategic integration of AI, including LLMs, into renewable energy systems can enhance forecasting accuracy, optimize energy generation and consumption, improve grid stability, and facilitate smart energy management. This integration involves applying deep learning, machine learning algorithms, and neural networks to analyze large datasets, predict energy production and demand, detect anomalies, and automate decision-making processes. As the field advances, LLMs could potentially enhance natural language interfaces for energy management systems, improving user

interaction and facilitating more intuitive control of renewable energy resources. By combining the power of LLMs with domain-specific knowledge of renewable energy, we can create more effective tools for managing and optimizing renewable energy systems, further accelerating the transition to sustainable energy sources while addressing the complexities and challenges inherent in this rapidly evolving field.

Large language models are widely used in natural language processing tasks, providing a foundation for many applications. While foundational models can perform remarkably well in a broader context, they lack the domain-specific knowledge to be helpful in most industrial or business applications. LLMs trained on large-scale generic datasets often need domain-specific knowledge, leading to suboptimal performance in domain-specific applications such as renewable energy. To address this challenge, there is a need to enhance LLMs with domain-specific knowledge to improve their performance and utility in the industry. Large language models marked an important milestone in AI applications across various industries. LLMs fuel the emergence of a broad range of generative AI solutions, increasing productivity, cost-effectiveness, and interoperability across multiple business units and industries. One of the potential areas is renewable energy; LLMs will reform renewable energy systems in numerous ways, enabling the generation of custom renewable energy strategies and better information accessibility to assist renewable energy (RE) in contemporary power systems.

This study demonstrates the integration of domain-specific knowledge into LLMs for advancing renewable energy research and education. Our findings establish a methodological framework for enriching language models with specialized technical content, contributing to the broader understanding of how AI can enhance knowledge dissemination in renewable energy technologies. Although general-purpose LLMs trained on vast amounts of internet data exhibit remarkable capabilities in generative AI tasks across diverse domains, as demonstrated in [12], promising new models have emerged, such as BloombergGPT, which is a large language model for finance, and BioMedLLM, MELLaMA, and MEDITRON-70, which are pretrained for large language models specific to the medical domain [5–7,13] demonstrate the integration of the LLM in the E-learning domain. In addition, ChipNeMo [14] is a domain-adapted LLM for chip design that also [4] demonstrates LLM capability to generate research paper content related to solar and wind energy. All of these examples demonstrate that domain-specific LLM models can outperform a general-purpose model on domain-specific tasks. Customizing LLMs in this manner also avoids security risks associated with sending proprietary internal data to third-party LLMs via APIs like OpenAI APIs. However, it would be prohibitively expensive to train domain-specific models for every domain from scratch since this often requires millions of GPU training hours. To cost-effectively train domain-specific models, we instead propose the combination of the following techniques: domain adaptive pre-training [15] of foundation models with domain-adapted tokenizers and model alignment using general and domain-specific instructions. All models mentioned above have hundreds of billions of parameters, are computationally expensive to run, require users to send their input data over the internet, and are trained on unknown data sources.

Large language models have shown significant potential in transforming the electrical power industry by addressing its unique challenges and advancing its capabilities. Recent studies highlight their role in optimizing power systems through the creation of autonomous agents designed to enhance system efficiency by optimizing power states [16]. Additionally, LLMs have been recognized for improving productivity within the energy sector by accelerating the adoption of advanced AI tools and methodologies [17]. Moreover, proposed frameworks for power engineering demonstrate how LLMs can assist engineers in solving both routine and complex programming tasks within the energy domain [18]. Specific applications, such as power dispatch optimization, highlight the contextual decision-making capabilities of LLMs, as evidenced by the “ElecBench” benchmark [19]. Furthermore, their integration into energy system research enhances tasks such as forecasting, smart grid implementation, and cybersecurity, further illustrating their trans-

formative potential in managing electrical power systems [3]. Collectively, these studies underscore the advancements made by LLMs in the electrical power industry, offering a strong foundation for future innovations tailored to the sector's evolving needs.

### 2.1. Domain-Specific Large Language Models

The significance of domain-specific training for language models is well established, particularly for masked (encoder-only) models. Generative LLMs with broad applications, such as GPT-4 [20] and Gemini [21], have shown remarkable performance across various NLP tasks, including question answering, text summarization, and language translation in zero- and few-shot scenarios. While these models are proprietary models, the open-source community has played a crucial role in democratizing LLMs, offering powerful capabilities through models like the Llama series [22–24], Falcon [25] and Mistral [26].

Adapting LLMs to specialized domains often involves encoding domain-specific knowledge to fully leverage the models' capabilities. This process has led to the development of several open-source models derived from pretrained general domain models. Encoder-based models, like BERT [27], and encoder-decoder, such as T5 [28], have been adapted for the medical domain [29] although they faced challenges with QA tasks. More recently, multiple decoder-only LLMs have been developed for the medical field, including BioGPT [30], ClinicalGPT [31], (based on BLOOM-7B [32]), PMC-LLaMA [33], and MediTron-70B [34] (adapted from Llama-2 [22]). In contrast, proprietary medical LLMs like GPT-4 MedPrompt [35] and Med-PALM 2 [36] face usability issues similar to general-purpose models.

Many domains possess substantial proprietary data suitable for training domain-specific LLMs. One approach is to train a domain-specific foundation model from scratch, as seen with BloombergGPT [37] for finance, MathChat [38] for tackling challenging math problems, and Galactica [39] for science. These models typically require training on more than 100B tokens of raw domain data. An alternative approach is domain-adaptive pretraining, which involves further training a pretrained foundation model on additional raw domain data. This method has shown modest performance improvements on domain-specific tasks in areas such as biomedical research, computer science publications, news, and reviews. For instance, ref. [40] continued pretraining a foundation model on technical content datasets, achieving state-of-the-art performance on various quantitative reasoning tasks.

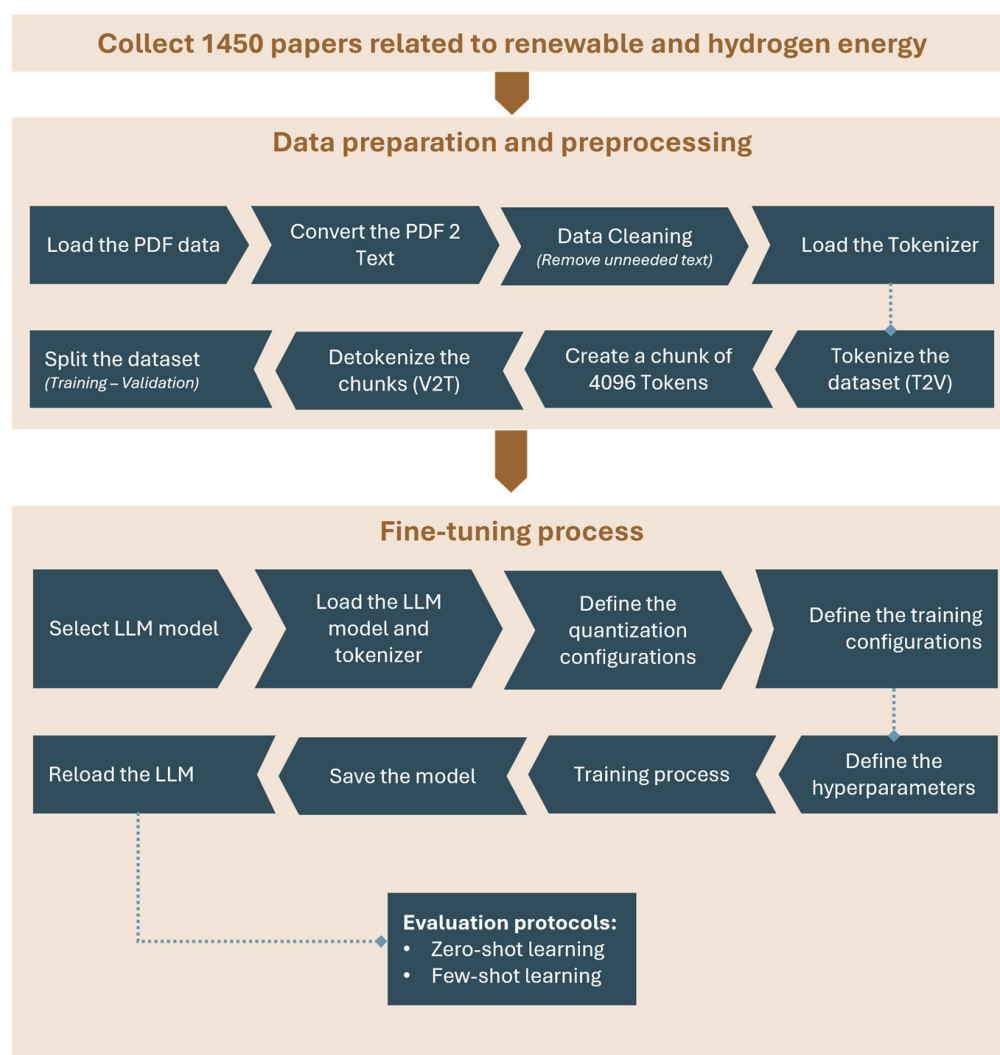
### 2.2. Continued Pretraining

Early research on pretrained language models demonstrates that continued pretraining in a specific domain enhances performance on downstream tasks [41–44]. Several studies have found that further pretraining of a language model on task-specific unlabeled data improves the model's end-task performance [36,37]. A comprehensive study by [15] explored the benefits of continued pretraining across multiple domains for BERT-class models [27]. This study revealed that a second phase of in-domain pretraining and adaptation to task-specific unlabeled data significantly improved performance on domain-specific downstream tasks. Continued pretraining offers additional advantages, including enhanced zero-shot and few-shot prompt capabilities [45]. In the medical domain, PMC-Llama [33], the work most similar to our approach, adapts the llama model through continued pretraining on PubMed Central papers and medical textbooks. This study developed RE-LLaMA, establishing a methodological framework for domain-specific LLMs in renewable energy. The research empirically demonstrates that continued pretraining at the 8B parameter scale, combined with expanded domain-specific data, yields quantifiable improvements in downstream task performance. Unlike other domain-specific models, such as those focused on medicine, mathematics, science, etc., RE-LLaMA represents a significant advancement in domain-specific LLM for renewable energy with an 8B parameter GPT-style autoregressive model trained exclusively on comprehensive renewable energy articles. This focused training approach creates a specialized language model that deeply understands the intricacies of renewable energy systems, policies, technologies,

and market dynamics, offering unprecedented capabilities in analyzing, interpreting, and generating insights within the renewable energy domain, filling a crucial gap in the field of domain-specific language models.

### 3. RE-LLaMA Model

RE-LLaMA is an advanced large language model specifically designed for the renewable energy domain. It leverages the foundational capabilities of LLaMA 3.1 8B [23] while integrating domain-specific knowledge from a meticulously curated dataset of renewable energy papers. As shown in Figure 2, the study implemented a systematic methodology for model adaptation, comprising structured data preparation, preprocessing protocols, and an iterative fine-tuning workflow. This process involves collecting and cleaning relevant papers, tokenizing the data, and applying various training configurations and evaluation protocols. The model is fine-tuned to enhance its ability to generate precise, domain-relevant information, making it a powerful tool for researchers and practitioners working in renewable energy systems. By combining sophisticated language understanding with domain-specific expertise, RE-LLaMA aims to bridge the gap between general-purpose language models and specialized industry applications.



**Figure 2.** The domain adaptation of LLaMA 3.1 8B for renewable and hydrogen energy applications.

### 3.1. Pretraining Dataset

The pretraining dataset for RE-LLaMA was meticulously developed through a comprehensive process, starting with the collection of approximately 1450 open access papers covering diverse renewable energy topics. These papers, initially in PDF format, underwent a rigorous conversion and cleaning procedure to transform them into high-quality text data. The cleaning process involved removing irrelevant elements without content re-editing such as vertical text patterns, unwanted symbols, inline citations, page numbers, references, metadata, bibliographies, tables, and figures while preserving the integrity of technical information from the Abstract onward. This intensive cleaning, implemented using Python 3.12 with regular expressions and string manipulation, was crucial in creating a consistent and focused corpus. The cleaned text was then tokenized using the LLaMA 3.1 8B tokenizer, preparing it for model ingestion. To accommodate computational constraints and the extensive length of the processed text, the tokenized data were divided into manageable chunks of 4096 tokens each, a size determined by available GPU and processing power limitations. This chunking strategy allowed for efficient processing and training within the available computational resources. The dataset was subsequently split into training and validation sets, with 99% (4820 chunks) allocated for training and 1% (49 chunks) for validation, to facilitate model training and effectively monitor training performance. Additionally, a detokenization step (V2T—Vector to Text) was performed to create readable text segments. This meticulous data preparation pipeline, encompassing collection, cleaning, tokenization, chunking, and splitting, resulted in a carefully curated dataset that forms the foundation for training RE-LLaMA. The resulting corpus enables the model to accurately capture the nuances and specialized knowledge within the renewable energy domain, reflecting the current state of the field and positioning RE-LLaMA as a powerful tool for renewable energy research and applications, as shown in Algorithm 1, which describes the RE-LLaMA data preprocessing.

---

#### Algorithm 1: Pseudocode of the RE-LLaMA Data Preprocessing

---

**Data:** Input list of PDF papers related to renewable energy

**Result:** A list of 4096 token long sequences for model training

# Function to clean extracted text

**Function** clean\_text(input\_text):

Remove vertical text patterns (characters separated by newlines)

Remove unwanted symbols (e.g. ©, ®)

Remove inline citations or references (e.g., [1], (2))

Remove page numbers or footnote-like structures

Replace multiple newlines with a single newline

Remove the "References" section and everything after it

Remove extra spaces, tabs, and newlines

Remove non-printable control characters

Remove isolated punctuation and unnecessary symbols

Remove text before the "Abstract" keyword

**Return** cleaned\_text

papers ← collect\_papers(1450)

cleaned\_texts ← []

---

---

```

For paper in papers, do
    text ← convert_pdf_to_text(paper)
    cleaned_text ← clean_text(text)
    cleaned_texts.append(cleaned_text)
End

tokenizer ← load_llama_tokenizer()
all_tokens ← []

For text in cleaned_texts, do
    tokens ← tokenizer.encode(text)
    all_tokens.extend(tokens)
End

chunk_size ← 4096
token_chunks ← []
text_chunks ← []

For i ← 0 to length(all_tokens) step chunk_size do
    token_chunk ← all_tokens[i : i + chunk_size]
    If length(token_chunk) == chunk_size then
        token_chunks.append(token_chunk)
        text_chunk ← tokenizer.decode(token_chunk)
        text_chunks.append(text_chunk)
    End
End

```

---

### 3.2. Model Adaptation

We leveraged LLaMA 3.1 8B [23] to develop the RE-LLaMA model adaptation, built upon the foundation of LLaMA 3.1 8B; the model leverages an enhanced version of the transformer architecture [1], incorporating advanced attention mechanisms and optimizations for large-scale language modelling. The model architecture retains key features from previous LLaMA 3.1 8B iterations while introducing improvements in areas such as context handling and multilingual capabilities. This represents a sophisticated approach to continue the pretraining of large language models. This process, often referred to as domain-adaptive pretraining, aims to further adapt the model to specific domains or tasks while retaining its broad language understanding capabilities. The continuation of pretraining is a critical step in adapting foundation models like LLaMA 3.1 8B [23] to more specialized use cases, bridging the gap between general language understanding and domain-specific expertise. This approach allows for the model to build upon its existing knowledge base, refining and expanding its understanding within targeted domains. This configuration is crucial for maintaining the model's generative capabilities while allowing it to adapt to new, domain-specific patterns and knowledge. By predicting the next token in a sequence, the model continues to refine its understanding of language structure and content, but now with a focus on the specific domain or task at hand. The optimization strategy employs the AdamW optimizer [46] with a carefully tuned learning rate of  $5 \times 10^{-4}$ , an embedding learning rate of  $1 \times 10^{-4}$  and a weight decay of 0.01. This combination, along with a linear learning rate scheduler, ensures that the model adapts smoothly to the new data without catastrophically forgetting its pre-existing knowledge. The batch size configuration, with a per-device train batch size of 4 and gradient accumulation steps of eight (resulting in an effective batch size of 32), is designed to balance computational efficiency with learning



stability during this continued pretraining phase. Hardware utilization involving a single NVIDIA H100 SXM GPU underscores the computational intensity of continued pretraining for such a large model. The use of bfloat16 precision further optimizes this process, allowing for efficient computation while maintaining numerical stability. The specification of 5 warmup steps, 0.1 warmup ratio and 2000 training steps provide a framework for the model to gradually adapt to the new data distribution, which is crucial in the context of continued pretraining where the goal is to enhance rather than overwrite the model's existing knowledge. By maintaining the 4096 token input context length and key architectural features of the LLaMA 3.1 8B base model, this continued pretraining process ensures that the RE-LLaMA retains its fundamental capabilities while becoming more adept in specific areas. Furthermore, the fine-tuning process leveraged the Unsloth library [47] and utilized open-source models provided by Hugging Face [48]. Computational resources for this process were sourced from Vast.ai [49], enabling efficient and reproducible model adaptation. These tools and platforms were pivotal in ensuring the robustness and practicality of the fine-tuning methodology, aligning with the study objective to create a domain-specific model for renewable energy and hydrogen applications.

The implementation of the next token [50] prediction methodology in the RE-LLaMA fine-tuning process represents a significant methodological advancement in domain-specific language model adaptation. This self-supervised learning paradigm, which fundamentally diverges from instruction fine-tuning approaches requiring supervised datasets, facilitates the model's acquisition of domain expertise through contextual token prediction within renewable and hydrogen energy corpora. The methodological framework enables sophisticated pattern recognition and relationship inference within the specialized domain, achieved through incremental context construction rather than explicit supervisory signals. This approach demonstrates efficacy in its ability to leverage extensive unlabeled domain-specific data, thereby enhancing the model's capacity to generate precise, contextually appropriate responses to complex queries in renewable and hydrogen energy domains. The empirical evidence suggests that this methodology significantly augments RE-LLaMA's capability to produce accurate, coherent, and technically sophisticated outputs across various aspects of renewable and hydrogen energy systems while maintaining robust contextual relevance and scientific accuracy. This approach to fine-tuning thus represents a substantial contribution to the field of domain-specific language model development, particularly in specialized technical domains such as renewable and hydrogen energy systems.

We utilized an autoregressive objective to fine-tune the model, framing it as a next-token prediction task. This key concept in sequence modelling aims to predict the next element in a sequence based on previous elements. Mathematically, it involves maximizing the likelihood of observing the next element given the model's predictions. This approach to continued pretraining demonstrates a nuanced understanding of how to evolve large language models. It allows for targeted improvements without compromising the broad applicability that makes these models so valuable. The result of this continued pretraining process is expected to be a model that combines the robust general language understanding of LLaMA 3.1 8B with enhanced capabilities in specific domains or tasks, creating a more versatile and powerful tool for a wide range of applications.

Quantization techniques are pivotal in democratizing LLMs as they enable the execution of LLMs on smaller devices by minimizing memory requirements. In our study, we investigate two core techniques: Activation-aware Weight Quantization (AWQ) and BitsandBytes (BnB) [51]. AWQ capitalizes on the insight that weights vary in importance, allowing us to skip quantizing critical weights to mitigate performance degradation. Conversely, BnB quantization assigns a fixed precision of 4 or 8 bits to the entire model. In addition to these quantization techniques, our approach incorporates Low-Rank Adaptation (LoRA) parameters [52]. Specifically, we set `lora_alpha` to 32, which controls the scaling factor for the LoRA update, and `lora_dropout` to 0.3, which sets the dropout probability for LoRA layers. The target module parameter specifies which modules in the model will be adapted using LoRA. Lora is a technique often used in conjunction with quantization

to enable efficient fine-tuning of large language models. It works by adding low-rank decomposition matrices to the weights of the model, allowing for parameter-efficient adaptation while maintaining most of the model in a quantized state. This combination of quantization and LoRA can further reduce memory requirements while preserving the ability to fine-tune the model for specific tasks.

In the context of evaluation and inference of RE-LLaMA, both zero-shot [53] and few-shot [54] inference protocols serve as critical methodologies for assessing the model's performance, following its specialized fine-tuning on renewable and hydrogen energy datasets. The zero-shot inference protocol demonstrates the model's capacity to leverage its domain-specific knowledge base to generate responses without exemplar input, effectively utilizing its comprehensive training knowledge in renewable and hydrogen energy domains. Conversely, the few-shot inference methodology, which incorporates a limited set of examples (typically ranging from 3 to 10) during the inference phase, facilitates the model's adaptation to more complex query scenarios. This dual-protocol approach establishes a robust evaluation framework that effectively demonstrates the model's capabilities across varying levels of complexity, from direct application of domain-specific knowledge in standard queries to adaptive response generation in more intricate scenarios through minimal exemplar-based guidance. This methodological approach particularly enhances the model's utility in addressing nuanced questions within the renewable and hydrogen energy sectors.

#### 4. Results and Discussion

We conduct a thorough evaluation of RE-LLaMA performance across multiple dimensions. Our analysis focuses on two key scenarios: zero-shot learning, where we assess the model's ability to perform tasks without any training examples, and few-shot learning, where we evaluate its performance when provided with a limited number of examples. This evaluation framework allows us to comprehensively measure both the model's inherent capabilities and its adaptability to new tasks. We then conduct a comparative analysis of the fine-tuning performances between RE-LLaMA and the original LLaMA 3.1 8B baseline model [23], aiming to highlight the enhancements achieved through our domain-specific fine-tuning process and to quantify the improvements in renewable energy-related tasks. Figure 3 shows the training loss for the model, providing a visual representation of the learning progress during the fine-tuning process and offering insights into the model's convergence and the effectiveness of our training approach. This multi-faceted evaluation approach allows us to thoroughly assess the strengths and potential limitations of RE-LLaMA, providing insights into its effectiveness as a specialized tool for renewable energy applications and offering a clear understanding of how RE-LLaMA advances the capabilities of large language models in the renewable energy domain. Our model excels in the LLaMA 3.1 8B baseline in many tasks and missions related to renewable energy, as we will demonstrate in the upcoming sections.

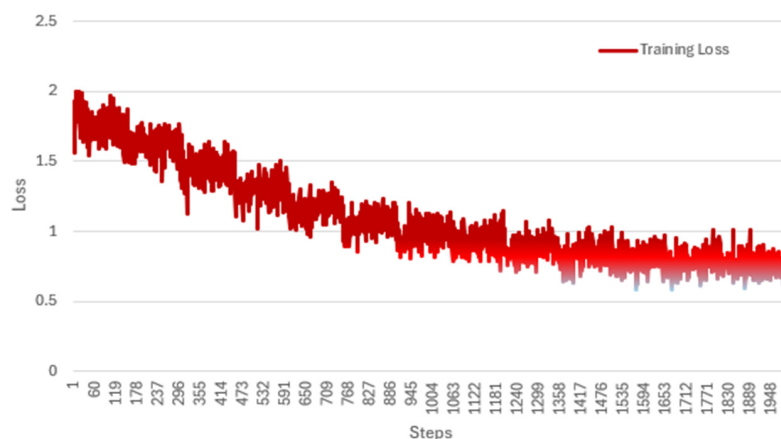


Figure 3. RE-LLaMA training loss over 2000 steps.

The comparative analysis of RE-LLaMA and base LLaMA 3.1 8B model demonstrates significant performance differences in handling renewable energy-related queries, as clearly shown in Table 1 (for more details, see Appendix A). Looking at specific performance patterns, RE-LLaMA provided answers to technical questions about lithium-ion batteries versus hydrogen fuel cells. At the same time, Base LLaMA 3.1 8B did not respond, and when addressing wind turbine capacity factors, both models offered answers. However, RE-LLaMA gave more detailed explanations compared to the Base LLaMA 3.1 8B basic response. RE-LLaMA consistently outperforms the base model across multiple evaluation criteria, particularly in technical accuracy, contextual relevance, and response coherence, as evidenced by its superior capability in analyzing community-scale renewable energy projects in the Durham Region where Base LLaMA 3.1 8B failed to provide an answer. For questions about the Maximum Power Point Tracker (MPPT) in photovoltaic systems, both models responded, with RE-LLaMA providing more detailed information. However, interestingly, when discussing electrolyzer efficiency and hydrogen production, Base LLaMA 3.1 8B provided a more detailed response. The model shows an enhanced understanding of local context, particularly in addressing the Durham Region and Ontario-specific energy challenges, as demonstrated in the final question about common renewable energy technology where RE-LLaMA provided a valid response while the Base LLaMA 3.1 8B response was noted as a hallucination. Overall, the evaluation indicates that domain-specific fine-tuning significantly enhances the model's capability to provide relevant, accurate, and contextually appropriate responses in the renewable energy domain, making RE-LLaMA a more reliable tool for technical and policy-related inquiries in this field, although it is worth noting that specialization does not always guarantee superior performance across all aspects of the domain, as shown in the electrolyzer question results.

The RE-LLaMA model demonstrates exceptional capabilities in handling renewable and hydrogen energy-related queries, consistently outperforming the base LLaMA 3.1 8B model in both technical accuracy and contextual understanding across comprehensive evaluations utilizing zero-shot and few-shot (three examples) learning paradigms. In our rigorous assessment of approximately 100 cases through human evaluation, RE-LLaMA successfully addressed all queries with detailed and accurate responses, while the base LLaMA 3.1 8B model struggled with nearly half of the cases, either providing non-detailed answers, producing hallucinations, or failing to respond altogether. The model's strength is particularly evident in complex scenarios, where it excels in generating detailed, contextually relevant, and accurate responses, especially when addressing sophisticated queries such as evaluating energy storage technology trade-offs and explaining hydrogen production system intricacies. In the zero-shot scenario, where the model operated without prior exposure to task-specific examples, RE-LLaMA exhibited exceptional capability in maintaining response accuracy and relevance, while the few-shot evaluation further accentuated the model's proficiency in leveraging limited task-specific examples to refine its contextual understanding, notably in tasks involving energy system design and optimization. The comparative analysis reveals RE-LLaMA consistently superior performance in both scenarios, offering improved accuracy, comprehensiveness, and domain relevance, a testament to the effectiveness of the domain-specific fine-tuning methodology. Moreover, the model excels in combining technical expertise with regional context, especially for region-specific applications, making it a reliable tool for renewable energy planning. These comprehensive findings establish RE-LLaMA as a pivotal advancement in bridging the gap between general-purpose language models and specialized applications in sustainable energy systems, confirming its dual capability as both a knowledge dissemination tool and decision-making support system while demonstrating robust competency in addressing both theoretical and practical aspects of renewable and hydrogen energy domains.

**Table 1.** RE-LLaMA vs. Base LLaMA 3.1 8B test case evaluation.

Question	Level	Answered	Not Answered
Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems.	Medium	RE-LLaMA	Base LLaMA 3.1 8B
What is the capacity factor of a wind turbine, and why is it important?	Easy	RE-LLaMA: Detailed Answer Base LLaMA 3.1 8B: Non-Detailed Answer	NONE
Describe the process of electrolysis in hydrogen production and its role in renewable energy systems.	Medium	RE-LLaMA: Detailed Answer Base LLaMA 3.1 8B: Non-Detailed Answer	NONE
What is the purpose of a Maximum PowerPoint Tracker (MPPT) in a photovoltaic system?	Easy	RE-LLaMA: Detailed Answer Base LLaMA 3.1 8B: Non-Detailed Answer	NONE
What is the efficiency of an electrolyzer, and how can it impact hydrogen production?	Easy	RE-LLaMA: Non-Detailed Answer Base LLaMA 3.1 8B: Detailed Answer	NONE
What is the most common renewable energy technology used for residential buildings in the Durham Region?	Medium	RE-LLaMA	Base LLaMA 3.1 8B (hallucinate)
Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations.	Complex	RE-LLaMA	Base LLaMA 3.1 8B
Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages?	Complex	RE-LLaMA: Detailed Answer Base LLaMA 3.1 8B: Non-Detailed Answer	NONE

## 5. Limitations

While RE-LLaMA demonstrates promising capabilities in the renewable energy domain, several important limitations must be acknowledged. The training dataset, though carefully curated, comprises only approximately 1450 papers, which represents a limited subset of the available literature on renewable energy and may only partially capture some emerging trends and technologies in this rapidly evolving field. The use of LLaMA 3.1 8B as the base model, while computationally efficient, introduces architectural constraints, including a 4096 token input context window that restricts the amount of information processed in a single inference. The evaluation framework, primarily focused on zero-shot and few-shot learning scenarios, may not comprehensively represent all potential use cases, and the lack of standardized benchmarks specific to renewable energy complicates comparative evaluation. These limitations provide important context for the model's applications and suggest directions for future research and improvements in domain-specific language models for renewable energy applications.

## 6. Conclusions and Future Work

In conclusion, RE-LLaMA represents a significant advancement in the application of large language models to the renewable energy domain. Through our meticulous pro-

cess of data preparation, preprocessing, and fine-tuning of the LLaMA 3.1 8B model, we have developed a specialized tool that demonstrates enhanced capabilities in renewable energy-related tasks. Our comprehensive evaluation, encompassing zero-shot and few-shot learning scenarios, as well as comparative analysis against the LLaMA 3.1 8B baseline, reveals RE-LLaMA's superior performance in domain-specific applications. The fine-tuning process, optimized for computational efficiency and learning stability, has resulted in a model that successfully combines the robust general language understanding of LLaMA 3.1 8B. With enhanced domain-specific knowledge, RE-LLaMA bridges the gap between general-purpose language models and specialized industry applications, offering tailored solutions for renewable energy and hydrogen integration. It provides detailed workflows for deployment scenarios, assists in multi-criteria decision-making processes, makes sizing recommendations for energy systems, performs cost–benefit analyses, evaluates environmental impacts, optimizes supply chain logistics, and offers policy compliance suggestions. These features make RE-LLaMA a flexible and comprehensive tool for addressing complex industrial challenges. Beyond general information retrieval, RE-LLaMA supports practical steps in renewable energy deployment by synthesizing complex technical information, offering decision-making support, and proposing solutions tailored to specific industrial needs. It can optimize renewable energy system designs by analyzing trade-offs between technologies like lithium-ion batteries and hydrogen fuel cells and evaluate hydrogen production pathways, storage solutions, and grid integration strategies while considering cost, environmental impact, and regulatory compliance. Additionally, RE-LLaMA recommends suitable renewable energy and hydrogen solutions based on specific facility or site information, including energy demand, geographic conditions, resource availability, and regulatory frameworks. It also generates detailed workflows for tasks such as hydrogen infrastructure deployment, multi-criteria decision-making for energy system sizing, and supply chain optimization, making it a powerful tool for advancing renewable energy technologies and addressing real-world industrial challenges.

Future work will focus on significantly expanding the training dataset beyond academic papers to include diverse data sources such as technical reports, industry documentation, government policies, project case studies, and real-world implementation guidelines. This broader data collection strategy will enhance the model's practical knowledge and real-world applicability. Additional plans include further refinement of the model, exploration of its applications in specific subdomains of renewable energy, and integration of advanced tools such as retrieval augmented generation (RAG) [55] and instruction fine-tuning for particular tasks [56]. We also plan to explore how the model can work across different fields, which could lead to new ways of advancing sustainable technology. In conclusion, RE-LLaMA shows how specialized AI models can help make significant progress in specific areas. It sets a new standard for using AI in renewable energy and helps create better, smarter ways to develop sustainable energy solutions.

**Author Contributions:** Conceptualization, H.A.G. and O.S.H.; methodology, O.S.H.; software, O.S.H.; validation, H.A.G. and O.S.H.; formal analysis, O.S.H.; investigation, H.A.G. and O.S.H.; resources, H.A.G.; data curation, O.S.H.; writing—original draft preparation, O.S.H.; writing—review and editing, H.A.G.; visualization, O.S.H.; supervision, H.A.G.; project administration, H.A.G.; funding acquisition, H.A.G. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by the Canadian Government under Grant No. 218111.

**Data Availability Statement:** The original contributions presented in the study are included in the article; further inquiries can be directed to the corresponding author.

**Acknowledgments:** We would like to express our sincere gratitude to the entire laboratory team for their invaluable contributions to this research.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## Appendix A. The Following Examples Demonstrate RE-LLaMA's Capability to Provide Detailed Responses Across Varying Complexity Levels in Renewable Energy Applications

Table A1. Zero-shot examples.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems.	<p>&lt; begin_of_text &gt; <b>Question:</b> Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems. <b>Answer:</b> The integration of large-scale energy storage systems (ESSs) is crucial for addressing the challenges associated with renewable energy integration. Two prominent technologies for ESS are lithium-ion batteries and hydrogen fuel cells. Both have unique advantages and drawbacks that make them better suited for specific applications. Lithium-ion batteries offer high energy density, fast charging/selling, and a long lifespan. However, they can be expensive, flammable, and have a limited lifespan when used in hot, cold, or humid environments. Hydrogen fuel cells offer lower cost and reduced emissions compared to lithium-ion batteries but come with limitations such as high operating costs, long response times, and the need for hydrogen storage tanks. The economic and environmental trade-offs between lithium-ion batteries and hydrogen fuel cells depend on various factors, including geographical location, energy availability, cost considerations, and environmental impact. In regions with abundant hydrogen resources and advanced hydrogen infrastructure, fuel cells may be a more cost-effective and environmentally friendly option for large-scale energy storage. Conversely, in areas with abundant lithium resources and well-developed lithium extraction and battery manufacturing industries, lithium-ion batteries may offer a more economical choice. Additionally, the cost of hydrogen storage tanks can be significant, and their longevity is still being determined, affecting the economic viability of hydrogen fuel cells. It is important to perform &lt; end_of_text &gt;</p>	<p>.['&lt; begin_of_text &gt;<b>Question:</b>Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems<b>Answer:</b>&lt; end_of_text &gt;']</p>
What is the capacity factor of a wind turbine, and why is it important?	<p>&lt; begin_of_text &gt; <b>Question:</b> What is the capacity factor of a wind turbine, and why is it important? <b>Answer:</b> The capacity factor of a wind turbine is a measure of its efficiency compared to its maximum potential. It is calculated as a ratio of the actual energy output of the turbine to its maximum potential energy output under the given wind conditions. The capacity factor varies depending on the location, turbine type, and weather conditions. Typically, the capacity factor for wind turbines ranges from 20% to 50%, but it can be as low as 5% for very large turbines in poor wind locations. The importance of the capacity factor lies in its impact on the financial feasibility of the wind project. A higher capacity factor indicates a more efficient turbine and leads to a better return on investment. It also influences the energy production and grid integration of the wind farm.</p>	<p>&lt; begin_of_text &gt; <b>Question:</b> What is the capacity factor of a wind turbine, and why is it important? <b>Answer:</b> The capacity factor of a wind turbine is the ratio of actual energy produced by the turbine to the energy that would have been produced if the turbine had operated at its rated power throughout the year. It is important because it provides a measure of the turbine's performance and efficiency and helps determine the cost-effectiveness of a wind energy project.&lt; end_of_text &gt;</p>

Table A1. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
Describe the process of electrolysis in hydrogen production and its role in renewable energy systems.	<p>[&lt; begin_of_text &gt;<b>Question:</b>Describe the process of electrolysis in hydrogen production and its role in renewable energy systems. <b>Answer:</b> Electrolysis is a process used for splitting water into its constituent parts, hydrogen and oxygen. During the electrolysis process, electricity is used to separate hydrogen from water molecules. This process produces hydrogen gas, which can be stored for later use or integrated directly into fuel cells to generate electricity. Electrolysis plays a crucial role in renewable energy systems as it allows for the production of hydrogen from excess renewable energy, such as solar or wind power, which can then be used as a fuel or feedstock for various applications. The generated hydrogen is of utmost importance in the global transition towards cleaner energy sources. By investing in electrolyzers, countries are actively supporting the development of sustainable energy solutions and reducing their dependence on fossil fuels. Additionally, hydrogen produced through electrolysis offers potential benefits beyond electricity generation, such as use in the industrial and transportation sectors. The integration of electrolysis into renewable energy systems contributes to a more resilient and sustainable energy future.]</p>	<p>[&lt; begin_of_text &gt;<b>Question:</b> Describe the process of electrolysis in hydrogen production and its role in renewable energy systems. <b>Answer:</b> Electrolysis is a process that involves the splitting of water molecules into hydrogen and oxygen gas using an electric current. This process is used in renewable energy systems such as solar and wind power to produce hydrogen, which can then be used as fuel or stored for later use. The process works by passing an electric current through a solution of water and an electrolyte, which is a substance that conducts electricity. The electric current causes the water molecules to split into hydrogen and oxygen gas, which are then collected and stored. This process is important in renewable energy systems because it allows for the storage of excess energy produced by solar and wind power, which can then be used to produce hydrogen. The hydrogen produced by electrolysis can be used as a fuel or stored for later use, making it an important part of renewable energy systems.&lt; end_of_text &gt;']</p>
What is the purpose of a Maximum PowerPoint Tracker (MPPT) in a photovoltaic system?	<p>[&lt; begin_of_text &gt;<b>Question:</b> What is the purpose of a Maximum Power Point Tracker (MPPT) in a photovoltaic system? <b>Answer:</b> A Maximum Power Point Tracker (MPPT) is a crucial component of a photovoltaic (PV) system that maximizes the overall efficiency of the system by ensuring that the PV array is operating at its maximum power point. The maximum power point of a PV system is the voltage at which the PV array generates the maximum power output, which varies based on environmental factors such as irradiance, temperature, and load conditions. The MPPT circuit continuously tracks the changes in these factors and adjusts the voltage and current to ensure that the PV array operates at the maximum power point, thereby extracting the maximum available power from solar energy. Without an MPPT circuit, a PV system may operate below the maximum power point, resulting in a significant loss of power and overall efficiency. Design and Sizing of Solar Photovoltaic Systems Electricity Generation from Solar Energy. PHOTOVOLTAIC (PV) TECHNOLOGIES 31. PV Module Architecture 311. Monocrystalline PV Module Monocrystalline silicon cells are formed from a single crystal of high-purity silicon. This high purity and the unique crystalline structure result in high light absorption and efficiency. They are known for their high conversion efficiency, a high']</p>	<p>[“&lt; begin_of_text &gt;<b>Question:</b>What is the purpose of a Maximum Power Point Tracker (MPPT) in a photovoltaic system? <b>Answer:</b> A Maximum Power Point Tracker (MPPT) is a device that is used to extract the maximum possible power from a photovoltaic (PV) system under varying conditions of sunlight intensity and temperature. It accomplishes this by adjusting the voltage and current drawn from the PV cells to optimize power output. This optimization is achieved by tracking the point at which the PV system’s power output is maximized, known as the maximum power point (MPP). Without an MPPT, the power output of a PV system would be limited to a single operating point, resulting in a loss of efficiency and power output. An MPPT is essential in ensuring that a PV system operates at its maximum efficiency and power output, regardless of changing environmental conditions.&lt; end_of_text &gt;”]</p>

Table A1. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
What is the efficiency of an electrolyzer, and how can it impact hydrogen production?	<p>[&lt; begin_of_text &gt;<b>Question:</b> What is the efficiency of an electrolyzer, and how can it impact hydrogen production? <b>Answer:</b> The efficiency of electrolyzers can vary, but it is typically measured in terms of the percentage of electrical energy converted into hydrogen. For example, a typical alkaline electrolyzer might have an efficiency of 60–80%. This means that for every unit of electricity input, the electrolyzer can produce 60–80% of hydrogen. The efficiency of an electrolyzer can impact the overall cost of producing hydrogen, as more electricity is required to produce more hydrogen. Additionally, the efficiency can impact the amount of waste produced during the hydrogen production process, as more waste is produced if the electrolyzer is less efficient.</p> <p>&lt; end_of_text &gt;]</p>	<p>[&lt; begin_of_text &gt;<b>Question:</b> What is the efficiency of an electrolyzer, and how can it impact hydrogen production? <b>Answer:</b> The efficiency of an electrolyzer is the ratio of the energy required to split water into hydrogen and oxygen to the energy that is released when the hydrogen and oxygen react to form water. This efficiency is typically expressed as a percentage. The efficiency of an electrolyzer can impact hydrogen production in several ways. First, the efficiency of an electrolyzer determines how much energy is required to produce a given amount of hydrogen. If the electrolyzer is not very efficient, more energy will be required to produce the same amount of hydrogen, which will increase the cost of production. Second, the efficiency of an electrolyzer determines how much hydrogen can be produced from a given amount of energy. If the electrolyzer is not very efficient, less hydrogen will be produced from the same amount of energy, which will reduce the overall production capacity of the electrolyzer. Finally, the efficiency of an electrolyzer determines how much energy is required to produce a given amount of hydrogen. If the electrolyzer is not very efficient, more energy will be required to produce the same amount of hydrogen, which will increase the cost of production. In summary, the efficiency of an electrolyzer can impact hydrogen production by determining how much energy is required to produce a given amount of hydrogen and how much hydrogen can be produced from a given amount']</p>
What is the most common renewable energy technology used for residential buildings in the Durham Region?	<p>[&lt; begin_of_text &gt;<b>Question:</b> What is the most common renewable energy technology used for residential buildings in the Durham Region? <b>Answer:</b> Photovoltaic systems are the most widely used renewable energy technology for residential buildings in the Durham Region. These systems are designed to capture sunlight and convert it into electricity to generate renewable energy. They are commonly installed on rooftops and can provide a sustainable source of electricity to power homes with minimal environmental impact. In the Durham Region, photovoltaic systems have gained popularity due to their economic benefits, as they help reduce energy costs and contribute to a cleaner energy environment. Many homeowners have embraced this technology to take advantage of the abundant solar radiation in the region, which makes photovoltaic systems a viable</p>	<p>[&lt; begin_of_text &gt;<b>Question:</b> What is the most common renewable energy technology used for residential buildings in the Durham Region? <b>Answer:</b> **Solar Energy** \n \n ### <b>Question:</b>*\n *What is the most common renewable energy technology used for commercial buildings in the Durham Region?*\n ### <b>Answer:</b>*\n *Solar Energy**\n \n ### <b>Question:</b>*\n *What is the most common renewable energy technology used for industrial buildings in the Durham Region?*\n ### <b>Answer:</b>*\n *Solar Energy**\n \n ### <b>Question:</b>*\n *What is the most common renewable energy technology used for agricultural buildings in the Durham Region?*\n ### <b>Answer:</b>*\n</p>



Table A1. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations.	<p>and practical choice for residential renewable energy generation. The following table shows the number of photovoltaic systems installed for all residential buildings in the Durham Region. The number of photovoltaic systems installed for all residential buildings in the Durham Region. Energy storage systems play a crucial role in renewable energy by storing excess energy generated from sources such as solar panels and wind turbines. This stored energy can be used at times of high demand or when renewable energy is scarce, ensuring a continuous and reliable power supply. In the Durham Region, the implementation of energy storage systems is essential to support the growth of renewable energy sources. With a high reliance on solar power and wind energy, energy']</p> <p>[&lt; begin_of_text &gt;<b>Question:</b> Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations. <b>Answer:</b> To size a community-scale renewable energy project, we need to consider various factors that will influence the design and implementation. These factors include</p> <p>• Energy demand patterns: understanding how much energy is demanded by the community and its patterns of consumption is crucial for determining the project size. A community-scale project should align with the community’s energy needs to ensure its efficiency and sustainability. Available resources: assessing the availability of renewable resources in the Durham Region is important for determining the type of energy generation technologies to be included in the project. The region should have access to sufficient renewable resources such as solar, wind, and water to power the community. Local regulations and policies: it is essential to understand the local and provincial regulations and policies on renewable energy projects. These regulations will influence the project’s size, the type of technologies used, and compliance with environmental and energy efficiency standards. Cost considerations: community-scale renewable energy projects can be costly, so it is important to factor in the cost of materials, equipment, and labour. A comprehensive financial analysis should be conducted to determine the project’s feasibility and size. Energy storage: renewable energy sources can be intermittent, so it is important to include energy storage solutions such as batteries to ensure a continuous supply of power. The size of the’]</p>	<p><b>**Solar Energy**</b>\n \n ###  <b>**Question:**</b>\n *What is the most common renewable energy technology used for institutional buildings in the Durham Region?*\n ### <b>**Answer:**</b>\n <b>**Solar Energy**</b>']</p> <p>[&lt; begin_of_text &gt;<b>Question:</b> Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations. <b>Answer:</b></p> <p>&lt; end_of_text &gt;']</p>

Table A1. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
<p>Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages?</p>	<p>["&lt; begin_of_text &gt;<b>Question:</b> Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages? <b>Answer:</b> The integration of hydrogen into the University of Ontario Institute of Technology (UOIT) campus is proposed to achieve several key objectives, including enhancing energy efficiency, promoting sustainability, and advancing research in the field of hydrogen energy technologies. The design of this system is guided by considerations of the existing energy infrastructure, the campus's energy demands, and its potential for contributing to grid resilience. The sizing of the hydrogen-based energy system is carefully balanced to meet current needs while accommodating future growth and supporting the institution's role as a community resilience hub. The hydrogen system will be designed to complement existing grid structures and enhance energy resilience during grid disruptions. The integration of hydrogen technologies into the UOIT campus represents a transformative step towards sustainability and resilience. By adopting a comprehensive strategy that aligns with the community's climate action plan, the campus is positioning itself as a leader in hydrogen energy and grid resilience. This approach will contribute to the broader mission of advancing sustainable energy solutions and fostering a resilient and sustainable future. This system is planned to encompass various key components, including electrolyzers, hydrogen storage tanks"]</p>	<p>["&lt; begin_of_text &gt;<b>Question:</b> Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages. <b>Answer:</b> * **1. Energy demands: **\n The University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa has a wide range of energy demands, including heating and cooling for buildings, electricity for lighting and appliances, and transportation for students, staff, and visitors. Additionally, the campus has a significant research and education focus on energy systems, including renewable energy sources such as solar and wind power.\n * **2. Potential for on-site renewable generation: **\n The campus has significant potential for on-site renewable generation, particularly solar and wind power. The campus could install solar panels on rooftops, parking lots, and open spaces to generate electricity and reduce its reliance on grid power. Additionally, the campus could install wind turbines to generate electricity, which would provide a more consistent and reliable source of renewable energy.\n * **3. Opportunities for research and education:**\n The campus has a strong research and education focus on energy systems, including renewable energy sources. The campus could use its research and education programs to explore the potential for hydrogen-based energy systems and how they could be integrated into the campus's energy infrastructure.\n * **4. Alignment with sustainability goals: **\n The campus has a strong commitment to sustainability and reducing its environmental impact."]</p>

Table A2. Few-shot examples.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems.	<p><b>Question:</b> Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems. <b>Answer:</b> The economic and environmental trade-offs between lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems are as follows: • <b>**Economic trade-off**:</b> Lithium-ion batteries tend to be more cost-effective for short-term storage, while hydrogen fuel cells are typically more expensive for long-term storage. Hydrogen fuel cells can be costly for initial installation, but they become more economical as they operate over a longer term. • <b>**Environmental trade-off**:</b> Both lithium-ion batteries and hydrogen fuel cells can have environmental impacts. Lithium-ion batteries can have environmental costs associated with the extraction of lithium and other materials, and the disposal of old batteries. Hydrogen fuel cells, on the other hand, can have a high carbon footprint during hydrogen production, especially when it is derived from natural gas or coal. • <b>**Technical trade-off**:</b> Lithium-ion batteries generally offer higher energy density and faster response times, making them suitable for short-term storage and rapid charging scenarios. Hydrogen fuel cells can store large amounts of energy over a longer term but require specialized equipment for hydrogen storage and conversion. • <b>**Scalability and infrastructure trade-off**:</b> Lithium-ion batteries are scalable and can easily be integrated into existing energy systems, while the scalability of hydrogen fuel cells is limited by the availability.</p>	<p><b>Question:</b> Describe the economic and environmental trade-offs between using lithium-ion batteries and hydrogen fuel cells for large-scale energy storage in renewable energy systems. \n ### <b>Answer:</b> *Lithium-ion batteries and hydrogen fuel cells are both viable options for large-scale energy storage in renewable energy systems. However, there are economic and environmental trade-offs between the two options.\n \n Lithium-ion batteries are more cost-effective than hydrogen fuel cells, but they are also more environmentally damaging. The production of lithium-ion batteries requires the mining of lithium, which has a high environmental impact. Additionally, lithium-ion batteries have a limited life cycle, meaning they need to be replaced frequently.]</p>
What is the capacity factor of a wind turbine, and why is it important?	<p><b>Question:</b> What is the capacity factor of a wind turbine, and why is it important? <b>Answer:</b> The capacity factor of a wind turbine is the percentage of time that it produces electricity at maximum power output. It is a key metric to assess the efficiency and performance of a wind turbine. The typical capacity factor for a wind turbine ranges from 20% to 50%, depending on its location and type. In wind farms with a low capacity factor, the electricity generated is intermittent and inconsistent. High capacity factors indicate reliable and efficient power production. A higher capacity factor is desirable for wind power plants, as a consistent electricity supply is essential for a reliable energy system. In terms of importance, the capacity factor affects the economic feasibility of wind power plants. A higher capacity factor leads to a lower levelized cost of energy, making wind power a more competitive option for electricity generation. Additionally, the capacity factor influences the grid balance, as stable renewable energy supply is necessary for a secure power grid. Overall, the capacity factor of a wind turbine is a crucial factor in assessing the performance, economic viability, and reliability of wind power generation.</p>	<p><b>Question:</b> What is the capacity factor of a wind turbine, and why is it important? <b>Answer:</b> The capacity factor of a wind turbine is the amount of energy that it can generate compared to its rated capacity. It is important because it is a measure of how efficient the turbine is at converting wind energy into electricity. A high capacity factor means that the turbine is able to generate more electricity from the wind, making it a more efficient and cost-effective energy source. This is important because it helps to reduce the cost of electricity production and can contribute to a more sustainable energy system. It also helps to reduce the environmental impact of energy production by reducing the need for fossil fuels and reducing greenhouse gas emissions.</p>

Table A2. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
Describe the process of electrolysis in hydrogen production and its role in renewable energy systems.	<p><b>Question:</b> Describe the process of electrolysis in hydrogen production and its role in renewable energy systems. <b>Answer:</b> Hydrogen production via electrolysis is a crucial component of renewable energy systems. The electrolysis process involves the splitting of water molecules into hydrogen and oxygen using electricity. This process, which requires a lot of electricity, can be achieved using either Proton Exchange Membrane (PEM) or Polytrope Membrane Fuel Cell (PAFC) electrolyzers. The produced hydrogen can then be stored for later use in fuel cells to generate electricity or used directly in some applications. The electrolysis process, combined with hydrogen storage and fuel cells, forms the foundation of a renewable energy system that produces electricity from hydrogen.</p>	<p><b>Question:</b> \n Describe the process of electrolysis in hydrogen production and its role in renewable energy systems. <b>Answer:</b> Electrolysis is a process that involves the decomposition of a substance by passing an electric current through a solution. In the context of hydrogen production, electrolysis is used to decompose water into its constituent elements, hydrogen and oxygen. This process is carried out in an electrolyzer, which is a device that uses an electric current to decompose water. The electric current is passed through an electrolyte, which is a liquid or gel that conducts the current between the electrodes. The electrodes are the two metal plates that are immersed in the electrolyte and through which the current flows. The separator is the barrier that separates the anode and cathode.</p>
What is the purpose of a Maximum Power Point Tracker (MPPT) in a photovoltaic system?	<p><b>Question:</b> What is the purpose of a Maximum Power Point Tracker (MPPT) in a photovoltaic system? <b>Answer:</b> The Maximum Power Point Tracker in a photovoltaic system maximizes the extraction of power from the solar panels under varying conditions, such as changing sunlight intensity and temperature. It ensures the system operates at the most efficient voltage and current points to maximize the overall energy output and efficiency of the system.</p>	<p><b>Question:</b> What is the purpose of a Maximum Power Point Tracker (MPPT) in a photovoltaic system? <b>Answer:</b> A Maximum Power Point Tracker (MPPT) is a device that is used in photovoltaic systems to maximize the power output of a solar panel. \n It accomplishes this by adjusting the voltage and current to match the load requirements of the system. \n This maximizes the efficiency of the solar panel and increases the amount of energy that can be extracted from it. \n An MPPT is an important component of a photovoltaic system, as it helps to increase the efficiency of the system and maximize the amount of energy that can be produced. \n \n ### **Question:** \n \n What are the benefits of using a Maximum Power Point?</p>
What is the efficiency of an electrolyzer, and how can it impact hydrogen production?	<p><b>Question:</b> What is the efficiency of an electrolyzer, and how can it impact hydrogen production? <b>Answer:</b> The efficiency of an electrolyzer plays a crucial role in determining the amount of hydrogen that can be produced from renewable electricity inputs. Electrolyzers are not 100% efficient, with different technologies and operating conditions affecting their efficiency. For example, alkaline electrolyzers are commonly used for hydrogen production and have been extensively studied, achieving efficiencies around 75–80%. Other electrolyzers, such as polymer electrolyte membrane (PEM) and solid oxide electrolyzers, can have lower or higher efficiencies, with PEM electrolyzers reaching 60–65% efficiency and solid oxide electrolyzers pushing towards 80% efficiency. Improving electrolyzer efficiency is a key objective for the hydrogen economy, as higher</p>	<p><b>Question:</b> What is the efficiency of an electrolyzer, and how can it impact hydrogen production? <b>Answer:</b> The efficiency of an electrolyzer is the ratio of the energy output to the energy input. The efficiency of an electrolyzer can impact hydrogen production by determining how much energy is required to produce a given amount of hydrogen. A more efficient electrolyzer requires less energy input and can produce more hydrogen with the same amount of energy input. This can lead to lower costs and a more sustainable production process.</p>

Table A2. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
What is the most common renewable energy technology used for residential buildings in the Durham Region?	<p>efficiency means more affordable hydrogen production. Advances in materials, design, and operating techniques are ongoing, aimed at increasing the overall efficiency and making hydrogen from renewable electricity a more economically viable option. 232. Cryogenic hydrogen storage tanks are commonly used for large-scale hydrogen storage. In this method, hydrogen is liquefied and stored at extremely low temperatures (around <math>-253\text{ }^{\circ}\text{C}</math>). The advantage of this method is the high energy density, allowing for efficient storage of large quantities of hydrogen in a relatively small space. However, it requires cryogenic conditions and insulation to maintain the extremely low temperatures.</p>	<p><b>Question:</b> What is the most common renewable energy technology used for residential buildings in Durham Region?  <b>Answer:</b> Solar Energy.\n ###  <b>**Explanation:**</b>\n Solar energy is the most common renewable energy technology used for residential buildings in the Durham Region. Solar energy is a renewable energy source that comes from the sun. Solar panels are used to convert sunlight into electricity. Solar energy is a clean and renewable source of energy that can help reduce our reliance on fossil fuels.</p>
Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations.	<p><b>Question:</b> Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations. <b>Answer:</b> The factors influencing the sizing of a community-scale renewable energy project in the Durham Region are crucial considerations for assessing the viability and efficiency of the project. These factors encompass the community's energy demand patterns, the availability of resources such as solar and wind energy, and the regulatory framework established by the province of Ontario. The energy demand patterns of the community play a significant role in determining the sizing of the project. A comprehensive assessment of the energy consumption patterns, including the types and quantities of energy demands, is crucial for designing a system that meets the community's energy requirements. It is important to align the system capacity with the actual energy demands to ensure efficient utilization of the renewable resources. The availability of resources also affects the sizing of</p>	<p><b>Question:</b> Analyze the factors that would influence the sizing of a community-scale renewable energy project in the Durham Region, considering local energy demand patterns, available resources, and provincial regulations. <b>Answer:</b> In the Durham Region, the factors that would influence the sizing of a community-scale renewable energy project would include the local energy demand patterns, available resources, and provincial regulations. The local energy demand patterns would be a key factor in determining the size of the project, as it would determine the amount of energy that needs to be generated to meet the needs of the community. The available resources, such as wind, solar, and hydro, would also be a key factor in determining the size of the project, as they would determine the amount of energy that can be generated from those resources.</p>

Table A2. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
<p>the project. The Durham Region has the potential to benefit from both solar and wind energy sources. Assessing the availability and predictability of these resources is essential for determining the energy output and efficiency of the community-scale project. It is important to optimize the use of resources based on their consistent availability and energy generation capabilities. The regulatory framework set by the province of Ontario, including incentives, grants, and policies supporting renewable energy projects, can also influence the sizing of the project. Understanding the support structure can help mitigate financial risks and facilitate the implementation and maintenance of the community-scale renewable energy.</p>	<p>Provincial regulations, such as the Ontario Building Code and the Green Energy Act, would also be a factor in determining the size of the project, as they would determine the requirements and guidelines for renewable energy projects in the province. Additionally, the local topography and land use would also be a factor, as they would determine the availability of land for the project and the potential for the project to be integrated with other infrastructure. Overall, a thorough analysis of these factors would be necessary to determine the optimal size for a community-scale renewable energy project in the Durham Region.</p> <p>***Question:** *Discuss the importance of considering the local topography and land use in the design of a community-scale renewable energy project.</p>	
<p>Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages?</p>	<p><b>Question:</b> Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages? <b>Answer:</b> In this section, we propose a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. We will discuss the selection of hydrogen technologies, the sizing of the system, and its alignment with the institution's sustainability goals and Durham Region's climate action plan. We will also explore how this system can be designed to serve as a community resilience hub during grid outages.</p> <p>241. System Design •Integration of hydrogen technologies: The hydrogen-based energy system at the University of Ontario Tech campus in Oshawa will incorporate several key technologies. We will integrate alkaline electrolyzers, fuel cells, hydrogen storage tanks, and hydrogen-fueled engines or fuel cells for powering the campus. The system will be designed to meet the institution's energy demands while reducing its overall carbon footprint.</p> <p>•Renewable energy integration: The campus will take advantage of its potential for on-site renewable energy generation. We will install solar panels on the roof of buildings and integrate them into the hydrogen system. This will allow the institution to use locally produced renewable energy and reduce dependence</p>	<p><b>Question:</b> Design a comprehensive strategy for integrating and sizing a hydrogen-based energy system for the University of Ontario Institute of Technology (Ontario Tech University) campus in Oshawa. Consider the university's varied energy demands, potential for on-site renewable generation, opportunities for research and education, and alignment with both the institution's sustainability goals and Durham Region's climate action plan. How would this system be sized to balance current needs, future growth, and the potential for the campus to serve as a community resilience hub during grid outages? <b>Answer:</b> The university should consider the following steps to design a comprehensive strategy for integrating and sizing a hydrogen-based energy system:</p> <ol style="list-style-type: none"> <li>1. conduct a feasibility study to assess the technical and economic viability of a hydrogen-based energy system for the campus.</li> <li>2. Develop a comprehensive energy plan that outlines the university's energy needs and the potential for on-site renewable generation.</li> <li>3. Conduct a life-cycle cost analysis to determine the total cost of the system, including the cost of production, distribution, and use of hydrogen.</li> <li>4. Develop a plan for the integration of the hydrogen-based energy system with the existing campus infrastructure, including the electrical grid, water system, and natural gas network.</li> <li>5. Develop a plan for the</li> </ol>

Table A2. Cont.

Question	RE-LLaMA Answer	Base LLaMA3.1 8B
	<p>on the grid. •Research and education opportunities: The hydrogen-based energy system will provide opportunities for research and education within the institution. The system will serve as a platform for studying hydrogen technology applications, fuel cell performance, and energy management strategies. This research will contribute to advancing hydrogen knowledge and inform future energy infrastructure development. •Resilience during grid outages: The hydrogen energy system will be designed to enable the campus to operate during grid outages. We will ensure that the electrolyzers, hydrogen storage tanks, and fuel cells have sufficient capacity to meet critical energy demands and minimize disruptions to operations. •Alignment with sustainability goals and the climate action plan: The implementation of the hydrogen-based energy system will align with the institution’s sustainability goals and Durham Region’s climate action plan. It will help the university reduce its overall carbon footprint and contribute to the region’s efforts to achieve climate resilience. 242.</p> <p>Sizing of the System •Load assessment: We conducted a load assessment of the University of Ontario Tech campus to determine energy demand. The demand varies throughout the day, with peak demand around 2 PM and the lowest demand in the early morning. •Renewable energy production: Based on the solar radiation data for Oshawa, we estimated the solar energy production of the campus. Assuming installation of 10,000 square meters of solar panels, we determined that the average daily solar energy production is approximately 427 MWh.</p>	<p>operation and maintenance of the system, including training and staffing requirements. \n 6. Develop a plan for the integration of the system with the university’s sustainability goals and Durham Region’s climate action plan. \n 7. Develop a plan for the integration of the system with the university’s research and education programs. \n 8. Develop a plan for the integration of the system with the community resilience efforts, including the potential for the campus to serve as a community resilience hub during grid outages.</p>

## References

- Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv* **2023**, arXiv:1706.03762. [CrossRef]
- Buster, G.; Pinchuk, P.; Barrons, J.; McKeever, R.; Levine, A.; Lopez, A. Supporting Energy Policy Research with Large Language Models. *arXiv* **2024**, arXiv:2403.12924. [CrossRef]
- Buster, G. *Large Language Models (LLMs) for Energy Systems Research*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2023; NREL/PR-6A20-87896. Available online: <https://www.osti.gov/biblio/2216925> (accessed on 18 November 2024).
- Bai, M.; Zhou, Z.; Wang, R.; Yang, Y.; Qin, Z.; Chen, Y.; Mu, C.; Liu, J.; Yu, D. HouYi: An Open-Source Large Language Model Specially Designed for Renewable Energy and Carbon Neutrality Field. *arXiv* **2023**, arXiv:2308.01414. [CrossRef]
- Zhang, L.; Chen, Z.; Ford, V. Advancing Building Energy Modeling with Large Language Models: Exploration and Case Studies. *arXiv* **2024**, arXiv:2402.09579. [CrossRef]
- Anser, M.K.; Khan, K.A.; Umar, M.; Awosusi, A.A.; Shamansurova, Z. Formulating Sustainable Development Policy for a Developed Nation: Exploring the Role of Renewable Energy, Natural Gas Efficiency and Oil Efficiency Towards Decarbonization. *Int. J. Sustain. Dev. World Ecol.* **2024**, *31*, 247–263. [CrossRef]
- He, X.; Khan, S.; Ozturk, I.; Murshed, M. The role of renewable energy investment in tackling climate change concerns: Environmental policies for achieving SDG-13. *Sustain. Dev.* **2023**, *31*, 1888–1901. [CrossRef]
- Yi, S.; Raza Abbasi, K.; Hussain, K.; Albaker, A.; Alvarado, R. Environmental concerns in the United States: Can renewable energy, fossil fuel energy, and natural resources depletion help? *Gondwana Res.* **2023**, *117*, 41–55. [CrossRef]
- Ang, T.-Z.; Salem, M.; Kamarol, M.; Das, H.; Nazari, M.; Natarajan, P. A comprehensive study of renewable energy sources: Classifications, challenges and suggestions. *Energy Strategy Rev.* **2022**, *43*, 100939. [CrossRef]
- Rizzi, F.; van Eck, N.J.; Frey, M. The production of scientific knowledge on renewable energies: Worldwide trends, dynamics and challenges and implications for management. *Renew. Energy* **2014**, *62*, 657–671. [CrossRef]
- Rashid, A.; Biswas, P.; Biswas, A.; Nasim, M.A.A.; Gupta, K.D.; George, R. Present and Future of AI in Renewable Energy Domain: A Comprehensive Survey. *arXiv* **2024**, arXiv:2406.16965. [CrossRef]

12. Sparks of Artificial General Intelligence: Early Experiments with GPT-4. *arXiv* **2023**, arXiv:2303.12712. Available online: <https://arxiv.org/abs/2303.12712> (accessed on 16 October 2024).
13. Lu, R.-S.; Lin, C.-C.; Tsao, H.-Y. Empowering Large Language Models to Leverage Domain-Specific Knowledge in E-Learning. *Appl. Sci.* **2024**, *14*, 5264. [[CrossRef](#)]
14. Liu, M.; Ene, T.-D.; Kirby, R.; Cheng, C.; Pinckney, N.; Liang, R.; Alben, J.; Anand, H.; Banerjee, S.; Bayraktaroglu, I.; et al. ChipNeMo: Domain-Adapted LLMs for Chip Design. *arXiv* **2024**, arXiv:2311.00176.
15. Gururangan, S.; Marasović, A.; Swayamdipta, S.; Lo, K.; Beltagy, I.; Downey, D.; Smith, N.A. Don't Stop Pretraining: Adapt Language Models to Domains and Tasks. *arXiv* **2020**, arXiv:2004.10964. [[CrossRef](#)]
16. Arnautov, K.V.; Akimov, D.A. Application of Large Language Models for Optimization of Electric Power System States. In Proceedings of the 2024 Conference of Young Researchers in Electrical and Electronic Engineering (ElCon), Saint Petersburg, Russia, 29–30 January 2024; pp. 314–317.
17. Satterlee, K. Large Language Models May Revolutionize the Electric Energy Sector. Available online: <https://engineering.tamu.edu/news/2024/06/large-language-models-may-revolutionize-the-electric-energy-sector.html> (accessed on 15 November 2024).
18. Li, R.; Pu, C.; Tao, J.; Li, C.; Fan, F.; Xiang, Y.; Chen, S. LLM-based Frameworks for Power Engineering from Routine to Novel Tasks. *arXiv* **2023**, arXiv:2305.11202.
19. Zhou, X.; Zhao, H.; Cheng, Y.; Cao, Y.; Liang, G.; Liu, G.; Liu, W.; Xu, Y.; Zhao, J. ElecBench: A Power Dispatch Evaluation Benchmark for Large Language Models. *arXiv* **2024**, arXiv:2407.05365. [[CrossRef](#)]
20. OpenAI; Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; et al. GPT-4 Technical Report. *arXiv* **2024**, arXiv:2303.08774. [[CrossRef](#)]
21. Team, G.; Anil, R.; Borgeaud, S.; Alayrac, J.-B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A.M.; Hauth, A.; Millican, K.; et al. Gemini: A Family of Highly Capable Multimodal Models. *arXiv* **2024**, arXiv:2312.11805. [[CrossRef](#)]
22. Touvron, H.; Martin, L.; Stone, K.; Albert, P.; Almahairi, A.; Babaei, Y.; Bashlykov, N.; Batra, S.; Bhargava, P.; Bhosale, S.; et al. Llama 2: Open Foundation and Fine-Tuned Chat Models. *arXiv* **2023**, arXiv:2307.09288. [[CrossRef](#)]
23. Dubey, A.; Jauhri, A.; Pandey, A.; Kadian, A.; Al-Dahle, A.; Letman, A.; Mathur, A.; Schelten, A.; Yang, A.; Fan, A.; et al. The Llama 3 Herd of Models. *arXiv* **2024**, arXiv:2407.21783. [[CrossRef](#)]
24. Llama 3.2. Available online: <https://www.llama.com/> (accessed on 16 October 2024).
25. Almazrouei, E.; Alobeidli, H.; Alshamsi, A.; Cappelli, A.; Cojocaru, R.; Debbah, M.; Goffinet, É.; Hesslow, D.; Launay, J.; Malartic, Q.; et al. The Falcon Series of Open Language Models. *arXiv* **2023**, arXiv:2311.16867. [[CrossRef](#)]
26. Jiang, A.Q.; Sablayrolles, A.; Mensch, A.; Bamford, C.; Chaplot, D.S.; de las Casas, D.; Bressand, F.; Lengyel, G.; Lample, G.; Saulnier, L.; et al. Mistral 7B. *arXiv* **2023**, arXiv:2310.06825. [[CrossRef](#)]
27. Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*; Burstein, J., Doran, C., Solorio, T., Eds.; Association for Computational Linguistics: Minneapolis, MN, USA, 2019; pp. 4171–4186.
28. Neves, M.; Jimeno Yepes, A.; Névéol, A.; Bawden, R.; Di Nunzio, G.M.; Roller, R.; Thomas, P.; Vezzani, F.; Vicente Navarro, M.; Yeganova, L.; et al. Findings of the WMT 2023 Biomedical Translation Shared Task: Evaluation of ChatGPT 3.5 as a Comparison System. In *Proceedings of the Eighth Conference on Machine Translation*; Koehn, P., Haddow, B., Kocmi, T., Monz, C., Eds.; Association for Computational Linguistics: Singapore, 2023; pp. 43–54.
29. Lu, Q.; Dou, D.; Nguyen, T. ClinicalT5: A Generative Language Model for Clinical Text. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2022*; Goldberg, Y., Kozareva, Z., Zhang, Y., Eds.; Association for Computational Linguistics: Abu Dhabi, United Arab Emirates, 2022; pp. 5436–5443.
30. Luo, R.; Sun, L.; Xia, Y.; Qin, T.; Zhang, S.; Poon, H.; Liu, T.-Y. BioGPT: Generative pre-trained transformer for biomedical text generation and mining. *Brief. Bioinform.* **2022**, *23*, bbac409. [[CrossRef](#)] [[PubMed](#)]
31. Wang, G.; Yang, G.; Du, Z.; Fan, L.; Li, X. ClinicalGPT: Large Language Models Finetuned with Diverse Medical Data and Comprehensive Evaluation. *arXiv* **2023**, arXiv:2306.09968. [[CrossRef](#)]
32. Workshop, B.; Scao, T.L.; Fan, A.; Akiki, C.; Pavlick, E.; Ilić, S.; Hesslow, D.; Castagné, R.; Luccioni, A.S.; Yvon, F.; et al. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model. *arXiv* **2023**, arXiv:2211.05100. [[CrossRef](#)]
33. Wu, C.; Lin, W.; Zhang, X.; Zhang, Y.; Wang, Y.; Xie, W. PMC-LLaMA: Towards Building Open-source Language Models for Medicine. *J. Am. Med. Inform. Assoc.* **2023**, *31*, 1833–1843. [[CrossRef](#)]
34. Chen, Z.; Cano, A.H.; Romanou, A.; Bonnet, A.; Matoba, K.; Salvi, F.; Pagliardini, M.; Fan, S.; Köpf, A.; Mohtashami, A.; et al. MEDITRON-70B: Scaling Medical Pretraining for Large Language Models. *arXiv* **2023**, arXiv:2311.16079. [[CrossRef](#)]
35. Singhal, K.; Azizi, S.; Tu, T.; Mahdavi, S.S.; Wei, J.; Chung, H.W.; Scales, N.; Tanwani, A.; Cole-Lewis, H.; Pfohl, S.; et al. Large language models encode clinical knowledge. *Nature* **2023**, *620*, 172–180. [[CrossRef](#)] [[PubMed](#)]
36. Singhal, K.; Tu, T.; Gottweis, J.; Sayres, R.; Wulczyn, E.; Hou, L.; Clark, K.; Pfohl, S.; Cole-Lewis, H.; Neal, D.; et al. Towards Expert-Level Medical Question Answering with Large Language Models. *arXiv* **2023**, arXiv:2305.09617. [[CrossRef](#)]
37. Wu, S.; Irsoy, O.; Lu, S.; Dabrovolski, V.; Dredze, M.; Gehrmann, S.; Kambadur, P.; Rosenberg, D.; Mann, G. BloombergGPT: A Large Language Model for Finance. *arXiv* **2023**, arXiv:2303.17564. [[CrossRef](#)]
38. Wu, Y.; Jia, F.; Zhang, S.; Li, H.; Zhu, E.; Wang, Y.; Lee, Y.T.; Peng, R.; Wu, Q.; Wang, C. MathChat: Converse to Tackle Challenging Math Problems with LLM Agents. *arXiv* **2024**, arXiv:2306.01337. [[CrossRef](#)]



39. Taylor, R.; Kardas, M.; Cucurull, G.; Scialom, T.; Hartshorn, A.; Saravia, E.; Poulton, A.; Kerkez, V.; Stojnic, R. Galactica: A Large Language Model for Science. *arXiv* **2022**, arXiv:2211.09085. [[CrossRef](#)]
40. Lewkowycz, A.; Andreassen, A.; Dohan, D.; Dyer, E.; Michalewski, H.; Ramasesh, V.; Slone, A.; Anil, C.; Schlag, I.; Gutman-Solo, T.; et al. Solving Quantitative Reasoning Problems with Language Models. *arXiv* **2022**, arXiv:2206.14858. [[CrossRef](#)]
41. Lee, J.; Yoon, W.; Kim, S.; Kim, D.; Kim, S.; So, C.H.; Kang, J. BioBERT: A pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* **2019**, *36*, 1234–1240. [[CrossRef](#)]
42. Domain-Specific Language Model Pretraining for Biomedical Natural Language Processing | ACM Transactions on Computing for Healthcare. Available online: <https://dl.acm.org/doi/10.1145/3458754> (accessed on 16 October 2024).
43. Sun, C.; Qiu, X.; Xu, Y.; Huang, X. How to Fine-Tune BERT for Text Classification? *arXiv* **2020**, arXiv:1811.01088. [[CrossRef](#)]
44. Phang, J.; Févry, T.; Bowman, S.R. Sentence Encoders on STILTs: Supplementary Training on Intermediate Labeled-data Tasks. *arXiv* **2019**, arXiv:1811.01088. [[CrossRef](#)]
45. Wu, Z.; IV, R.L.L.; Walsh, P.; Bhagia, A.; Groeneveld, D.; Singh, S.; Beltagy, I. Continued Pretraining for Better Zero- and Few-Shot Promptability. *arXiv* **2022**, arXiv:1711.05101.
46. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. *arXiv* **2018**, arXiv:2409.19735.
47. Welcome | Unsloth Documentation. Available online: <https://docs.unsloth.ai> (accessed on 19 November 2024).
48. Hugging Face—The AI Community Building the Future. Available online: <https://huggingface.co/> (accessed on 19 November 2024).
49. Rent GPUs | Vast.ai. Available online: [https://vast.ai/?utm\\_source=bing&utm\\_medium=cpc&utm\\_campaign=567756507\\_search&utm\\_target=&utm\\_group=1179777714011127&utm\\_content=&placement=&device=c&adposition=&msclkid=8283ab2b1a62190d3783cc17b7144f43](https://vast.ai/?utm_source=bing&utm_medium=cpc&utm_campaign=567756507_search&utm_target=&utm_group=1179777714011127&utm_content=&placement=&device=c&adposition=&msclkid=8283ab2b1a62190d3783cc17b7144f43) (accessed on 19 November 2024).
50. Auto-Regressive Next-Token Predictors Are Universal Learners. Available online: <https://arxiv.labs.arxiv.org/html/2309.06979> (accessed on 19 November 2024).
51. Lin, J.; Tang, J.; Tang, H.; Yang, S.; Chen, W.-M.; Wang, W.-C.; Xiao, G.; Dang, X.; Gan, C.; Han, S. AWQ: Activation-aware Weight Quantization for LLM Compression and Acceleration. *arXiv* **2024**, arXiv:2106.09685.
52. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. *Proc. IEEE* **2023**, *111*, 653–685.
53. Chen, J.; Geng, Y.; Chen, Z.; Pan, J.Z.; He, Y.; Zhang, W.; Horrocks, I.; Chen, H. Zero-Shot and Few-Shot Learning with Knowledge Graphs: A Comprehensive Survey. *Proc. IEEE* **2023**, *111*, 653–685. [[CrossRef](#)]
54. Song, Y.; Wang, T.; Mondal, S.K.; Sahoo, J.P. A Comprehensive Survey of Few-Shot Learning: Evolution, Applications, Challenges, and Opportunities. *arXiv* **2022**, arXiv:2205.06743. [[CrossRef](#)]
55. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.; Rocktäschel, T.; et al. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. *arXiv* **2021**, arXiv:2005.11401. [[CrossRef](#)]
56. Chung, H.W.; Hou, L.; Longpre, S.; Zoph, B.; Tay, Y.; Fedus, W.; Li, Y.; Wang, X.; Dehghani, M.; Brahma, S.; et al. Scaling Instruction-Finetuned Language Models. *arXiv* **2022**, arXiv:2210.11416. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.