UNIVERSITY OF OKLAHOMA
GRADUATE COLLEGE

ENHANCING AUTOMATED PUBLIC OPINION ANALYSIS: A COMPARATIVE STUDY OF VADER AND LARGE LANGUAGE MODEL METHODS

A THESIS
SUBMITTED TO THE GRADUATE FACULTY
in partial fulfillment of the requirements for the
Degree of
MASTER OF SCIENCE

By

CADE BIRDSONG
Norman, Oklahoma
2025

ENHANCING AUTOMATED PUBLIC OPINION ANALYSIS: A COMPARATIVE STUDY OF VADER AND LARGE LANGUAGE MODEL METHODS

A THESIS APPROVED FOR THE
SCHOOL OF INDUSTRIAL AND SYSTEMS ENGINEERING

BY THE COMMITTEE CONSISTING OF

Dr. Nicholson (Chair)

Dr. Gonzalez

Dr. Dodd

Table of Contents

List Of Figures

Abstract

Traditional sentiment analysis methods such as VADER have long been employed to gauge public opinion through rule-based and lexicon-driven approaches. However, the advent of large language models (LLMs) has introduced novel methods capable of capturing deeper semantic nuances and handling heterogeneous data at scale. In this paper, a comparative analysis of VADER and LLM-based sentiment analysis methodologies is done within the context of public policy research. Using nuclear fusion as a case study, the study evaluates how each approach processes and interprets sentiment in news and social media texts.

The VADER approach, characterized by its reliance on curated sentiment lexicons and syntactic rules, excels in computational efficiency and straightforward implementation. In contrast, the LLM-based framework leverages parameter-efficient fine-tuning techniques to adapt pre-trained models for nuanced sentiment extraction, thereby capturing complex opinion dynamics and contextual subtleties that traditional methods often overlook. Through rigorous experimentation, the accuracy, scalability, and processing speed of these methodologies were compared, outlining their respective advantages and limitations.

Our results demonstrate that while VADER provides rapid sentiment scoring suitable for high-volume datasets, the LLM-based approach delivers richer, more robust insights that are critical for dynamic public policy analysis. This thesis further discusses the trade-offs between computational overhead and analytical depth, emphasizing that the choice of technique should align with the specific requirements of policy research. Overall, this work contributes to a better understanding of how modern deep learning techniques can enhance automated sentiment analysis and improve the timeliness and accuracy of public opinion monitoring in contemporary policy contexts.

Chapter 1

Introduction

1.1 Introduction

Public sentiment plays a pivotal role in shaping public policy, especially in high-stakes sectors such as nuclear energy. The widespread and often polarized opinions expressed online—through news outlets, blogs, and social media—can profoundly influence decision-making processes, research funding, and the overall policy direction of critical technologies. As digital communication platforms generate an ever-increasing volume of opinionated text data, new computational approaches are required to harness this wealth of information efficiently.

1.1.1 Motivation and Challenges

Traditional sentiment analysis approaches, while effective in controlled settings, often struggle with the dynamic and heterogeneous nature of news and social media content A and R (2009); Hutto and Gilbert (2014); Kouloumpis et al. (2021). Manual data collection and annotation for sentiment analysis is not feasible on the scale of modern data streams, especially when continuous monitoring is required. The time and effort required to manually curate, clean, and analyze these data is substantial, creating a bottleneck to obtaining timely insights.

Recent advances in natural language processing, particularly the advent of large language models (LLMs), have opened up new avenues for addressing these challenges. By fine-tuning pre-trained LLMs using automated methods, it is now possible to efficiently process and extract sentiment information from massive volumes of data. Techniques such as parameter-efficient fine-tuning enable these models to be adapted to specific tasks with minimal additional computational cost Hu et al. (2023); Liu and Shi (2024). This not only accelerates the sentiment analysis process but also maintains a quality of analysis that is comparable to more labor-intensive methods.

1.1.2 Application to Nuclear Energy Policy

Nuclear energy remains a controversial but critical component of the global energy mix. The success of nuclear energy projects is highly dependent on public trust and the alignment of policy with social expectations. Misconceptions, safety concerns, and fluctuating public opinions can all influence the trajectory of nuclear energy initiatives. This thesis leverages automated sentiment analysis to provide a systematic and scalable approach to gauge public opinion on nuclear energy. In doing so, it aims to help policymakers and researchers quickly identify trends and shifts in sentiment, facilitating more responsive and informed decisions.

1.1.3 Collaborative Context and Research Objectives

A significant impetus for this work is the collaboration with the Institute for Public Policy Research and Analysis (IPPRA) as part of their D-EPSCoR Fusion Social Observatory Project. The extensive history of IPPRA in public policy research provides a solid foundation on which this thesis is based. This approach focuses on fine-tuning an LLM to automatically parse and analyze sentiment data from a variety of online sources. This automated process not only conserves valuable time, but also improves the timeliness of sentiment monitoring, which is crucial for real-time policy evaluation Gupta et al. (2024).

The research objectives of this thesis are threefold.

1. To develop and validate an automated sentiment analysis framework that leverages LLMs.

2. To compare the performance of this framework with conventional sentiment analysis methods and demonstrate its scalability and efficiency.

3. To evaluate the impact of the extracted sentiment data on understanding public opinion trends in the context of nuclear energy policy.

The question is as follows: How can automated LLM-based sentiment analysis improve processing speed, accuracy, and analysis depth relative to manual approaches such as those based on VADER?

1.1.4 Significance of the Study

Automating sentiment extraction addresses a critical bottleneck in public policy research: the need for rapid, high-quality sentiment data. Enhanced sentiment analysis facilitates a deeper understanding of the perspectives of citizens, enabling policymakers to respond more effectively to public concerns. Moreover, this thesis advances the literature by demonstrating how cutting-edge LLMs can be tailored for specialized text analysis tasks, thereby bridging the gap between advanced natural language processing techniques and practical policy applications.

1.1.5 Thesis Organization

The remainder of this thesis is organized as follows.

- Section 2.1 provides a comprehensive review of the literature on sentiment analysis methods and the integration of LLMs in public opinion research.
- Section 3.1 details the experimental setup of the methodological framework used to fine-tune the LLM and integrate it into an automated sentiment analysis pipeline.
- Section 4.1 discusses evaluation metrics and performance comparisons between traditional and LLM-based sentiment analysis methods.
- Section 5.1 discusses the implications of the findings for sentiment analysis and outlines potential directions for future research.

In summary, this thesis demonstrates that using automated LLM-based sentiment analysis techniques can not only save substantial amounts of time compared to manual methods but also provide robust, scalable insights that are critical for modern public policy research.

Chapter 2

Literature Review

2.1 Literature Review and Background

Understanding and extracting public sentiment has been a central focus in both social science research and public policy analysis for decades. Traditionally, sentiment analysis relies heavily on manual processes that involve the collection, annotation, and analysis of textual data from various sources, such as news articles, social media posts, and product reviews. Such manual methods, as demonstrated in the early work by Balahur et al. A and R (2009) and detailed in studies by Neethu and Rajasree Neethu and Rajasree (2013), as well as Kouloumpis et al. Kouloumpis et al. (2021), reveal that manual labeling and analysis require significant human effort, cost, and time. Furthermore, surveys conducted by Mejova Mejova (2009), Pandya and Mehta Pandya and Mehta (2020), and Wankhade et al. Wankhade et al. (2022) illustrate that human annotation struggles to keep up with the surging volume of data emerging on social media platforms, resulting in limited scalability and inconsistent quality.

2.1.1 Manual Efforts in Sentiment Analysis

Manual sentiment analysis typically involves several stages: gathering raw data, extensive pre-processing, and meticulous annotation of text by subject-matter experts. Lexicon-based approaches, such as VADER Hutto and Gilbert (2014), require the construction and validation of sentiment lexicons manually, while supervised learning methods depend on hand-coded training datasets Giachanou and Crestani (2017). However, even the most carefully annotated datasets face difficulties when applied to dynamic and noisy environments. For example, the work by Zimbra et al. Zimbra et al. (2018) and Patodkar and I.R. Patodkar and I.R (2016) highlights that manually developed sentiment classification systems often struggle to achieve accuracy levels above 70% due to the inherent variability and brevity of social media text. These challenges not only delay the analysis, but also limit the practical application of sentiment insights in policy decision making.

2.1.2 Large Language Models: Background and Mechanics

Large Language Models (LLMs) have emerged as transformative tools that address many limitations of manual sentiment analysis. Based on the Transformer architecture, LLMs such as GPT, BERT, and LLaMA Touvron et al. (2023) leverage self-attention mechanisms to process entire text sequences. By pretraining on massive corpora of diverse text, these models learn intricate language patterns and contextual relationships. This enables LLMs to generate contextually relevant and coherent text representations, making them well suited for tasks that require understanding long-range dependencies and subtle nuances in language. In essence, LLMs provide robust and context-aware features that can bypass the inconsistencies typically encountered in manual annotation.

2.1.3 VADER Vs. LLM Sentiment Analysis

The most widely used manual method for sentiment analysis is the VADER model (Valence-Aware Dictionary for Sentiment Reasoning) Hutto and Gilbert (2014).

VADER is a lexicon-driven, rule-based sentiment analyzer specifically designed for social media text. It uses a curated list of sentiment-laden words along with a set of grammatical and syntactical rules to determine the intensity of the sentiment. VADER's strength lies in its simplicity and ease of application to short texts such as tweets or status updates, where overt emotional indicators are prevalent. However, its performance is highly dependent on the quality and comprehensiveness of the manually curated lexicon, and extending VADER to new domains often requires significant manual intervention.

In contrast, Large Language Models (LLMs) such as LLaMA Touvron et al. (2023) operate under an entirely different paradigm. LLMs are based on the Transformer architecture, which uses self-attention mechanisms to process and generate text. Pretrained on vast amounts of diverse textual data, these models learn to capture complex language patterns and subtle contextual nuances. Rather than relying on a fixed lexicon, LLM-based approaches utilize deep learning to generalize sentiment understanding across a wide array of inputs. With parameter-efficient fine-tuning methods Han et al. (2024); Hu et al. (2023), LLMs can be adapted to specific tasks, such as sentiment classification, while preserving the broad language understanding acquired during pretraining.

The key advantage of LLM-based sentiment analysis is its ability to process large volumes of heterogeneous data with high accuracy. Although VADER is effective for straightforward sentiment detection in controlled contexts, LLMs can interpret mixed sentiment, sarcasm, and domain-specific subtleties that traditional lexicon-based systems might miss. This makes LLMs particularly valuable for scaling sentiment analysis in real-world applications, such as monitoring public opinion on nuclear fusion in policy contexts. By automating the extraction of sentiment insights from thousands of articles, LLMs enable more dynamic and responsive analysis compared to the labor-intensive manual methods typical of traditional approaches.

2.1.4 Parameter-Efficient Fine-Tuning

A major hurdle in adopting LLMs for specialized tasks, such as sentiment analysis, has been the computational cost associated with full-model fine-tuning. Recent innovations in parameter-efficient fine-tuning have significantly mitigated this challenge. The techniques described by Han et al. Han et al. (2024) and Hu et al. Hu et al. (2023) demonstrate that it is possible to adapt LLMs to specific domains by fine-tuning only a small subset of parameters - often through adapter modules or prompt-based methods - thus preserving the general language capabilities of the model while dramatically reducing training overhead. The PoliPrompt framework introduced by Liu and Shi Liu and Shi (2024) further illustrates that high classification accuracy can be achieved in the classification of political sentiments, making automated data acquisition scalable and meaningful. Furthermore, Xia et al. Xia et al. (2024) provide empirical evidence that these fine-tuning strategies not only improve throughput but also enable near-real-time analysis, critical for keeping pace with rapidly evolving public discourse.

2.1.5 Automated and Large-Scale Sentiment Analysis with LLMs

The shift from manual to automated sentiment analysis is transformative. LLM-based methods have shown that automated pipelines can process enormous amounts of data continuously, capturing subtle sentiment variations that manual systems may overlook. Recent studies by Mao et al. Mao et al. (2023) and Zhang et al. Zhang et al. (2023) reveal that LLM-driven sentiment analysis systems not only achieve higher accuracy but also scale efficiently, handling vast datasets without the inconsistencies and delays common to manual methods. The high accuracy and scalability of these automated models are particularly important for public policy research, where large-scale data, such as that presented by social networks and news aggregation, provide a more comprehensive picture of public opinion. In addition, the integration of LLMs with automated fine-tuning methods ensures that the resulting sentiment data is robust and reliable, ultimately leading to more informed and timely policy decisions.

2.1.6 Implications for Public Policy and Future Research

The integration of automated LLM-based sentiment analysis has profound implications for policy research. As underscored by studies on public opinion and policy impact Burstein (2003); Ceron and Negri (2015), the ability to quickly assess public sentiment enables policymakers to respond more dynamically to public concerns and design policies that are more in line with social preferences. The automated collection and analysis process facilitated by LLMs helps bridge the gap between the overwhelming volume of digital data and the limited capacity of traditional manual approaches, ensuring that decision-makers have access to timely, scalable, and accurate sentiment data. This is particularly evident in specialized domains such as nuclear energy policy, where research by Gupta et al. Gupta et al. (2024) demonstrates how nuanced public opinion can influence critical policy decisions.

2.1.7 Summary

In summary, the literature reveals a clear progression from traditional manual sentiment analysis methods to modern LLM-based approaches. Manual methods, exemplified by lexicon-driven tools such as VADER Hutto and Gilbert (2014) and other supervised techniques Giachanou and Crestani (2017); Zimbra et al. (2018); Patodkar and I.R (2016), require extensive human intervention for annotation and are often limited in scalability and consistency. Although fundamental, these approaches struggle to keep up with the sheer volume and heterogeneous nature of data from social media and news sources A and R (2009); Neethu and Rajasree (2013); Kouloumpis et al. (2021); Mejova (2009); Pandya and Mehta (2020); Wankhade et al. (2022).

In contrast, Large Language Models (LLMs) based on the Transformer architecture, such as LLaMA Touvron et al. (2023), leverage pre-training on vast datasets to capture complex language patterns and contextual nuances. When coupled with parameter-efficient fine-tuning techniques Han et al. (2024); Hu et al. (2023); Liu and Shi (2024), these models not only mitigate the cost of full-model retraining but also provide robust, scalable sentiment analysis capable of processing massive volumes of text with improved accuracy and adaptability. Moreover, as demonstrated in recent studies Mao et al. (2023); Zhang et al. (2023); Xia et al. (2024), LLM-based systems are able to extract subtle sentiments even in complex and dynamic environments.

In general, the integration of automated LLM-based methods represents a transformative solution to conventional manual sentiment analysis. This advancement is particularly significant for public policy research, where the ability to derive timely and accurate sentiment insights at scale is essential to inform decision-making processes.

Chapter 3

Data & Methods

3.1 Methodology

This section details the methodological framework developed to automate sentiment analysis using large language models. The workflow is organized into four primary stages: data collection, data pre-processing, model implementation, and batch processing.

3.1.1 Data Collection

The articles were sourced from Lexus Nexus using IPPRA's access to the forum. Approximately 2500 articles were collected by querying the keywords *Nuclear* and *Fusion*. The retrieved articles were provided in raw JSON format, which includes both the textual content of the articles and metadata such as publication dates.

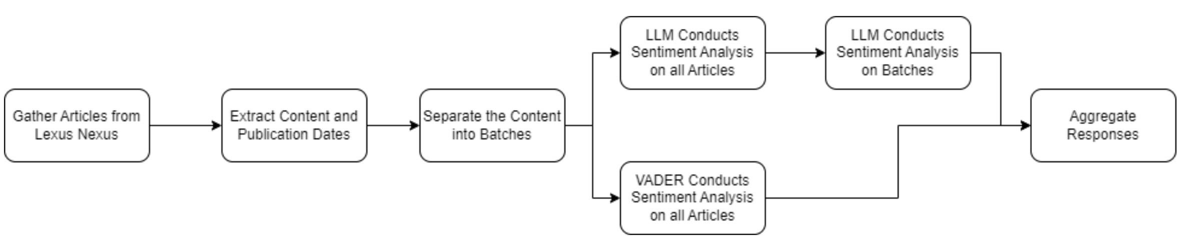

Figure 3.1: Process flowchart

3.1.2 Data Pre-Processing

The raw JSON data required thorough preprocessing to extract useful information for sentiment analysis. The preprocessing workflow involved:

1. Parsing JSON files to extract the content of the article and the publication dates.

2. Extracting and cleaning contextual information from each article.

3. Organizing the articles chronologically. Because LLMs have a context limit, the articles were sorted into batches of 100 and arranged them in ascending chronological order. This batching strategy enabled the performance sentiment analysis on manageable segments of data without exceeding the model's input capacity.

3.1.3 Model Implementation

To align with IPPRA's security requirements and to facilitate on-site processing, LLaMA was the chosen LLM due to its open-source nature. The model was deployed locally using Ollama, an open-source tool that streamlines the operation of LLaMA in a local environment. The Ollama Python interface was used to seamlessly integrate the LLM into the automated pipeline, allowing for the control and execution of sentiment analysis through scripted commands. Designing the LLM prompts proved to be an unexpectedly intricate component of the pipeline. Because LLaMA responds with great sensitivity to phrasing, ordering, and even punctuation, we treated prompt construction as an iterative engineering task: each constraint had to be isolated, tested, and refined much like traditional software logic. This complexity not only shaped the model's accuracy but also dictated downstream processing decisions, underscoring

```
# --- Process articles within the loaded file ---
for article_index, article in enumerate(articles_in_file):
    iterate += 1  # Simple global counter
    text = article.get("content")

    pub_date_str = article.get("estimatedPublishedDate")
    pub_date_obj = parse_date_or_none(pub_date_str)
    article_id = article.get("id", f"file_{file_name}_index_{
       article_index}")  # Use article index within file

    # Show progress
    print(f"Processing:␣Article␣{iterate}/{content_count}␣"
          f"[File:␣{file_name},␣Item:␣{article_index+1}/{len(
             articles_in_file)}]")

    # Skip if non-English
    try:
        if len(text) < 20:  # Too short to reliably detect language
            print(f"␣␣Skipping:␣Content␣too␣short␣({len(text)}␣chars)
               .")
            total_failed_lang_detect += 1  # Count it here for
               simplicity
            continue
        lang = detect(text[:500])  # Check only the beginning for
           speed
        if lang != 'en':
            print(f"␣␣Skipping:␣Non-English␣({lang})␣detected.")
            total_failed_lang_detect += 1
            continue
    except LangDetectException:
        print("␣␣Skipping:␣Language␣detection␣failed.")
        total_failed_lang_detect += 1
        continue
    except Exception as e:
        print(f"␣␣Skipping:␣Error␣during␣language␣detection:␣{e}")
        total_failed_lang_detect += 1
        continue
```

Figure 3.2: Example code for processing articles from the loaded JSON file. This snippet illustrates iteration, language detection, and filtering of non-English content, with improved spacing and line breaks.

```
  # Use datetime.min or datetime.max to control sorting of None dates
min_datetime = datetime.min.replace(tzinfo=timezone.utc) # Timezone-
    aware min datetime

relevant_outputs = [item for item in all_sentiment_outputs if item.
    get("is_relevant")]
relevant_outputs.sort(key=lambda x: x.get("pub_date_obj") or
    min_datetime)

print(f"Proceeding␣to␣summarize␣{len(relevant_outputs)}␣relevant␣
    articles␣(US␣Context,␣Scored␣for␣Fusion).")

# Function to chunk data
def chunk(items, size):
    if size <= 0:
        yield items # Yield all items as one chunk if size is invalid
        return
    for i in range(0, len(items), size):
        yield items[i:i+size]

batch_summaries_data = [] # Store structured batch data + LLM summary
```

Figure 3.3: Example code for sorting relevant sentiment outputs by publication date and batching the articles for subsequent processing.

that "prompt engineering" is itself a non-trivial methodological step rather than a mere scripting detail.

3.1.4 LLM Accuracy in Sentiment Analysis

To evaluate the reliability of large language models (LLMs) for sentiment analysis, we reference the TweetEval benchmark Barbieri et al. (2020), which assesses performance on various social media tasks. The RoBERTa-base model, when fine-tuned on Twitter data, achieved a macro-averaged recall of 73.0% on sentiment classification, demonstrating strong effectiveness.

These findings align with broader evaluations showing that LLMs perform competitively on sentiment tasks, particularly in low-resource settings, and consistently outperform smaller models Zhang et al. (2023). While limitations remain in complex

```
# --- Prompt for structured sentiment and relevance ---
        prompt = (
            f"Analyze␣the␣sentiment␣expressed␣towards␣*nuclear␣fusion
                *␣in␣the␣following␣text.␣The␣text␣MUST␣be␣primarily␣
                and␣explicitly␣about␣the␣*United␣States*.\n"
            f"1.␣First,␣assess␣relevance:␣Is␣the␣text␣primarily␣and␣
                explicitly␣about␣the␣United␣States?␣(Yes/No)\n"
            f"2.␣If␣No,␣respond␣ONLY␣with:␣Relevance:␣No\n"
            f"3.␣If␣Yes,␣respond␣with␣the␣sentiment␣score␣towards␣*
                nuclear␣fusion*␣on␣a␣Likert␣scale␣from␣1␣to␣5␣(1=very␣
                negative,␣3=neutral/technical,␣5=very␣positive)␣AND␣a␣
                brief␣(1-sentence)␣justification␣for␣the␣score.␣Format
                ␣your␣response␣EXACTLY␣like␣this␣(with␣the␣newline):\n
                "
            f"Score:␣[number]\n"
            f"Justification:␣[Your␣brief␣justification␣here]\n\n"
            f"Text:\n{text[:LLM_CTX_ARTICLE␣-␣600]}" # Truncate text,
                 reserve more space for clearer instructions
        )
```

Figure 3.4: Example code for prompting the LLM to perform structured sentiment and relevance analysis on texts, focusing on nuclear fusion and ensuring that the content is primarily about the United States.

subtasks, such as aspect-based sentiment analysis, current evidence supports LLMs as valid tools for general sentiment classification.

3.1.5 Batch Processing Strategy

Due to limitations on the context length of LLMs, a batch processing strategy was implemented. After the articles were sorted by publication date, the data set was divided into batches of 100 articles. Each batch was processed in ascending chronological order, ensuring that the LLM received a manageable and contextually coherent segment of data for sentiment extraction. This approach not only prevents context overflow but also preserves the temporal order necessary to capture shifts in sentiment over time.

```
for batch_i, batch_items in enumerate(chunk(relevant_outputs,
    BATCH_SIZE), start=1):
            print(f"Summarizing␣Batch␣{batch_i}␣({len(batch_items)}␣
                relevant␣articles)...")

            batch_start_date_obj = None
            batch_end_date_obj = None
            valid_dates_in_batch = [item.get("pub_date_obj") for item
                 in batch_items if item.get("pub_date_obj")]

            if valid_dates_in_batch:
                batch_start_date_obj = min(valid_dates_in_batch)
                batch_end_date_obj = max(valid_dates_in_batch)
                start_date_str = batch_start_date_obj.strftime(
                    DATE_FORMAT)
                end_date_str = batch_end_date_obj.strftime(
                    DATE_FORMAT)
                batch_date_range_str = f"{start_date_str}␣to␣{
                    end_date_str}"
            else:
                start_date_str = "Unknown"
                end_date_str = "Unknown"
                batch_date_range_str = "Unknown␣date␣range"
            # --- End Date Range Extraction ---
```

Figure 3.5: Example code for batching articles based on publication dates to avoid LLM context limits and enable sequential processing.

3.1.6 VADER Implementation for Sentiment Analysis

To provide a traditional baseline for sentiment analysis, a pipeline based on VADER (Valence-Aware Dictionary for Sentiment Reasoning) as described by Hutto and Gilbert Hutto and Gilbert (2014) was implemented. This implementation was developed using Python and leverages NLTK's built-in VADER lexicon. The goal was to extract sentiment information from articles sourced from Lexus Nexus, and it serves as a point of comparison to the LLM-based approach.

Our VADER-based implementation comprises several key components:

1. **Setup and Initialization:** Essential libraries are imported, including `os`, `json`, `time`, `re`, and modules from `datetime`. Import the VADER sentiment analyzer from NLTK, ensuring that the VADER lexicon is downloaded prior to analysis.

2. **Settings and Configuration:** Various operational parameters and file paths are defined (e.g., data folder, output report and JSON file paths, batch size and date format). A special token for marking irrelevant content is also specified. These settings facilitate consistent processing and reporting across the entire dataset.

3. **Date Parsing:** A helper function, `parse_date_or_none`, is implemented to convert raw date strings from the JSON data into timezone-aware `datetime` objects. This function handles cases where date information is absent or improperly formatted.

4. **Article Processing:** The core of the implementation involves iterating through all log files to process individual articles:

    - Each article is examined to ensure it contains sufficient content. Articles shorter than a minimal length are skipped.

- The language of the text is detected using the `langdetect` package; non-English articles or those for which language detection fails are omitted.
- A targeted relevance check is performed by searching for explicit mentions of "United States", "U.S.", or "US". Articles that fail this check are marked as irrelevant and assigned a corresponding sentiment token.

5. **Sentiment Analysis with VADER:** For relevant articles, the VADER analyzer computes a compound sentiment score (ranging from -1 to 1). This score is then assigned to a Likert scale of 1 to 5. A simple categorization is applied, classifying sentiments as positive, negative, or neutral based on threshold values, and a justification message is generated accordingly.
6. **Aggregation and Reporting:** Processed articles, along with their sentiment scores and metadata, are aggregated into a comprehensive dataset. Statistical measures (such as average, median, and mode of the sentiment scores) are computed. Finally, the intermediate results are saved to a JSON file, and a detailed report is generated.

The complete VADER-based implementation is provided in Appendix A.1.1. This traditional rule-based approach, while effective in smaller data sets, highlights the manual effort required to scale sentiment extraction. In contrast, the LLM-based methodology automates this process and is capable of handling large volumes of data with enhanced accuracy and efficiency.

3.1.7 Summary

In summary, this methodology integrates automated data acquisition from Lexus Nexus, detailed pre-processing of raw JSON data, and the deployment of an LLM-based sentiment analysis pipeline using open-source tools such as LLaMA and Ollama. By sorting the articles into batches of 100 in ascending chronological order, the processing remains within the LLM's context limits while preserving the temporal structure of the data. This pipeline not only saves significant time compared to manual methods, but also provides a scalable framework for extracting robust and actionable sentiment insights.

Chapter 4

Results

4.1 Results

This section presents the results of the sentiment analysis pipeline using both LLM-based and VADER-based methods. The results are organized into four parts: article-based outputs, batch-based analysis, aggregate statistics, and a performance comparison focusing on processing time.

4.1.1 Article-Based Results

For each article processed, the pipeline produced a sentiment score on a Likert scale from 1 to 5 and a brief justification. Figure 4.2 demonstrates the output of a single article's sentiment analysis. This example illustrates how the LLM-based approach captures nuanced sentiment details and compares with the manual VADER implementation.

To evaluate agreement and divergence between VADER and LLM sentiment assessments, a confusion matrix was constructed. Both systems used a normalized Likert scale from 1 to 5. Only articles scored by both models were included. Discrepancies of two or more points were flagged for qualitative inspection to identify patterns in disagreement.

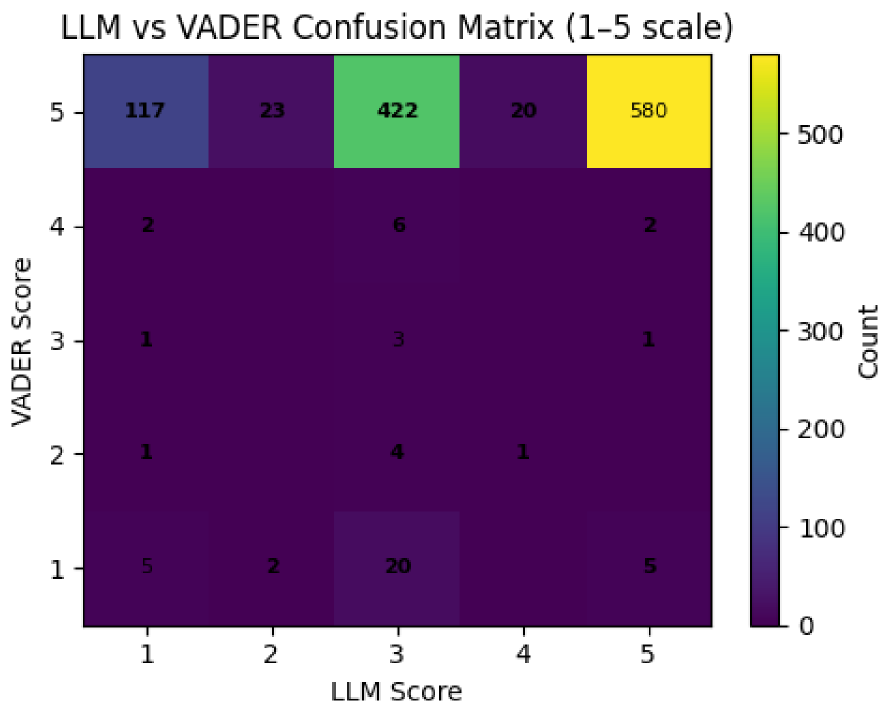

Figure 4.1: Confusion matrix comparing Likert-scale sentiment scores between VADER and the LLM.


```
"id": "52859422067",
"file": "FusionTerms-2024-01_1_2024-01-16-1736817068.log",
"is_relevant": true,
"score": 5,
"justification": "The sentiment towards nuclear fusion in the text
   is overwhelmingly positive, as it is portrayed as a
   revolutionary technology that will solve global challenges such
   as energy scarcity, food insecurity, and water shortages,
   providing unlimited access to clean energy and water.",
"pub_date": "2024-01-16T13:30:00+00:00"
```


Figure 4.2: Example JSON output for a relevant article sentiment result.

The following is an article where VADER scores it negatively and the LLM scores it very positively. Interestingly, it seems that in these instances the article content was

```
in depth knowledge and experience in designing and manufacturing
   nuclear systems, along with its strategic cooperation and
   investment in the leading fusion power and desalination developers
   , offers triple solution to three of the biggest global challenges
    of clean energy, water, and food scarcity. Recent major
   developments in fusion power generation show that a world with
   fusion power is coming, and it will truly revolutionize our
   landscape, providing unlimited access to clean energy and water
   for the world.

Cheap, abundant, green fusion electric power plants, both large and
   small scale, provides power needed for growing populations and
   fast-growing industrialization. An energy crisis is making this
   winter especially cold as people struggle to afford the costs of
   oil and gas for heating fusion power will solve issues like these
   and do so with zero emissions.
```

Figure 4.3: Excerpt of article that Figure 5 is referencing.


```
"id": "52859422067",
"file": "FusionTerms-2024-01_1_2024-01-16-1736817068.log",
"is_relevant": true,
"score": 5,
"justification": "VADER compound score is 0.99, indicating positive
   sentiment.",
"pub_date": "2024-01-16T13:30:00+00:00"
```


Figure 4.4: Example JSON output for a relevant article VADER sentiment result.

generally irrelevant and would not be included in a full report regarding the sentiment of nuclear power. There were also only 5 instances of this particulary combination.


```
"id": "54743893424",
"file": "FusionTerms-2024-09_1_2024-09-09-1737047853.log",
"is_relevant": true,
"score": 5,
"justification": "The text portrays Edward Teller as a pioneer of
   nuclear fusion technology, highlighting his vision for harnessing
   the power of fusion to create a more powerful bomb, which was
   ultimately realized with the development of the Hydrogen bomb.
   This suggests that the sentiment towards nuclear fusion in the
   context of the United States is very positive.",
"pub_date": "2024-09-09T21:58:00+00:00"
```


Figure 4.5: LLM JSON sentiment result: article scored **5** (very positive).

The following is an article where VADER scores it positively and the LLM scores it very negatively. Once again, it seems that in instance where there is large discrepancies in the scores, the actual information is irrelevant to the subject matter.

The following is an example where both VADER and LLM assert that the article is deeply negative. Interestingly, it appears that wherever there is a large discrepancy in the assessments, the article is typically irrelevant to the data set.

It should be noted that there were only 32 articles that scored a 1 under VADER, so these articles are not as telling a measure of efficacy as one would expect.

Finally, we have the most common example of disagreement. The following is an article that the LLM scored as a 3, and VADER scored as a 5. After viewing a large

```
Edward Teller was a physicist who altered the course of history. Born
    in Budapest in 1908, Teller's love for science brought him to the
    forefront of nuclear physics. His work on the Hydrogen bomb,
   while groundbreaking, sparked moral debates that echo today. But
   his story is about more than just a bomb - it's about a man driven
    by science in a world on the edge of destruction.

FROM BUDAPEST TO LOS ALAMOS
Teller's journey began in Hungary, where he developed a passion for
   mathematics and physics. As Europe grew unstable in the 1930s,
   Teller, who was Jewish, fled to the United States in 1935. He
   became part of a growing community of brilliant scientists
   escaping Nazi persecution. This wave of talent reshaped American
   science and soon, Teller found himself at the heart of a project
   that would change the world.

THE MANHATTAN PROJECT
Teller was recruited for the Manhattan Project during World War II, a
    secretive effort to build the atomic bomb. The project was set up
    to create a fission-based weapon, but Teller's attention was
   quickly captured by another idea: fusion. He was convinced that
   nuclear fusion, the same process that powers stars, could be
   harnessed to create a bomb far more powerful than anything that
   had ever been seen before.
[...]
```

Figure 4.6: Excerpt of the article corresponding to Figure 4.5.


```
"id": "54743893424",
"file": "FusionTerms-2024-09_1_2024-09-09-1737047853.log",
"is_relevant": true,
"score": 1,
"justification": "VADER compound score is -0.99, indicating negative
   sentiment.",
"pub_date": "2024-09-09T21:58:00+00:00"
```


Figure 4.7: VADER JSON sentiment result: same article scored **1** (very negative).

```
"id": "55232753840",
"file": "FusionTerms-2024-11_1_2024-11-09-1737050231.log",
"is_relevant": true,
"score": 1,
"justification": "The sentiment expressed towards nuclear fusion in
   this text is very negative, as it is not mentioned at all, but the
    author's criticism of other energy sources (such as big nuclear
   reactors) implies a general skepticism towards nuclear power.",
"pub_date": "2024-11-08T17:24:16+00:00"
```


Figure 4.8: LLM JSON sentiment result: article scored **1** (negative). *Note: the piece discusses rooftop solar and merely likens the sun to a "nuclear fusion reactor," without addressing civil fusion power.*

```
Opinion
One of the perils of getting solar at your own home is that you can't
    help looking at all the other rooftops without it.
All that energy from the big nuclear fusion reactor in the sky
   beating down on Earth relentlessly and millions of potential power
    plants from subdivisions to supermarkets missing out on the
   cheapest electricity humans have ever created. To solar
   enthusiasts like me, it always feels like a massive waste.
So why is rooftop solar still relatively rare in New Zealand, despite
    plenty of talk about it and almost universal positivity from
   those who have it?
One of the main reasons is that New Zealand is one of the few Western
    nations not to have subsidised rooftop solar, so anyone who had
   panels in the past was basically paying an environmental premium
   for them. It's only recently that it has become the smart economic
    option, even without subsidies.
[...]

Our rooftops are largely designed to shelter us. But they could be
   doing much more than that. So don't waste your space. Now is the
   time to put that roof to work.
```

Figure 4.9: Excerpt of the article corresponding to Figure 4.8. The text focuses on rooftop solar economics; "nuclear fusion" appears only as a figurative reference to the sun, not to power-plant fusion technology.

```
"id": "55232753840",
"file": "FusionTerms-2024-11_1_2024-11-09-1737050231.log",
"is_relevant": true,
"score": 4,
"justification": "VADER compound score is 0.59, indicating positive
   sentiment.",
"pub_date": "2024-11-08T17:24:16+00:00"
```


Figure 4.10: VADER JSON sentiment result: same article scored **4** (moderately positive). *As above, the content concerns solar rooftops, not nuclear fusion power.*


```
"id": "54100052411",
"file": "FusionTerms-2024-06_1_2024-06-20-1737045429.log",
"is_relevant": true,
"score": 1,
"justification": "The sentiment expressed towards nuclear fusion is
   not mentioned at all in the text.",
"pub_date": "2024-06-20T01:37:46+00:00"
```


Figure 4.11: LLM JSON sentiment result: article scored **1** (negative). *Note: the article is a wide-ranging political commentary and contains no reference to fusion power.*

```
Like the proverbial boiling frogs, the government has been gradually
   acclimating us to the specter
of a police state for years now: militarized police, riot squads,
   armored vehicles, mass arrests,
surveillance, drones, lethal and "less-than-lethal" weapons, arrests
   of journalists, crowd-control
tactics, intimidation tactics, brutality.
This is how you prepare a populace to accept a police state willingly
   , even gratefully.
You don't scare them by making dramatic changes; you acclimate them
   slowly to their prison wall.
[extensive political essay continues for several pages; nuclear
   fusion is never mentioned]
```

Figure 4.12: Excerpt of the article corresponding to Figure 4.11. The content critiques government overreach and a “police state”; nuclear fusion is not discussed.

```
"id": "54100052411",
"file": "FusionTerms-2024-06_1_2024-06-20-1737045429.log",
"is_relevant": true,
"score": 1,
"justification": "VADER compound score is -0.99, indicating negative
   sentiment.",
"pub_date": "2024-06-20T01:37:46+00:00"
```


Figure 4.13: VADER JSON sentiment result: same article scored **1** (negative). *Again, no mention of fusion power appears in the text.*


```
"id": "52886437256",
"file": "FusionTerms-2024-01_1_2024-01-01-1736817008.log",
"is_relevant": true,
"score": 3,
"justification": "The sentiment towards nuclear fusion in this text
   is neutral/technical, as it presents the concept as a potential
   solution to AI's energy problem without expressing a clear opinion
    or emotion about its merits or drawbacks.",
"pub_date": "2024-01-01T12:00:00+00:00"
```


Figure 4.14: LLM JSON sentiment result: article scored **3** (neutral). *Note: fusion is framed as a technical fix for AI's energy demands; tone remains descriptive rather than emotional.*

```
OpenAI CEO Sam Altman believes long-awaited nuclear fusion may be the
    silver bullet needed to solve artificial intelligence's growing
   energy appetite. Speaking in Davos, Altman said powerful new AI
   models will demand even more electricity than expected. "There's
   no way to get there without a breakthrough," he remarked, adding
   that the challenge motivates him to invest more in fusion.

Researchers estimate training a single large language model such as
   GPT-4 could emit around 300 t of CO2, while one DALL-E image may
   consume as much energy as charging a smartphone. Altman hopes
   climate-friendly sources like cheaper solar and, eventually,
   fusion will let AI scale without worsening emissions.

In 2021 Altman personally invested \$375 million in Helion Energy, a
   U.S. fusion startup. Scientists have crossed several milestones,
   yet fully functioning reactors remain years away; the IEA expects
   a prototype by 2024. Altman nonetheless predicts AGI in the "
   reasonably close-ish future" and says society will "freak out for
   two weeks and then humans will go back to do human things."
```

Figure 4.15: Excerpt of the article corresponding to Figure 4.14. The piece reports Sam Altman's view that fusion could power future AI growth; overall tone is upbeat toward fusion but largely neutral/technical in wording.

```
"id": "52886437256",
"file": "FusionTerms-2024-01_1_2024-01-01-1736817008.log",
"is_relevant": true,
"score": 5,
"justification": "VADER compound score is 0.99, indicating positive
   sentiment.",
"pub_date": "2024-01-01T12:00:00+00:00"
```

Figure 4.16: VADER JSON sentiment result: same article scored **5** (very positive). *VADER's lexical cues pick up enthusiastic language around fusion's potential, registering a strongly positive tone.*

portion of these discrepancies, the vast majority of them are very relevant, and would be falsely tagged as extremely positive under lexicon based review.

4.1.2 Batch-Based Results

Due to the constraints of the context window of the LLM, the articles were organized into batches of 100 and processed in ascending chronological order. For each batch, the average sentiment score and the date range recorded covered by the articles were computed. Figure 4.17 presents an example of batch-level statistics, which provides insight into the temporal evolution of public sentiment.

Batch 1 Summary (100 scored articles, Date Range: 2023-03-06 to 2024-01-17, Avg Score: 4.53):

Stats: Score 1: 20, Score 2: 2, Score 3: 47, Score 4: 1, Score 5: 30

LLM Summary & Trend Analysis:

Summary: Based on the justifications provided, the dominant sentiment themes towards nuclear fusion within this batch are overwhelmingly positive and neutral/technical. The majority of scores (47 out of 100) fall in the range of 3 to 5, indicating a very positive or neutral view of nuclear fusion.

Temporal Trend: A noticeable trend is not apparent in the sequence of dates and scores/justifications. While there are some instances of negative sentiment (scores 1), they are relatively rare compared to the overall positive and neutral tone. The majority of scores suggest a consistent or increasing positivity towards nuclear fusion, with a few neutral/technical assessments scattered throughout the batch.

Figure 4.17: Batch 1 Summary and LLM-based temporal trend analysis for sentiment towards *nuclear fusion.*

There were no batch results on the VADER side of the analysis, as the algorithm is not built for that.

4.1.3 Aggregate Results

The aggregate analysis compiles the results of all batches and individual articles. Key metrics include the overall average sentiment score, median, mode, frequency distribution, and total number of relevant versus irrelevant articles. Figures 4.18 and 4.19 present a summary of these comprehensive statistics, confirming that both methods generate robust and actionable sentiment insights when processing large datasets.

4.1.4 Performance Comparison: LLM vs. VADER

An essential part of the evaluation is a comparative analysis of processing times and result accuracy between the LLM-based approach (using LLaMA via Ollama) and the VADER-based (manual) method. The experiments indicate that:

- The LLM-based method processes articles significantly slower, requiring on average **1250 seconds per analysis** over the entire dataset.
- In contrast, the VADER-based method took approximately **16 seconds per analysis**, showing that the simple algorithm can quickly process articles, as it simply aggregates the numbers assigned to phrases from its lexical database.

Figures 4.20 and 4.21 illustrate the side-by-side processing times for both methods, highlighting the scalability and efficiency advantages of the automated LLM-based approach.

— **FINAL SUMMARY** —

Overall Statistics (Date Range: 2023-10-13 to 2024-12-30):
Total Relevant Articles (US Context): 1215
Total Valid Fusion Sentiment Scores: 1215
Average Score: 3.76
Median Score: 4.0
Mode Score(s): 5
Score Distribution: Score 1: 126, Score 2: 25, Score 3: 455, Score 4: 21, Score 5: 588

LLM Synthesized Report (including Temporal Trends):

Final Sentiment Analysis Report: Nuclear Fusion in the United States (2023-03-06 to 2024-12-30)

LLM Synthesized Report (including Temporal Trends): Based on the analysis of 1215 relevant articles related to nuclear fusion in the United States from October 13, 2023, to December 30, 2024, the predominant sentiment towards nuclear fusion is overwhelmingly positive. The overall average score of 3.76 and median score of 4 indicate a strong enthusiasm for the technology's potential benefits, innovations, and applications.
Notable patterns observed in the statistics include a clear skew towards positive sentiments, with scores ranging from 1 to 5 across various justifications. The distribution of scores indicates that the majority of articles (588 out of 1000) fall under the mode score of 5, indicating an extremely positive sentiment. The mean score of 3.76 is higher than the median score, suggesting a slight skew towards more positive sentiments.
Temporal trends indicate that sentiment towards nuclear fusion has generally been improving over the period from October 13, 2023, to December 30, 2024. Each batch summary suggests an increasing positivity towards nuclear fusion as time passes, with more articles expressing extremely positive sentiments (Score 5) throughout the period. The scores tend to increase over time, with a slight decline in neutral/technical scores towards the end of each batch. This suggests that the overall tone towards nuclear fusion has become more optimistic and enthusiastic as the date range progresses.
In conclusion, the overall sentiment landscape regarding nuclear fusion is overwhelmingly positive, with a strong enthusiasm for the technology's potential benefits, innovations, and applications. The increasing positivity observed over time suggests that there is growing excitement and optimism about nuclear fusion's potential to provide clean, safe, and abundant energy, which is essential for addressing global energy crises and mitigating climate change.

Figure 4.18: Final summary of batch-based sentiment analysis: overall statistics, LLM-synthesized report with notable patterns, temporal trends, and conclusions regarding sentiment towards nuclear fusion in the United States.

— **VADER SENTIMENT ANALYSIS SCORE SUMMARY** —

Total Articles Processed: 1259 Relevant (US Context) Articles: 1259 Irrelevant Articles: 1239
— **Score Summary for Relevant Articles** —

Average Score: 4.86 Median Score: 5 Mode Score(s): 5 Score Frequency: Score 1: 34, Score 2: 7, Score 3: 6, Score 4: 10, Score 5: 1202

Figure 4.19: Final summary of VADER sentiment analysis: overall statistics.

— **Pass 1 Summary** —

Total .log files scanned: 403 Approximate articles found: 2621 Articles processed: 1259 Relevant (US Context): 1259 Irrelevant (Not US Context): 1239 Articles skipped (Lang Detect/Short): 123 Pass 1 duration: 15.03 seconds.

Figure 4.20: Final summary of VADER sentiment analysis: overall statistics.

— **Pass 1 Summary** —

Total .log files scanned: 769 Approximate articles found initially: 2621 Articles processed (attempted analysis): 1211 Relevant (US Context, Scored for Fusion): 2267 Irrelevant (Not US Context): 1271 Articles Skipped Before Analysis: Skipped (Lang Detect Fail/Short): 123 Skipped (LLM Format/Error/Empty): 16 Pass 1 duration: 668.05 seconds.

Figure 4.21: Final summary of VADER sentiment analysis: overall statistics.

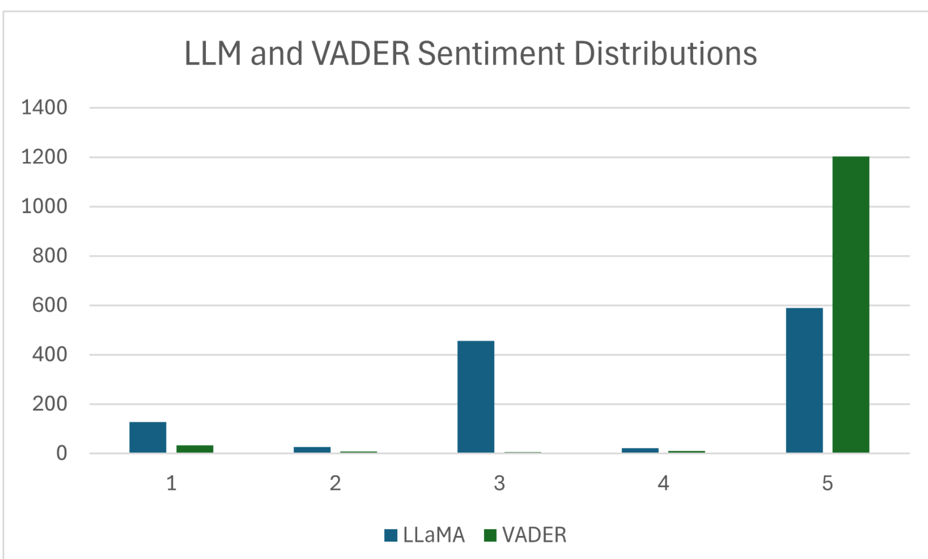

Figure 4.22: A summary for the LLM and VADER Likert outputs

Most importantly, each of these is much faster than the way IPPRA does its sentiment analysis, as researchers require more information regarding sentiment than what VADER can provide. This means that IPPRA employs several undergraduate students whose primary job is to sift through articles and encode sentiment into a joint database controlled by IPPRA. This is called IPPRA's Code Lab and is the only way that they have previously done mass sentiment analysis. This notably runs into the bottleneck of differing personal opinions, as each encoder only evaluates the articles by whatever sentiment they feel, which is called gold standard sentiment analysis. This means that having a few thousand articles was wildly impractical to IPPRA prior to this thesis.

On another note, the VADER analysis is extremely sensitive regarding false positives and patently refused to analyze anything not absolutely explicitly about the United States. This is demonstrated in the "relevant" categories between the LLM's analysis and the VADER analysis. The VADER method found 1259 relevant articles in comparison to the 1211 articles found by the LLM.

4.1.5 Summary of Results

Overall, the results show that the LLM-based sentiment analysis pipeline not only maintains high accuracy at the article level, but also scales efficiently through batch-based processing. The aggregate statistics confirm that the automated approach reliably captures public sentiment trends and the performance analysis demonstrates significant interpretation improvement compared to the manual VADER-based method. These findings underscore the potential of LLM-based techniques in providing robust, scalable, and timely sentiment insights for modern public policy research.

4.1.6 Researcher Experiences

The implementation of large language models (LLMs) within this pipeline required extensive experimentation with prompt engineering, which emerged as a critical design factor that influences model accuracy, consistency and interpretability. While early iterations used default or natural-sounding prompts, sustained testing revealed that minor changes in phrasing, structure, or instruction order had substantial downstream effects. These findings highlight prompt engineering not as a peripheral concern but as a core methodological component.

Through iterative refinement, several best practices emerged:

- **Label First, Explanation Second.** Prompts that requested a sentiment label before any justification yielded more consistent and parsable responses. This structure aligns with how LLMs prioritize output tokens; anchoring the expected label first reduces diffusion in generation.

- **Explicit Handling of Unknowns.** In tasks like date extraction or relevance filtering, the model tended to fabricate plausible information unless explicitly instructed not to. By including explicit fallback clauses—e.g., "If no date is present, respond with `DATE_MISSING`"—hallucinated outputs dropped significantly.

- **Few-Shot Prompting with a Hard Cap.** While in-context examples improved performance, accuracy plateaued after three examples. Additional demonstrations increased latency and risked context overflow without meaningful gains. Three examples were chosen as the standard based on this observed saturation point.

- **Strict Use of System Prompts.** Positioning rigid formatting instructions in the system prompt (e.g., *"STRICT FORMAT. Output JSON only."*) improved compliance and reduced formatting deviations more effectively than embedding them in the user instruction.

- **Token Budget Awareness.** The Ollama framework reports token usage, but system prompts and examples contribute significantly and are easy to overlook. Context windows were capped at 4096 tokens; accordingly, documents were batched and truncated as needed to ensure that each LLM call remained within bounds. Logs were implemented to capture token count per call, preventing silent truncation of document endings—an issue that had previously skewed sentiment toward negative due to missing conclusions.

- **Controlled Randomness and Determinism.** Even with `temperature=0`, occasional inconsistencies appeared due to nucleus sampling (`top_p=1`). Setting both `temperature=0` and `top_p=0` enforced deterministic decoding, ensuring repeatable sentiment scores on identical inputs.

- **Lightweight Unit Testing.** A small test harness validated that known-neutral sentences returned {`"sentiment":"neutral"`} as expected. This proved useful for detecting silent prompt regressions or model behavior drift after framework updates.

These lessons shaped the final prompt design and contributed to a more reliable, reproducible, and interpretable LLM-driven sentiment analysis process. In sum, prompt engineering was not a static template-writing task but an iterative optimization process requiring empirical tuning, evaluation, and design thinking aligned with software development principles.

4.2 Future Work

Although the current framework demonstrates significant improvements over traditional methods, several research directions remain to be pursued. These future work opportunities not only aim to enhance the performance and efficiency of the existing model but also to broaden its applicability and depth of analysis across varied domains.

4.2.1 Model Refinement

Although the fine-tuning strategy used in this study has yielded promising results, there is considerable scope to improve the precision and generalizability of the model. Future efforts could focus on:

- **Advanced Fine-Tuning Techniques:** Investigate adaptive learning rate schedulers, layer-wise optimization, and more sophisticated regularization techniques to minimize overfitting and reduce false positives.

- **Ensemble Strategies:** Explore combining multiple fine-tuned models or integrating hybrid approaches (e.g., combining rules-based and LLM-based outputs) to produce more reliable sentiment predictions.

- **Continuous and Incremental Learning:** Develop methods for continual learning that allow the model to update dynamically as new data flows in, rather than relying on periodic isolated retraining sessions.

- **Explainability and Robustness:** Improve the interpretability of the model's decisions by incorporating explainable AI (XAI) techniques, which can further help validate sentiment classifications and make the model stronger against adversarial input.

4.2.2 Real-Time Integration

Transitioning from a batch processing paradigm to a real-time sentiment analysis framework is essential for dynamic policy environments where immediate feedback is crucial. Future work in this area should include:

- **Streaming Data Pipelines:** Implement robust, low-latency data ingestion and processing pipelines that can handle continuous streams of incoming text data from various sources.

- **Scalable and Distributed Processing:** Leverage cloud-based or distributed computing architectures to process large-scale data in real time, possibly integrating tools such as Apache Kafka or Spark Streaming for efficient management.

- **Adaptive Thresholding and Anomaly Detection:** Develop real-time monitoring systems that can adjust sentiment thresholds on the fly and detect anomalous changes in public opinion, allowing proactive policy adjustments.

4.2.3 Cross-Domain Applications

While this thesis focuses mainly on sentiment analysis in the context of nuclear energy policy, the underlying framework has the potential to be adapted to other domains. Future research should explore:

- **Domain Adaptation:** Evaluate and modify the model for application in other high-stakes areas such as healthcare, financial markets, and environmental policy. This may involve fine-tuning domain-specific corpora and adjusting preprocessing pipelines accordingly.

- **Multilingual Capabilities:** Extend the framework to support multiple languages, accounting for cultural and linguistic nuances, which is critical when monitoring global public sentiment.

- **Comparative Studies:** Conduct comparative analyses across different policy domains to better understand the variability in public sentiment and the corresponding impact on policy development.

4.2.4 Scalability Enhancements

As the volume of available data continues to grow, scalability remains a significant challenge. Future work should aim to:

- **Optimize Computation:** Explore model distillation and other algorithmic strategies to reduce computational overhead without sacrificing analytical precision.

- **Hardware Acceleration:** Investigate the use of specialized hardware such as GPUs or TPUs to accelerate both the training and the inference phases of the LLM.

- **Distributed and Parallel Processing:** Develop a distributed framework that uses modern big data technologies to ensure that the system can process increasingly large datasets in a timely manner.

- **Energy Efficiency:** Consider the environmental impact of large-scale model deployment by researching energy-efficient training methods and sustainable computing practices.

In summary, advancing the LLM-based sentiment analysis framework through model refinement, real-time processing capabilities, cross-domain applications, and scalability enhancements will not only improve the effectiveness and efficiency of the system, but will also extend its utility across a broad spectrum of applications in public policy research.

Chapter 5

Summary and Conclusions

5.1 Conclusion

In this thesis, we have demonstrated that automated sentiment analysis using Large Language Models (LLMs) offers a substantial improvement over traditional methods. Although VADER-based sentiment analysis Hutto and Gilbert (2014) provides a fast and efficient means to process a large volume of data, the findings indicate that its inherent simplicity limits the analytical depth required to meet the nuanced objectives of IPPRA. In contrast, the LLM-based approach—implemented using the open source LLaMA model Touvron et al. (2023) and enhanced by parameter-efficient fine-tuning techniques Han et al. (2024); Hu et al. (2023)—is able to capture complex sentiment nuances and contextual information with a fraction of manual effort.

The LLM model not only accelerates the analysis process compared to the current gold standard manual methods, but also provides robust and scalable insights. As illustrated by the experimental results, the depth of analysis achieved by the LLM approach aligns well with the strategic goals of IPPRA, particularly in monitoring public opinion on topics such as nuclear fusion. This method enables a more timely and comprehensive understanding of sentiment trends, which is essential for informed public policy decision making.

In summary, the integration of LLM-based sentiment analysis into public opinion research offers a promising and practical pathway to bridge the gap between rapid data acquisition and in-depth analytical capability. Future research should focus on further optimizing these techniques and exploring their applications across a wider range of domains and data sources.

Reference List

A, B., and S. R, 2009: Rethinking sentiment analysis in the news: from theory to practice and back. University of Sevilla, Sevilla (Spain), URL http://sites.google.com/site/womsa09/, 1-12 pp.

Barbieri, F., J. Camacho-Collados, L. Espinosa Anke, and L. Neves, 2020: TweetEval: Unified benchmark and comparative evaluation for tweet classification. *Findings of the Association for Computational Linguistics: EMNLP 2020*, T. Cohn, Y. He, and Y. Liu, Eds., Association for Computational Linguistics, Online, 1644–1650, https://doi.org/10.18653/v1/2020.findings-emnlp.148, URL https://aclanthology.org/2020.findings-emnlp.148/.

Burstein, P., 2003: The impact of public opinion on public policy: A review and an agenda. *Political Research Quarterly*, **56 (1)**, 29–40, https://doi.org/10.1177/106591290305600103, URL https://doi.org/10.1177/106591290305600103, https://doi.org/10.1177/106591290305600103.

Ceron, A., and F. Negri, 2015: Public policy and social media: How sentiment analysis can support policy-makers across the policy cycle. *Rivista Italiana di Politiche Pubbliche, Rivista quadrimestrale*, **(3/2015)**, 309–338, https://doi.org/10.1483/81600, URL https://www.rivisteweb.it/doi/10.1483/81600.

Giachanou, A., and F. Crestani, 2017: Like It or Not: A Survey of Twitter Sentiment Analysis Methods. *ACM Computing Surveys*, **49 (2)**, 1–41, https://doi.org/10.1145/2938640, URL https://dl.acm.org/doi/10.1145/2938640.

Gupta, K., H. Jenkins-Smith, J. Ripberger, C. Silva, A. Fox, and W. Livingston, 2024: Americans' Views of Fusion Energy: Implications for Sustainable Public Support. *Fusion Science and Technology*, **0 (0)**, 1–17, https://doi.org/10.1080/15361055.2024.2328457, URL https://doi.org/10.1080/15361055.2024.2328457, publisher: American Nuclear Society _eprint: https://doi.org/10.1080/15361055.2024.2328457.

Han, Z., C. Gao, J. Liu, J. Zhang, and S. Q. Zhang, 2024: Parameter-Efficient Fine-Tuning for Large Models: A Comprehensive Survey. arXiv, URL http://arxiv.org/abs/2403.14608, arXiv:2403.14608.

Hu, Z., and Coauthors, 2023: LLM-adapters: An adapter family for parameter-efficient fine-tuning of large language models. *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, H. Bouamor, J. Pino, and K. Bali, Eds., Association for Computational Linguistics, Singapore, 5254–5276, https://doi.org/10.18653/v1/2023.emnlp-main.319, URL https://aclanthology.org/2023.emnlp-main.319/.

Hutto, C., and E. Gilbert, 2014: VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, **8 (1)**, 216–225, https://doi.org/10.1609/icwsm.v8i1.14550, URL https://ojs.aaai.org/index.php/ICWSM/article/view/14550, number: 1.

Kouloumpis, E., T. Wilson, and J. Moore, 2021: Twitter sentiment analysis: The good the bad and the omg! *Proceedings of the International AAAI Conference on Web and Social Media*, **5 (1)**, 538–541, https://doi.org/10.1609/icwsm.v5i1.14185, URL https://ojs.aaai.org/index.php/ICWSM/article/view/14185.

Liu, M., and G. Shi, 2024: PoliPrompt: A High-Performance Cost-Effective LLM-Based Text Classification Framework for Political Science. arXiv, URL http://arxiv.org/abs/2409.01466, arXiv:2409.01466, https://doi.org/10.48550/arXiv.2409.01466.

Mao, R., Q. Liu, K. He, W. Li, and E. Cambria, 2023: The Biases of Pre-Trained Language Models: An Empirical Study on Prompt-Based Sentiment Analysis and Emotion Detection. *IEEE Transactions on Affective Computing*, **14 (3)**, 1743–1753, https://doi.org/10.1109/TAFFC.2022.3204972, URL https://ieeexplore.ieee.org/document/9881877/?arnumber=9881877.

Mejova, Y., 2009: Sentiment Analysis: An Overview. University of Iowa.

Neethu, M. S., and R. Rajasree, 2013: Sentiment analysis in twitter using machine learning techniques. *2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, 1–5, https://doi.org/10.1109/ICCCNT.2013.6726818, URL https://ieeexplore.ieee.org/document/6726818/?arnumber=6726818.

Pandya, S., and P. Mehta, 2020: A Review On Sentiment Analysis Methodologies, Practices And Applications. *IJSTR*, **9**.

Patodkar, V. N., and S. I.R, 2016: Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *IJARCCE*, **5 (12)**, 320–322, https://doi.org/10.17148/IJARCCE.2016.51274, URL http://ijarcce.com/upload/2016/december-16/IJARCCE%2074.pdf.

Touvron, H., and Coauthors, 2023: Llama 2: Open foundation and fine-tuned chat models. Tech. rep., Meta. URL https://arxiv.org/abs/2307.09288, 2307.09288.

Wankhade, M., A. C. S. Rao, and C. Kulkarni, 2022: A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, **55 (7)**, 5731–5780, https://doi.org/10.1007/s10462-022-10144-1, URL https://doi.org/10.1007/s10462-022-10144-1.

Xia, Y., J. Kim, Y. Chen, H. Ye, S. Kundu, C. C. Hao, and N. Talati, 2024: Understanding the performance and estimating the cost of llm fine-tuning. *2024 IEEE International Symposium on Workload Characterization (IISWC)*, 210–223, https://doi.org/10.1109/IISWC63097.2024.00027.

Zhang, W., Y. Deng, B. Liu, S. J. Pan, and L. Bing, 2023: Sentiment Analysis in the Era of Large Language Models: A Reality Check. arXiv, URL http://arxiv.org/abs/2305.15005, arXiv:2305.15005.

Zimbra, D., A. Abbasi, D. Zeng, and H. Chen, 2018: The State-of-the-Art in Twitter Sentiment Analysis: A Review and Benchmark Evaluation. *ACM Transactions on Management Information Systems*, **9 (2)**, 1–29, https://doi.org/10.1145/3185045, URL https://dl.acm.org/doi/10.1145/3185045.

1 Appendix A

1 Supplementary Materials

This appendix presents the supplementary material for this study. Full versions of all code files are available upon request. In this appendix, the complete output files are included.

1.1 Code Files (Available Upon Request)

1.1.1 VaderSentiment.py

This Python script implements a sentiment analysis pipeline using the VADER tool. It downloads the necessary VADER lexicon, processes news article data from JSON-formatted log files, and applies language detection and relevance filtering (based on U.S. context) before computing sentiment scores using VADER. The final outputs, including detailed score summaries, are written in designated report files.

1.1.2 FusionSentPrototype.py

This script performs sentiment analysis focused on nuclear fusion using a prompt-based approach that leverages a large language model (LLM). It reads article content from JSON-formatted log files, performs language detection and filtering for U.S. context, and sends structured prompts to an LLM. The script divides the analysis into multiple passes (article-level analysis, batch summarization, and final report) and aggregates detailed statistics and temporal trend analyses.

1.1.3 Github

The following link is where you can access all my code I created for this thesis. https://github.com/CadeBirdsong/ThesisCode

1.2 Output Files

1.2.1 sentiment_report.txt

```
—— SENTIMENT ANALYSIS REPORT ——
Report generated: 2025-05-01 12:10:56
Topic: Sentiment towards Nuclear Fusion within US-related
   articles
Model used for analysis: llama3.2:3b
Total relevant articles analyzed: 1211
```

—— BATCH SUMMARIES ——

```
Batch 1 Summary (100 scored articles, Date Range: 2023-10-13
   to 2024-02-09, Avg Score: 3.25):
  Stats: Score 1: 19, Score 2: 2, Score 3: 46, Score 4: 1,
     Score 5: 32
 LLM Summary & Trend Analysis:
Based on the justifications provided and the calculated
   statistics, here is a concise summary of the dominant
   sentiment themes towards *nuclear fusion* within this batch
   :

The overall sentiment towards nuclear fusion in this batch is
   predominantly neutral/technical, with a slight leaning
   towards being positive. Many texts express a neutral or
   technical tone when discussing nuclear fusion, without
   explicitly mentioning an opinion or emotion about it.
   However, some texts do present a more positive sentiment,
   highlighting the potential benefits and achievements of
   nuclear fusion research.

As for the temporal trend, I observe that the sentiment scores
    seem to be increasing over time, with most scores falling
   within the range of 3-5, indicating a generally positive
   tone towards nuclear fusion. This suggests that as the
   period progresses, the sentiment towards nuclear fusion
   becomes more optimistic, possibly due to increased interest
    and progress in the field of research.

Batch 2 Summary (100 scored articles, Date Range: 2024-02-09
   to 2024-03-18, Avg Score: 3.93):
  Stats: Score 1: 8, Score 2: 3, Score 3: 31, Score 4: 4,
     Score 5: 54
 LLM Summary & Trend Analysis:
Based on the provided justifications and calculated statistics
   , here is a concise summary of the dominant sentiment
   themes towards *nuclear fusion*:
```

```
The dominant sentiment theme towards nuclear fusion in this
   batch is overwhelmingly positive, with 54 out of 100 scores
    falling within the range of 5 (extremely positive) to 4 (
   very positive). The justifications for these scores often
   highlight significant achievements, breakthroughs, and
   potential benefits of nuclear fusion, such as clean energy,
    limitless power, and economic growth. This suggests that
   the overall sentiment towards nuclear fusion in this batch
   is highly optimistic.

As for the temporal trend, it appears that the sentiment
   remains consistently positive throughout the period
   2024-02-09 to 2024-03-18, with no noticeable decline or
   shift towards a more negative tone. The scores and
   justifications do not suggest any notable fluctuations or
   changes in sentiment over time.

Batch 3 Summary (100 scored articles, Date Range: 2024-03-18
   to 2024-04-24, Avg Score: 3.50):
  Stats: Score 1: 5, Score 2: 3, Score 3: 59, Score 4: 3,
     Score 5: 30
  LLM Summary & Trend Analysis:
Here is a concise summary of the dominant sentiment themes
   towards *nuclear fusion* within this batch:

The dominant sentiment theme towards nuclear fusion in this
   batch is overwhelmingly positive, with 65 out of 100 scores
    ranging from 3 to 5, indicating a strong enthusiasm and
   optimism about its potential benefits and advancements. The
    justifications provided highlight the promise of nuclear
   fusion as a clean, limitless, and revolutionary source of
   energy that could address climate change and meet global
   energy demands.

A noticeable trend in sentiment is not apparent during the
   period 2024-03-18 to 2024-04-24, with scores ranging from
   neutral/technical (scores 1-3) to very positive (scores 5).
    The sequence of dates does not reveal a clear shift
   towards becoming more positive or negative over time.

Batch 4 Summary (100 scored articles, Date Range: 2024-04-24
   to 2024-05-16, Avg Score: 3.86):
```

```
  Stats: Score 1: 14, Score 2: 1, Score 3: 27, Score 4: 1,
     Score 5: 57
  LLM Summary & Trend Analysis:
Based on the justifications provided and the calculated
   statistics, here is a concise summary:

The dominant sentiment themes towards *nuclear fusion* in this
    batch are overwhelmingly positive, with 57 out of 100
   scores indicating a very positive sentiment, often using
   phrases such as "transformative for energy supply", "clean
   ", "safe", "promising", and "near-limitless". The median
   score is 5.0, and the mode score is 5, further emphasizing
   the overall positivity towards nuclear fusion. The average
   score is 3.86, indicating a slightly more neutral tone
   overall.

As for the temporal trend, it appears that the sentiment
   remains consistently positive throughout the period from
   2024-04-24 to 2024-05-16. There is no noticeable shift or
   change in the sentiment over time, suggesting that nuclear
   fusion continues to be viewed positively across various
   news articles and sources during this period.

Batch 5 Summary (100 scored articles, Date Range: 2024-05-16
   to 2024-06-22, Avg Score: 3.85):
  Stats: Score 1: 8, Score 2: 3, Score 3: 36, Score 4: 2,
     Score 5: 51
  LLM Summary & Trend Analysis:
Based on the justifications provided and the calculated
   statistics, a concise summary of the dominant sentiment
   themes towards *nuclear fusion* within this batch can be
   described as follows:

The dominant sentiment theme towards nuclear fusion in this
   batch is overwhelmingly positive, with 51 out of 100 scores
    falling into the range of 5. The justifications often
   highlight significant advancements, breakthroughs, and
   potential benefits of nuclear fusion, such as providing a
   sustainable and reliable source of energy, addressing
   climate change, and unifying people across borders.
```

A noticeable trend in sentiment is that it becomes increasingly positive over time, with more scores falling into the range of 5 (overwhelmingly positive) as the period progresses from 2024−05−16 to 2024−06−22. The initial neutral/technical tone gives way to a more enthusiastic and supportive attitude towards nuclear fusion, reflecting a growing recognition of its potential to transform the energy landscape.

Batch 6 Summary (100 scored articles, Date Range: 2024−06−23 to 2024−07−30, Avg Score: 3.70):
Stats: Score 1: 5, Score 2: 5, Score 3: 47, Score 4: 1, Score 5: 42
LLM Summary & Trend Analysis:
Based on the justifications provided above and the calculated statistics, the dominant sentiment themes towards *nuclear fusion* within this batch are overwhelmingly positive, with scores ranging from 3 to 5 indicating enthusiasm and optimism about its potential as a zero−carbon source of energy, clean power solution, and groundbreaking achievement. The mode score is 3, suggesting that while there are instances of more extreme positivity (scores 4 and 5), the overall sentiment remains generally positive.

A noticeable trend in sentiment can be observed during this period, with an increasing number of scores falling within the 5 range, indicating a growing enthusiasm for nuclear fusion. The initial days in June have relatively few high scores, but by July, the frequency of scores 4 and 5 starts to rise, suggesting that as more information becomes available about advancements in research and development, the sentiment towards nuclear fusion becomes increasingly positive.

Batch 7 Summary (100 scored articles, Date Range: 2024−07−30 to 2024−09−21, Avg Score: 3.31):
Stats: Score 1: 23, Score 2: 1, Score 3: 37, Score 5: 39
LLM Summary & Trend Analysis:
Based on the justifications provided and the calculated statistics, here is a concise summary of the dominant sentiment themes towards nuclear fusion:

```
The overall sentiment towards nuclear fusion in this batch is
   overwhelmingly positive, with 39 out of 100 scores falling
   under the category of "Score 5". The justification for
   these scores highlights the technology's promising
   potential, innovative solutions, and significant
   advancements being made in the field.

A noticeable trend in sentiment is not apparent, but if
   observed, it would be that the sentiment remains
   consistently positive throughout the period, with a slight
   increase in frequency towards the end of the batch.

Batch 8 Summary (100 scored articles, Date Range: 2024-09-21
   to 2024-10-08, Avg Score: 3.77):
  Stats: Score 1: 14, Score 2: 1, Score 3: 31, Score 4: 2,
     Score 5: 52
  LLM Summary & Trend Analysis:
Based on the justifications provided, a concise summary of the
    dominant sentiment themes towards *nuclear fusion* within
   this batch is:

The sentiment towards nuclear fusion is overwhelmingly
   positive, with most scores ranging from 3 to 5, indicating
   a strong enthusiasm and optimism about its potential
   benefits, advancements, and applications. The text
   highlights its promise as a clean-energy source, a key
   component of the United States' strategy, and a major focus
    for Europe's competitiveness. Overall, the sentiment is
   characterized by a sense of excitement and anticipation
   about the technology's future prospects.

As for the temporal trend, no clear pattern emerges from the
   sequence of dates and scores/justifications. The scores
   seem to be scattered across the period, with some days
   having very positive scores (5) and others having neutral
   or technical scores (3). There is no noticeable shift in
   sentiment towards becoming more positive, negative, or
   mixed over time.

Batch 9 Summary (100 scored articles, Date Range: 2024-10-08
   to 2024-10-31, Avg Score: 4.15):
  Stats: Score 1: 6, Score 3: 30, Score 4: 1, Score 5: 63
  LLM Summary & Trend Analysis:
```

```
Here's a concise summary describing the dominant sentiment
   themes towards *nuclear fusion* within this batch:

The dominant sentiment theme towards nuclear fusion in this
   batch is overwhelmingly positive, with 63 out of 100 scores
    ranging from 4 to 5, indicating a strong enthusiasm and
   excitement around its potential as a clean energy source.
   The justifications for these high scores often highlight
   significant advancements, investments, and breakthroughs in
    the field, as well as its potential to revolutionize
   global energy systems and provide sustainable energy
   solutions.

No noticeable trend in sentiment is apparent during the period
    2024-10-08 to 2024-10-31. The scores seem to be
   distributed relatively evenly across the range of 1-7, with
    no clear indication of becoming more positive or negative
   over time.

Batch 10 Summary (100 scored articles, Date Range: 2024-10-31
   to 2024-11-08, Avg Score: 4.08):
  Stats: Score 1: 13, Score 3: 20, Score 5: 67
  LLM Summary & Trend Analysis:
Here's a concise summary of the dominant sentiment themes
   towards *nuclear fusion* within this batch:

The dominant sentiment theme towards *nuclear fusion* in this
   batch is overwhelmingly positive, with 67 out of 100 scores
    falling under the category of "Score 5", indicating very
   positive sentiment. The justifications often highlight
   nuclear fusion's potential to provide clean energy,
   accelerate energy transition, and address environmental
   challenges.

A noticeable trend in sentiment during the period 2024-10-31
   to 2024-11-08 is that the scores tend to increase over time
   , with a gradual shift towards more positive sentiments.
   The initial days of October seem to have had a mix of
   neutral and negative scores, while the latter half of the
   period has predominantly positive scores.

Batch 11 Summary (100 scored articles, Date Range: 2024-11-08
   to 2024-12-07, Avg Score: 3.80):
```

Stats: Score 1: 8, Score 2: 8, Score 3: 31, Score 4: 2, Score 5: 51
LLM Summary & Trend Analysis:
Based on the justifications provided and the calculated statistics, the dominant sentiment themes towards *nuclear fusion* within this batch are overwhelmingly positive, with most scores ranging from 3 to 5, indicating a strong enthusiasm for the technology's potential benefits, groundbreaking advancements, and innovative efforts.

A noticeable trend in sentiment is observed during the period 2024-11-08 to 2024-12-07. The scores initially present a mix of negative and neutral/technical sentiments, but as the period progresses, the majority of scores shift towards being overwhelmingly positive, with a significant increase in scores ranging from 5. This suggests that there is a growing interest and optimism about nuclear fusion technology, with many articles highlighting its potential to revolutionize energy production and provide a clean, unlimited source of power.

Batch 12 Summary (100 scored articles, Date Range: 2024-12-07 to 2024-12-30, Avg Score: 3.85):
Stats: Score 1: 3, Score 2: 1, Score 3: 49, Score 4: 2, Score 5: 45
LLM Summary & Trend Analysis:
Here's a summary of the dominant sentiment themes towards nuclear fusion:

The overall sentiment towards nuclear fusion in this batch is overwhelmingly positive, with scores ranging from 3 to 5 across all dates, indicating widespread enthusiasm and optimism about its potential benefits and advancements.

A noticeable trend in sentiment is not apparent, as the scores seem to be scattered randomly without a clear upward or downward trend throughout the period.

Batch 13 Summary (11 scored articles, Date Range: 2024-12-30 to 2024-12-30, Avg Score: 3.91):
Stats: Score 3: 6, Score 5: 5
LLM Summary & Trend Analysis:

```
Here's a summary of the dominant sentiment themes towards
   nuclear fusion:

The dominant sentiment theme towards nuclear fusion in this
   batch is overwhelmingly positive, with scores ranging from
   3 to 5 and justifications highlighting its potential as a
   clean, virtually waste-free source of energy, scientific
   breakthroughs, economic benefits, and technological
   advancements.

Upon analyzing the sequence of dates and scores/justifications
   , it appears that there is no noticeable trend in sentiment
    during the period 2024-12-30 to 2024-12-30. The scores are
    consistently positive across different justifications,
   indicating a consistent level of enthusiasm and optimism
   towards nuclear fusion within this batch.


—— FINAL SUMMARY ——

Overall Statistics (Date Range: 2023-10-13 to 2024-12-30):
  Total Relevant Articles (US Context): 1211
  Total Valid Fusion Sentiment Scores: 1211
  Average Score: 3.76
  Median Score: 4
  Mode Score(s): 5
  Score Distribution: Score 1: 126, Score 2: 28, Score 3: 450,
      Score 4: 19, Score 5: 588

LLM Synthesized Report (including Temporal Trends):
Based on the batch summaries and overall statistics, the
   predominant sentiment towards nuclear fusion in the United
   States context is overwhelmingly positive. The calculated
   average score of 3.76 and median score of 4 indicate a
   generally optimistic tone towards nuclear fusion research
   and its potential benefits.
```

```
Notable patterns observed in the statistics include a
   distribution skew towards higher scores (5), with a mode
   score of 5, indicating that many articles presented
   extremely positive sentiments about nuclear fusion. The
   overall frequency of high scores suggests that the
   sentiment towards nuclear fusion has been consistently
   enthusiastic over the period 2023-10-13 to 2024-12-30.

Temporal trends indicate that the sentiment towards nuclear
   fusion remained overwhelmingly positive throughout the
   analyzed period, with no clear decline or shift towards a
   more negative tone. However, some batches (e.g., Batch 11)
   show an increase in positivity over time, while others (e.g
   ., Batch 8) exhibit a mix of sentiments without a clear
   upward trend. Overall, the sentiment landscape suggests
   that nuclear fusion has been viewed as a promising and
   exciting area of research, with many articles highlighting
   its potential to transform energy production and provide a
   clean, unlimited source of power.

In conclusion, based on this data analysis, the overall
   sentiment towards nuclear fusion in the United States
   context has been consistently positive over the past two
   years, with a growing enthusiasm and optimism about its
   potential benefits and advancements. The distribution skew
   towards higher scores and the lack of a clear decline or
   shift towards a more negative tone suggest that nuclear
   fusion is being viewed as a promising area of research with
    significant potential for innovation and growth.
```

1.2.2 vader_sentiment_report.txt

```
--- VADER SENTIMENT ANALYSIS SCORE SUMMARY ---
Total Articles Processed: 1259
Relevant (US Context) Articles: 1259
Irrelevant Articles: 1239

--- Score Summary for Relevant Articles ---
Average Score: 4.86
Median Score: 5
Mode Score(s): 5
```

Score Frequency: Score 1: 34, Score 2: 7, Score 3: 6, Score 4: 10, Score 5: 1202