


## Article

# Using Generative Pre-Trained Transformers (GPT) for Electricity Price Trend Forecasting in the Spanish Market

Alberto Menéndez Medina and José Antonio Heredia Álvaro \* 

Cátedra Industria 4.0, Universitat Jaume I, 12071 Castellón, Spain

\* Correspondence: heredia@uji.es

**Abstract:** The electricity market in Spain holds significant importance in the nation's economy and sustainability efforts due to its diverse energy mix that encompasses renewables, fossil fuels, and nuclear power. Accurate energy price prediction is crucial in Spain, influencing the country's ability to meet its climate goals and ensure energy security and affecting economic stakeholders. We have explored how leveraging advanced GPT tools like OpenAI's ChatGPT to analyze energy news and expert reports can extract valuable insights and generate additional variables for electricity price trend prediction in the Spanish market. Our research proposes two different training and modelling approaches of generative pre-trained transformers (GPT) with specialized news feeds specific to the Spanish market: in-context example prompts and fine-tuned GPT models. We aim to shed light on the capabilities of GPT solutions and demonstrate how they can augment prediction models by introducing additional variables. Our findings suggest that insights derived from GPT analysis of electricity news and specialized reports align closely with price fluctuations post-publication, indicating their potential to improve predictions and offer deeper insights into market dynamics. This endeavor can support informed decision-making for stakeholders in the Spanish electricity market and companies reliant on electricity costs and price volatility for their margins.

**Keywords:** electricity market price; Spain; Generative AI; GPT; sentiment analysis



**Citation:** Menéndez Medina, A.; Heredia Álvaro, J.A. Using Generative Pre-Trained Transformers (GPT) for Electricity Price Trend Forecasting in the Spanish Market. *Energies* **2024**, *17*, 2338. <https://doi.org/10.3390/en17102338>

Academic Editors: Mateus Mendes and Inácio Fonseca

Received: 19 March 2024

Revised: 6 May 2024

Accepted: 7 May 2024

Published: 13 May 2024

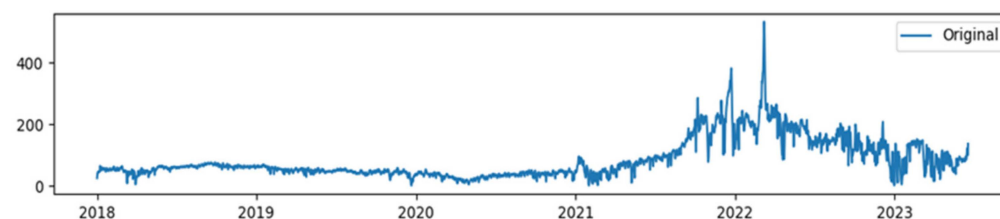


**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The forecast of the medium- to long-term price trend of the electricity market is subject to considerable uncertainty due to the influence of multiple complex factors—geopolitical events, climatic phenomena, social developments, regulations, technical factors, economic cycles, etc. [1].

Analyzing the evolution of the Spanish electricity market (OMIE) from 2018 to 2023 [2] unravels key factors behind heightened volatility. Understanding the intricate evolution of energy prices between 2018 and 2023 (Figure 1) sheds light on the multifaceted factors that have influenced this journey. According to the International Energy Agency [3], during the initial years, from 2018 to 2021, energy prices were shaped by a confluence of variables, including rising costs across various fuels and technologies, supply chain pressures, labor market constraints, and fluctuations in critical mineral supplies and construction materials. Notably, clean energy costs, which had previously witnessed steady declines, exhibited a distinct uptrend during this period.



**Figure 1.** Evolution of energy price in the electricity Spanish market OMIE (2018–2023).

However, the narrative dramatically turned after 2020, with energy prices experiencing heightened volatility. As reported, the primary catalyst for this volatility was the growing disparity between surging demand and a limited pipeline of new conventional projects within the oil industry. This imbalance introduced a significant risk of price spikes, casting shadows on the global economy's stability. Moreover, competitive electricity markets grappled with a widening gap between revenue from electricity sales and total generation costs, further contributing to price fluctuations.

The COVID-19 pandemic, a global disruptor, played its part by dampening overall energy demand, particularly in the case of carbon-intensive fuels like coal and oil. Conversely, renewable energy sources proved more resilient in the face of the pandemic's impacts. CO<sub>2</sub> emissions were notably reduced, and the energy sector witnessed a decline in capital investment, primarily affecting oil and natural gas supply projects. These repercussions are expected to reverberate through energy markets in the years to come.

Furthermore, the world's current energy crisis, initiated by Russia's invasion of Ukraine, has added another layer of complexity to the energy landscape. High energy prices, particularly in the natural gas sector, have led to wealth transfers from consumers to producers, affecting electricity generation costs globally. The crisis has also posed challenges in ensuring access to modern energy for many, with lingering uncertainties about its duration and fossil fuel price trends. While short-term shifts have seen increased demand for oil and coal as alternatives to costly gas, the long-term trajectory points toward low-emissions sources such as renewables, nuclear energy, and heightened efficiency measures.

Previous time series forecasting approach models used to determine energy prices prior to the COVID-19 pandemic [4–6] faced significant challenges during and after the crisis. They struggled to adapt to the abrupt disruptions in energy demand, increased price volatility, supply chain pressures, and changes in the energy mix brought about by the pandemic. Additionally, uncertainties in the economic and policy landscapes further hindered their accuracy. The pandemic exposed the limitations of these forecasting models in coping with such unforeseen disruptions, emphasizing the need for more adaptable and robust forecasting approaches to navigate the evolving energy market dynamics effectively.

### *1.1. Research Gap Analysis*

Scientific evidence suggests that having prior information about these socio-economic factors can be of great importance in understanding the possible future evolution of the market. For example, understanding what affects consumption patterns [7,8], demand density [9], and electricity generation [10,11] can be critical. In fact, many market specialists publish their understanding of recent market developments and possible trends based on factors such as social phenomena affecting demand, conflicts, government agreements, natural disasters, legislative actions, nuclear plant shutdowns, weather conditions, etc., either through journals or company reports available on the web.

These reports can be processed to apply sentiment analysis techniques. Typically, sentiment analysis techniques applied to markets (stock markets, oil markets, electricity markets) capture only headlines and classify the possible price direction. Research insights suggest that including sentiment analysis results improves quantitative predictions of price prediction models.

Several authors have applied some form of sentiment analysis as an intermediate tool for the purpose of forecasting markets; however, to our knowledge there are no published studies on the application of sentiment analysis to the electricity market.

The closest application relates to the oil market. There is a consensus among the authors that using the results of sentiment analysis models as input improves the quality of the results of the forecasting models [12–14].

Three categories of techniques recently used to analyze sentiment, also called opinion mining, in market news can be distinguished: lexicon-based, machine learning, and LLM. The aim of applying these tools has been to generate an additional input variable to time series forecasting models that concisely captures the influence of the multitude of

socio-economic factors that can affect the evolution of prices according to the opinion of market specialists.

The lexicon-based methods rely on a predefined set of words, called a lexicon, that are associated with a specific sentiment or opinion, e.g., positive, negative or neutral [15]. Essentially, in these methods, the sentiment of a text is determined by evaluating the number of positive, negative, and neutral words it contains. Zhao et al. [16] use the VADER algorithm [17], considered a reference lexicon-based algorithm with very good accuracy results.

Because manual creation and validation of a complete sentiment lexicon is labor-intensive and time-consuming, machine learning algorithms have been applied to “learn” the sentiment-relevant features of the text. Examples of such algorithms are those based on support vector machines (SVM) classifiers, [18], naive Bayes [17], and convolutional and recurrent neural networks [12,19]. Santos et al. [20] perform a state-of-the-art review of these types of techniques applied to the oil market up to 2022. Among the approaches used to extract public opinion, they show that machine learning-based methods have increased in prevalence in recent years because they improve classification accuracy over dictionary-based ones.

The fact that energy markets are very sensitive to factors outside the market itself, most of which are difficult to quantify and consider comprehensively, is a major bottleneck in analyses of electricity price news sentiment using methods that only consider market news in isolation such as those above. Large pre-trained linguistic models (LLM), such as BERT or ChatGPT, have shown remarkable potential for the interpretation of the intended message of any text.

On the one hand, since these models are general purpose language models and are not explicitly trained for prediction, they do not seem to be useful for predicting the evolution of a numerical time series such as price. Thus, Xie, Han, Lai, et al. [21] find that ChatGPT is no better than simple linear regression when using numerical data in prediction tasks, and Ko and Lee [22] attempt to use ChatGPT to help with a stock portfolio selection problem, but find no positive performance.

On the other hand, to the extent that these models are trained using truly large volumes of text and are able to understand the context of natural language, it could be argued that they could be valuable for processing textual information dealing precisely with possible market developments [23]. Several studies have employed BERT to forecast stock returns [24,25] and ChatGPT with good results [23].

Therefore, the performance of LLMs in predicting market movements is an open question. Our paper aims to help fill this gap by studying the potential of LLMs to help predict the evolution of the electricity market. The approach we propose is to use LLMs to interpret expert opinion on the possible evolution of the trend, and not to try to forecast a specific value of price.

In the literature we note that sentiment studies in stock or oil market have mostly focused on the analysis of news headlines. We understand that this is a limitation inherited from previous techniques and can be improved with the ability of LLMs to process more extended texts. For this reason, we propose to include in the specific training specialist articles and reports so that the model can better interpret expert opinion and also be able to explain the classification in an argumentative way.

Large language models (LLMs) are based on transformer architecture trained on a large volume of data (GPT), which are then fine-tuned for specific contexts [26]. One of the features of GPT neural network architecture is its use of transformer layers, which enable the model to handle long sequences of text by using self-attention mechanisms to focus on the most relevant parts of the input. This attention mechanism allows the model to understand the input context better and generate more accurate and coherent responses. Current LLMs have been trained on large and varied datasets, giving them the capability to understand the functioning of the economy and markets. To evaluate sentiment in a particular market, they need to be fine-tuned with specific new datasets derived from

specialist reports and related news to derive sentiments of the specific market [18,20]. Breitung et al. [27] apply this approach to the oil market by generating a specific dataset with 1600 records.

Advanced natural language processing methods have transformed AI interaction and generative pre-trained transformers (GPT) are ground-breaking for generating human-like text and comprehending natural language [28]. GPT models can be extended and refined with specialized news and reports analysis [29], which can significantly enhance energy price prediction models, in addition to sentiment analysis, in several ways:

- **Data Enrichment:** GPT-based models can analyze and extract valuable insights from a vast corpus of specialized news articles and reports on the energy market. This data enrichment provides a broader context for energy price forecasting models.
- **Event Detection:** GPT models can detect and highlight significant events [30], such as geopolitical developments, supply disruptions, or regulatory changes, that may impact energy markets. These detected events can be used as input variables for forecasting models.
- **Market News Summarization:** GPT can generate concise summaries of complex news articles and reports [31] making it easier for analysts and traders to stay informed about market developments. These summaries can serve as valuable inputs for forecasting models.
- **Identifying Influential Factors:** GPT can identify and rank factors mentioned in the news and reports likely to influence energy prices. This information can guide feature selection and help prioritize variables in forecasting models.
- **Customized Reports:** In the case of OpenAI's GPT, users can provide customized prompts to extract specific information or insights from news and reports. This allows for tailored analysis based on the unique requirements of the forecasting model.

Another important gap detected in the current state of the art on the application of sentiment analysis models to market performance is the lack of consideration of the time horizon of the prediction. For these tools to help with decision-making about the electricity market, we propose to qualify and distinguish whether the opinion is about the short, medium, or long term. This consideration requires the use of metrics that go beyond a one-dimensional measure of the impact of the news and also identify its time horizon.

Our assumption is that by incorporating GPT-based specialized news and reports analysis into energy price prediction models allows for a more comprehensive and timely understanding of market dynamics [23]. This, in turn, may improve the models' accuracy and helps energy market participants make informed decisions in an increasingly complex and dynamic environment. The application of GPT technology in predicting energy prices in the Spanish market is the focus of our research.

### 1.2. Research Objectives

In this paper, we analyze how the reasoning capabilities of large language models (LLMs) can be used to derive context-specific trend predictions from specialized news and expert reports. To this end, we propose a new multidimensional metric to evaluate the expert opinion gathered in the news that in addition to considering the direction of the trend also evaluates the intensity of the possible impact and its foreseeable time horizon.

We construct a dataset of electricity news and reports and compare the trend forecasting performance of the GPT with two baseline methods: VADER and BERT. Our findings indicate that LLMs can successfully leverage their reasoning capabilities to contextualize the evolution of electricity market prices.

We contribute to the literature on advanced natural language processing (NLP) tools applied to management science, illustrating how LLMs can be effectively used to extract business experts' perspectives on the future and consolidate their consensus, and provide tactical guidance to drive these models towards achieving optimal results.

In what follows, we present the methods used to adjust the GPT, the implementation details (the original code and the dataset are openly available (<https://github.com/AMM->

[UJI/energy-price-prediction-OpenAI](#))), the metrics that we have defined to evaluate the different dimensions of the sentiment extracted from the analysis of the texts, the results obtained, the future developments that would be interesting to address, and the main conclusions of this article.

## 2. Materials and Methods

OpenAI's GPT models, renowned for their natural language understanding and generation capabilities, provide a robust foundation for various applications. Two primary paradigms emerge regarding integrating private data and tailoring these models to specific tasks: context learning and fine-tuning [29]. These paradigms unlock the potential to enhance model performance, adapt to domain-specific nuances, and leverage the unique characteristics of proprietary datasets, all while preserving the underlying capabilities of the pre-trained GPT model.

### 2.1. Paradigm 1: In-Context Learning

In-context learning, also called prompt engineering or prompt design, is a paradigm that allows users to interact with GPT models by providing context-specific instructions or queries [32]. This approach enables the customizations of model responses without directly modifying the model. By crafting tailored prompts, users can elicit responses that align with their specific requirements, making it a versatile tool for leveraging the pre-trained knowledge of GPT models while adding a layer of domain specificity. In-context learning is precious when quick, context-aware responses are needed, such as in chatbots, customer support, or generating domain-specific content. However, it relies on the skillful design of prompts and may require iterative adjustments to achieve optimal results, as understanding and controlling model behavior through prompts can be nuanced.

### 2.2. Paradigm 2: Fine-Tuning

Fine-tuning represents a deeper level of customization, where the pre-trained GPT model is adapted to perform specific tasks or Excel domains [33]. The model is further trained on domain-specific data, including proprietary datasets, specialized text, or even structured information in this paradigm. Fine-tuning allows the model to learn from this additional data while preserving its general language understanding capabilities. This approach results in a more refined and specialized model capable of providing nuanced and context-aware responses in a particular domain. Fine-tuned models have applications in medical diagnosis, legal document analysis, and financial forecasting, where precise and domain-specific insights are essential. However, fine-tuning requires careful data curation, domain expertise, and a thorough understanding of the trade-offs involved in specialization, as excessive fine-tuning can risk overfitting to specific datasets and limit model generality.

### 2.3. Implementation Details

To conduct a comprehensive comparative analysis and measure the efficacy of OpenAI technologies in supporting time series trend forecasting, we created two sets of predictions with the in-context examples and fine-tuning approaches of OpenAI's customization, using a collection of news articles [34,35] and expert analysis reports specifically focused on the Spanish energy markets [36,37].

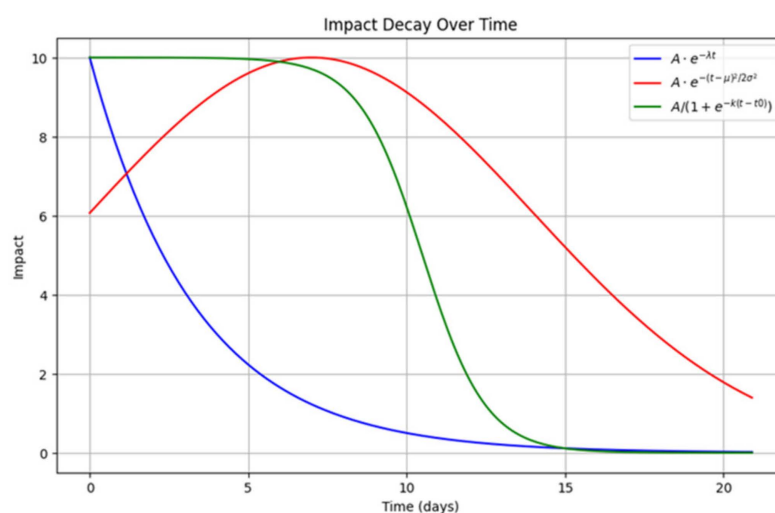
The OpenAI models (GPT engine: text-davinci-003) calculated three variables for each news article's content:

1. **Impact on Electricity Price (Scale 0–10):** The first variable quantifies the perceived impact of each news article on the price of electricity within a scale ranging from 0 (no impact) to 10 (high impact). This quantification allows us to discern the potential influence of each piece of news on energy prices, a critical factor in sentiment analysis used for forecasting models.



2. Direction of Impact (Up, Down, None): We evaluated whether the news articles indicated a potential price impact in the form of an increase (“up”), a decrease (“down”), or no discernible impact (“none”). Understanding the direction of influence is paramount for making informed predictions in the dynamic energy market.
3. Impact Period (Past, Short-term, Mid-term, Long-term, None): The third variable delves into the temporal aspect of impact, categorizing it into various periods—past, short-term, mid-term, long-term, or none. This temporal classification aids in determining when the anticipated price effects are likely to materialize, further enhancing the precision of our models.

Simulating the impact of news on the energy market involves understanding market dynamics and investor behavior when they receive updates [38]. While no formula perfectly captures this, we used a simplified “exponential decay” model to represent how the impact of news might fade over time for short-term news and a “Gaussian model” for mid- to long-term impacts (Figure 2). This model assumes that the impact of the news will be most significant immediately after or close to its release and will gradually diminish over time.



**Figure 2.** Three different approaches to simulate news impact decay over time.

At no point are dates provided to GPT, nor is it asked about specific moments in the past, to determine impact. The purpose of these three variables is to quantify the LLM’s ability to discern from news text what influence it will have on energy prices, effectively conducting sentiment analysis with the LLM.

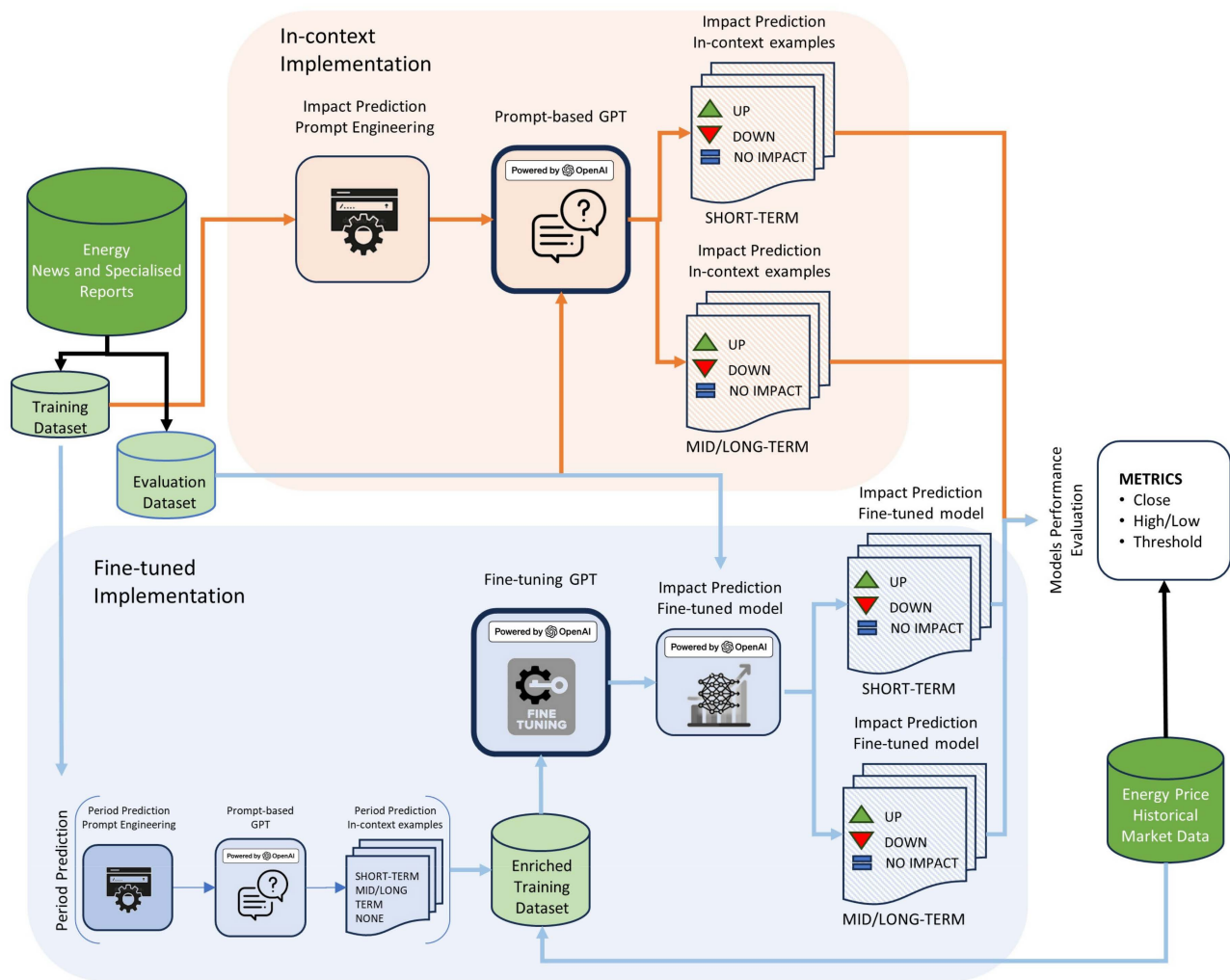
To refine this sentiment accuracy and transform GPT into an energy expert, we employ two paradigms or approaches (in-context learning and fine-tuning) to broaden its specific knowledge on the subject (Figure 3).

Once the model is trained, we feed it news to find similarities in language use, concepts, etc., to assign a value. Then, we aggregate and analyze using the proposed metrics, this time using the news’s specific date, to gauge the predictions’ accuracy.

### 2.3.1. In-Context Implementation

The in-context GPT prompt includes as training dataset eight examples of news with the expected impact on price, the timeframe, and the direction. Using these examples and its text analysis capability, the LLM assigns a value between 0 and 10.

Once trained, it is tasked with analyzing each news item (irrespective of its generation time). After an iteration process, the crafted prompt finally used follows this structure (Figure 4):



**Figure 3.** High-level solution design.

*Forget all your previous instructions. You are an expert in financial markets, especially in the analysis of the electricity market and industry.*

*I'm going to pass a series of news headlines, and you have to determine the following information:*

- *The impact on the price of electricity, on a scale from 0 to 10.*
- *The direction of the impact: Up (increase), Down (decrease), or None.*
- *The period when the impact will occur: Past, Short-term, Mid-term, Long-term, or None.*

*Please provide the information in the following format:*

*<Impact: (0–10)>*

*<Direction: (Up, Down, None)>*

*<Period: (Past, Short, Mid, Long, None)>*

**Figure 4.** Prompt for in-context training.

After obtaining responses for the selected news, we aggregate and analyze using the proposed metrics, using the specific date of the news, to assess the accuracy of predictions.

Taking the “impact” value of the news on the date of release, we calculated the average impact over the following period depending on the identified “duration”: for the short term, up to 3 weeks; for the mid/long term, up to 3 months. Then we grouped all these values, calculating the average of each interval (weeks for short-term, months for mid/long term). The symbol of the value indicated if it was a positive or negative impact (increase or decrease of the energy price).

### 2.3.2. Fine-Tuned Implementation

To create the fine-tuned model, for each news or article, we completed a first run with OpenAI using the “in-context examples prompt” to identify its assessment of “impact period” for each record (short-term, mid/long-term, past, none), keeping aside 25 randomly selected records for each possible value for the fine-tuning.

In parallel, we calculated for each date of the time series range (2018–2023) what difference (%) the price has suffered within the previous time interval to the news event (weeks for short term, months for mid/long term) as per the following formulas (the weights in the formula are a simplified approximation to the area under the curve for “exponential decay” and “Gaussian model” per Figure 2):

$$Impact_{week} = 0.5 \cdot \Delta Price_{week\ t} + 0.3 \cdot \Delta Price_{week\ t+1} + 0.2 \cdot \Delta Price_{week\ t+2}$$

$$Impact_{month} = 0.5 \cdot \Delta Price_{month\ t} + 0.3 \cdot \Delta Price_{month\ t+1} + 0.2 \cdot \Delta Price_{month\ t+2}$$

With this information, we populated each news sample record to be used for fine-tuning, including the “direction” field (sign of the impact), obtaining a dataset as the one shown below (Figure 5):

	Time	Headline	Price	RealImpact	Direction	Period
333	2019-12-27	Energía El sistema eléctrico registró un super...	32.518750	6.7	Up	Long
952	2022-09-03	GAS y PETROLEO REPowerEU: acción europea conju...	118.456250	-18.2	Down	Long
205	2018-12-13	Interrumpibilidad Alcoa se adjudica 21 millone...	60.335833	-5.2	Down	Long
709	2021-10-22	MERCADO ELÉCTRICO Iberdrola se compromete a no...	203.084583	3.3	Up	Long
564	2021-06-22	Energía El Ejecutivo dará 900 millones en ayud...	89.838750	25.3	Up	Long
...	...	...	...	...	...	...
889	2022-05-31	INFORMES El Barómetro energético de AEGE señal...	213.801667	9.4	Up	Short
623	2021-08-26	Combustible La gasolina baja por segunda seman...	122.575833	20.6	Up	Short
617	2021-08-16	Energía El precio de la luz sigue por debajo d...	90.188750	-1.2	Down	Short
370	2020-04-01	Coronavirus La crisis de las pymes asesta un d...	25.994583	-17.8	Down	Short
717	2021-10-26	Crónica de Bolsa Iberdrola da impulso al Ibex ...	215.337083	-22.1	Down	Short

125 rows x 6 columns

Figure 5. Pre-fine-tune data frame.

From here we were ready to follow the fine-tuning steps explained by OpenAI to create a custom model [33]:

- Dataset preparation: Every instance within the dataset should represent a conversation structured in a manner consistent with OpenAI’s Chat Completions API. This structure entails organizing the conversation as a list of messages, where each message comprises a role, content, and the possibility of including a name.
- Validate data formatting and divide training and testing datasets.
- Upload dataset file and create the fine-tuning job using the OpenAI SDK.
- Use the new fine-tuned model with the rest of the news and articles to enrich the dataset with calculated variables.



After applying the new model to the rest of the news and article dataset, we calculated the average impact per time interval (short-term: weekly average; mid/long-term: monthly average) before evaluating the performance with the metrics.

### 3. Results

From the results of analyzing the news and articles, we extract information like its impact on the price, the direction of that impact, and the period in which it will occur.

We have grouped the insights provided by this analysis in the short term and mid/long term, calculating the average impact for each interval and the sign indicating direction, and we evaluate the accuracy of that prediction based on the variation of price over the following period after the news occurred.

We have defined three different types of custom metrics to measure the accuracy of the direction of the impact prediction:

- **Close Price:** If the direction indicates that the price will go UP (Figure 6), the OPEN PRICE at the beginning of the first interval when the news is published (interval  $t$ ) should be LOWER than at least 1 of the CLOSE PRICE values of the current or the following two intervals ( $t$ ,  $t + 1$ ,  $t + 2$ ). The intervals will be weeks for short term and months for mid/long term.

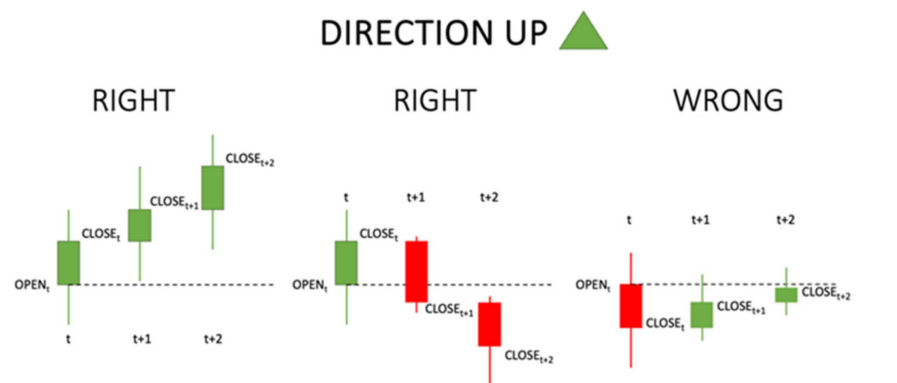


Figure 6. OpenAI close price metric: direction UP.

If the direction indicates that the price will go DOWN (Figure 7), the OPEN PRICE at the beginning of the first interval when the news is published ( $t$ ) should be HIGHER than at least 1 of the CLOSE PRICE values of the current or the following two intervals ( $t$ ,  $t + 1$ ,  $t + 2$ ).

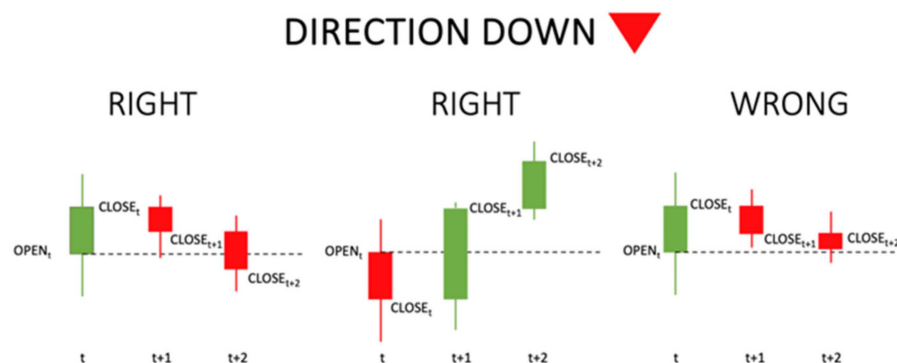
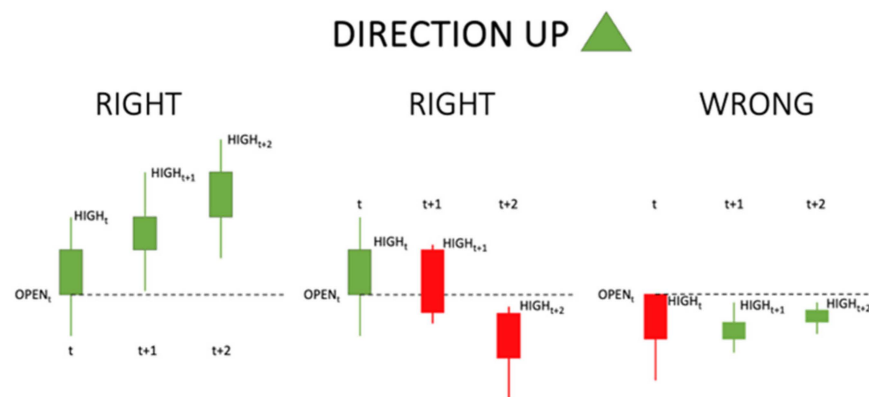


Figure 7. OpenAI close price metric: direction DOWN.

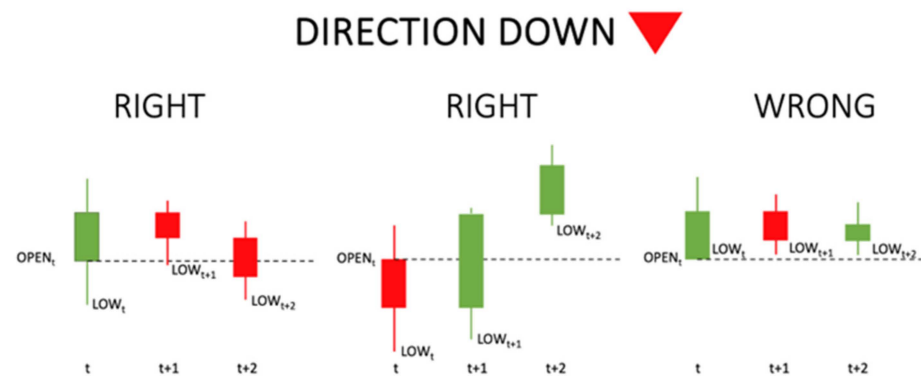
- **High/Low:** If the direction indicates that the price will go UP (Figure 8), the OPEN PRICE at the beginning of the first interval when the news is published (interval  $t$ ) should be LOWER than at least 1 of the HIGH PRICE values of the current or the

following two intervals ( $t, t + 1, t + 2$ ). The intervals will be weeks for short term and months for mid/long term.



**Figure 8.** OpenAI high/low price metric: direction UP.

If the direction indicates that the price will go DOWN (Figure 9), the OPEN PRICE at the beginning of the first interval when the news is published ( $t$ ) should be HIGHER than at least 1 of the LOW PRICE values of the current or the following two intervals ( $t, t + 1, t + 2$ ).



**Figure 9.** OpenAI high/low price metric: direction DOWN.

Threshold: Same as high or low, but there should be a minimum difference between the open price value and the high or low, depending on the direction of 2% for short-term and 5% for mid/long term.

### 3.1. Short-Term Analysis

The following figures (Figures 10 and 11) and tables (Tables 1 and 2) show the analysis results of short-term identified impacts on price using OpenAI and comparing both approaches: in-context examples and fine-tuning. The color of the arrows indicates OpenAI's prediction that prices will go UP (green) or DOWN (red), next to the price time series.

Tables 1 and 2 represents two metrics for evaluating classification models: accuracy and Matthews correlation coefficient (MCC), as these two metrics can cover balanced and imbalanced datasets. We have also added to the table the results of two baseline sentiment analysis methods for comparison: VADER (lexicon) and BERT (deep learning transformer).

Upon reviewing the short-term results for all models, the fine-tuned GPT model emerges as the most effective across various metrics. Its consistently higher accuracy and MCC scores, particularly in comparison to VADER and BERT models, signify its ability to better balance true positives and negatives, crucial for sentiment analysis tasks. Also, note that the BERT model's NaN MCC values are due to only be able to predict negative sentiment with the news dataset provided.

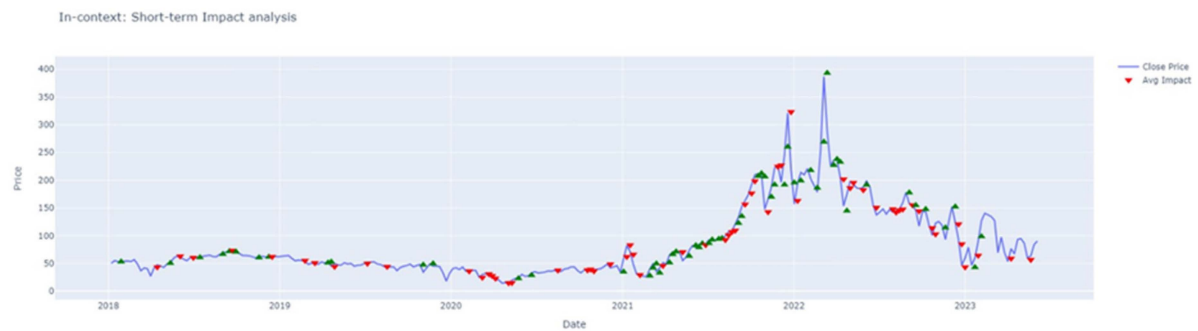


Figure 10. In-context short-term graph.



Figure 11. Fine-tuned short-term graph.

Table 1. Short-term metrics—accuracy.

Accuracy	In-Context	Fine-Tunned	VADER	BERT
Close Price	0.67	0.71	0.68	0.70
High/Low	0.76	0.81	0.77	0.79
Threshold 2%	0.59	0.65	0.55	0.57

Table 2. Short-term metrics—MCC.

MCC	In-Context	Fine-Tunned	VADER	BERT
Close Price	0.35	0.49	0.33	NaN
High/Low	0.53	0.64	0.52	NaN
Threshold 2%	0.20	0.32	0.07	NaN

### 3.2. Mid/Long-Term Analysis

The following figures (Figures 12 and 13) and Tables 3 and 4 show the results of the analysis of mid/long-term identified impacts on price using OpenAI and comparing both approaches (in-context examples and fine-tuning) and the two baseline sentiment analysis methods (VADER and BERT).



Figure 12. In-context mid/long-term graph.



Figure 13. Fine-tuned mid/long-term graph.

Table 3. Mid/long-term metrics—accuracy.

Accuracy	In-Context	Fine-Tuned	VADER	BERT
Close Price	0.69	0.81	0.71	0.69
High/Low	0.90	0.93	0.95	0.94
Threshold 5%	0.81	0.86	0.83	0.79

Table 4. Mid/long-term metrics—MCC.

MCC	In-Context	Fine-Tuned	VADER	BERT
Close Price	0.35	0.63	0.36	NaN
High/Low	0.61	0.87	0.89	NaN
Threshold 5%	0.47	0.74	0.64	NaN

The fine-tuned GPT model achieves similar or higher accuracy scores across all three evaluated scenarios: Close Price, High/Low, and Threshold 5%. Moreover, its MCC scores also show significant enhancements, indicating a better balance between true positives and false positives.

#### 4. Discussion

Based on the results for both short- and mid/long-term analyses, leveraging advanced language models like GPT, especially when fine-tuned for specific tasks and domains, yields superior results compared with state-of-the-art sentiment analysis methods. The fine-tuned GPT model consistently outperforms across various metrics, including accuracy and MCC, showcasing its ability to capture nuanced sentiment trends in energy market news over different time horizons.

In our model generation, we do not provide specific dates or historical moments to GPT, nor do we ask it about past events to gauge their impact. Instead, we leverage its capability to analyze news text and extract the opinion about how it may influence energy prices, essentially performing sentiment analysis with the LLM. Therefore, it is not a concern that the GPT model was trained using data from before September 2021 because it is not providing direct answers based on historical knowledge but rather utilizing its language analyzing capabilities. This assumption agrees with [16]; nevertheless, we checked it using news data from 2022, achieving similar performance results as when evaluating the model using data spanning from 2018 to 2022.

Building on the insights gained from this study, the contextual understanding and data sources used as input for the GPT models should be enhanced by incorporating additional relevant news and reports sources and increased periodicity and optimizing the accurate impact calculation and decay over time methods to translate into real impact.

The key benefit of increasing LLM awareness about market evolution with the in-context approach is that it avoids the need to re-modify LLM parameters for this specific task application. Instead, users can append an external knowledge repository, enriching the input and thus refining the output accuracy of the model. Therefore, in-context is seen as a more practical and economical approach, with a lower barrier to entry and independent

of the specific LLM model. However, for more extensive and professional applications, it is necessary to automate the workflows that allow for scalability and better precision of responses. A RAG architecture [39] can be developed for this purpose.

RAG has become one of the most popular architectures in LLM systems, combining automated information retrieval mechanisms and in-context learning to bolster LLM performance. In this framework, a query initiated by a user requests the retrieval of relevant information through search algorithms. This information is then integrated into the LLM indications, providing additional context for the generation process [40].

On the basis of a RAG architecture, we plan to develop the following additional functionalities:

- the incorporation of GPT-calculated features into multivariate time series prediction models as input variables.
- influential event detection as early warning signals (natural disasters, geopolitical conflicts, regulatory changes).
- automatic generation of reports that describe the recent evolution of the electricity market price and the prediction of price trends.

## 5. Conclusions

Our research explored the crucial realm of energy price prediction within the Spanish electricity market, which has significant implications for the nation's economy, sustainability goals, and energy security. We utilized GPT models, specifically OpenAI's ChatGPT, to create a new approach to create price trend variables based on news and specialized text insight analysis. This new approach can be used to create innovative hybrid models merging GPT insights with domain-specific knowledge and real-time data feeds to address the unique complexities of the electricity market, and more particularly the Spanish market.

Our results indicate the potential of GPT models to provide valuable insights in understanding mid-term price trends. The superiority of ChatGPT with augmented knowledge in predicting electricity trends can be attributed to its advanced language understanding capabilities, which allow it to capture the nuances and subtleties within specialized news and reports.

Therefore, we conclude that continued exploration and optimization of OpenAI's capabilities are essential to unlock their full potential in energy price forecasting.

**Author Contributions:** Conceptualization, A.M.M. and J.A.H.Á.; software, A.M.M.; writing A.M.M. and J.A.H.Á.; visualization, A.M.M.; supervision, J.A.H.Á. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original code and the dataset supporting reported results are openly available and can be found in the repository: <https://github.com/AMM-UJI/energy-price-prediction-OpenAI>.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Pezzutto, S.; Grilli, G.; Zambotti, S.; Dunjic, S. Forecasting Electricity Market Price for End Users in EU28 until 2020—Main Factors of Influence. *Energies* **2018**, *11*, 1460. [CrossRef]
2. OMI, Polo Español S.A. (OMIE). Market Results. 2018–2023 Retrieved from OMIE. Available online: <https://www.omie.es/en> (accessed on 1 February 2024).
3. International Energy Agency. *World Energy Outlook Annual Report. 2018–2023*; International Energy Agency: Paris, France, 2024.
4. Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *Int. J. Forecast.* **2014**, *30*, 1030–1081. [CrossRef]
5. Nowotarski, J.; Weron, R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. *Renew. Sustain. Energy Rev.* **2018**, *81*, 1548–1568. [CrossRef]
6. Qin, Q.X. An effective and robust decomposition-ensemble energy price forecasting paradigm with local linear prediction. *Energy Econ.* **2019**, *83*, 402–414. [CrossRef]



7. Bianco, V.; Manca, O.; Nardini, S. Electricity consumption forecasting in Italy using linear regression models. *Energy* **2009**, *34*, 1413–1421. [\[CrossRef\]](#)
8. Kumar, U.; Jain, V.K. Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* **2010**, *35*, 1709–1716. [\[CrossRef\]](#)
9. Hyndman, R.J.; Fan, S. Density forecasting for long-term peak electricity demand. *IEEE Trans. Power Syst.* **2010**, *25*, 1142–1153. [\[CrossRef\]](#)
10. Kamalov, F.; Sulieman, H.; Moussa, S.; Avante Reyes, J.; Safaraliev, M. Powering Electricity Forecasting with Transfer Learning. *Energies* **2024**, *17*, 626. [\[CrossRef\]](#)
11. Kok, M.; Lootsma, F.A. Pairwise-comparison methods in multiple objective programming, with applications in a long-term energy-planning model. *Eur. J. Oper. Res.* **1985**, *22*, 44–55. [\[CrossRef\]](#)
12. Gong, X.; Guan, K.; Chen, Q. The role of textual analysis in oil futures price forecasting based on machine learning approach. *J. Future Mark* **2022**, *42*, 1987–2017. [\[CrossRef\]](#)
13. Li, X.; Shang, W.; Wang, S. Text-based crude oil price forecasting: A deep learning approach. *Int. J. Forecast.* **2019**, *35*, 1548–1560. [\[CrossRef\]](#)
14. Jiang, Z.; Zhang, L.; Zhang, L.; Wen, B. Investor sentiment and machine learning: Predicting the price of China's crude oil futures market. *Energy* **2022**, *247*, 123471. [\[CrossRef\]](#)
15. Liu, B. Sentiment Analysis and Subjectivity. In *Handbook of Natural Language Processing*, 2nd ed.; Indurkha, N., Damerau, F., Eds.; Chapman & Hall: Boca Raton, FL, USA, 2010.
16. Zhao, L.T.; Zeng, G.R.; Wang, W.J.; Zhang, Z.G. Forecasting oil price using web-based sentiment analysis. *Energies* **2019**, *12*, 4291. [\[CrossRef\]](#)
17. Hutto, C.J.; Gilbert, E. VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. In Proceedings of the International AAAI Conference on Web and Social Media 2014, Ann Arbor, MI, USA, 1–4 June 2014.
18. Nguyen, T.H.; Shirai, K.; Velcin, J. Sentiment analysis on social media for stock movement prediction. *Expert Syst. Appl.* **2015**, *42*, 9603–9611. [\[CrossRef\]](#)
19. Socher, R.; Perelygin, A.; Wu, J.; Chuang, J.; Manning, C.D.; Ng, A.Y.; Potts, C. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, Seattle, WA, USA, 18–21 October 2013; pp. 1631–1642.
20. Santos, M.V.; Morgado-Dias, F.; Silva, T.C. Oil Sector and Sentiment Analysis—A Review. *Energies* **2023**, *16*, 4824. [\[CrossRef\]](#)
21. Xie, Q.; Han, W.; Lai, Y.; Peng, M.; Huang, J. The wall street neophyte: A zero-shot analysis of chatgpt over multimodal stock movement prediction challenges. *arXiv preprint* **2023**, arXiv:2304.05351.
22. Ko, H.; Lee, J. Can ChatGPT improve investment decisions? From a portfolio management perspective. *Financ. Res. Lett.* **2024**, *64*, 105433. [\[CrossRef\]](#)
23. Lopez-Lira, A.; Tang, Y. Can chatgpt forecast stock price movements? return predictability and large language models. *arXiv preprint* **2023**, arXiv:2304.07619. [\[CrossRef\]](#)
24. Li, M.; Chen, L.; Zhao, J.; Li, Q. Sentiment analysis of Chinese stock reviews based on BERT model. *Appl. Intell.* **2021**, *51*, 5016–5024. [\[CrossRef\]](#)
25. Li, M.; Li, W.; Wang, F.; Jia, X.; Rui, G. Applying BERT to analyze investor sentiment in stock market. *Neural Comput. Appl.* **2021**, *33*, 4663–4676. [\[CrossRef\]](#)
26. Kheiri, K.; Karimi, H. Sentimentgpt: Exploiting GPT for advanced sentiment analysis and its departure from current machine learning. *arXiv preprint* **2023**, arXiv:2307.10234.
27. Breitung, C.; Kruthof, G.; Müller, S. *Contextualized Sentiment Analysis using Large Language Models*; SSRN: Rochester, NY, USA, 2023. [\[CrossRef\]](#)
28. Lund, B.D.; Wang, T. Chatting about ChatGPT: How may AI and GPT impact academia and libraries? *Libr. Hi Tech News* **2023**, *40*, 26–29. [\[CrossRef\]](#)
29. Kamnis, S. Generative pre-trained transformers (GPT) for surface engineering. *Surf. Coat. Technol.* **2023**, *466*, 129680. [\[CrossRef\]](#)
30. Veyseh, A.P. Unleash GPT-2 power for event detection. In Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing, Bangkok, Thailand, 1–6 August 2021; Volume 1, pp. 6271–6282.
31. Goyal, T.L. News summarization and evaluation in the era of gpt-3. *arXiv preprint* **2022**, arXiv:2209.12356.
32. Liu, J.S. What Makes Good In-Context Examples for GPT-3? *arXiv* **2021**, arXiv:2101.06804.
33. OpenAI. Fine-Tuning. Retrieved from OpenAI platform. 2023. Available online: <https://platform.openai.com/docs/guides/fine-tuning> (accessed on 1 November 2023).
34. CincoDías-ElPaís. CincoDías Energía. Retrieved from CincoDías-ElPaís. 2018–2023. Available online: <https://cincodias.elpais.com/noticias/energia/> (accessed on 1 December 2023).
35. EnergyNews. EnergyNews Mercado Eléctrico. Retrieved from EnergyNews. 2018–2023—Todo Energía. Available online: <https://www.energynews.es/mercadoelectrico/> (accessed on 1 December 2023).
36. Grupo ASE. Informe Mercado. 2018–2023. Retrieved from Grupo ASE. Available online: <https://informesdemercado.grupoase.net/en/inicio-2/> (accessed on 1 December 2023).

37. Exclusivas Energéticas. Informes Mindee. 2018–2023. Retrieved from Exclusivas Energéticas. Available online: <https://exclusivas-energeticas.com/> (accessed on 1 December 2023).
38. Engle, R.F. Measuring and testing the impact of news on volatility. *J. Financ.* **1993**, *48*, 1749–1778. [[CrossRef](#)]
39. Lewis, P.; Perez, E.; Piktus, A.; Petroni, F.; Karpukhin, V.; Goyal, N.; Küttler, H.; Lewis, M.; Yih, W.T.; Rocktäschel, T.; et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 9459–9474.
40. Gao, Y.; Xiong, Y.; Gao, X.; Jia, K.; Pan, J.; Bi, Y.; Dai, Y.; Sun, J.; Wang, H. Retrieval-augmented generation for large language models: A survey. *arXiv preprint* **2023**, arXiv:2312.10997.

**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.