

Energy Price Forecasting in the Spanish Market using Generative Pre-trained Transformers (GPT)

Masters in Data Engineering and Artificial Intelligence
2022-2023

Masters Final Project

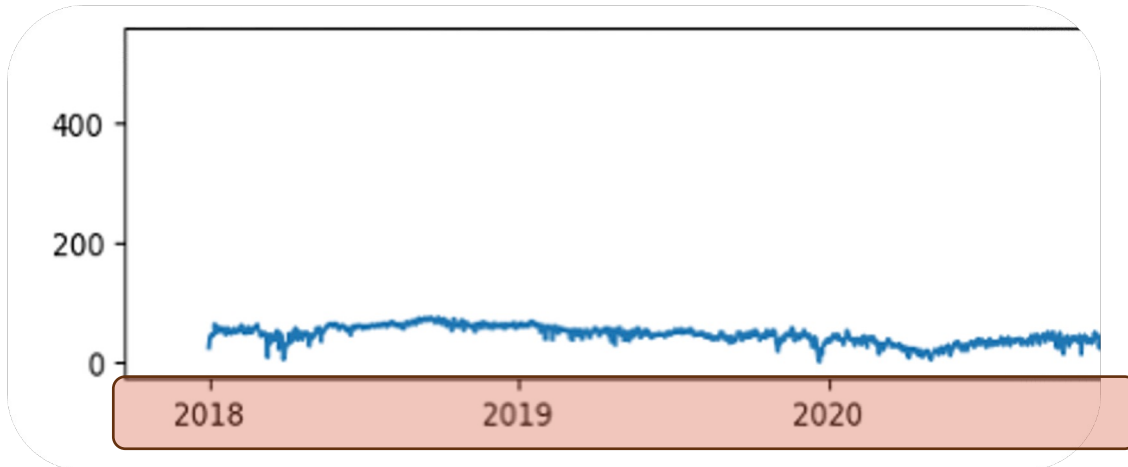


Author: Alberto Menéndez Medina

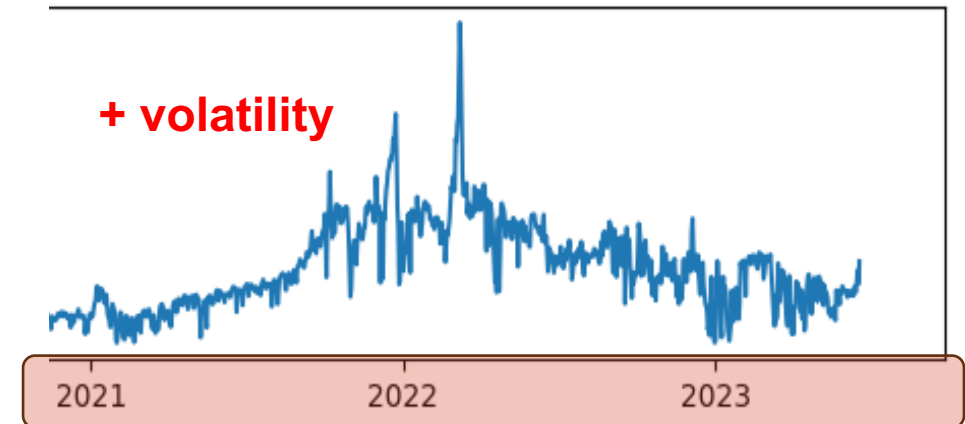
Tutor/Supervisor: José Antonio Heredia Álvaro

Background and Context

Analysing the Evolution of Energy Prices from 2018 to 2023



- Rising costs across fossil fuels and technologies
- Supply chain pressures and cost fluctuation
- Labour market constraints



- Disparity between surging demand and limited offer
- Less investment in oil and gas supply projects
- Increase of fossil fuels price due to Ukraine's war

How to overcome impact on existing prediction models

Updating and enhancing existing energy price forecasting models to improve their accuracy in the post-COVID:

- **Incorporate Real-time Data** to enhance model responsiveness to market changes
- **Scenario analysis** and **flexible model architectures** that adapt to changing market conditions
- **Advanced Machine Learning** and **AI techniques**

GPT-BASED MODELS

1. **Data enrichment:** extract insights from news and reports
2. **Sentiment analysis:** help forecast investor movements
3. **Event detection:** to be used as model inputs
4. **Market news summarisation:** insight generation
5. **Identify influential factors:** feature selection and prioritisation
6. **Customised prompts:** extract specific information
7. **Historical context:** summarising past news and impacts
8. ...

Datasets and sources

Target variable: **Energy price in the Spanish Market** (OMIE)

Independent variables:

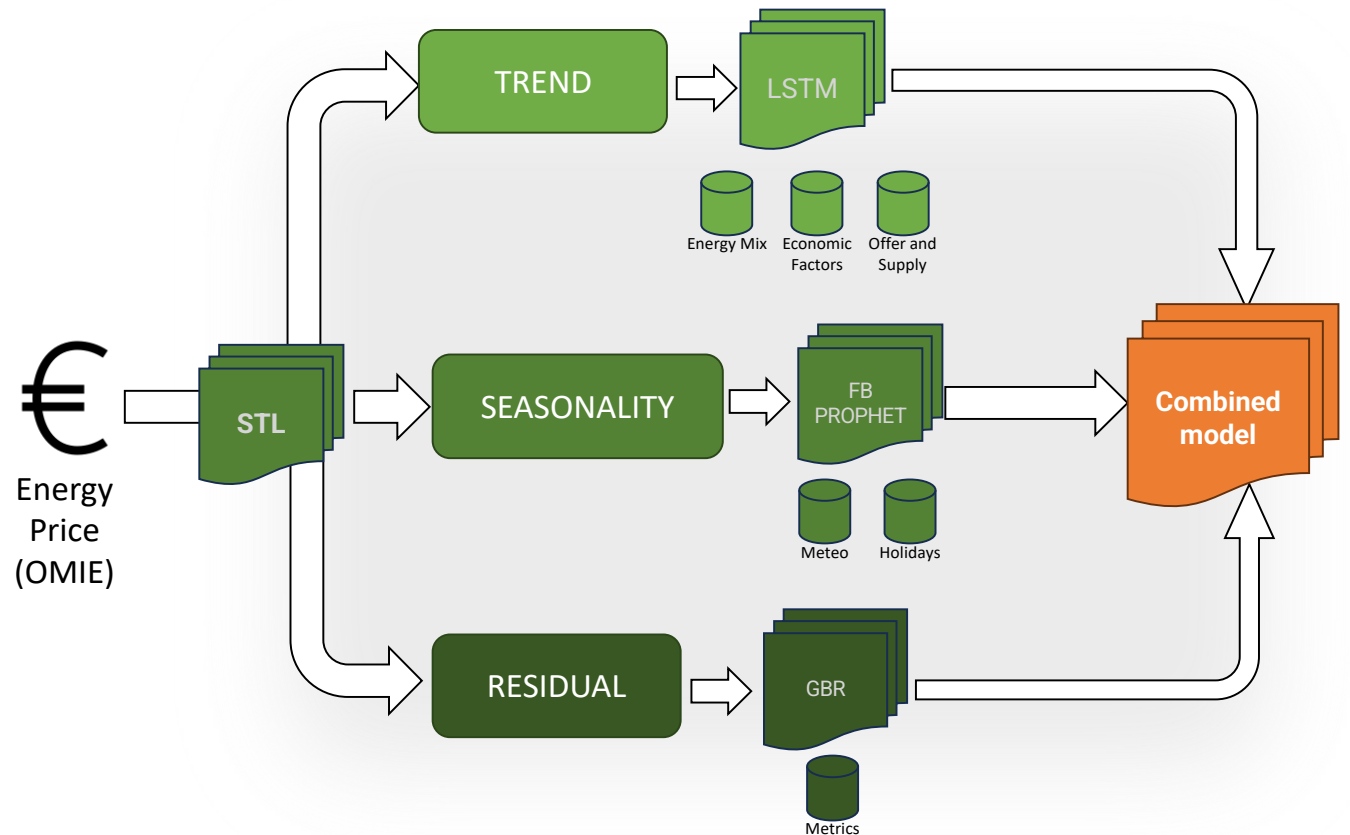
1. **Fossil fuels/Renewable Energy mix:** Production, usage, costs.
2. **Economic Factors:** Overall economic conditions, such as GDP growth, inflation rates, and interest rates
3. **Energy Demand and Supply:** Changes in energy demand and supply, including factors like population growth, industrial activities
4. **Seasonal Factors:** Weather forecast, holidays and special events.
5. **Specialised news and reports**

Online sources: OMIE, Red Eléctrica, Exclusivas Energéticas, INE, GrupoAse, CincoDias, EnergyNews

Machine Learning Algorithms and Techniques

Time series forecasting offers diverse approaches to capture and predict complex temporal patterns:

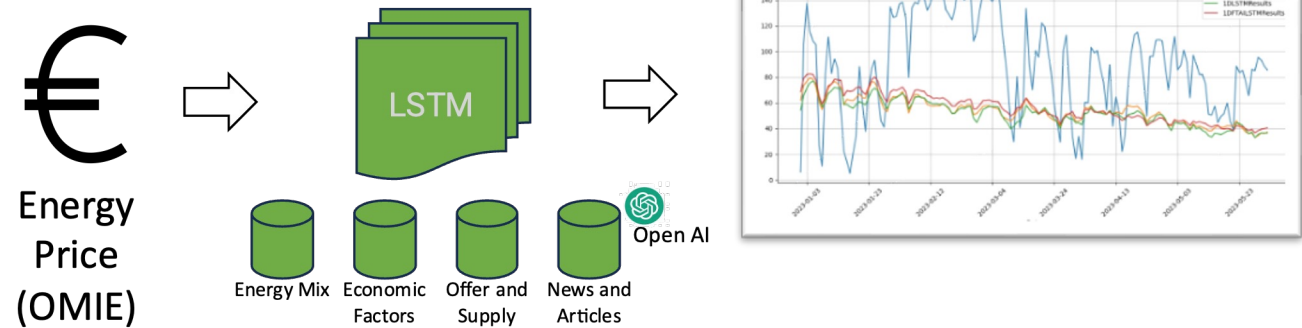
1. Univariate modelling: **Neural Prophet**
2. Multivariate Modelling: **LSTM**
3. Time Series Decomposition (see image)



Integration of OpenAI Technologies

We created new features using OpenAI tools to analyse specialised news and reports that were incorporated into the multivariate model:

1. **Impact** on Electricity Price
2. **Direction** of Impact
3. Impact **Period**

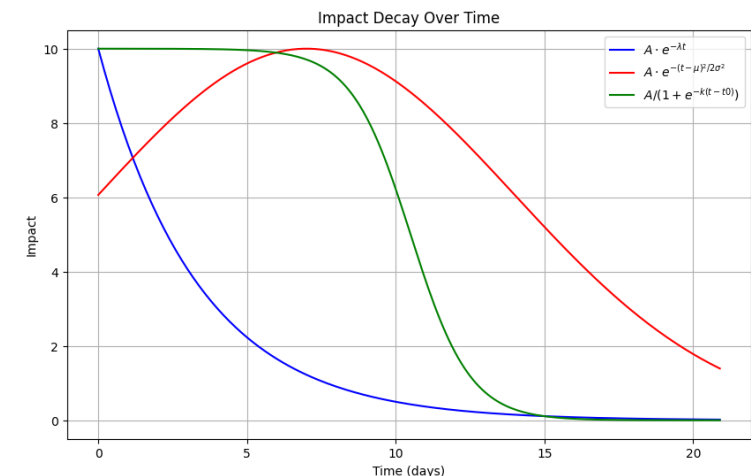


With 2 different approaches:

- **In-context examples prompt**
- **Fine-tuning**

Simulating the impact of news and articles over time following a decay function

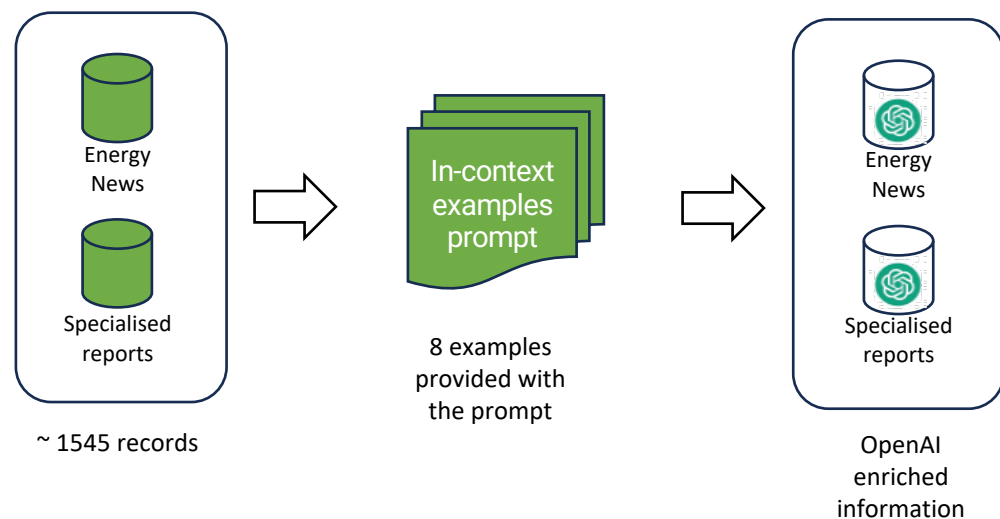
- **Exponential decay** for short-term records
- **Gauss model** for mid/long-term records



Calculation of daily, weekly or monthly average impact and addition to the multivariate model

In-context examples prompt

8 examples of the market news and articles dataset and provided the values of Impact, Direction and Period from an energy expert point of view



Enriched news file

	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

```
prompt = '''
    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)>

    Example 1:
    Headline: "Editorial - Moderar el coste de la factura de la luz y mantener la
    inflación bajo control .El Gobierno congelará de nuevo en 2018 la parte regulada del recibo"
    Expected output: <Impact: 8> <Direction: Down> <Period: Long>

    Example 2:
    Headline: "Petróleo - El transporte teme que la OPEP facilite hoy una subida de los
    carburantes. El precio de gasolina o gasóleo no ha recogido el alza del crudo desde los
    mínimos de 2016. El coste en el surtidor podría situarse en 1,6 y 1,4 euros por litro"
    Expected output: <Impact: 9> <Direction: Up> <Period: Short>

    Example 3:
    Headline: "Energía - El PNV reclama al Gobierno la rebaja de tarifa prometida a la
    industria vasca. En 2014, Industria cambió el umbral de tensión de las redes para rebajar la
    factura vasca\Los nacionalistas le piden que cumpla su promesa, que no figuran en la orden
    de peajes"
    Expected output: <Impact: 2> <Direction: Down> <Period: Mid>

    Example 4:
    Headline: "El Foco - ¿Podemos desterrar la pobreza energética? - La futura Ley de
    Cambio Climático debería incluir medidas para atajar el problema"
    Expected output: <Impact: 0> <Direction: None> <Period: None>

    Example 5:
    Headline: "Energía - La CNMC obligará a las eléctricas a cambiar sus marcas para no
    confundir al usuario - EDP pide un único nombre para todas las comercializadoras reguladas"
    Expected output: <Impact: 0> <Direction: None> <Period: (None)>

    Example 6:
    Headline: "Energía - Nadal culpa de la fuerte subida de la luz a la parada de tres
    nucleares = Asegura que a pesar de las fuertes lluvias no ha habido generación hidroeléctrica
    suficiente\"
    Expected output: <Impact: 8> <Direction: Up> <Period: Past>

    Example 7:
    Headline: "ENERGÍAS RENOVABLES - El autoconsumo eléctrico despegó en 2020: claves de
    la nueva energía - La potencia fotovoltaica doméstica suma unos 113 MW en 2020"
    Expected output: <Impact: 5> <Direction: Down> <Period: Long>

    Example 8:
    Headline: "Coronavirus - La demanda eléctrica cae un 10% frente al alza del 2% de
    principios de marzo - El consumo cae esta mañana a algo más de 30.000 MWh, 3.200 MWh menos
    que lunes pasado."
    Expected output: <Impact: 7> <Direction: Down> <Period: Short>
    '''
```

Fine-tuned implementation

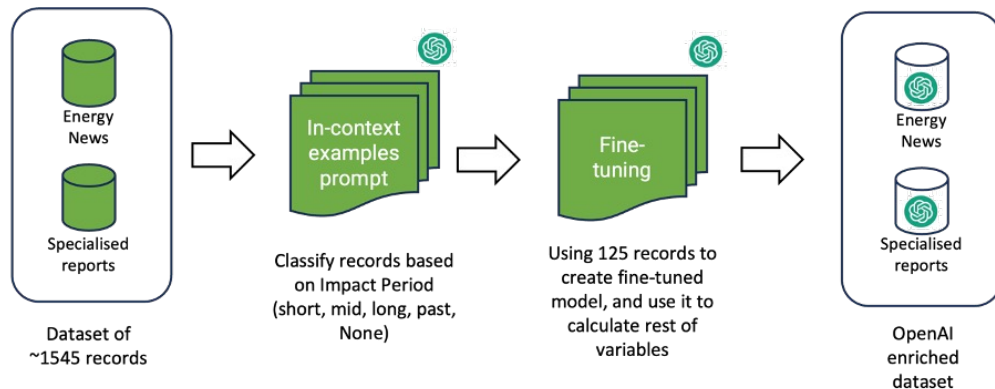
```
completion = openai.ChatCompletion.create(  
    model="ft:gpt-3.5-turbo-0613:personal::7s6AzNA5",  
    temperature=0.6,  
    messages=[  
        {"role": "system", "content": "Given an electricity market and industry headline  
from the specialised analysts, provide the following fields in a JSON dict, where applicable:  
\"Impact\" (on the price of electricity)\", \"Direction\" (of the impact), \"Volatility\"  
(that the news will generate), and \"Period\" (when the impact will occur)."},  
        {"role": "user", "content": "Headline: " + headline + "\""}  
    ]  
)
```

Calculated for each date of the time-series range (2018-2023) what difference (%) the price has suffered with the previous time interval to the news event (weeks for short-term, months for mid/long term) as per the following formulas:

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

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

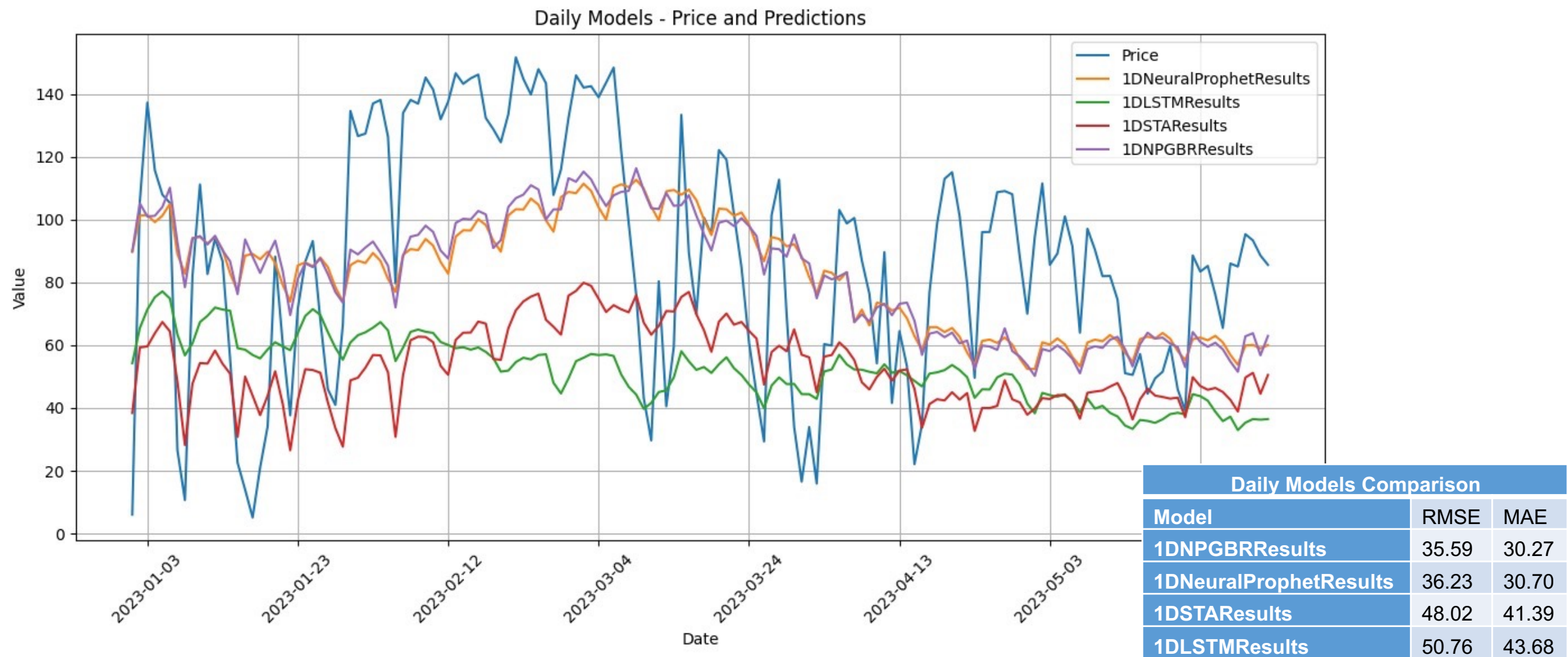
Completed a first run with OpenAI using the “In-context examples prompt” to identify its assessment of “Impact Period” for each record (short-term, mid/long-term, past, None), keeping aside 25 randomly selected records for each possible value for the fine-tuning



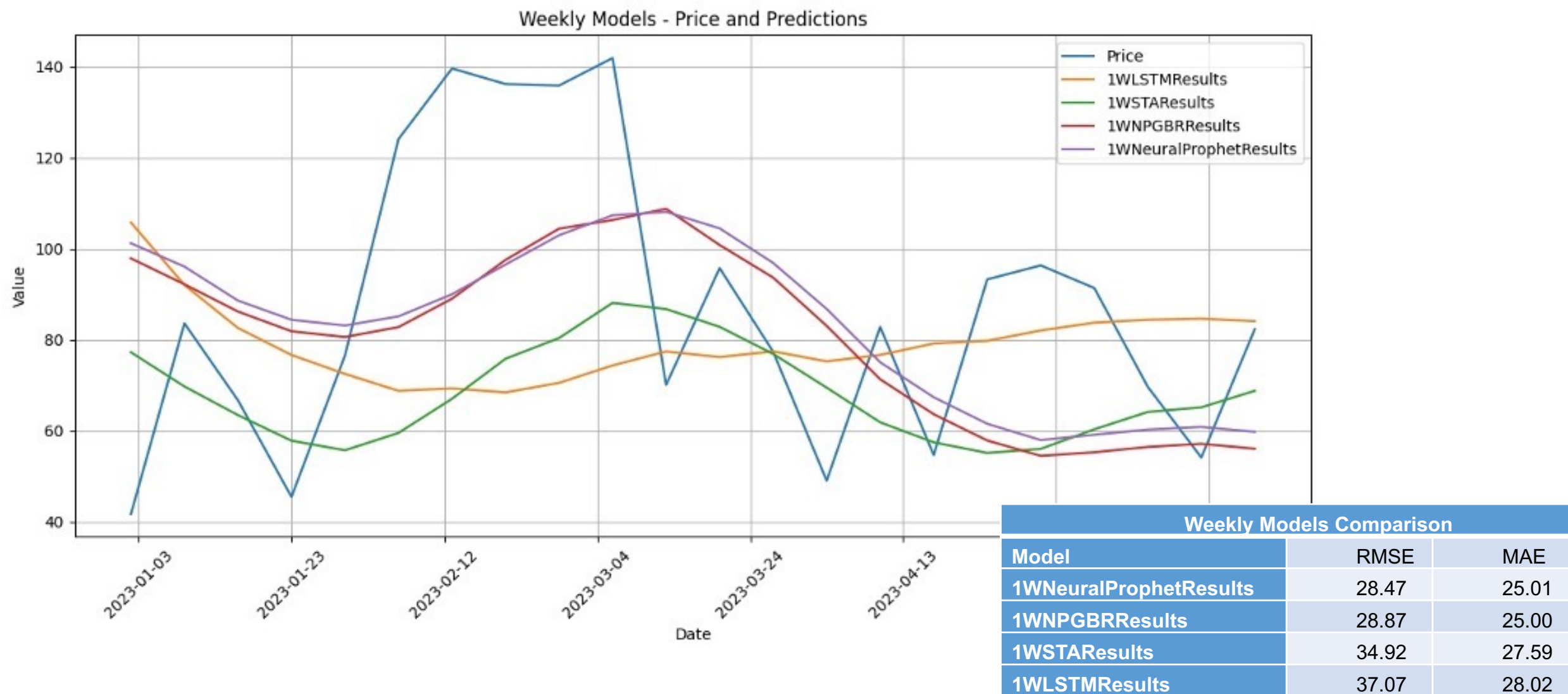
Fine-tuning steps (OpenAI):

- Dataset preparation: conversation structured (role, content)
- Validate data formatting
- Divide training and testing datasets
- Upload file and create the fine-tuning job
- Use the new fine-tuned model with the rest of the news and articles to enrich the dataset with calculated variables

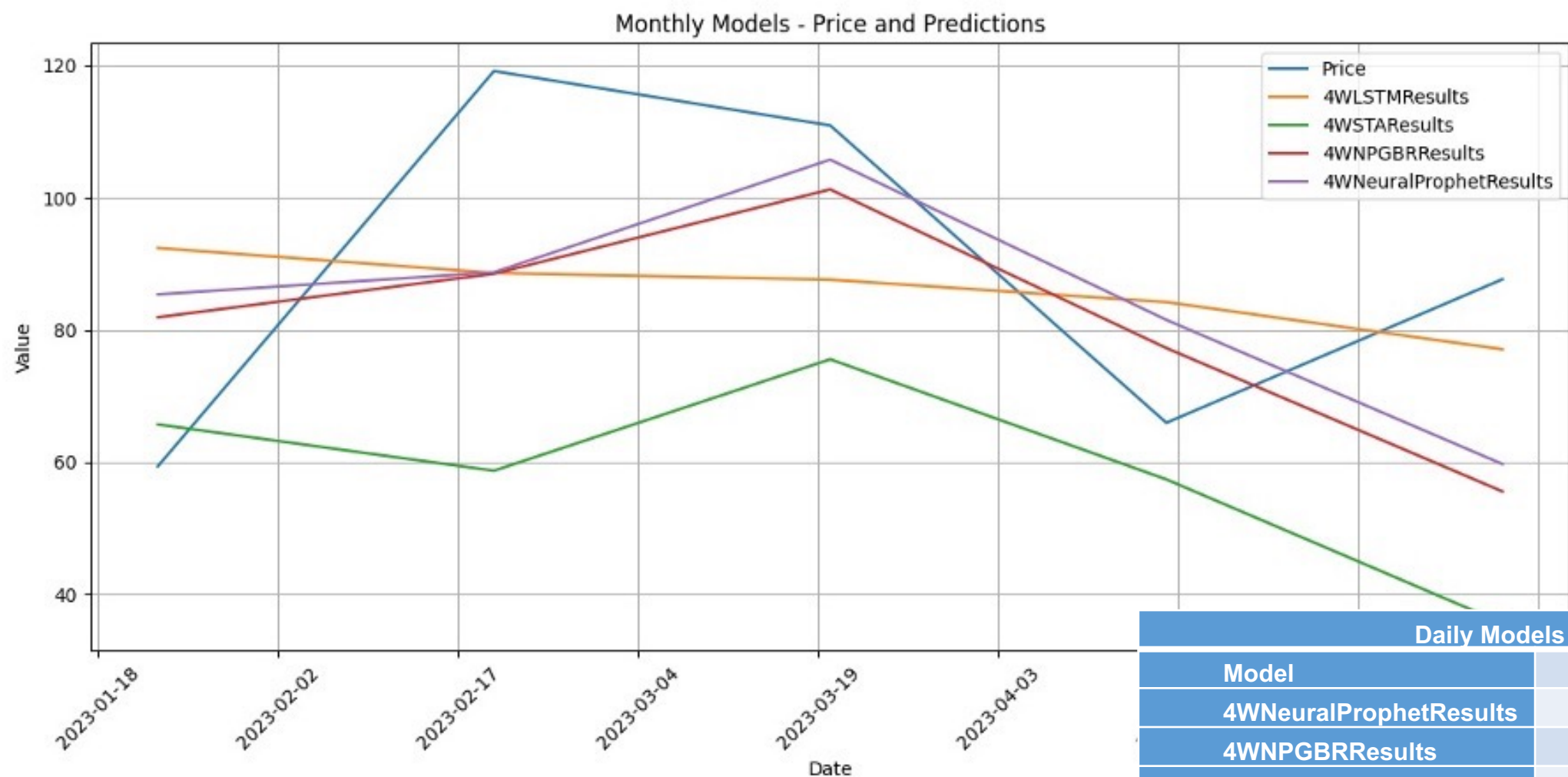
Raw models results – Daily Models



Raw models results – Weekly Models



Raw models results – Monthly Models

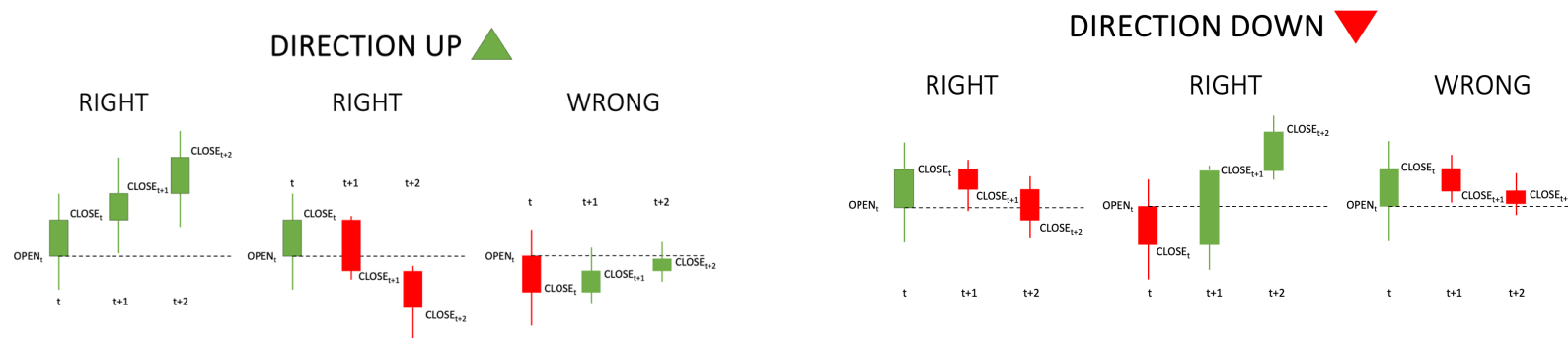


Daily Models Comparison		
Model	RMSE	MAE
4WNeuralProphetResults	20.82	18.60
4WNPGBRResults	21.44	18.82
4WLSTMResults	24.54	23.14
4WSTARResults	37.66	31.80

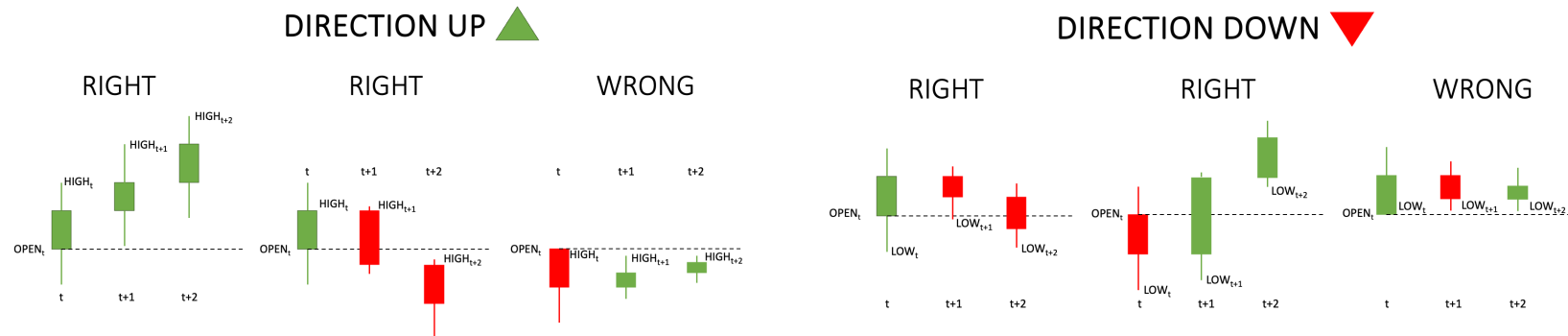
OpenAI models: Metrics

We have defined three different types of metrics to measure the accuracy of the Direction of the Impact prediction:

1. Close Price



2. High/Low



3. Threshold

OpenAI models: Short-term

	In-context examples	Fine-tuned
Close Price	0.64	0.71
High/Low	0.74	0.81
Threshold 2%	0.64	0.64

In-context



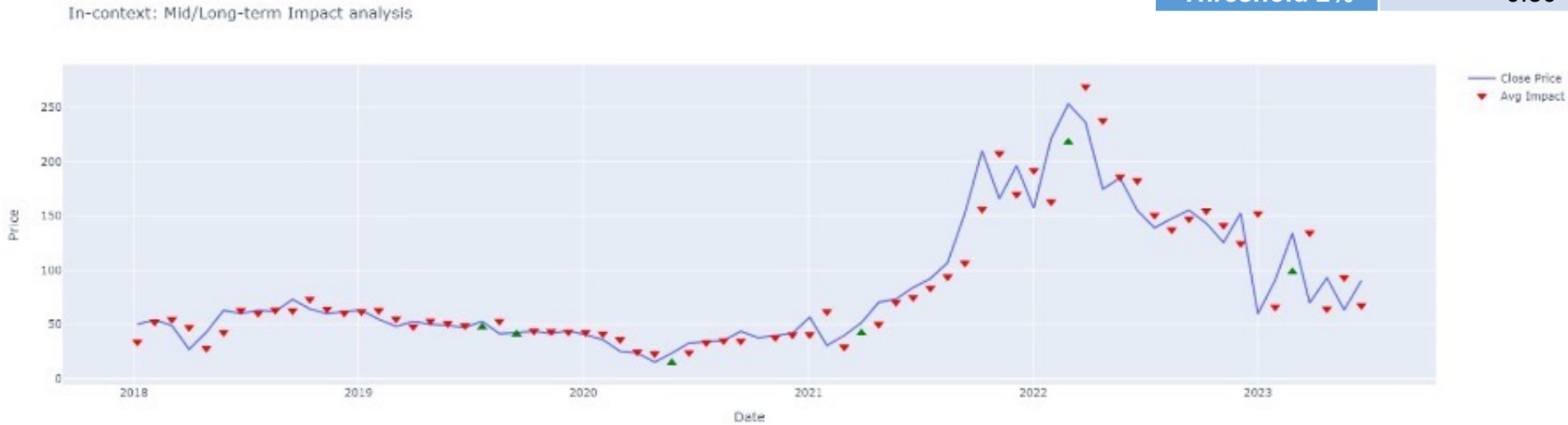
Fine-tuned



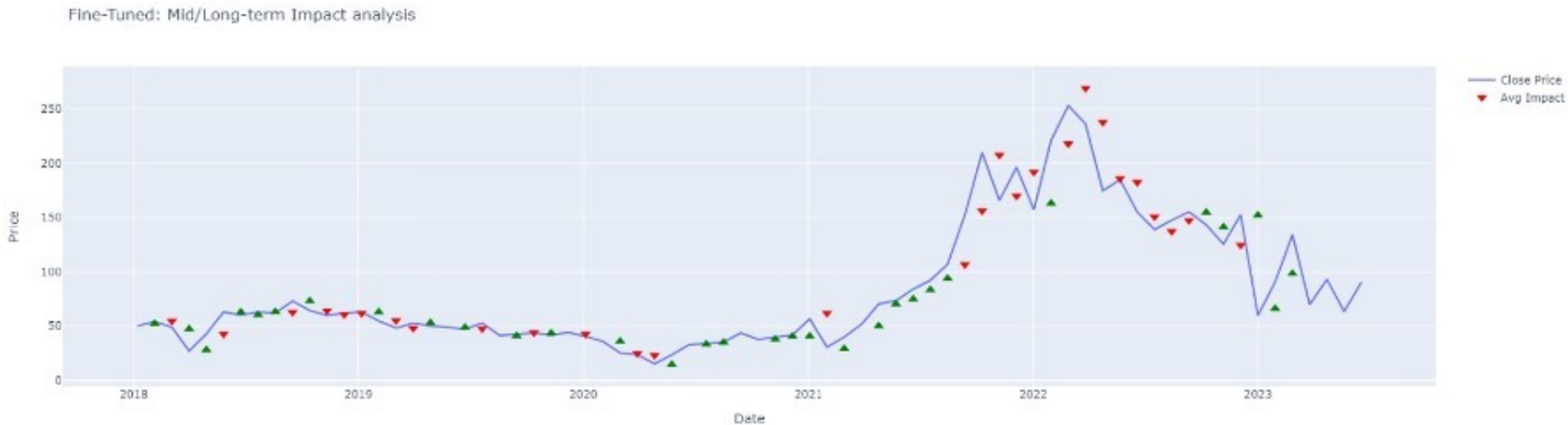
OpenAI models: Mid/long-term

	In-context examples	Fine-tuned
Close Price	0.68	0.81
High/Low	0.89	0.93
Threshold 2%	0.80	0.86

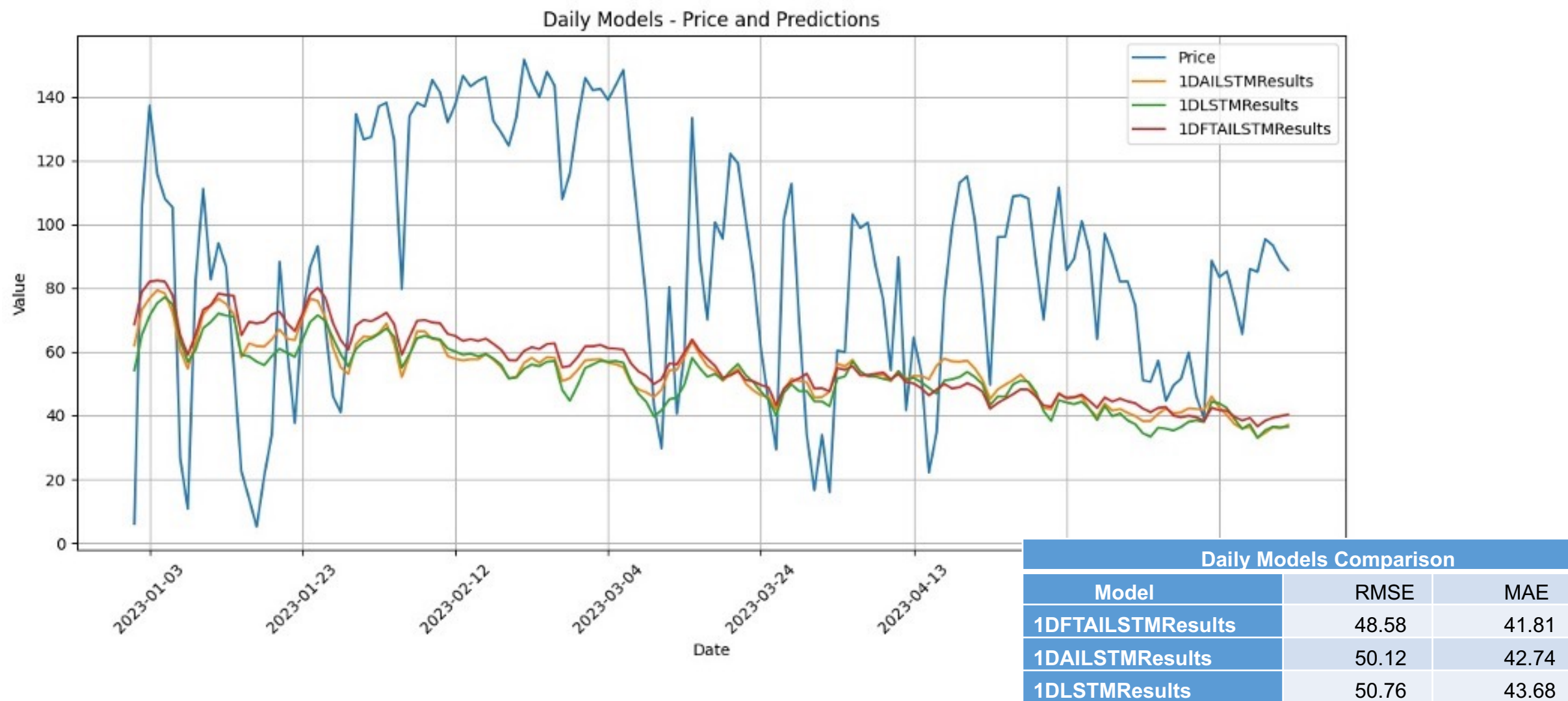
In-context



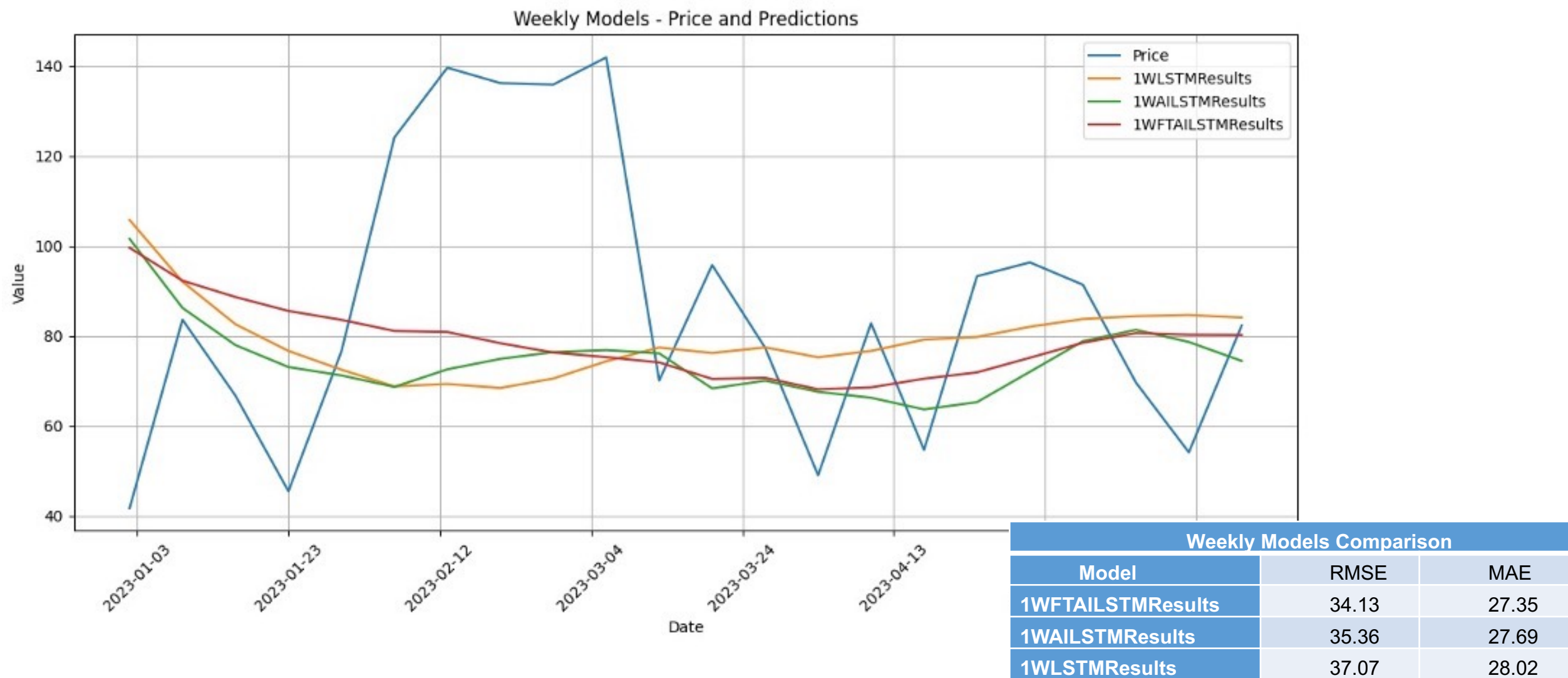
Fine-tuned



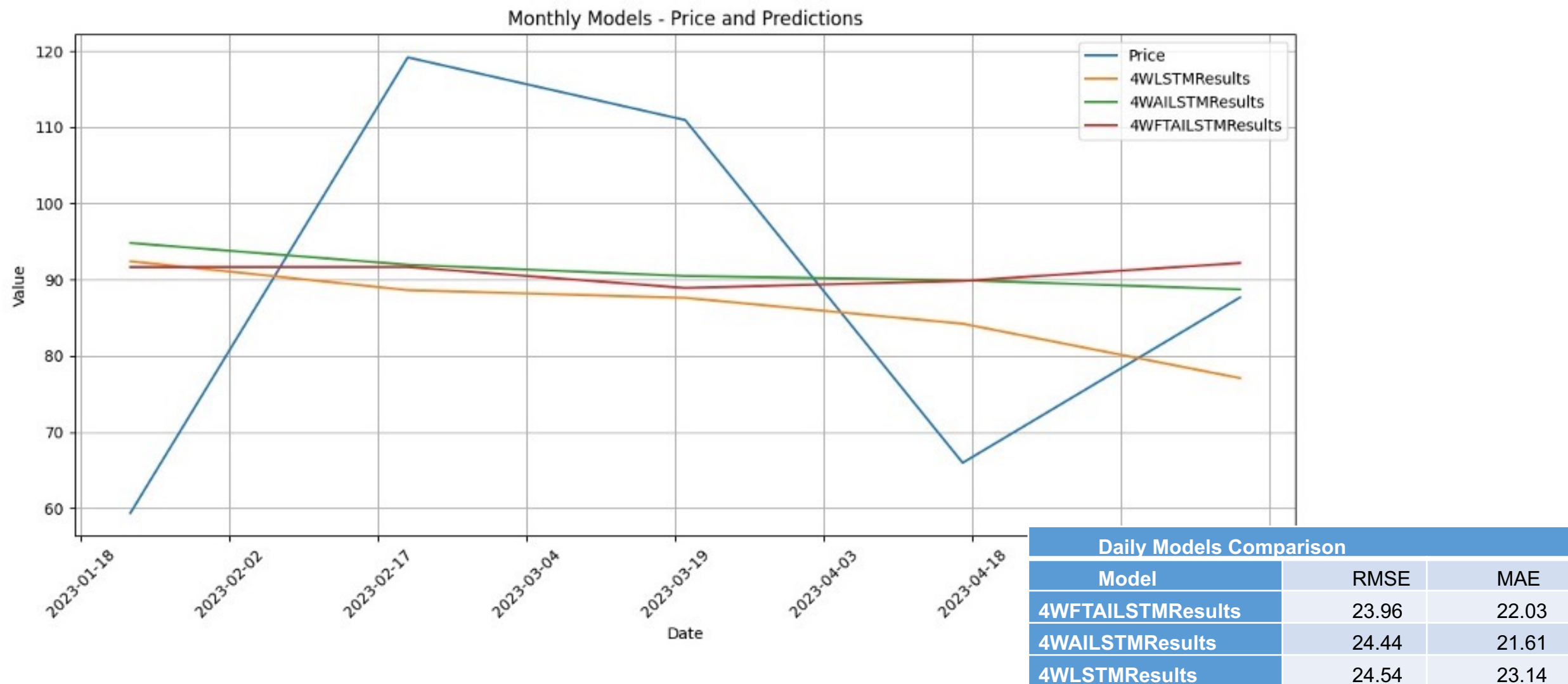
OpenAI models results – Daily Models



OpenAI models results – Weekly Models



OpenAI models results – Monthly Models



Future research directions

Improvement of model features:

- Data from the futures market electric energy (OMIP)
- Geopolitical factors, such as conflicts fuel production reviews
- Localised weather prediction, including the likelihood of events related to climate change

Continued exploration of OpenAI technologies can lead to improved contextual understanding of external factors influencing energy prices:

- Quality of news sources and increased periodicity
- Accurate impact calculation and decay over time methods to translate real impact
- Additional number of energy experts' reports
- Automatic generation of reports analysing recent
- Incorporating real-time data sources

Conclusion

Our research project explored the application of GPT models, particularly OpenAI's ChatGPT, to generate features adding a new approach to the energy price prediction within the Spanish market.

Our findings indicate the potential of GPT models to provide valuable insights and improve predictions, particularly in understanding short to mid-term price trends. However, we acknowledge that further refinement is necessary when integrating GPT into multivariate time series forecasting models. The differences observed needed to be sufficiently significant to outperform the raw model.

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

Q&A