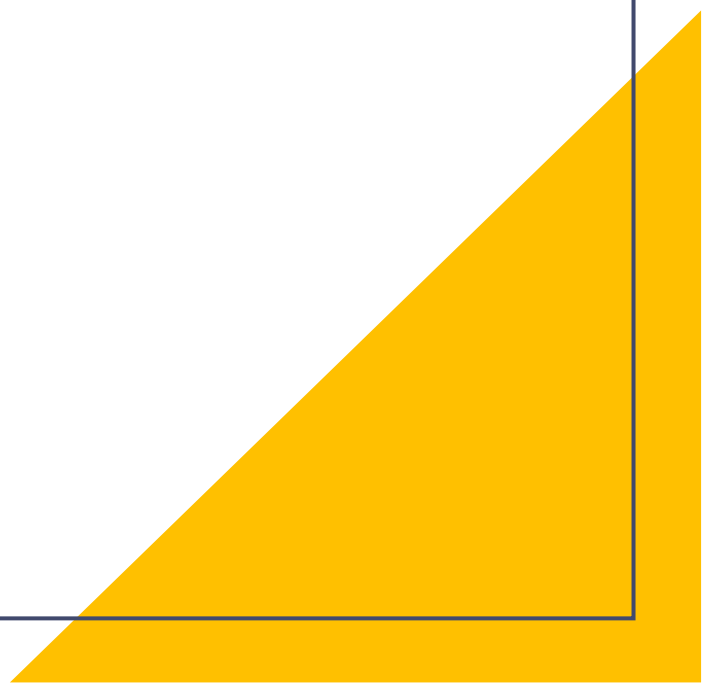


# LLM

Jovan Torio



# Task

Using Llama/Gemini API to help  
“recreate a more communicative  
ground.news”

# Ground.news

201 Articles

All

Left 41

Center 46

Right 51



Blavity News & Politics

Mixed Factuality

Left



## Virginia Allowed To Proceed With Purge Of 1,600 Voter Registrations, Per Supreme Court Decision

The Supreme Court's conservative justices handed Republicans a temporary victory Wednesday, allowing Virginia to purge about 1,600 voter registrations just days before the presidential election. According to The New York Times, the court's ruling aims to prevent people who are not U.S. citizens from voting, a move that state Republican officials and former President Donald Trump support after claiming that it is a concern. The case was place on ...

20 hours ago

[Read Full Article](#)

# Bias Comparison Example

The left emphasizes the seriousness of North Korean troops potentially entering combat, framing it as a direct concern for Europe and the Indo-Pacific, while the center focuses more on the numbers and military training with less urgency.

The right presents a broader view of international and NATO responses to North Korea's involvement, highlighting strategic ramifications, whereas the other sides provide different details on troop movements.

The left highlights South Korean fears regarding weapons technology from Russia, while the center does not emphasize these geopolitical implications to the same extent.

## LEFT

North Korea has deployed troops to Russia, marking a significant development in their increasingly friendly ties, according to Secretary of Defense Lloyd Austin.

The U.S. Confirmed that 3,000 North Korean troops are training in Russia and warned they would be considered 'fair game' in combat.

South Korean officials express concern that Russia might provide North Korea with advanced weapon technologies to enhance its nuclear and missile capabilities.

<https://www.newstatesman.com/world/europe/ukraine/2024/10/north-korean-troops-in-russia-is-a-dangerous-escalation>

## CENTER

U.S. Defense Secretary Lloyd Austin confirmed that North Korea has sent troops to Russia on Monday.

South Korea's spy chief stated that 3,000 North Korean troops are in Russia receiving training for deployment.

The NIS chief believes North Korea aims to send a total of 10,000 troops to Russia by December.

<https://www.hawaiitribune-herald.com/2024/10/24/nation-world-news/us-says-evidence-shows-north-korea-has-troops-in-russia-possibly-for-ukraine-war/>

## RIGHT

U.S. Defense Secretary Lloyd Austin announced on Oct. 23 that there is evidence that North Korea has deployed troops to Russia.

Allies have confirmed evidence of a DPRK troop deployment to Russia.

<https://www.latestly.com/agency-news/world-news-us-says-north-korea-sent-at-least-3000-soldiers-into-eastern-russia-in-october-6365438.html>

# LEFT-CENTER- RIGHT SUMMARY



# Bias Rating

## Rating Description

### Far Left

These publications tend to reflect the policy positions of the most extreme left-leaning party members. They are strongly biased toward liberal causes through story selection and/or political affiliation. They use strong wording that attempts to influence an audience by appealing to emotion or stereotypes. They may publish misleading reports or leave out information that could damage liberal causes.

### Left

These publications tend to reflect the current positions held by leaders of left-leaning parties. They are moderately biased toward liberal causes through story selection and/or political affiliation. They may use loaded words, publish misleading reports or leave out information that doesn't support liberal causes.

### Lean Left

These publications have a slight to moderate liberal bias. They often publish factual information, but still may use loaded words that favor liberal causes.

### Center

These publications have no discernable political position. They use very few loaded words and the reporting is well sourced. On a given issue, they present a relatively complete survey of key competing positions. This rating does not necessarily represent "balance" or "neutrality." A Center rating does not imply that the position is best or most valid.

### Lean Right

These publications are slightly to moderately conservative in bias. They often publish factual information, but still may use loaded words that favor conservative causes.

### Right

These publications tend to reflect the current positions held by leaders of right-leaning parties. They are moderately biased toward conservative causes through story selection and/or political affiliation. They may publish misleading reports or leave out information that doesn't support conservative causes.

### Far Right

These publications tend to reflect the policy positions of the most extreme right-leaning party members. They are strongly biased toward conservative causes through story selection and/or political affiliation. They use strong wording that attempts to influence an audience by appealing to emotion or stereotypes. They may use loaded words, publish misleading reports or leave out information that could damage conservative causes.




# Bias Characteristics

## Caution!

*If you notice the following, the source may be biased:*

- \* Heavily opinionated or one-sided
- \* Relies on unsupported or unsubstantiated claims
- \* Presents highly selected facts that lean to a certain outcome
- \* Pretends to present facts, but offers only opinion
- \* Uses extreme or inappropriate language
- \* Tries to persuade you to think a certain way with no regard for factual evidence
- \* The author is unidentifiable, lacks expertise, or writes on unrelated topics
- \* Is entertainment-based or a form of parody or satire
- \* Tries to sell you something in disguise

# Experiment

- LlamaIndex  LlamaIndex
    - LLM Framework
    - Flexible layers of customizability
    - Allows to build with RAG techniques
    - Integrations across data loading, indexing, and retrieval.
    - Active community
  - LlamaIndex + HuggingFace
  - LlamaIndex + Local Llama using LM Studio
-



# Prompt

You are a news article analyzer capable of determining political bias by analyzing what is being highlighted. Please provide an explanation of why you chose a particular side, citing specific examples from the text that demonstrate its alignment. Do not speak in first person.

Assess the article using the following bias ratings:

Far Left: Strongly biased toward liberal causes. Uses extreme, emotional language, misleading reports, or omits unfavorable information.

Left: Moderately biased toward liberal causes. May use loaded words or omit certain details.

Lean Left: Slight to moderate liberal bias. Generally factual but may favor liberal causes.

Center: Neutral. Few loaded words, well-sourced reporting, and balanced presentation of views.

Lean Right: Slight to moderate conservative bias. Generally factual but may favor conservative causes.

Right: Moderately biased toward conservative causes. May use loaded words, omit unfavorable information, or mislead.

Far Right: Strongly biased toward conservative causes. Uses extreme, emotional language, misleading reports, or omits unfavorable information.

Expected Answer:

Bias Rating: Rate the article and explain your choice with brief examples from the text.

Key Indicators of Bias: Opinionated content, selective facts, extreme language, persuasive tactics, unsupported claims, anonymity, entertainment-based content, or disguised sales pitches.

Conclusion: Summarize findings, mentioning if there were any nuances or mixed signals.

# Dataset

indonesian\_news\_dataset

▲ 10

New Notebook

Download

Data Card

Code (5)

Discussion (0)

Suggestions (0)

About Dataset

Indonesian News Datasets

The dataset compiles information from seven prominent Indonesian news platforms: **Tempo**, **CNN Indonesia**, **CNBC Indonesia**, **Okezone**, **Suara**, **Kumparan**, and **JawaPos**. Each source contributes a diverse range of articles, collectively forming a comprehensive repository of Indonesian news content. The dataset consists of diverse columns: 'id' denotes the unique identifier for each news article, 'title' contains the headline, 'image' refers to the associated visual content, 'url' specifies the source link, 'content' encompasses the full text of the news article, 'embedding' houses the text embeddings extracted using the OpenAI `text-embedding-ada-002` model, 'created\_at' marks the creation timestamp, 'updated\_at' signifies the last update timestamp, and 'summary' encapsulates the concise article summary generated via the ChatGPT API. These columns collectively represent a comprehensive set of information for each news article, facilitating various natural language processing tasks and analysis while providing enriched contextual details for effective exploration and modeling.

Usability

10.00

License

CC BY-NC-SA 4.0

Expected update frequency

Never

Tags

Computer Science

Text

Text Classification

Summarization

Data Explorer

Version 1 (737.92 MB)

data.csv

Summary

1 file

11 columns

data.csv (737.92 MB)

Download

Fullscreen

Close

Detail

Compact

Column

10 of 11 columns

About this file

Add Suggestion

The dataset compiles information from seven prominent Indonesian news platforms: Tempo, CNN Indonesia, CNBC Indonesia, Okezone, Suara, Kumparan, and JawaPos. Each source contributes a diverse range of articles, collectively forming a comprehensive repository of Indonesian news content.

id	source	title	image	url	content
denotes the unique identifier for each news article	news site source	contains the headline	refers to the associated visual content	specifies the source link	encompasses the full text of the news article
<div><div></div><div>33</div><div>63.6k</div></div>	<div>kumparan28%</div> <div>okezone22%</div> <div>Other (17090)52%</div>	<div>32651</div> <div>unique values</div>	<div>26450</div> <div>unique values</div>	<div>32735</div> <div>unique values</div>	<div>[null]</div> <div>Jakarta, 1</div> <div>Other (32</div>
83	tempo	Depo Plumpang Terbakar, Anggota DPR Minta Pertamina Pastikan Pasokan BBM Tak Terganggu	<a href="https://statik.tempo.co/data/2023/03/04/id_1186868/1186868_728.jpg">https://statik.tempo.co/data/2023/03/04/id_1186868/1186868_728.jpg</a>	<a href="https://nasional.tempo.co/read/1698528/depo-plumpang-terbakar-anggota-dpr-minta-pertamina-pastikan-pasokan-bbm-tak-terganggu">https://nasional.tempo.co/read/1698528/depo-plumpang-terbakar-anggota-dpr-minta-pertamina-pastikan-pasokan-bbm-tak-terganggu</a>	TEMPO.CO - Anggota DPR RI meminta Pertamina memastikan pasokan BBM tak terganggu setelah depo minyak di Plumpang, Bekasi, terbakar.

```

[19] llm = HuggingFaceInferenceAPI(model_name=model, token=hfToken)

llm
HuggingFaceInferenceAPI(callback_manager=<llama_index.core.callbacks.base.CallbackManager object at 0x7cac42b0c220>, system_prompt=None, messages_to_prompt=<function messages_to_prompt at 0x7cac44e48f70>,
default_completion_to_prompt at 0x7cac44d4a4d0>, output_parser=None, pydantic_program_mode=<PydanticProgramMode.DEFAULT: 'default'>, query_wrapper_prompt=None, model_name='mistralai/Mixtral-8x7B-Instruct-
timeout=None, headers=None, cookies=None, task=None, context_window=3900, num_output=256, is_chat_model=False, is_function_calling_model=False)

[21] template = (
    "The context information is below. \n"
    "-----\n"
    "{context_str}"
    "\n-----\n"
    "Given this information, please answer the question: {query_str}\n"
)

[22] qa_template = PromptTemplate(template)

[23] qa_template
PromptTemplate(metadata={'prompt_type': <PromptType.CUSTOM: 'custom'>}, template_vars=['context_str', 'query_str'], kwargs={}, output_parser=None, template_var_mappings=None, function_mappings=None, temp
-----\n{context_str}\n-----\nGiven this information, please answer the question: {query_str}\n')

```

# LlamaIndex + Huggingface

Model used: mistralai/Mixtral-8x7B-Instruct-v0.1

[25] q = """

You are a news article analyzer capable of determining political bias by analyzing what is being highlighted. Please provide an explanation of why you chose a particular side, citing specific examples from the text that demonstrate its alignment. Do not speak in first person.

Assess the article using the following bias ratings:

Far Left: Strongly biased toward liberal causes. Uses extreme, emotional language, misleading reports, or omits unfavorable information.

Left: Moderately biased toward liberal causes. May use loaded words or omit certain details.

Lean Left: Slight to moderate liberal bias. Generally factual but may favor liberal causes.

Center: Neutral. Few loaded words, well-sourced reporting, and balanced presentation of views.

Lean Right: Slight to moderate conservative bias. Generally factual but may favor conservative causes.

Right: Moderately biased toward conservative causes. May use loaded words, omit unfavorable information, or mislead.

Far Right: Strongly biased toward conservative causes. Uses extreme, emotional language, misleading reports, or omits unfavorable information.

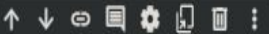
Expected Answer:

Bias Rating: Rate the article and explain your choice with brief examples from the text.


Key Indicators of Bias: Opinionated content, selective facts, extreme language, persuasive tactics, unsupported claims, anonymity, entertainment-based content, or disguised sales pitches.

Conclusion: Summarize findings, mentioning if there were any nuances or mixed signals.

"""



```
prompt = qa_template.format(context_str = text, query_str=q)
response = llm.complete(prompt)
print(response.text)
```

 Bias Rating: Lean Right

Key Indicators of Bias: The article does not use extreme language or misleading reports, but it does highlight Israel's actions and downplays Iran's response.

Conclusion: The article is generally factual and presents a balanced view of the situation, but it slightly favors Israel's perspective by emphasizing its retaliation and downplaying Iran's response. This could be due to the source, which may have a pro-Israel bias, or the focus on Israeli actions.





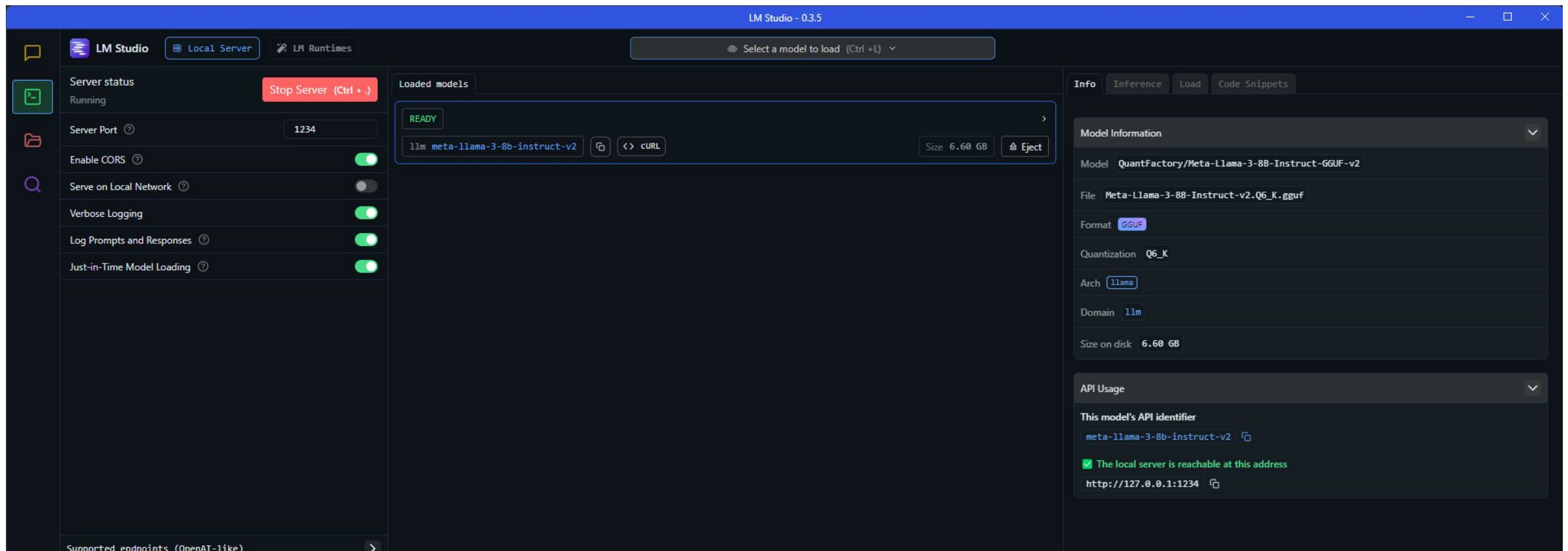
## Result

- Bias Rating: Lean Right
- Key Indicators of Bias: The article does not use extreme language or misleading reports, but it does highlight Israel's actions and downplays Iran's response.
- Conclusion: The article is generally factual and presents a balanced view of the situation, but it slightly favors Israel's perspective by emphasizing its retaliation and downplaying Iran's response. This could be due to the source, which may have a pro-Israel bias, or the focus on Israeli actions, which could be interpreted as favoring one side over the other. However, the article does not use loaded words or omit unfavorable information, so it is not strongly biased.

•

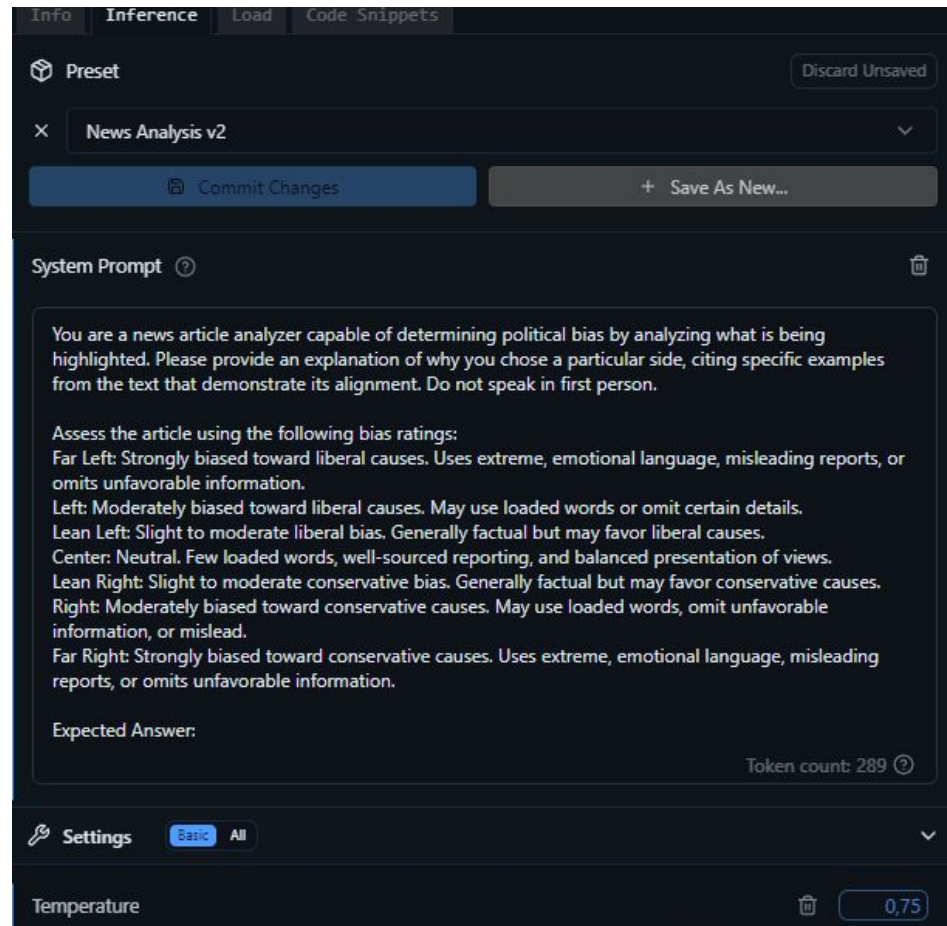
# LlamaIndex + Local Llama using LM Studio

Model: QuantFactory/Meta-Llama-3.1-8B-Instruct-GGUF





# System Prompt



```
from llama_index.llms.lmstudio import LMStudio
import pandas as pd
# from llama_index.core.llms import ChatMessage, MessageRole
# from llama_index.core import ChatPromptTemplate, PromptTemplate
```

c:\Users\USER\AppData\Local\Programs\Python\Python312\Lib\site-packages\pydantic\\_internal\\_fields.py:132: UserWarning: Field "model\_name" in LMStudio has conflict with protected namespace "model\_".

You may be able to resolve this warning by setting `model\_config['protected\_namespaces'] = ()`.

```
warnings.warn(
```

```
# df = pd.read_csv(path + "/data.csv")
df = pd.read_csv('data.csv')
df = df.dropna(subset=['content'])
```

```
llm = LMStudio(
    model_name="meta-llama-3-8b-instruct-v2",
    base_url="http://localhost:1234/v1",
    temperature=0.8,
)
```

```
template = "Here is the news: {news}"
```

```
text = df['content'].iloc[7]
print(text)
```

TEMPO.CO, Jakarta -Wacana Presiden Joko Widodo atau Jokowi 3 periode tak kunjung surut. Keputusan Pengadilan Negeri Jakarta Pusat mengetok palu yang berarti penundaan Pemilu 2024, mengindikasikan Jokowi kudu menjabat hingga 2025.

```
response = llm.complete(template.format(news=text))
print(str(response))
```

# Result

- Based on the news article provided by TEMPO.CO, I would rate its political bias as Center or Lean Left.
- The article focuses on a factual report of a fire incident at Depo Plumpang in Jakarta and the response efforts by the Indonesian Police (Polri) to evacuate victims, identify the dead, and investigate the cause of the fire. The language used is neutral, and there is no apparent emotional bias towards any particular political ideology.
- The article quotes the Chief of Public Relations Division of Polri, Irjen Dedi Prasetyo, who provides detailed information about the incident and the response efforts by Polri. The quote does not contain any loaded words or emotionally charged language that might suggest a specific political bias.
- Additionally, the article mentions the involvement of various agencies such as Bareskrim, Ditreskrim Polda Metro Jaya, Labfor, Inafis, Sat Brimobda PMJ, and BPBD in the response efforts, which indicates a neutral and factual reporting style.
- Overall, the article maintains a professional tone and provides accurate information about the incident without taking sides or promoting any specific political agenda. Therefore, I would rate its political bias as Center or Lean Left.
- As for key topics such as immigration, healthcare, economy, they are not mentioned in this article. The article focuses on a local news story related to public safety and disaster response efforts by Polri.



# To-Do List

- Experiment with RAG
  - Improve the prompt (less wordy, more accurate, less AI-like, sound more professional)
-

THANK YOU

