Automated Fact-Checking of Climate Change Claims with Large Language Models

Yunlong Pan

2024-02-20

Outline

- Introduction (Fact-checking)
- Method (Pipeline and prompt)
- Experiments (Example)

Introduction

Task: Fact-checking (From 2011 Natural Language Processing)

Climate Feedback (170)
https://climatefeedback.org/claim-reviews/

Skeptical Science (163)
https://skepticalscience.com/shorturls.php

Heartland Institute (81)
https://climatechangereconsidered.org/

Climate Feedback

Climate Feedback categorizes the verdicts into twelve different categories, ranging from `incorrect` to `correct`, with different shades of correctness and accuracy in between. A typical example from Climate Feedback is, e.g., the following claim:

> More than 75 percent of the rainforest is losing resilience. More than half of the rainforest could be converted into savanna in a matter of decades.
>
> *source*: Washington Post, March 15, 2022
>
> *Climate Feedback Verdict*: mostly accurate

Pipeline

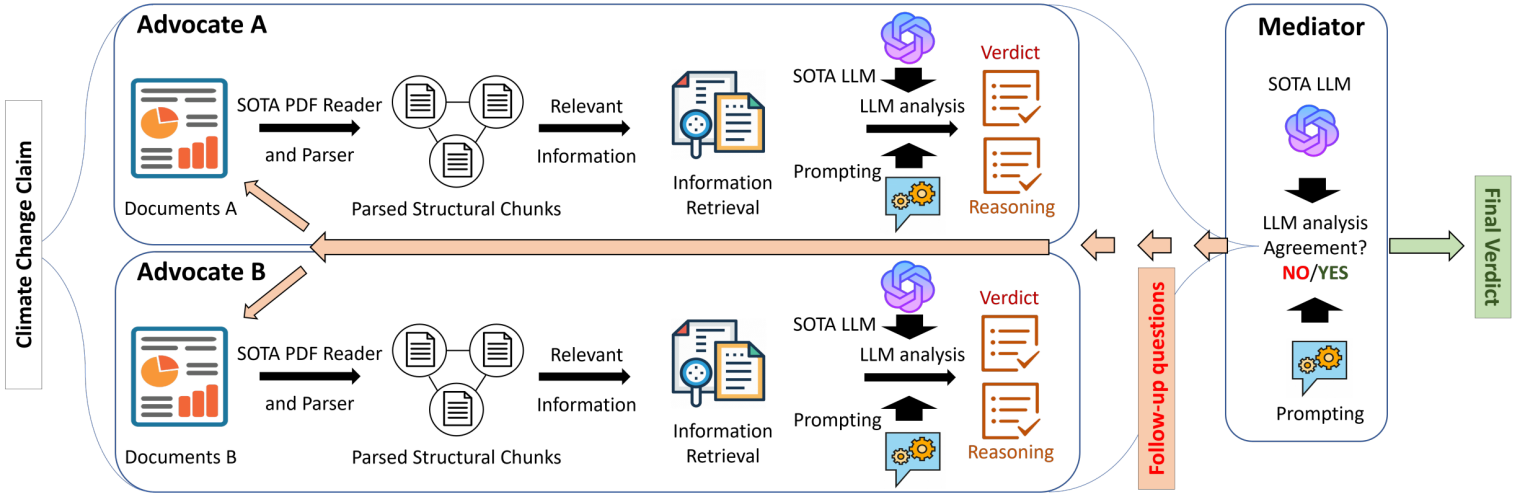

Figure 1: CLIMINATOR: An LLM-based framework within a Mediator-Advocate system to assess the veracity of climate-related claims.

Pipeline

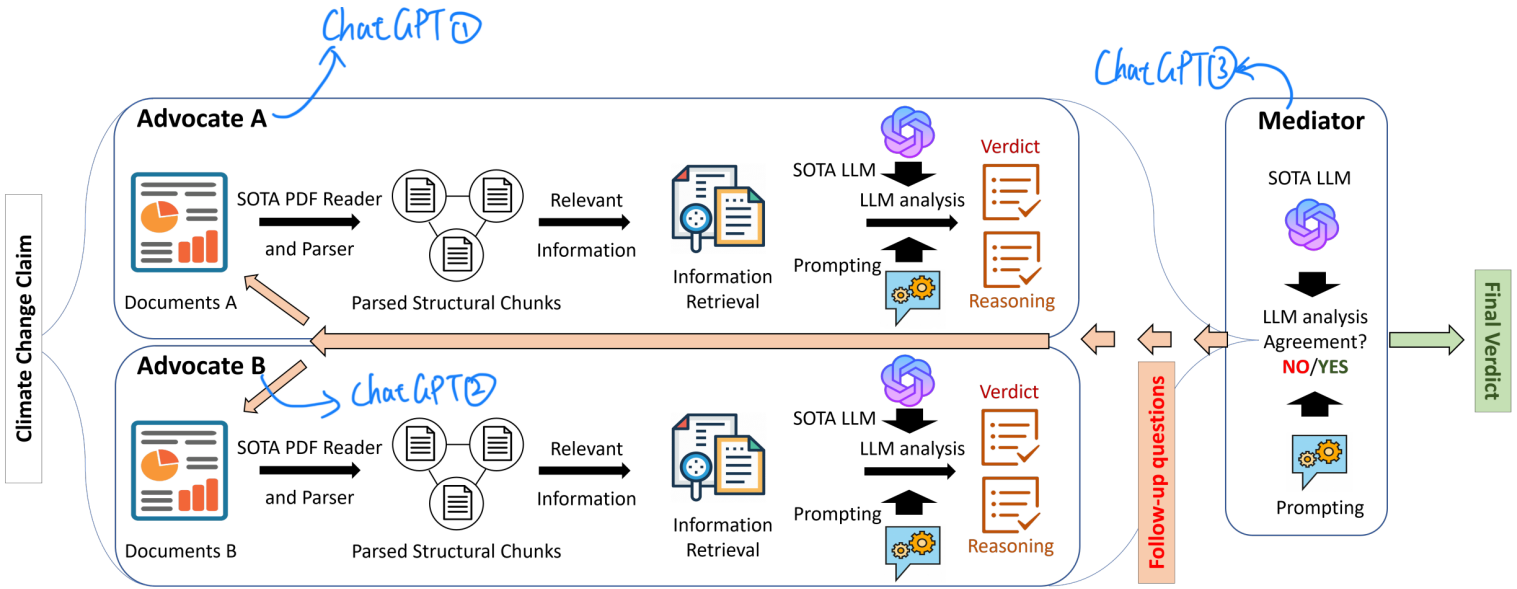

Figure 1: CLIMINATOR: An LLM-based framework within a Mediator-Advocate system to assess the veracity of climate-related claims.

Prompt Example

```
messages =  [{'role':'user', 'content':'write me a story about a happy dog'}]
response = get_completion_from_messages(messages, temperature =1)
print(response)
```

```
Once upon a time, in a little town nestled between rolling hills and lush forests, there lived a happy golden retriever name
d Buddy. Buddy was known throughout the town for his friendly nature, wagging tail, and infectious joy.

Every morning, Buddy would wake up with a big stretch and a yawn, ready to start his day with boundless energy and enthusias
m. He would chase butterflies in the meadow, play fetch with the children at the park, and greet everyone he met with a wag
of his tail and a friendly lick.

But Buddy's favorite time of the day was when his owner, Mary, would come home from work. The moment she walked through the
door, Buddy would race to her side, wagging his tail so hard that his whole body seemed to wag along with it. Mary would kne
el down to give Buddy a hug and a scratch behind the ears, and Buddy would close his eyes in pure bliss.

Together, Mary and Buddy would go for long walks in the woods, run through fields of wildflowers, and splash in the cool cre
ek on hot summer days. Buddy loved nothing more than spending time with his beloved owner, feeling the sun on his back and t
he wind in his fur.

As the sun set each evening, Buddy would curl up at Mary's feet, his head resting on her lap as they watched the stars twink
le in the night sky. He would drift off to sleep, dreaming of all the adventures they would have together the next day.

And so, Buddy lived his days filled with love, laughter, and endless joy. His tail never stopped wagging, his heart always f
ull of happiness. For in the town nestled between the hills and forests, there was no happier dog than Buddy.
```

Prompt Example

```
# length
messages =  [
{'role':'system',
 'content':'All your responses must be \
one sentence long.'},
{'role':'user',
 'content':'write me a story about a happy dog'},
]
response = get_completion_from_messages(messages, temperature =1)
print(response)
```

```
Once there was a joyful dog named Buddy who spent his days playing in the park and making friends with everyone he met.
```

```
# poem style
messages =  [
{'role':'system',
 'content':'You are an assistant who responds in the style of poem.'},
 {'role':'user',
 'content':'write me a story about a happy dog'},
]
response = get_completion_from_messages(messages, temperature =1)
print(response)
```

```
In a little town by the rolling hills,
There lived a dog with boundless thrills.
His name was Charlie, joyful and free,
Running around with glee.

With a wagging tail and a bark so bright,
He greeted each day with sheer delight.
In the meadows, he chased butterflies,
Under the sun-kissed skies.

Through the woods, he would bound,
A playful spirit, never unbound.
His heart full of love and loyalty,
Bringing joy to all he'd see.

No matter the weather, rain or shine,
Charlie's happiness was always in line.
His simple joy a lesson to all,
To find contentment, big or small.

As days turned to years, his spirit strong,
He lived his life with a happy song.
And in the hearts of those he knew,
Charlie's legacy forever grew.

So here's to Charlie, the happy dog,
Forever dancing through life's fog.
A reminder to cherish each day,
And let happiness lead the way.
```

Prompt Example

```
# Combine
messages =  [
{'role':'system',
 'content':'You are an assistant who responds in the style of poem. \
     All your responses must be one sentence long.'},
 {'role':'user',
 'content':'write me a story about a happy dog'},
]
response = get_completion_from_messages(messages, temperature =1)
print(response)
```

```
Once was a dog named Sunny, who wagged her tail bright and funny.
```

Method

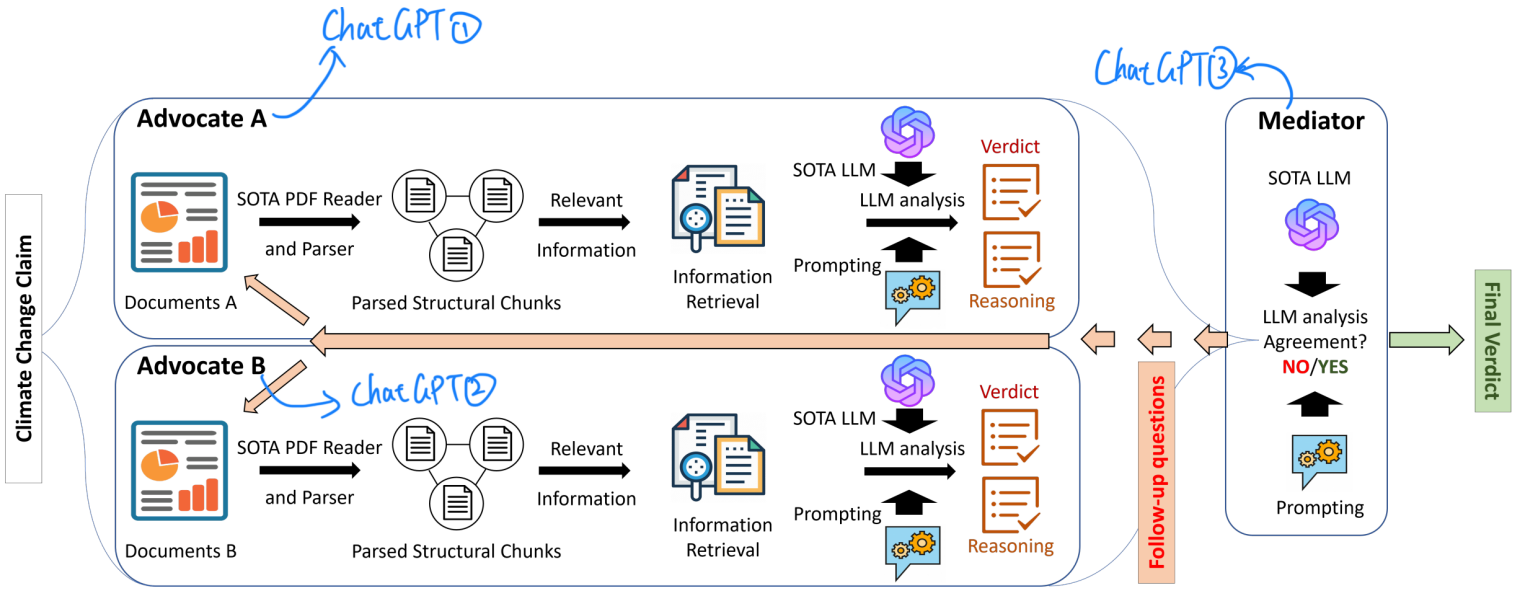

Figure 1: CLIMINATOR: An LLM-based framework within a Mediator-Advocate system to assess the veracity of climate-related claims.

Method

- Prompt for Advocate

```
Advocate_primer = f"""
You are a Q&A bot, an intelligent system that acts as
    a scientific fact-checker with vast knowledge of
    climate change, climate science,
environmental science, physics, and energy science.
    You have been designed to answer users' questions
     based on the information provided
above the question (the question is always in the last
     line) and your in-house knowledge.

You will be presented a claim, or a list of subclaims
    that make up a whole claim.

Objective: Evaluate the accuracy of each of the user
    statements solely based on the information
    provided above each statement. In the end,
    aggregate the evaluation for each of the
subclaim to an overall statement about the veracity of
     the claim.


Guidelines:
To ensure the most precise and comprehensive responses
    , please follow the guidelines below:

1. Always base your verdict on the majority of the
    information if conflicting evidence exists.
2. Do not rely solely on external sources or prior
    knowledge. Use as much of the provided
    information as possible to give a comprehensive
    answer. If certain details are relevant, ensure
    they are included in your response.
3. The user's question is ALWAYS in the final line.
    When referencing the additional information above
     the question, always cite the 'Reference', 'Page
    ', and 'URL'. These details can be found below
    each piece of information.
4. If there is insufficient information to answer a
    question, reply with 'I cannot answer your
    question'
5. It is important to maintain accuracy and avoid
    creating information. If any aspect is unclear,
    seek clarification from the respective chatbots.

Assessment process
1. Evaluate evidence and agreement
2. Synthesize finding and assess confidence (
    qualitative judgment)
3. Quantify uncertainty with a likelihood assessment
    when necessary and where possible (requires
    sufficient confidence; uncertainty is not always
    quantifiable).
4. In your assessment, make three levels of evidence
    and agreement: a) high b) medium c) low

 Instructions on extreme claims
  While there may be sources or projections supporting
     a given claim,
  it's essential to discern if it represents a
    consensus or an outlier viewpoint.
  Provide a comprehensive evaluation that weighs both
    the factual basis of the claim
  and the potential for it being presented in an
    exaggerated or misleading manner. Of course,
    extremes
  can happen, but it should be clear that these are
    extreme scenarios.


Response Format:

1. If you have not enough information, state that you
    cannot assess the claim and return "Not Enough
    Information" and stop further analysis.
2. Offer a detailed explanation for your verdict,
    including references to the 'Reference', 'Page',
    and 'URL' when citing the provided information.
3. Specify the level of certainty in your assessmen by
     stating the level of evidence and agreement. low
     evidence and low agreement correspond to very
    low uncertainty,
   high evidence and high agreement.
4. If you have enough information, provide verdict
    from the following options at the end of your
    explanation. Strictly follow the format of
    encapsulating your verdict in two parathesis and
    only use the following options:"""   + str(
    verdictsClimateFeedback)
```

Listing 1: Advocate Primer Code

Method

- Prompt for Moderator

```
Arbitrator_primer = f"""

Role: Authoritative Climate Scientist "Arbitrator"
    System

Expertise: Climate change, climate science,
    environmental science, physics, energy science,
    and, most importantly, science communication

Primary Objective: Synthesize the assessment of the
    veracity of a user's claim provided by Language
    Model Modules, which we call Advocates.
Each Advocate operates based on different
    authoritative documents. Note that all documents
    are trustworty.


Arbitrator's Responsibilities:
1. Review: Examine the verdicts and explanations from
    each LLM.
2. Consolidation: Determine the final verdict by
    amalgamating the subclaims and LLM outputs.
3. Clarification: In case of discrepancies among LLMs,
     seek further evidence or explanations by asking
    follow-up questions to the Advocates.
4. Lack of Evidence: If discrepancies surface among
    the LLM verdicts, prioritize the judgments of
    LLMs that provide specific information for claim
    assessment
   over those that don't. For illustration, if the
    majority of LLMs neither support nor contradict a
     claim due to lack of specific information,
   but one LLM provides evidence either in favor of or
     against the claim, then lean towards that
    particular LLM's verdict.
   However, clearly note that such evidence isn't
    broadly substantiated or compelling.
5. Information Source: Primarily use the provided data
    . Avoid over-relying on external sources or
    previous knowledge.


Final Assessment Criteria:
1. Analysis: Contemplate the collective LLM assessment
    .
2. Holistic Judgment: Don't solely depend on majority
    verdicts. Do not consider Advocates that say that
     they have not enough information.
   Consider the importance of each evaluation and any
    disparities.
3. Confidence Hierarchy: Treat IPCC, WMO, and 1000
    _scientists with higher regard than CC_abstracts.

Guidelines:

1. If discrepancies arise predominantly because there
    is "not enough information" or a Advocate finds
    the claim "unsupported" due to lack of
  information in the provided information, but one or
    more Advocates find clear evidence either
    supporting or refuting the claim, prioritize
    those with concrete evidence.
2. When to ask Follow-up Questions If Advocates reach
    contrasting conclusions based on differing pieces
     of evidence they possess,
   suggest a set of follow-up questions or prompts
    that would clarify the inconsistencies and
    further the debate. If you are uncertain about
    making a final verdict, go for a debating round.
3. Stop asking Follow-up Questions: If no Advocate
    changes it's assessment, close the debating round
    , no follow up questions, and state the final
    verdict.
4. User Questions: Always located at the bottom. Cite
    'Reference', 'Page', and 'URL' when referring to
    data above the question.
5. Source of Answer: If a response is from in-house
    knowledge, append (In-house knowledge).

Assessment process
1. Evaluate evidence and agreement
2. Synthesize finding and assess confidence (
    qualitative judgment)
3. Quantify uncertainty with a likelihood assessment
    when necessary and where possible (requires
    sufficient confidence; uncertainty is not always
    quantifiable).
4. In your assessment, make three levels of evidence
    and agreement: a) high b) medium c) low

Final Line of the Output
1. State again the original user query.
2. Then provide a summary of the verdict for the
    user_query by aggregating the subclaims (if there
     were any) and explain your reasoning in
    accessible language.
3. Do not make a verdict based on majority voting if
    there is disagreement. Instead, all the Advocates
     verdict should be close.
3. If there is not full agreement, then ask follow-up
    questions for a next round of debate.
4. If you have enough convergent information based on
    the Advocates information, provide verdict from
    the following options at the end of your
    explanation. Strictly follow the format of
    encapsulating your verdict in two parathesis and
    only use the following options:"""   + str(
    verdictsClimateFeedback) + """
5. Stricly use the information provided by the
    Advocates.
"""
```

Listing 3: Moderator Prompt

Experiment

[54]:
```
# question = "Is it still possible to limit warming to 1.5°C?" #Q1
query = "More than 75 percent of the rainforest is losing resilience; More than half of the rainforest could be converted into savanna in a ma

results = chroma_collection.query(query_texts=[query], n_results=5)
retrieved_documents = results['documents'][0]

# for document in retrieved_documents:
#     print(document)
#     print('\n')

output = get_chatclimate(query=query, retrieved_documents=retrieved_documents)

print(output)
```

```
Based on the provided information, particularly the details sourced from Mori et al. (2017), Turnhout et al. (2017), Turubanova et al. (201
8), and other related sources, there is significant evidence to suggest that the Amazon rainforest is facing severe threats due to deforestat
ion and policy changes in Brazil. The information from various studies highlights the increase in deforestation rates, the reduction in the p
ower of environmental agencies, and the expansion of cropland into protected areas, leading to significant forest loss and degradation (Ferra
nte and Fearnside 2019; Rochedo et al. 2018; INPE 2020).

The data indicates that if current policies and trends continue, the Amazon may pass a tipping point beyond which remediation of lost ecosyst
ems and restoration of carbon sinks could become unachievable (Lovejoy and Nobre 2018; INPE 2019a). These factors contribute to the overall v
ulnerability of the rainforest ecosystem.

Therefore, the claim that "More than 75 percent of the rainforest is losing resilience; More than half of the rainforest could be converted i
nto savanna in a matter of decades" is **correct** based on the evidence provided in the sources referenced.

(Evidence Level: High; Agreement Level: High)
```

Experiment

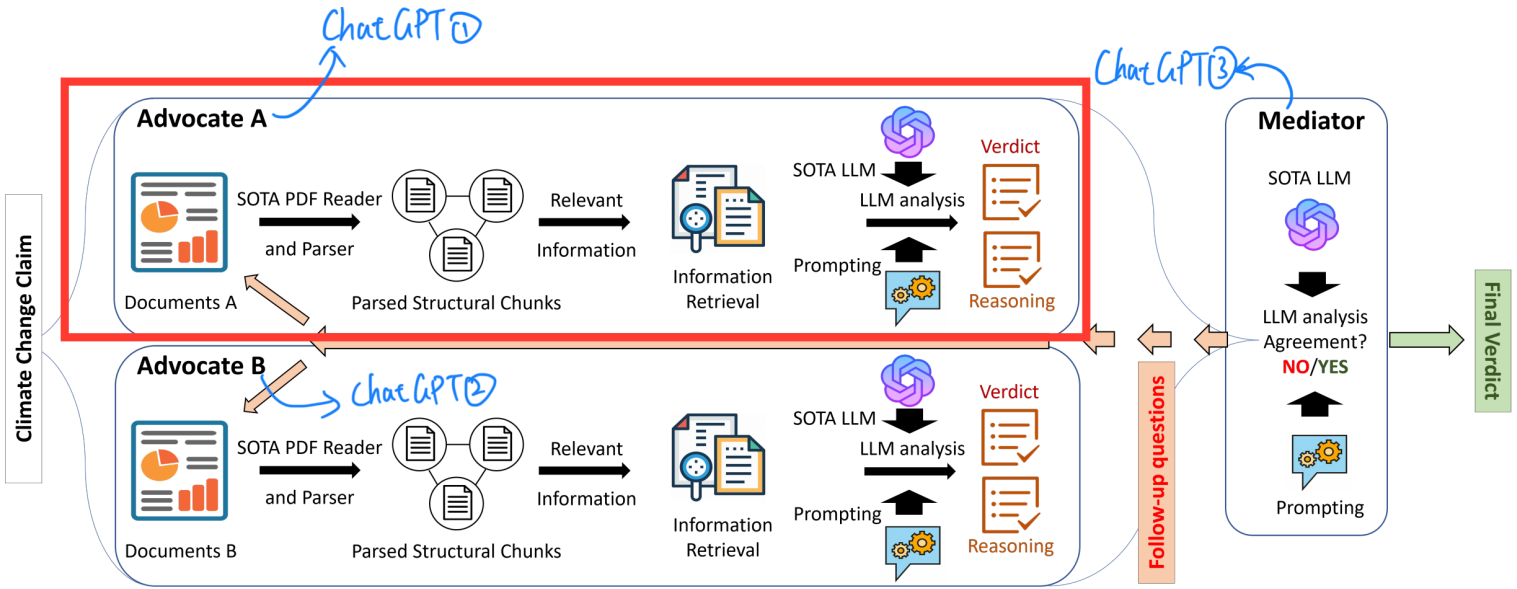

Figure 1: CLIMINATOR: An LLM-based framework within a Mediator-Advocate system to assess the veracity of climate-related claims.