

1 Appendix

2 Example of Using D2F for Fact-checking


<p>Claim:</p> <p>A significant number of internet users 'attacked' Andrew Pollack, a man mourning the death of his daughter, on the basis of his support for Donald Trump.</p> <p>Rating:</p> <div>  <p>Mixture</p> <p>About this rating</p> </div> <p>What's True</p> <p>A relatively very small number of people on Twitter made insulting and personal criticisms of Andrew Pollack, with one or two even blaming him for his own daughter's death, by way of his support for President Donald Trump.</p> <p>What's False</p> <p>Those negative voices were overwhelmingly outnumbered by those offering condolences and support to Andrew Pollack and his family, and we found no evidence of anyone addressing a criticism or attack directly at Pollack.</p>	<p>Questions:</p> <ol style="list-style-type: none"> 1. How many internet users criticized Andrew Pollack? {answer_1} 2. Who is Andrew Pollack? {answer_2} 3. Does Andrew Pollack support Donald Trump? {answer_3} <p>Subclaims:</p> <ol style="list-style-type: none"> 1. A significant number of internet users criticized Andrew Pollack. 2. Andrew Pollack is a man who mourned the death of his daughter. 3. Andrew Pollack supports Donald Trump. <p>Evidence:</p> <ol style="list-style-type: none"> 1. Databricks to Buy Startup Tabular in Effort to Broaden AI Reach · Salesforce Headquarters As Earnings Figures Released. -- bloomberg.com 2. Andrew Scott Pollack (born February 18, 1966) is an American author, school safety activist, and entrepreneur whose daughter Meadow was one of the 17 murdered victims in the... --wikipedia.org 3. Andrew Pollack spoke in support of President Donald Trump on the first night of the Republican National Convention. He was the 13th speaker ... --marquettewire.org <p>Answers:</p> <ol style="list-style-type: none"> 1. {answer_1}: unknown; 2. {answer_2}: Andrew Pollack is an American author, school safety activist, and entrepreneur who mourned the death of his daughter, Meadow, one of the 17 victims killed in a school shooting; 3. {answer_3}: true. <p>Final rating: Mixture.</p>
--	--

Figure 1: An example of using D2F for fact-checking. Left: a screenshot from Snopes.com; Right: the fact-checking process by D2F.

Figure 1 presents a real example of using D2F for fact-checking. This claim comes from Snopes.com. On the left side of Figure 1 is the fact-checking process conducted by the official fact-checker; on the right side is the process using D2F. As shown in Figure 1, D2F asks three questions about the claim and decomposes it into three subclaims. By using these questions as queries for search engine, three corresponding pieces of evidence are obtained. D2F then extracts answers to each question from the evidence, evaluates each subclaim, and provides a final rating: Mixture. This process mirrors the official explanation from Snopes.com, demonstrating the effectiveness of D2F in real-world scenarios.

10 Prompts

This subsection provides the detailed prompts used in our experiments to interact with the large language model. Each prompt was carefully designed to elicit the desired responses for our tasks. Table 1 presents the prompt for identifying ambiguities in claims using D2F, while Table 2 presents the prompt for fact-checking claims using D2F. The examples (including evidence) used in the prompts can be found in the RAWFC and LIAR datasets.

You are a fact-checker. Determine whether the subject in the claim contains ambiguity based on the following criteria. If there is, explain the type of ambiguity; if not, return None.

AMBIGUITY CRITERIA: Ambiguity manifests in diverse forms, including:

- Distinct Entities: Similar names denoting distinct entities.
- Lack of Context: Varied interpretations stemming from insufficient information.
- Vague or Unclear: Multiple understandings arising from vague or unclear information.

INSTRUCTIONS:

- Identify the main SUBJECT within the claim.
- Determine if the SUBJECT is ambiguous according to the provided AMBIGUITY CRITERIA.
- Utilize your world knowledge to enumerate potential DISAMBIGUATIONS for the identified SUBJECT.
- Specify the TYPE of ambiguity based on the list of AMBIGUITY CRITERIA.
- If the SUBJECT does not have ambiguous interpretations, return None.
- Provide an explanation of the method used to arrive at the final response.

Format your response as a combination of explanation and a dictionary with the following structure:

##EXPLANATION##:

<step-by-step-explanations>

##RESPONSE##:

```
{"subject": <subject>, "disambiguations": [ <instance-1>, <instance-2>..], "ambiguity_type":  
<type>}
```

Example 1:

##CLAIM##: Obama has more czars than the Romanovs.

##EXPLANATION##: The term "czars" is used to describe specific officials appointed by the president, such as an "energy czar" or "drug czar." However, this term has a completely different historical context and meaning compared to the Russian "czars" (rulers of the Romanov dynasty). This metaphor may cause misunderstanding by equating modern political positions with historical emperors.

##RESPONSE##:

```
{"subject": "czars", "disambiguations": ["czars - specific officials appointed by the  
president", "czars - Russian czars (rulers of the Romanov dynasty)"], "ambiguity_type":  
"Distinct Entities"}
```

Example 2:

##CLAIM##: "Nearly 20% of our residents" are born abroad.

##EXPLANATION##: The SUBJECT is "our residents". It is ambiguous because it lacks the context of what it refers to. For example, "our residents" could mean residents of a particular city, state, or country, but the claim does not specify which location is being referred to.

##RESPONSE##:

```
{"subject": "our residents", "disambiguations": ["Austin residents", "Ohio residents"],  
"ambiguity_type": "Lack of Context"}
```

Example 3:

##CLAIM##: The Fed created \$1.2 trillion out of nothing, gave it to banks, and some of them foreign banks, so that they could stabilize their operations.

##EXPLANATION##: The term "foreign banks" is ambiguous. Because it can refer to either foreign-funded banks located locally or banks worldwide. But this claim does not specify clearly.

##RESPONSE##:

```
"subject": "foreign banks", "disambiguations": ["foreign-funded banks", "banks worldwide"],  
"ambiguity_type": "Vague or Unclear"
```

Example 4:

##CLAIM##: Ruth Bader Ginsburg served as a Supreme Court justice.

##EXPLANATION##: The SUBJECT is "Ruth Bader Ginsburg". According to my world knowledge, this is a unique individual and I am not aware of any other individuals/entities with a similar name. Hence, there are no ambiguous interpretations of this SUBJECT and the claim requires no further disambiguation.

##RESPONSE##:

```
{"subject": "Ruth Bader Ginsburg", "disambiguations": "None"}
```

Similarly, disambiguate the following claim by detecting the main SUBJECT and disambiguation information for the SUBJECT using your world knowledge. Generate an EXPLANATION followed by dictionary-formatted RESPONSE.

##CLAIM##: [claim]

##EXPLANATION##:

Table 1: Prompt for identifying ambiguities in claims Using D2F

You are a fact-checking assistant. Your task is to analyze a claim and the corresponding evidence, and perform the following:

1. Disambiguation: Identify and clarify any ambiguous terms, references, or phrases in the claim. If multiple interpretations are possible, explicitly list them and choose the most plausible one based on context.
2. Decomposition: Break down the claim into smaller, independent subclaims or components that can be individually verified. Ensure each subclaim is simple and unambiguous.
3. Decontextualization: Restate each subclaim in a way that removes any dependency on specific contexts or situations, ensuring it can be understood and verified as a standalone statement.
4. Verify the claim: Based on the evidence, label the claim as TRUE, FALSE, or MIXTURE. Provide reasoning for your decision.

Example 1:

Claim: In 2020, PepsiCo gave \$100 million to entities or projects associated with the Black Lives Matter movement.

Evidence:

- Fast forward to 2020, and in the face of renewed widespread Black Lives Matter protest, Pepsi's CEO said the company will spend \$400 million to "lift up Black community."
- PepsiCo has pledged \$400 million to a set of initiatives over five years to "lift up Black community and increase Black representation at PepsiCo."
- PepsiCo's CEO announced a commitment of \$400 million over five years to "lift up Black community and increase Black representation."

Label: FALSE

Steps:

1. Disambiguate: Clarify the claim's focus on the specific \$100 million amount and its association with Black Lives Matter.
2. Decompose:
 - Did PepsiCo commit \$100 million?
 - Was the commitment specifically tied to Black Lives Matter?
3. Decontextualize: Did PepsiCo pledge a specific amount toward supporting Black communities?
4. Verify:
 - The evidence indicates a \$400 million commitment over five years for broader initiatives, not specifically \$100 million or directly tied to Black Lives Matter.

Example 2:

Claim: A 9-year-old girl was paralyzed as a result of receiving a flu shot.

Evidence:

- A 9-year-old received a flu shot, triggering acute disseminated encephalomyelitis (ADEM).
- ADEM is a rare immune-mediated condition often associated with infections or, in rare cases, vaccines.
- Reports indicate that a vaccine, including flu shots, can be associated with ADEM in rare cases.

Label: MIXTURE

Steps:

1. Disambiguate: Define "paralyzed" as the medical impact of ADEM and its link to the flu shot.
2. Decompose:
 - Did the girl receive a flu shot?
 - Did the flu shot cause ADEM?
 - Did ADEM result in paralysis?
3. Decontextualize: Could a flu shot result in paralysis due to ADEM in rare cases?
4. Verify:
 - The evidence links the flu shot to ADEM in rare cases, but causation is not definitive for this specific instance.

Example 3:

Claim: A social media meme accurately lists U.S. government assets deployed by President Trump to aid Puerto Ricans in the aftermath of Hurricane Maria.

Evidence:

- Trump announced \$11.6 billion and later \$13 billion in aid for Puerto Rico.
- Claims of \$91 billion in aid were false, as only \$41 billion was allocated.
- Bureaucratic obstacles delayed \$20 billion in relief funds.

Label: MIXTURE

Steps:

1. Disambiguate: Clarify "accurately lists" as referring to the meme's claims about assets and aid amounts.
2. Decompose:
 - Did the government deploy aid?
 - Were the amounts in the meme accurate?
 - Were there delays or obstacles?
3. Decontextualize: Did the U.S. government provide substantial aid to Puerto Rico under Trump, and was it accurately depicted?
4. Verify:
 - Evidence shows inaccuracies in the claimed amounts and delays in aid distribution.

Claim: [Claim]

Evidence: [Evidence]

Label: