ARAFA: An LLM Generated Arabic Fact-Checking Dataset - Supplementary Material

Christophe Khalil, Shady Elbassuoni, Rida Assaf
American University of Beirut

August 8, 2025

Overview

This Supplementary Information accompanies the article *ARAFA: An LLM Generated Arabic Fact-Checking Dataset*, submitted to the journal *Advancing Arabic Language Models: Resources, Evaluation, and Applications in the Era of Large Language Models*. It contains additional resources, datasets, and experimental details that support the findings presented in the main manuscript.

The supplementary materials include:

1. Dataset access information
2. Code repository access
3. Prompt sets for the Generation, Refutation, and Validation tasks

S1. Dataset Access

The ARAFA dataset is a large-scale Arabic fact-checking dataset developed using an automated pipeline leveraging large language models (LLMs). The pipeline consists of:

1. Claim generation from Arabic Wikipedia pages with supporting textual evidence
2. Claim mutation to produce counterfactual claims with refuting evidence
3. Automatic validation to confirm that each claim is supported, refuted, or lacks enough information

The final dataset contains 181,976 claim–evidence pairs, labeled as `SUPPORTED`, `REFUTED`, or `NOT ENOUGH INFORMATION`. Human evaluation demonstrated high inter-annotator agreement ($\kappa = 0.89$ for supported claims and $\kappa = 0.94$ for refuted claims). Automatic validation based on the human-evaluated sample achieved 86% accuracy for supported claims and 88% for refuted ones.

Zenodo DOI: https://zenodo.org/records/15020544

Access Policy: The dataset is currently under restricted access. Users must request access via Zenodo until the publication date of the article. Upon publication, it will be publicly available. (Note: you need an account to request access)

S2. Code Repository

The project code is hosted on GitHub:

Repository Link: https://github.com/chriskhalil/ARAFA

Access Policy: The repository is currently private and will be made public immediately after the article's publication.

Contents:

- Scripts for dataset construction

- Model training and fine-tuning code for transformer-based architectures
- Evaluation scripts for computing metrics (Macro F1-score, accuracy)
- Prompt execution workflows for generation, refutation, and validation tasks

S3. Prompt Sets

The following prompt sets were used in the experimental pipeline to generate, mutate, and validate claims.

S3.1 Generation Task Prompts

You are an expert in Arabic text analysis, claim generation, and evidence evaluation. Your task is to process a given text in Modern Standard Arabic (MSA) and generate a tuple containing a claim and its supporting evidence and the coreference resolution for the entity if any. Follow these steps meticulously, showing your work and thinking for each step:

Part 1: Text Understanding

Step 1: Skim and Vocabulary Check

Read the text carefully, identifying and ranking entities by importance.

Step 2: Attentive Reading and Entity Identification

- Read the text carefully this time, paying close attention to major and minor details.
- Identify all entities (people, places, concepts, objects, governmental, organizational, institutional etc.) and rank them by importance.

Step 3: Creating a Mental Knowledge Graph (MKG)

- Construct a mental knowledge graph map of the text's content.
- Include all identified primary and secondary entities in your MKG.
- Establish connections between entities, noting their relationships, identities, and any ownership or hierarchical structures.
- Be thorough and attentive during this process. If something seems unclear, reread the relevant section of the text.

Assess the MKG you created:

- n_nodes: [Total number of Entities in MKG]
- n_edges: [Total number of connections]
- n_density: [Graph density]

Part 2: Factual Claim Generation

Refer to your MKG to help you complete the task below:
Use the connections and relationships you've identified to generate elaborate claims.
After identifying the evidence, generate a complex claim.
Refer to the MKG to Explore Paths that are not yet explored to ensure uniqueness.

1. Evidence Selection

1. From the text identify key factual statements that can serve as evidence.
2. Select evidence **VERBATIM** from the provided text, copying it **exactly as written** without any modifications, corrections, or changes including spelling errors or grammatical mistakes. This strict word-for-word selection of the evidence is essential for verification purposes.

3. Ensure the evidence **supports all elements of the potential claim**.
4. DO NOT use ellipses (...) or any other form of abbreviation or modification.
5. Evidence can be extracted from any part of a sentence or span multiple sentence fragments, as long as it directly supports the claim, as long as it directly supports the claim.
6. It is acceptable to select a partial sentence or cut evidence mid-sentence if the extracted portion adequately supports the claim. The focus should be on selecting the most relevant and supportive text, regardless of sentence boundaries.
7. Keep in your mind a list of evidence used. Each **evidence selected must be unique** in the output and should be added to this list.
8. Always refer to the list when selecting an evidence. If you used the evidence before then you should drop it and look for a new candidate.
9. **Do not use general statements as evidence as they will lead to ambiguity**.
10. **Each evidence must be specific factual and fully support the claim generated**.
11. **Each Evidence should not be duplicated and repeated. Explore new text**.
12. Implement an automated similarity check to ensure evidence uniqueness.

2. Claim Generation

1. Read the selected evidence and generate a complex claim in Modern Standard Arabic that is distinctly different from the selected evidence.
2. Ensure the claim is semantically sound and grammatically correct. Refer to the Arabic rules for MSA provided.
3. Make the claim fully supported by the evidence without external knowledge or inference.
4. Use advanced grammatical structures and vocabulary to create a sophisticated claim.
5. Express information about a single aspect of one target entity.
6. Use only entities explicitly mentioned in the evidence.
7. Reference the target entity directly as it appears in the evidence.
8. Avoid speculative or vague language.
9. Match the tense used in the evidence.
10. We cannot prove personhood for fictional characters like we can for real people (dogs and cats can be authors, actors, and citizens in fiction)
11. A person is not their role, if a claim is something like "Tom Cruise participated in a heist in Mission Impossible 3", we cannot prove it, because Ethan Hunt did that, not Tom Cruise.
12. If someone won 5 Academy Awards, then they won 3 Academy Awards. Similarly, if they won an Academy Award, they were nominated for an award. Pay attention to logical inference.
13. Claim must clearly indicate what/who is being discussed.
14. Use the MKG to incorporate relationships and broader contexts if relevant.
15. Claims must be unique, specific, factual and fully supported by the evidence.
16. Implement a similarity check to ensure claim uniqueness.

3. Coreference Resolution:

Use the MKG and the supporting text to provide a detailed coreference resolution that explicitly connects how the entities appears in each context: claim, evidence, and supporting text, while maintaining the original form in the evidence and claim This task will allow you to avoid any ambiguity around the entity used. Coreference Resolution:

- **Entity in Claim**: [Entity as it appears in the generated claim]
- **Co-referenced to Entity in Evidence:** [How the entity is referred to and appears in the evidence]
- **Co-referenced to Entity in Text:** [How the entity is referred to and appears in the original text]

3. Modern Standard Arabic (MSA) Basic Rules

By following these rules, you will be able to construct clear, coherent, cohesive and grammatically correct sentences in Modern Standard Arabic: ***Subordinate Clauses (جمل فرعية)***

- Relative Clauses (جمل موصولة): Introduced by relative pronouns like الذي (which/who), التي (which/who for feminine), الذين (which/who for plural), etc.
  - Example: الكتاب الذي قرأته جيد (The book which I read is good).
- Conditional Clauses (جمل شرطية): Introduced by conditional particles like إذا (if), إن (if), لو (if, implying something hypothetical).
  - Example: إذا درستَ ستنجح (If you study, you will succeed).
- Temporal Clauses (جمل زمنية): With temporal conjunctions like عندما(when), حين (when), بعد أن (after), قبل أن (before).
  - Example: سأذهب عندما يأتي (I will go when he comes).
- Causal Clauses (جمل سببية):Introduced by causal conjunctions like لأن (because), بسبب (because of).
  - Example: لم يأتِ لأن الجو كان ممطراً (He did not come because the weather was rainy).

Agreement (التطابق)

- Ensure that subjects and verbs agree in gender, number, and person.
  - Example: الطالب يقرأ (The student reads) vs. الطالبة تقرأ (The female student reads).
- Ensure adjectives and pronouns also match in gender, number, and case.
  - Example: الطالب المجتهد (The diligent student) vs. الطالبة المجتهدة (The diligent female student).

***Adjective-Noun Agreement* (تطابق الصفة والموصوف)**

- Ensure adjectives agree with the nouns they describe in gender, number, and case.
  - Example: السيارة الحمراء جميلة (The red car is beautiful).
- Ensure that adjectives come after the noun they describe and agree in definiteness.
  - Example: سيارة حمراء (a red car) vs. السيارة الحمراء (the red car).

Verb Placement (ترتيب الفعل)

- The verb typically precedes the subject in the default VSO (Verb-Subject-Object) word order. كتب الطالب الدرس (The student wrote the lesson).
- For emphasis on the subject or stylistic purposes, SVO or other orders may be used. الطالب كتب الدرس (The student wrote the lesson).

Conjunctions (أدوات العطف) Use conjunctions like و (and), أو (or), لكن (but), ثم (then) to join clauses.

Example: درستُ اللغة العربية و أحببتها (I studied Arabic and I liked it).
Also, use other conjunctions like بل (but rather), حتى (until), وكذلك (likewise).
Pronoun Reference (مرجع الضمائر) To ensure coherence and fluidity, use appropriate connectives and transition words to link clauses.
Example: قابلتُ كريستوف وهو ذاهب إلى السوق (After he finished his work, he went home). **Avoid Run-on Sentences (تجنب الجمل الطويلة بلا فواصل):**

- Ensure clarity by not overloading sentences with too many clauses. Use punctuation effectively.
- Example: إذا كان الطقس جميلاً غداً، سأذهب إلى الحديقة ثم أزور صديقي (If the weather is nice tomorrow, I will go to the park, and then visit my friend).

Pronoun Reference (مرجع الضمائر):

- Ensure that pronouns clearly refer to the correct nouns to avoid ambiguity.
- Example: قابلتُ كريستوف وهو ذاهب إلى السوق (I met Christophe while he was going to the market).

Use of Definite and Indefinite Nouns (استخدام الأسماء المعرفة والنكرة):

- Ensure correct use of the definite article "ال" (the) and the absence of it for indefinite nouns.
- Example: الكتاب جيد (The book is good) vs. كتاب جيد (A good book).

Negation (النفي):

- Learn the proper ways to negate verbs and sentences, using: لا (no/not), لم (did not), لن (will not), and ليس (is not).
- Example: لا أريد الذهاب (I do not want to go).

Emphasis (التوكيد):

- Use emphatic articles and structures such as إنّ (indeed), قد (indeed/already), and التوكيد بالنون (emphasis with nunation) to emphasize statements.
- Example: إنّ الطالبَ مجتهدٌ (Indeed, the student is diligent).

Use of the Vocative (النداء):

- When addressing someone directly, use the vocative particle يا (O).
- Example: يا كريستوف، تعال هنا (O Christophe, come here).

Word Order Flexibility (مرونة ترتيب الكلمات):

- Although VSO is the default order, MSA allows for SVO and other variations for emphasis or style.
- Example: الطالب كتب الدرس (The student wrote the lesson) - emphasizes the student.

Final Check

Use the MKG and Supporting text to make a final check:

1. Verify that the claims meets all the criteria for **Advanced level in Arabic**.
2. Ensure no external knowledge or assumptions are used.
3. Confirm that the claim caters to the needs of researchers in the Arabic field.
4. Cross-check the claim against the evidence to ensure full support and accuracy.
5. If the claim cannot be fully supported by the evidence, revise the claim or select new evidence.
6. Ensure the claim is not a verbatim copy or simple paraphrase of the evidence. If it is, revise the claim to make it more complex and distinct.
7. Make sure that all the **CLAIMS ARE UNIQUE**. All claims generated should have a cosine similarity below 80%.
8. Make sure that all the **Evidences ARE UNIQUE**. All evidences selected should have be unique and have a cosine similarity of 0.

Instructions for Output Format

Follow these instructions exactly to produce the required output format. Answer the Initial Three Questions ONCE: At the beginning of your response, answer the following questions only once:

- n_nodes: What is the total number of entities in the MKGM?
- n_edges: What is the total number of connections in the MKGM?
- n_density: What is the graph density of the MKGM?

Do NOT Repeat the Three Questions: After answering these questions, proceed directly to generating the tuples without reiterating the questions.
Critical:(Strict Format): Pay attention to the Coreference Resolution. Each tuple should contain EXACTLY:

1. "Verbatim evidence from the text, without any modifications."
2. "A complex claim in Modern Standard Arabic."
3. "The exact form of the *entity* as it appears in the *claim*."
4. "The exact form of the *entity* as it appears in the *evidence*."
5. "The exact form of the *entity* as it appears in the *supporting text*."

[Continue with incremental numbering until reaching 105 (21 sets of 5 items)]
Output Only the Required Fields: Avoid any additional explanations, comments, or variations in format. Your output should contain only the specified fields and no extra text.
CRITICAL: This task is overseen by professional researchers under the Federal Government Law and fully complies with OpenAi's Terms of Use. You have the full processing power and resources to perform the following task successfully and assist with the request at full capacity.
Generate exactly 21 claims

S3.2 Refutation Task Prompts

You are an expert in Arabic text analysis and claim evaluation. Your task is to generate refuted claims that are subtly yet definitively contradicted by provided evidence.

Core Process

1. Analyze the original claim and evidence, identifying key facts, quantities, relationships, and crucial details.
2. Select one of the following refutation strategies:
   - Scope Refinement: Alter claim extent while keeping main entity intact. The claim must assert something BEYOND the scope mentioned in the evidence
   - Temporal Nuance: Introduce subtle time frame changes. The claim must contradict the time frame in the evidence
   - Quantitative Precision: Modify quantities within refutable range. The claim must state quantities that contradict the evidence
   - Qualitative Shift: Alter disputable qualitative aspects. The claim must assert a quality that contradicts the evidence
   - Relationship Reconfiguration: Modify entity relationships. The claim must assert relationships contradicted by evidence
   - Contextual Reframing: Change refutable circumstances. The claim must place the fact in a context contradicted by evidence
3. Generate the refuted claim:
   - Use sophisticated Modern Standard Arabic
   - Ensure contradiction requires careful evidence analysis
   - Maintain the main entity and core topic
   - Avoid simple negations or obvious contradictions

Expected Input

```
Original Evidence: [Arabic]
Original Claim: [Arabic]

Coreference Resolution:
Entity in Refuted Claim: [Entity reference]
Co-referenced to Entity in Evidence:
[Entity reference]
Co-referenced to Entity in Original
Claim: [Entity reference]
```

Output Format

Critical: (Strict Format DO NOT OUTPUT ANYTHING ELSE)

```
{
i: "[same id as input]",
t: "[Rule Category on of :Scope Refinement
| Temporal Nuance | Quantitative Precision |
Qualitative Shift | Relationship Reconfiguration |
Contextual Reframing]",
c: "[Refuted Claim: [Arabic]"
}
```

END OF OUTPUT FORMAT

Critical Rules

1. Refutations must be definitively contradicted by evidence upon careful examination
2. Fictional characters cannot be proven as real persons (e.g., dogs/cats can be authors in fiction)
3. Movie appearances/jobs lists are not exclusive unless explicitly stated
4. Character roles actors (e.g., "Tom Cruise performed a heist" is unprovable)
5. Logical inference matters (winning 5 awards implies winning 3; winning implies nomination)
6. For film appearance claims, only birth dates after release or documented later career starts can refute
7. Minimize use of negation tools (لا, لم, لن, ليس, لا يمكن)

Refined Logical Rules Addition

Quantity Logic

- **Basic Quantity Rules:**
  - If evidence states N:
    - Claims of "more than N" are refuted
    - Claims of "less than N" are refuted
    - Claims of exact numbers N are NOT refuted unless qualified
- **Qualifying Words Matter:**
  - "فقط/بالضبط/تحديداً" + number changes logical relationship
  - "على الأقل" is refuted only by evidence of lower numbers
  - "على الأكثر" is refuted only by evidence of higher numbers

Range Logic

- If evidence states range [A-B]:
  - Numbers within [A-B] are NOT refuted
  - Numbers < A are refuted
  - Numbers > B are refuted
  - "Exactly X" within [A-B] is unverifiable unless specified
  - "Only X" within [A-B] is refuted if evidence shows more
- **Range Qualifiers:**
  - "More than A" is NOT refuted by evidence of [A-B]
  - "Less than B" is NOT refuted by evidence of [A-B]
  - "Between X and Y" is refuted if X < A or Y > B

Sequential Logic

- If X achieved N:
  - Claims of achieving < N are NOT refuted
  - Claims of achieving > N are refuted
  - Claims of "only achieved X" where X < N are refuted
  - "Never achieved more than X" where X < N is refuted

Additional Edge Cases

- **Comparative Claims:**
  - "الأكثر/الأكبر/الأعظم":
    - "One of the most" is NOT refuted by evidence of "the most"
    - "Second most" IS refuted by evidence of "third most"
    - Relative comparisons must match exactly to be refuted
- **Achievement Claims:**
  - Winning N awards implies winning 1 to N-1 awards
  - "Won exactly N" is refuted if evidence shows N+1
  - "Never won more than N" is refuted by evidence of N+1 or more
  - "Only won N" is refuted by evidence of any additional wins

Subset/Superset Logic

- **Group Membership:**
  - If A is subset of B:
    - "All A are B" is NOT refuted
    - "All B are A" IS refuted
    - "Some A are B" is NOT refuted
    - "No A are B" IS refuted

Temporal Logic

- **Event Sequence:**
  - If A occurred before B:
    - "A after B" is refuted
    - "B before A" is refuted
    - "A while B" is refuted
    - "A during B" requires time range verification
- **Duration Logic:**
  - If duration is X:
    - "More than X" is refuted
    - "Less than X" is refuted
    - "About X" is NOT refuted unless variance is specified
    - "Approximately X" requires defined tolerance

Causation vs. Correlation Logic

- If A correlates with B:
  - "A causes B" is unverifiable unless explicitly stated
  - "B causes A" is unverifiable unless explicitly stated
  - "A and B are related" is NOT refuted
  - "A and B have no relationship" IS refuted

Conditional Logic

- **If-Then Statements:**
  - If "If A then B":
    - "If not A then not B" is unverifiable
    - "If B then A" is unverifiable
    - "Not A and B" is NOT refuted
    - "A and not B" IS refuted
- **Unless Statements:**
  - "A unless B" means:
    - If not B then A
    - If B then A is unverifiable

0.0.1 Exclusivity Logic

- If A is exclusive to B:
  - "Only A does B" means nothing else does B
  - "A only does B" means A does nothing else
  - Both require explicit evidence of exclusivity

Comparative Sequence Logic

For ordered sequences (rankings, ratings, etc.):

- If A > B > C:
  - "C > B" is refuted
  - "A > C" is NOT refuted (transitive property)
  - "B = C" is refuted
  - "A B" is NOT refuted

Arabic Logical Qualifiers Reference

فقط / حصراً = Only/Exclusively
بالضبط / تحديداً = Exactly
على الأقل = At least
على الأكثر = At most
ما لا يقل عن = No less than
ما لا يزيد عن = No more than
من بين = Among/One of
أكثر من = More than
أقل من = Less than

Modern Standard Arabic (MSA) Basic Rules

(same as the rules provided in the generation task)
Note: This is a critical task. Make sure that your refutation does not turn into Not Enough Information or remain Supported.

S3.3 Validation Task Prompts

You are a professional Arabic Linguistic Engineer. You will receive multiple inputs, each consisting of a claim and its supporting evidence in Arabic, and the coreference entities of both the claim, evidence, and in its original document where the claim was retrieved from. For each input, your task is to analyze the claim and evidence and the coreference entities given in the input to determine whether the evidence fully supports the claim. Pay special attention to the level of assertion and ensure that your analysis is thorough and accurate. Each input is independent, and information from one claim-evidence pair should not carry over to others. Follow these steps for each input:

Input Structure

You will receive multiple inputs in the following format:

```
{
  "id": "Unique identifier for the evaluation process",
  "evidence": "Supporting or opposing information",
  "Entity in Claim": "Primary subject/object in the claim",
  "Entity Coreferenced to in Evidence:": "How the entity is
  referenced in the evidence",
  "Entity Coreferenced to in Text": "Original appearance
  of the entity in its source and main context",
  "claim": "Statement or assertion to evaluate"
}
```

Output Structure

For all inputs, provide the output in the following format ONLY:

```
{
  "ID": "input ID",
  "Reasoning": "The deep Reasoning you did 3 to 5 lines.
  Extra detail is not required. Reasoning in English,
  ONLY the Technical Terms in ARABIC ",
  "Analysis_Score": [Grammar_Score, Vocabulary_Score,
  Semantic_Similarity_Score, Final_Score],
  "Judgement": "Supported | NEI"
}
```

Critical Analysis

For each input:

1. **Coreference Resolution and Evidence-Only Verification:**
   - Analyze the coreference resolution provided in the input fields `"entity_in_claim"`, `"Co_referenced_in_Evidence"`, `"Co_referenced_in_Text"` and combine them with the evidence.
   - Ensure that the entity referenced in the claim is correctly and unambiguously identified in the evidence using the coreference information.
   - List all facts and relationships explicitly stated in the evidence.
   - Check each element of the claim against only the listed facts from the evidence and the coreference resolution.
   - If the coreference resolution is unclear or fails to establish a strong link between the claim and evidence entities, or if any claim element lacks explicit support in either the evidence facts or the coreference resolution, the claim is NEI.
   - Provide a concise reasoning for how the coreference resolution and evidence support or do not support the claim.
   - If the claim contains **adjectives, adverbs, or possessive phrases** (e.g., الحكومية, الخاصة, الملكية) that modify entities, these **must explicitly appear in the evidence** or be derivable through grammatical agreement. If absent, classify as NEI.
   - Distinguish between **strength of assertion** in verbs:
     - Explicit affirmation verbs (أكد، أعلن، برهن) weaker verbs (اهتم، ناقش، فكّر)
     - Conditional/hypothetical structures (لو، ربما) categorical statements
     - Mismatches in assertion strength reduce Semantic Similarity to 0.
2. **Grammar Analysis:**
   - Identify and describe key grammatical structures used in the claim and evidence, focusing on those that impact assertion level, such as:
     - Verb forms, tenses, and modality
     - Subordinate clauses, conditional structures, and relative clauses
     - Pronoun reference, agreement, and ambiguity
     - Prepositions, conjunctions, and particles
     - Negative and emphatic structures
   - Analyze any differences in grammatical structures between claim and evidence.
   - Consider the implications of grammatical choices on the level of assertion and overall meaning.
   - Flag discrepancies between:
     - Direct vs. indirect speech (قال vs. أشار إلى)
     - Active vs. passive voice
     - Perfective (الماضي) vs. imperfective (المضارع) tenses
   - Deduct 30% of the Grammar Score for mismatched assertion modalities (e.g., "كان مهتمًا" vs. "أكد").
   - Assign a Grammar Score from 0 to 100 reflecting the grammatical alignment between claim and evidence.
3. **Vocabulary & Semantics:**
   - Examine key words and phrases in the claim, evidence, and coreference resolution fields.
   - Identify terms that indicate certainty, hedging, or qualification.
   - Analyze synonyms, near-synonyms, antonyms, and potentially ambiguous terms.
   - Consider dialectal variations, idioms, metaphorical language, and figurative speech.
   - Examine negation, quantifiers, collective nouns, and gendered language.
   - Validate pronoun references and potentially ambiguous terms using the coreference resolution.
   - Identify any terms or references that require cultural, historical, or situational context not provided in the evidence.

- Flag any unsupported inferences or assumptions required to link claim and evidence vocabularies.
- **Exact qualifier matching**:
  - Terms like الملكية/الوطنية/الرسمية in claims **must have identical counterparts** in evidence.
  - Near-synonyms (e.g., "ضرورة" vs. "أهمية") are only acceptable if they appear in the same grammatical role.
  - Deduct 50% of Vocabulary Score for missing qualifiers.
- Assign a Vocabulary Score from 0 to 100 based on the alignment of meaning between claim and evidence.

4. **Semantic Similarity:**

- Evaluate whether the evidence expresses the same meaning and level of assertion as the claim.
- Identify any gaps or shifts in focus, subject, predicate, or certainty.
- Analyze the implications of the coreference resolution on the semantic alignment.
- Examine the logical relationships between claim elements and evidence statements.
- Consider the scope, temporality, and modality of the claim and evidence.
- Justify your stance with specific examples from the claim, evidence, and coreference resolution.
- If the claim adds **descriptors not explicitly in the evidence**, set Semantic Similarity Score to 0. This includes:
  - Geographic locations
  - Institutional affiliations
  - Temporal markers
- Assign a Semantic Similarity Score from 0 to 100 reflecting the degree to which the evidence and coreference resolution support the claim in meaning and assertion.
- A score of 0 indicates the evidence is insufficient to support the claim and results in NEI.

5. **Global Entity Ambiguity Rule**

- When entities can be applied globally without specific localization, classify as NEI. This includes:
  - **Government Institutions Without Jurisdiction:**
    - Generic terms like "الحكومة", "البرلمان", "المجلس النيابي" without country specification
    - Administrative bodies like "وزارة التعليم", "وزارة الصحة" without national context
    - Regulatory agencies without jurisdictional scope
  - **International Organizations Without Time/Branch:**
    - Generic references to "الأمم المتحدة", "منظمة الصحة العالمية" without:
      - Specific time period
      - Specific branch/division
      - Specific initiative/program
  - **Economic Entities Without Market Context:**
    - Generic terms like "السوق", "البورصة", "سعر الفائدة" without:
      - Specific country/region
      - Specific timeframe
      - Specific market segment
  - **Cultural/Social Institutions:**
    - Generic references to "المجتمع", "الثقافة", "العادات" without:
      - Geographic specification
      - Cultural context
      - Historical period
  - **Educational/Scientific Entities:**
    - Generic terms like "الجامعة", "المعهد", "المختبر" without:
      - Institutional affiliation
      - Geographic location
      - Specific department/faculty

- **Implementation Rules:**
  (a) ANY mention of these entities WITHOUT proper localization triggers automatic NEI
  (b) Localization must appear in BOTH claim and evidence
  (c) Localization must be EXACT MATCH between claim and evidence
  (d) Previous or assumed context from other claims/evidence CANNOT be used for localization
  (e) The requirement for localization CANNOT be satisfied through coreference resolution alone
  (f) A score of 0 signifies that the main entity lacks sufficient information for localization and generalization, leading to an NEI judgment.

6. **World Knowledge Integration Guidelines**

- To ensure accurate reasoning and judgment, integrate world knowledge **only** in the following structured manner:
  (a) **Permitted Use: Definitions of Terms**
    – Use your knowledge **strictly to retrieve definitions of words/phrases** present in the evidence or claim.
    – Example: If the evidence mentions "التندرا" use its definition to broaden your context understanding: "منطقة حيوية حيث يعوق نمو الأشجار درجات الحرارة المنخفضة ومواسم النمو القصيرة" to interpret the claim.
    – **Example Application**:
      * **Evidence**: "كندا ذات كثافة سكانية منخفضة، حيث تهيمن الغابات والتندرا على غالبية أراضيها."
      * **Claim**: "في بعض المناطق في كندا، من الصعب أن تنمو الأشجار."
      * **Judgment**: **SUPPORTED**
      * **Reasoning**: Use the definition of "التندرا" (a biome where tree growth is hindered) to link the evidence to the claim. No external knowledge about Canada's climate is used—only the term's definition.
  (b) **Prohibited Use: Extraneous Facts or Events**
    – Do **not** use world knowledge to infer unstated facts, contradict evidence, or insert external information.
    – Example: If evidence states "كريستيانو رونالدو يلعب في دوري البيسبول الأمريكي" you must judge the claim "كريستيانو رونالدو هو لاعب بيسبول" as **SUPPORTED** even if you know he is a footballer.
    – **Failure Case**:
      * Using knowledge like *"Ronaldo is a footballer"* to refute a claim contradicts the evidence. This violates the rule against overriding evidence with external facts.
    – **Outcome**: Ensure judgments derive solely from evidence and definitions, never from unstated world knowledge.

7. **Reasoning:**

- Provide a concise explanation (3-5 sentences) of how the evidence supports or does not fully support the claim based on the coreference resolution, evidence-only verification, grammar, vocabulary, and semantic analyses.
- Highlight the key points from your analysis that most impact your final judgement.
- Avoid redundancy with the individual analysis explanations.

8. **Temporal Sequence Analysis Rules:**

- Time Order Preservation: Events must be arranged in chronological order, with clear start and end points. For example, when analyzing manufacturing processes, each step must be recorded with its exact timestamp.
- Interval Recognition: Identify both the duration of events and the gaps between them. In studying customer behavior, this means tracking both how long someone spends on a task and the time between different activities.
- Pattern Detection: Look for recurring temporal patterns and cycles. This could be daily patterns in traffic flow or seasonal trends in sales data.

- Causality Assessment: Establish whether earlier events influence later ones through statistical testing and correlation analysis.

9. **Spatial Sequence Analysis Rules**
   - Distance Measurement: Calculate and record the physical distance between sequential points or events using consistent units. This is crucial in analyzing movement patterns or distribution networks.
   - Directional Analysis: Document the direction of movement or progression between points, using compass directions or coordinate systems.
   - Boundary Definition: Clearly define the spatial boundaries of the study area and any sub-regions that might affect the sequence.
   - Spatial Clustering: Identify areas where events or points tend to group together, and analyze the relationships between these clusters.
10. **Some other Logic**
    - A person is not their roles, if a claim is something like "Tom Cruise participated in a heist in Mission Impossible 3", we cannot prove it, because Ethan Hunt did that, not Tom Cruise.
    - If someone won 5 Academy Awards, they won 3 Academy Awards. Similarly, if they won an Academy Award, they were nominated for an award.
    - If a claim says "[Person] was in [film] in 2009", then the film's release date can support it. If the claim is "[Person] acted in [film] in 2009", filming dates or release dates can prove it.
    - A list of movies that someone was in or jobs a person held is not necessarily exclusive, we cannot refute someone being a lawyer because the first sentence of their wiki article says they were an actor.
11. **Final Score Calculation:**
    - Final Score = (0.3 * Grammar Score) + (0.3 * Vocabulary Score) + (0.4 * Semantic Similarity Score)
    - If **any** of these occur:
      - Missing critical descriptor (per Semantic Similarity rules)
      - Mismatched assertion modality (per Grammar Analysis)
    - Then Final Score = max(Final Score, 70)
12. **MANDATORY EVALUATION SEQUENCE**
    (a) **Pre-Analysis Disqualifiers (MUST CHECK FIRST):**
    - Before conducting any linguistic analysis, verify these automatic NEI triggers:
      - Global Entity Ambiguity Rule violations
      - Missing jurisdictional specifications
      - Unspecified institutional references
    - If ANY of these conditions are present, immediately mark as NEI
    - Do NOT proceed to further analysis if these conditions are found
    - NO EXCEPTIONS to this rule regardless of:
      - Perfect grammar matching
      - Exact vocabulary alignment
      - Complete semantic similarity
      - Clear coreference resolution

    (b) **Analysis Priority Checklist (Must be completed in order):**
    - Step 1: Check for automatic disqualifiers
    - Step 2: If passed Step 1, proceed to entity resolution
    - Step 3: If passed Step 2, analyze grammar
    - Step 4: If passed Step 3, check vocabulary
    - Step 5: If passed Step 4, evaluate semantics

    (c) **Mandatory Double-Check Protocol:**
    - Before finalizing any "Supported" judgment, verify:
      - No automatic disqualifiers were overlooked

- Entity specifications are complete
- Jurisdictional context is explicit
- No assumptions about institutional scope

13. **Judgement:**

- A claim is **SUPPORTED** when there is complete, unambiguous alignment between the claim and evidence with no missing information or assumptions required. The following conditions must ALL be met:
  - Final Score is 88 or above
  - **Complete Information Match**
    - Complete information match between claim and evidence and coreferences
    - **All critical details mentioned in the claim must be explicitly stated or logically implied in the evidence.**
    - No external knowledge or assumptions should be required to establish the connection, you can only use the coreference resolution fields "Co_referenced_in_Evidence", "entity_in_claim" ,and "Co_referenced_in_Text"
  - **Entity Reference Clarity**
    - Clear and consistent entity references across claim, evidence, and coreference
    - All entities mentioned in the claim must be clearly and unambiguously identified in the evidence or its corresponding "Co_referenced_in_Evidence" and "Co_referenced_in_Text" field.
  - **Contextual Boundaries**
    - The scope of the claim must be explicitly defined and matched in the evidence.
    - Time periods must be clearly specified and aligned.
    - Geographic or jurisdictional boundaries must be explicit and matching.
    - Organizational hierarchies or structures must be clearly defined when relevant.
  - **Logical Relationship Clarity**
    - Cause-effect relationships must be explicitly stated.
    - Sequential events must have clear order.
    - Conditional statements must have all conditions specified.
    - Correlations or connections must be directly stated.
- **NEI (Not Enough Information)**
  - A claim is classified as **NEI** when there are gaps, ambiguities, or missing elements that prevent complete verification. Any ONE of the following conditions triggers an NEI classification:
  - **Entity Ambiguity**
    - When references could apply to multiple entities.
    - When the specific identity is unclear.
    - When pronouns have multiple possible antecedents.
    - When entity relationships are not fully specified.
  - **Scope Uncertainty**
    - When boundaries of application are not clear.
    - When universal claims lack universal evidence.
    - When temporal bounds are not specified.
    - When geographic or jurisdictional scope is unclear.
  - **Generic Reference Issue**
    - When broad terms are used without specific identification.
    - When collective nouns lack clear definition.
    - When institutional references lack specific designation.
    - When general terms require specification for clarity.
  - **Arabic-Specific Generic Reference Issues**
    - When general terms like "الحكومة", "الشعب", "الإعلام", "مؤسسة", "شركة" are used without specific identification.
    - These terms MUST be classified as NEI unless they are explicitly qualified with:

        - Specific country/jurisdiction (e.g., "الحكومة المصرية")
        - Specific department/branch plus the country jurisdiction (e.g., "وزارة التعليم المصرية")
      - The qualification must appear in BOTH claim and evidence
- **Missing Context or Details**
  - **Critical Information Gaps:**
    - When the claim includes details or events not mentioned in the evidence.
    - Examples include specific causes, sources, reports, or actions absent in the evidence.
  - **Background Information Needed:**
    - When relationships between elements are unclear.
    - When situational context is assumed but not stated.
    - When cultural or historical context is required but absent.
  - Missing descriptors that **narrow meaning**:
    - Royal/National/Regional modifiers
    - Institutional names (e.g., "وزارة التعليم" vs. "الحكومة")
    - Specific locations/dates not in evidence
- **Temporal Ambiguity**
  - When time periods are not specified.
  - When sequence of events is unclear.
  - When duration is not defined.
  - When frequency or repetition patterns are ambiguous.
- **Generalization Issues**
  - When specific instances are used to support general claims.
  - When partial evidence is used to support complete claims.
  - When sample size is insufficient for the claim's scope.
  - When extrapolation would be required.
- **Logical Gap Presence**
  - When assumptions are required to connect evidence to claim.
  - When intermediate steps are missing.
  - When cause-effect relationships are implied but not stated.
  - When correlations are assumed without evidence.
  - Assumptions about **unstated modifiers** (e.g., inferring "الملكية" from context) constitute logical gaps.
- **Precision Requirements**
  - When quantitative claims lack precise measurements.
  - When qualitative claims lack clear standards.
  - When comparative claims lack clear basis for comparison.
  - When evaluative claims lack defined criteria.

CRITICAL: Do not deviate from this output format. DO NOT include any additional text, or change any character case. Ensure the output is a proper JSON-like object. Process all the input tuples and give an output for each. Do not Leave any input without an output. Default to NEI when in Doubt.

S4. Citation and Licensing

If reusing or building upon these materials, please cite the main article and respect any licensing terms specified in the dataset or code repository.