Supplementary Material

1. Prompt Template

1.1 LLM VERIFIER PROMPT TEMPLATE

```
You are a meticulous fact-checker specializing in academic citations.
Your task is to determine if a citation generated by an AI model is
real and exists by comparing it against a list of candidate references
retrieved from a bibliographic database.

Generated Citation:
BEGIN GENERATED CITATION
{generated_citation_text}
END GENERATED CITATION

Retrieved Candidate References:
BEGIN CANDIDATES
{candidate_references_list}
END CANDIDATES

Carefully compare the generated citation's title, authors, and year
against the retrieved candidates. Based on your analysis, classify the
generated citation into one of three categories: "Valid", "Partially
Valid", or "Hallucinated".

### Output Format:
Return your response as a single JSON object with the following keys:
- "label": Your classification ("Valid", "Partially Valid", or
"Hallucinated").
- "confidence": Your confidence in the classification, as a float
between 0.0 and 1.0.
- "reasoning": A brief, one-sentence explanation for your decision.
- "best_match": The full text of the best matching candidate reference,
or an empty string if none exists.
```

---

2. Implementation Details

2.1 Models Used

Benchmark Dataset Generation**: Literature-style prompts were** issued to three frontier LLMs to generate the initial set of citations.
M**odels used:** GPT-4o, Claude 3.5 Sonnet, Llama 4 Maverick.

LLM-Assisted Verification: As described in the paper, a smaller LLM is used to compare the

generated citation text with top-k candidates.
Model: Llama-3-8B-Instru**ct**
Parameters: 8B parameters, temperature = 0.2

---

3. Bibliographic Databases and Tools

3.1 Bibliographic Databases

1. CrossRef
2. OpenAlex
3. Semantic Scholar

3.2 Fuzzy Search Techniques

1. String Similarity: Levenshtein, Jaro-Winkler
2. Term-Based Retrieval: BM25
3. Semantic Retrieval: Embedding Similarity
