Qeios

Peer Review

Review of: "Quo Vadis, Artificial Intelligence? A Neuro-Symbolic Approach to Artificial Intuition"

Jean Sébastien Dessureault[1]

1. Université du Québec à Trois-Rivières, Trois-Rivières, Canada

The proposed neuro-symbolic approach (AI^2) presents an interesting and novel idea with significant potential for simulating intuitive reasoning. This is a promising research direction that warrants further exploration. However, the manuscript in its current form lacks the depth and rigor expected for a journal publication, particularly in methodological transparency, empirical validation, and scalability considerations. The introduction is underdeveloped for a full-length paper, and several critical sections require substantial expansion to meet scientific standards.

2.1. Knowledge Representation

This section is **severely under-documented** and raises fundamental questions:

1. Graph Structure & Tools:
    - Is the knowledge graph the *unique* representational tool used? If so, why? If not, what other structures (e.g., ontologies, rule-based systems) are employed?
    - Which library/framework was used to implement the graph? NetworkX, Neo4j, RDF/OWL, or a custom solution? This must be explicitly stated for reproducibility.
    - Were pre-trained knowledge graphs (e.g., Wikidata, ConceptNet) used as a foundation? If not, why reinvent the wheel?
2. **Data Sources & Construction**:
    - Node Population: What dataset(s) were used to populate the nodes? Were they extracted from text corpora, expert-curated knowledge bases, or manually constructed examples?
    - Edge Creation: How were the factual (G_k) and intuitive (G_i) edges established?

doi.org/10.32388/SA85XB

        - If derived from text data, which NLP tools (e.g., spaCy, Stanford CoreNLP) or embedding models (e.g., Word2Vec, BERT) were used?
        - If manually annotated, what criteria were used to classify edges as "factual" vs. "intuitive"? This distinction is critical and currently unclear.
    - Scalability Concerns:
        - If the current graph relies on handpicked examples, how do you plan to scale this to real-world datasets (e.g., millions of nodes)?
        - Will crowdsourcing or automated methods (e.g., LLM-assisted annotation) be used? This must be addressed.
3. Factual vs. Intuitive Distinction:
    - How is the boundary between factual and intuitive knowledge defined?
        - Is this manually labeled by domain experts?
        - Is there an automated mechanism (e.g., confidence thresholds from NLP models) to classify edges?
    - If human annotation is required, how will this scale to large-scale graphs (e.g., 100K+ nodes)? This is a major bottleneck that must be discussed.

2.2. Conceptual Focusing

This section is **vague** and lacks technical detail:

1. Semantic Extraction:
    - How are semantic components extracted from the input query?
        - Is tokenization + POS tagging (e.g., via spaCy) used?
        - Are pre-trained embeddings leveraged? If so, which dimensions are selected, and why?
    - Embeddings are high-dimensional vectors (e.g., 300+ dimensions in Word2Vec). How do you select relevant dimensions when mapping to graph nodes? This is non-trivial and must be justified.

2.3. Candidate Path Discovery

This subsection is critically underdeveloped (only 4 lines) and requires full algorithmic transparency:

1. Pathfinding Algorithm:
    - Which shortest-path algorithm is used? Dijkstra's, A*, BFS, or a custom heuristic?
    - Are weights assigned to edges? If so, how are they determined?

2. Plausibility vs. Creativity Trade-off:
    - The claim that "shorter paths = more plausible" is oversimplified.
        - Counterexample: *"I see a horse in a field with something on its head."*
            - The shortest path might suggest *"unicorn"* (intuitive but implausible).
            - The longer, more plausible path might be *"horse → bridle → rider nearby."*
        - How does the system avoid such pitfalls?

2.4 Path Assessment and Selection :

The scoring mechanism (WI = 0.7, WP = 0.3) is arbitrary and lacks justification:

What empirical or theoretical basis supports the 0.7/0.3 weight ratio? Is this tunable, or is it a fixed hyperparameter?

3.1 Results

The empirical validation is insufficient for a journal submission:

1. Limited Query Set:
    - Only 5 test queries are evaluated. This is far too few for a robust assessment.
    - Consider submitting to a workshop or conference that accepts preliminary results before targeting a journal.
2. Data Representation Issues:
    - Figure 1 claims to represent the graph, but:
        - Is this the full graph, or a subset? If the latter, what selection criteria were used?
        - The example query *"What happens when it is sunny?"* mentions a path: sunny → day → walk → mood, but this path is not clearly visible in the figure. Clarify the visualization logic.

Declarations

Potential competing interests: No potential competing interests to declare.