Quantifying the Shadow of Doubt: Neuro-Symbolic Causal DAGs for Verifying Evidentiary Standards in Court Verdicts

Minseong Kim*

(Dated: February 3, 2026)

Abstract

The legal principle of proof beyond a reasonable doubt is the highest standard of proof, yet it remains abstract and subjective in natural language verdicts. This paper argues that this standard can be operationalized through the integration of Probabilistic Causal Directed Acyclic Graphs (DAGs) and Neuro-Symbolic AI. We propose a framework where Neural Networks extract causal claims and their associated epistemic uncertainty from text, while Causal DAGs formally test the robustness of these claims against alternative hypotheses. We demonstrate that reasonable doubt can be topologically defined as the existence of a high-probability alternative causal path that explains the harm without implicating the defendant, providing a computational metric for appellate review.

* mkimacad@gmail.com

I. INTRODUCTION

In criminal jurisprudence, establishing causality is not merely a binary exercise; it is a probabilistic one constrained by the standard of proof beyond a reasonable doubt. A verdict must not only assert a causal link between the defendant's action and the harm but must also implicitly dismantle all other plausible explanations.

However, judicial opinions often obscure the weight of competing narratives. A judge may state, "The evidence overwhelmingly points to the defendant," without formally defining the 'weight' or the structure of the 'doubt' that was dismissed. This paper refines the application of Causal Directed Acyclic Graphs (DAGs) by introducing probabilistic weights to edges, arguing that Neuro-Symbolic systems can be used to audit verdicts for compliance with the reasonable doubt standard.

II. FORMALIZING REASONABLE DOUBT IN CAUSAL TOPOLOGY

Standard causal models determine if A causes B. To address the evidentiary burden, we must expand this to: *Is the probability that A caused B sufficiently high to exclude all reasonable alternatives?*

A. The Preponderance vs. The Doubt

In civil law, the standard is 'Preponderance of the Evidence' ($P > 0.5$). In DAG terms, this implies the causal path passing through the defendant is simply more likely than not. In criminal law, 'Beyond a Reasonable Doubt' requires a threshold $\theta \approx 0.95$ (symbolically).

B. Topological Definition of Doubt

We define 'Reasonable Doubt' in a DAG as the existence of an active path π_{alt} from an exogenous variable U (unknown or third-party factor) to the Harm H, such that:

$$P(H|do(\pi_{alt})) > \epsilon$$

where ϵ is the threshold of reasonableness. If such a path exists and cannot be 'blocked' (explained away) by the evidence presented, the standard of proof has not been met.

III. THE NEURO-SYMBOLIC ARCHITECTURE

To apply this to verdict texts, we propose a system that does not just extract facts, but quantifies the uncertainty of the extraction.

A. Neural Uncertainty Quantification

Large Language Models (LLMs) are used to parse the verdict. However, instead of standard extraction, we employ Bayesian Neural Networks or Temperature-Scaled LLMs to output a tuple for every causal assertion:

$$E_{i\to j} = (\text{Source}, \text{Target}, \text{Confidence Score})$$

The **Confidence Score** reflects the ambiguity in the evidence cited. If a witness testimony is described as 'shaky' or 'inconsistent' in the verdict, the NN assigns a lower weight to the causal edge derived from that testimony.

B. Symbolic Stress-Testing

Once the Weighted DAG is constructed, the symbolic engine performs a Stress Test:

1. **Identify the Prosecution Path:** The primary chain from Defendant $\to$ Harm.
2. **Identify Defense Paths:** Any path leading to Harm that bypasses the Defendant.
3. **The 'Reasonable Doubt' Calculation:** The system calculates the Probability of Necessity (PN). If $PN <$ Threshold, the system flags the verdict as potentially failing the evidentiary standard.

IV. CASE STUDY: THE 'REASONABLE ALTERNATIVE'

Consider a simplified arson case.

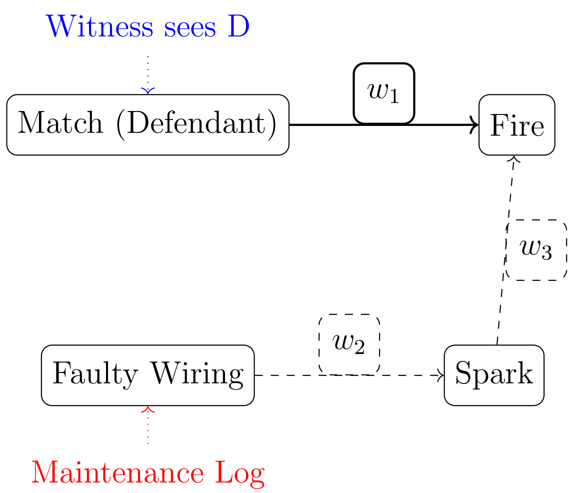
- **Prosecution:** Defendant dropped a match (M) $\to$ Fire (F).
- **Defense:** Wiring was faulty (W) $\to$ Spark (S) $\to$ Fire (F).

A. Verdict Narrative

"While the defense argued faulty wiring could be the cause, the defendant's presence at the scene leads us to conclude he started the fire."

B. DAG Analysis

A standard NLP parser might just extract $M \rightarrow F$. A Neuro-Symbolic system trained on reasonable doubt detects the dismissed hypothesis ($W \rightarrow F$).



C. Algorithmic Verification

The 'proof beyond a reasonable doubt' principle requires that the posterior probability of the Defense Path be negligible.

$$P(F|W, \text{Evidence}) \approx 0$$

If the verdict acknowledges the Maintenance Log (Ev2) confirms faulty wiring ($P(W) = 1$), but fails to explain *why* the spark didn't happen (breaking $W \rightarrow S$) or why the spark didn't cause the fire (breaking $S \rightarrow F$), the DAG reveals a structural hole in the verdict. The doubt (Faulty Wiring) remains reasonable because the path $W \rightarrow S \rightarrow F$ remains unbroken and probable ($> \epsilon$).

V. CONCLUSION

By incorporating reasonable doubt as a topological constraint, we elevate Causal DAGs from simple mapping tools to instruments of justice. This framework forces the explicit representation of alternative hypotheses. If a neural network detects that a verdict acknowledges a plausible alternative cause but fails to structurally negate it, the system can flag the decision for insufficient reasoning, ensuring that 'beyond a reasonable doubt' is a mathematical reality, not just a rhetorical phrase.