Supplementary Table S4: Canonical five-level rubric for node-scoring labels.

Level	Question-conditioned evidential role	Operational definition	Typical evidence-node categories used by the scorer	Representative soft-label value
L5	Direct answer-bearing node	Under the current question, the node directly provides the requested information and can serve as the primary answer-bearing evidence.	Explanation node for meaning questions; book/document node for “which book/document” questions; quotation/provenance node for explicit source-text questions	0.9
L4	Close semantic match	Under the current question, the node is not the canonical final answer node, but its content provides a highly similar semantic cue or a near-equivalent answer expression that strongly supports the target answer.	A semantically near-equivalent explanation node; a highly aligned interpretation cue; a closely matched alternative wording of the queried content	0.7
L3	Key provenance-support node	Under the current question, the node content does not itself constitute the final answer, but provides indispensable provenance or source-tracing support for verifying the answer.	Chapter/content-unit node, author node, dynasty/time-period node, quotation/provenance node when such information supports answer verification rather than directly constituting the target answer	0.5
L2	Auxiliary localization node	Under the current question, the node provides partial contextual, categorical, or localization support that helps narrow or organize the evidence space, but is not necessary as the main answer-bearing evidence.	Upper-/lower-level cultural content classification; secondary contextual metadata; synonym/antonym or related nodes when they are not the primary target of the current question and only provide weak auxiliary cues	0.3
L1	Irrelevant node	Under the current question, the node is unrelated to the information need, or its content does not provide meaningful support for answering the question.	Nodes from unrelated evidence chains; mismatched node types; contextually weak or off-target nodes	0.1