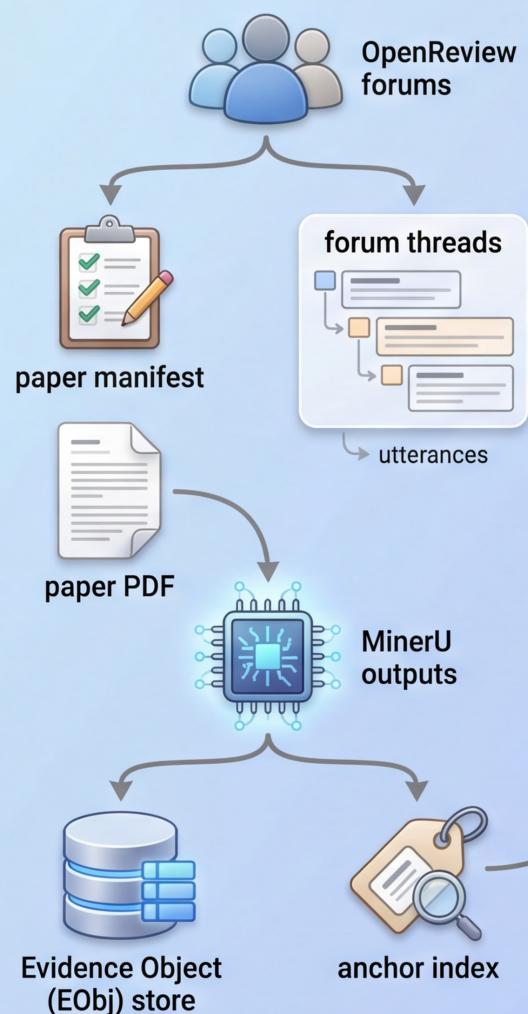


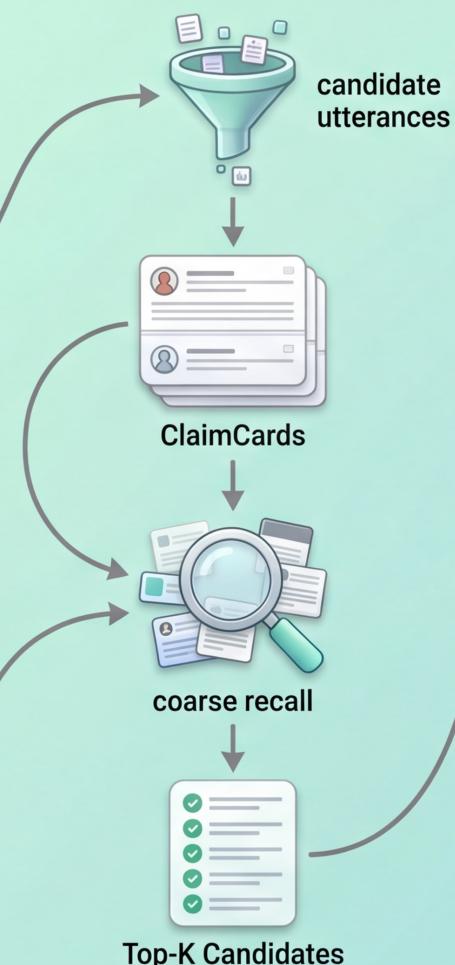
INPUTS

Step 1: Evidence-first extraction.



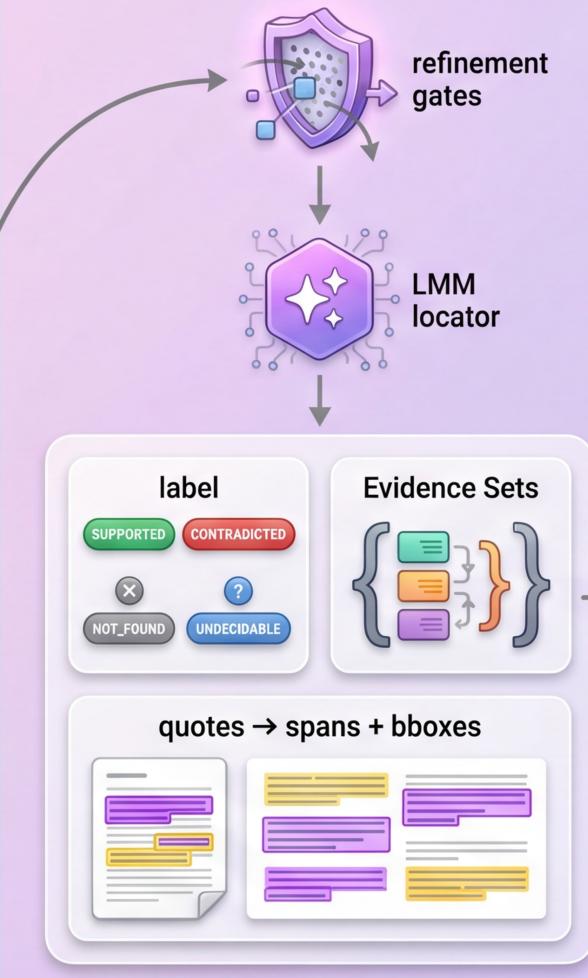
SCALE

Step 2: Scale via candidate selection and coarse recall.



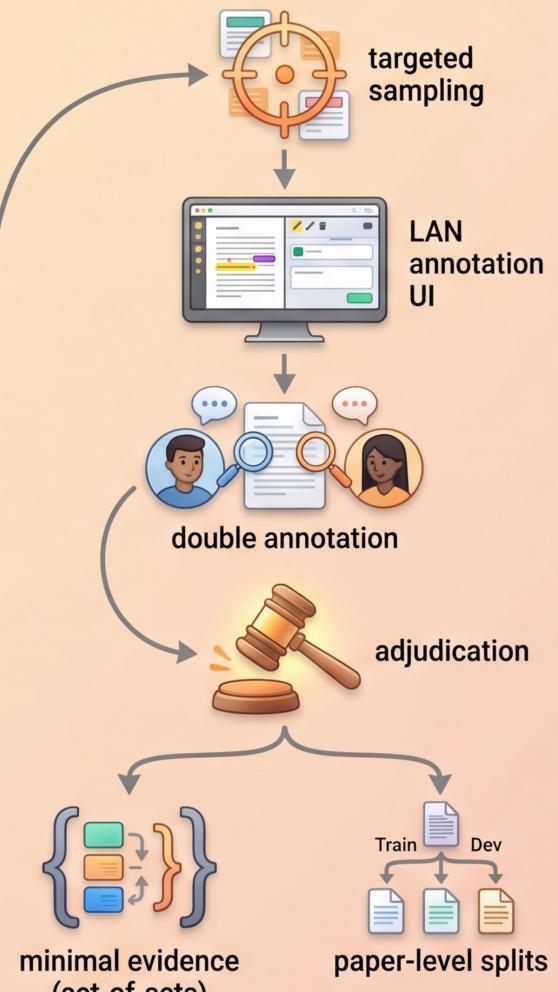
CORE

Step 3: Core via refinement gates and LMM pre-alignment.



GOLD

Step 4: Gold via targeted sampling and human adjudication.



Gold

$$\text{Scale} = \text{ClaimCards} + \text{Top-K Candidates}$$

$$\text{Core} = \text{Scale} + \text{label} + \text{Evidence Sets} + \text{spans/bboxes}$$

$$\text{Gold} = \text{Core} + \text{minimal evidence} + \text{splits}$$