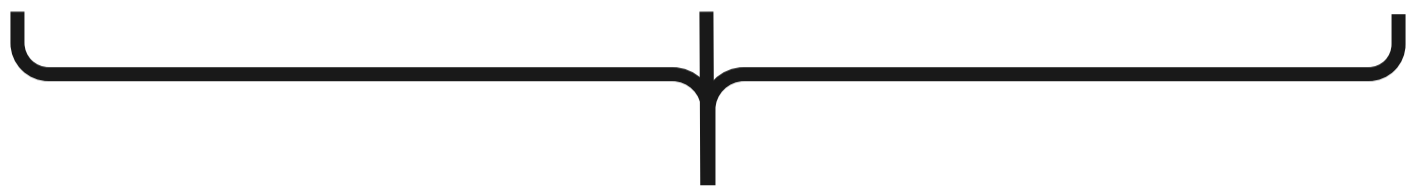


Qwen3.5-4B DFT

Qwen3.5-4B DFT
GRPO

Qwen3.5-4B DFT
Pseudo-labels



Analysis Per Document and Category



Pairwise Agreement Score

Span F1 agreement → Normalize to sum to 1.0



Cluster Overlapping Spans

Pooling spans from all models then clustering

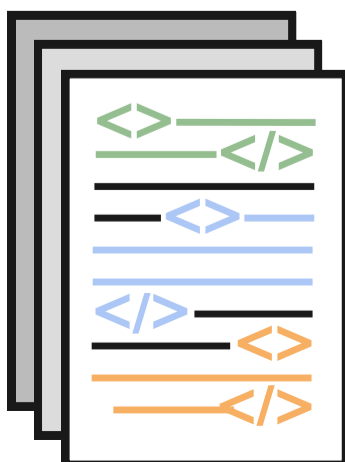


Refine Cluster Boundary

Weighted boundary vote



Deduplication & Sorting



Final Predictions