



"Extraction of questions and corresponding  
papers from OpenReview (Total: 151,000  
questions)."

#### Unfiltered Questions

QG1: Length

QG4: Granularity

QG3: Technicality

QG2: Uniqueness

Four-Step Progressive Filtering

**Probe-15k**

Questions from different sources



#### Research Paper

SOTA LLMs

Human

GPT-o3

Gemini 2.5 Pro

Qwen2.5-32B

Effort

Evidence

Groudedness

**ProbeVote-500**