

Initial Crawl ~47,000

GPT-40 Filter 3,186

Final Selection 970



eess.SP

cs.IT

2005-2025



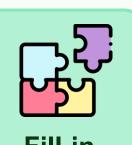
Mathematical Extraction

DeepSeek-R1 Processing

10-25 Formulas per Paper



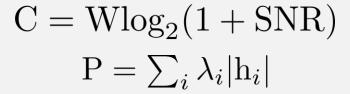
MCQ 3,800



Fill-in 11,400 (25%-50%-75%)



FEC 3,800





Quality Assurance

Dual-Layer Validation



GPT-40 Evaluation

Auto



6 Expert

Manual

Quality Threshold

≥ 3/5

22%



78%