Coverage-Accuracy Tradeoff: Selective Prediction

Accuracy (among answered questions) vs. Coverage (fraction of questions answered)

Legend:
- gpt-4o-mini / self_consistency
- gpt-4o-mini / verbalized
- qwen3:32b / verbalized
- CFA passing threshold (70%)