**Summary: Scaling Law Comparison Across Methods**

Method
- baseline
- filter_think
- filter_keep_last_think
- fiter_keep_last_tool
- long_context_filter_tool_p75
- long_context_filter_tool_p100

X axis: Round (Cost)
Y axis: Accuracy (%)