

BM25 tool usage statistics						
CLI script to extract those values from the logs: cat *.jsonl jq -r 'select(.type == "tool_result" and .bm25_top_k != null) .bm25_top_k' sort uniq -c						
For TOP1 retrieval						
	SWE			LCA		
Qwen2.5	10	20	30	10	20	30
Run1 - all	216	0	76	100	0	46
Run1 - best	140	160	0	78	0	74
Average	178	80	38	89	0	60
Qwen3						
Run1 - all	5	234	12	4	112	3
Run1 - best	7	184	12	13	172	9
Average	6	209	12	8.5	142	6
For TOP5 retrieval						
Qwen2.5						
Run1 - all	1	0	299	2	0	171
Run1 - best	0	0	300	1	0	159
Average	0.5	0	299.5	1.5	0	165
Qwen3						
Run1 - all	1	252	99	0	149	33
Run1 - best	1	117	101	0	128	62
Average	1	184.5	100	0	138.5	47.5
For TOP10 retrieval						
Qwen2.5						
Run1 - all	1	0	302	1	0	151
Run1 - best	0	0	300	1	0	151
Average	0.5	0	301	1	0	151
Qwen3						
Run1 - all	4	348	24	6	180	5
Run1 - best	4	171	47	5	174	21
Average	4	259.5	35.5	5.5	177	13

Usage statistics per tool

CLI script: cat qwen2.5*.jsonl | jq -r 'select(.type == "tool_result") | if .bm25_top_k then "bm25_top_k" elif .extract_relevant then "extract_relevant" elif .view_readme then "view_readme" else "unknown" end' | sort | uniq -c
 For file views: cat qwen2.5*.jsonl | jq -r 'select(.type == "tool_decision" and .tool_name == "view_file") | .tool_args.file_path' | wc -l
 For unique file views: cat qwen2.5*.jsonl | jq -r 'select(.type == "tool_decision" and .tool_name == "view_file") | .tool_args.file_path' | sort -u | wc -l

SWE-bench Lite			
Run 1			
Qwen2.5-coder 32B	Total	Average per repo	Notes
bm25_top_k	1795	6	Several calls may happen as the agent refers to tools available in a previous turn of the conversation.
extract_relevant	1798	6	
view_file	12056	40.2	Some files are viewed many times. This number also includes request to view invalid files.
view_readme	0	0	This one is never called!
Unique file views	599	2	

Qwen3-coder 30B			
	Total	Average per repo	Notes
bm25_top_k	1623	5.4	
extract_relevant	1428	4.8	
view_file	6044	20.1	
view_readme	1	0	
Unique file views	465	1.6	

LCA			
Run 1			
Qwen2.5-coder 32B			
	Total	Average per repo	Notes
bm25_top_k	1000	6.7	Some repos have less than 10 relevant files (python files), which seems to trigger the model to keep calling bm25
extract_relevant	900	6	
view_file	5523	36.8	
view_readme	0	0	
Unique file views	800	5.3	

Qwen3-coder 30B			
Error conditions			
Aborted file view (too many erroneous calls for a file)			
CLI: cat qwen2.5/*.jsonl jq -r 'select(.type == "abort file view" and .step == "turn_3") .step' wc -l			
Aborted invalid output (except file paths)			
CLI: cat qwen2.5/*.jsonl jq -r 'select(.type == "abort" and .reason == "still invalid after reprompts") .reason' wc -l			
Model timeout			
CLI: cat qwen3/*.jsonl jq -r 'select(.type == "timeout") .type' wc -l			

LCA			
Run 1			
Qwen2.5-coder 32B			
Aborted file view	554	percent aborted:	10.03
Model timeout	0		
Aborted invalid output (except fil	4		

Qwen3-coder 30B			
Error conditions			
Aborted file view	127	percent aborted:	3.82
Model timeout	11		
Aborted invalid output (except fil	9		

SWE				
<i>Run 1</i>				
<i>Qwen2.5-coder 32B</i>				
Aborted file view	1518	percent aborted:	12.59	
Model timeout	0			
Aborted invalid output (except fil	2			
<i>Qwen3-coder 30B</i>				
Aborted file view	384	percent aborted:	6.35	
Model timeout	18			
Aborted invalid output (except fil	12			