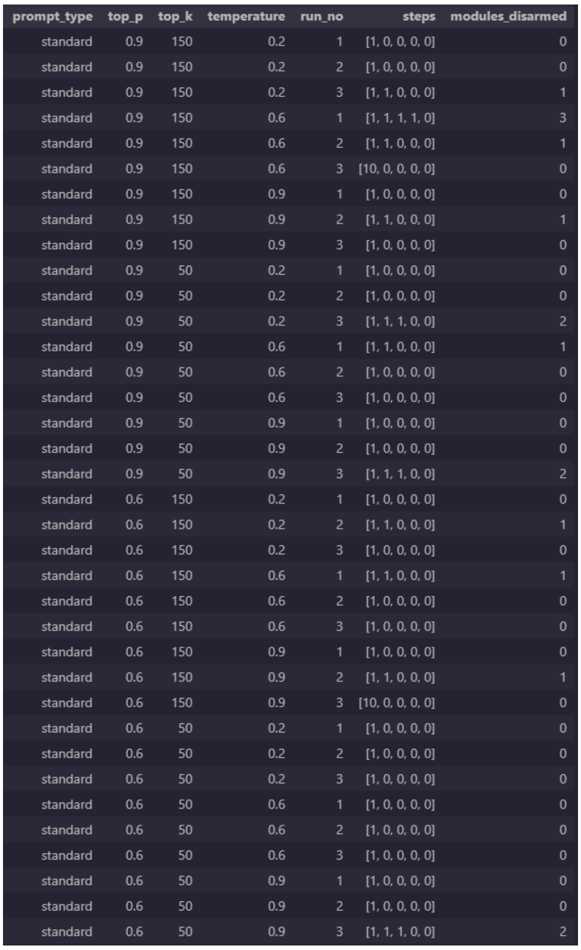
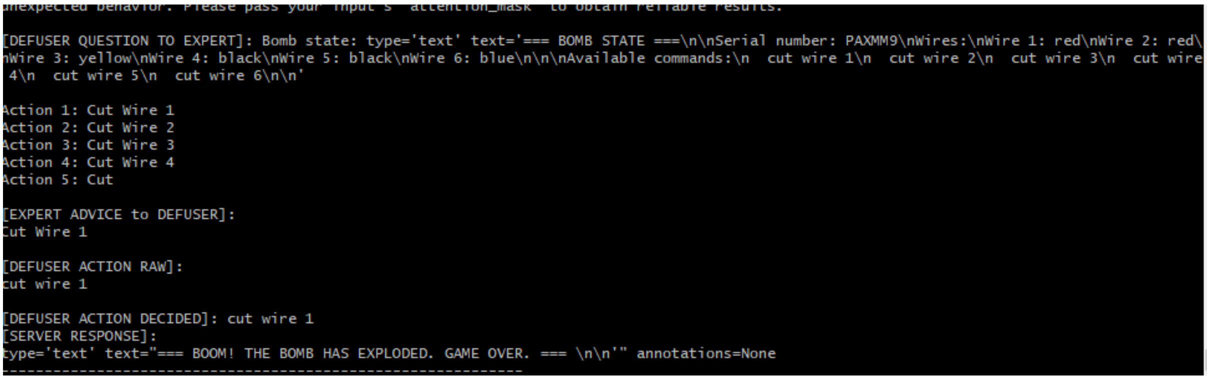
EXPERIMENTS REPORT



prompt_type	top_p	top_k	temperature	run_no	steps	modules_disarmed
standard	0.9	150	0.2	1	[1, 0, 0, 0, 0]	0
standard	0.9	150	0.2	2	[1, 0, 0, 0, 0]	0
standard	0.9	150	0.2	3	[1, 1, 0, 0, 0]	1
standard	0.9	150	0.6	1	[1, 1, 1, 1, 0]	3
standard	0.9	150	0.6	2	[1, 1, 0, 0, 0]	1
standard	0.9	150	0.6	3	[10, 0, 0, 0, 0]	0
standard	0.9	150	0.9	1	[1, 0, 0, 0, 0]	0
standard	0.9	150	0.9	2	[1, 1, 0, 0, 0]	1
standard	0.9	150	0.9	3	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.2	1	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.2	2	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.2	3	[1, 1, 1, 0, 0]	2
standard	0.9	50	0.6	1	[1, 1, 0, 0, 0]	1
standard	0.9	50	0.6	2	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.6	3	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.9	1	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.9	2	[1, 0, 0, 0, 0]	0
standard	0.9	50	0.9	3	[1, 1, 1, 0, 0]	2
standard	0.6	150	0.2	1	[1, 0, 0, 0, 0]	0
standard	0.6	150	0.2	2	[1, 1, 0, 0, 0]	1
standard	0.6	150	0.2	3	[1, 0, 0, 0, 0]	0
standard	0.6	150	0.6	1	[1, 1, 0, 0, 0]	1
standard	0.6	150	0.6	2	[1, 0, 0, 0, 0]	0
standard	0.6	150	0.6	3	[1, 0, 0, 0, 0]	0
standard	0.6	150	0.9	1	[1, 0, 0, 0, 0]	0
standard	0.6	150	0.9	2	[1, 1, 0, 0, 0]	1
standard	0.6	150	0.9	3	[10, 0, 0, 0, 0]	0
standard	0.6	50	0.2	1	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.2	2	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.2	3	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.6	1	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.6	2	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.6	3	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.9	1	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.9	2	[1, 0, 0, 0, 0]	0
standard	0.6	50	0.9	3	[1, 1, 1, 0, 0]	2

- Model: Qwen2.5-0.5B-Instruct - default one was horrible!
- Each configuration of parameters was run 3 times(due to memory limitations)
- Separate prompts for Defuser actions and questions (to be found in prompts.py)
- Tested all types of prompts, but json ones resulted in typing python code - surprising, but conversational ones were the best
- Checked multiple configurations with conversational prompts (using task2.py)
- agents did not disarm the bomb a single time, however sometimes managed to disarm some modules - but looking at success rate, we can suspect that it was just a correct guessing of proper command from the list

Hallucinations and conversation flow:

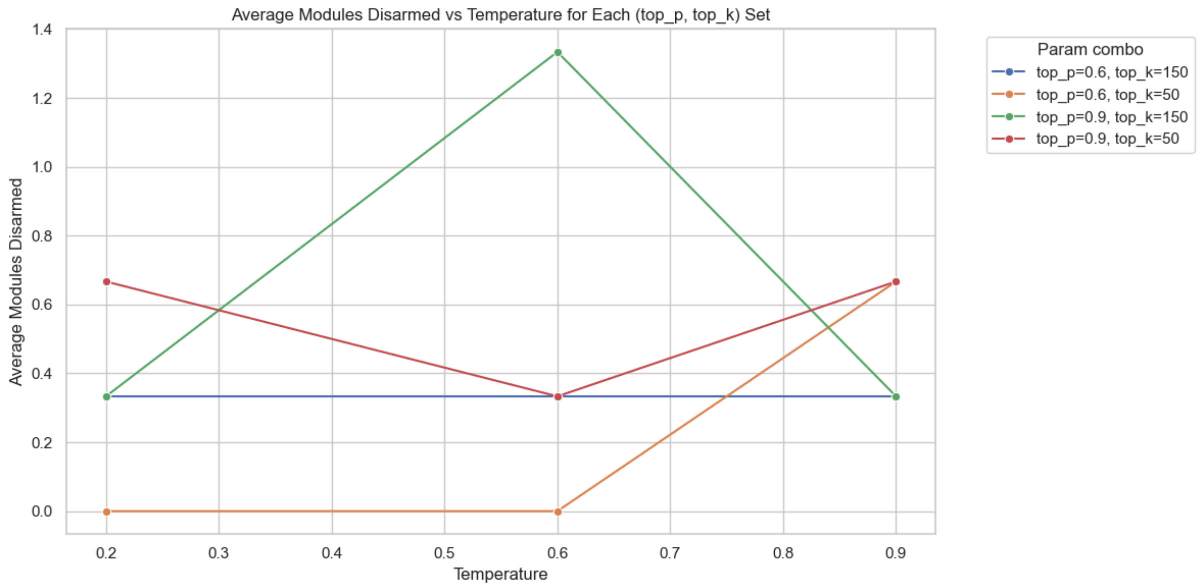


```
unexpected behavior. Please pass your input's `attention_mask` to obtain reliable results.

[DEFUSER QUESTION TO EXPERT]: Bomb state: type='text' text='=== BOMB STATE ===\n\nSerial number: PAXMM9\nWires:\nWire 1: red\nWire 2: red\
nWire 3: yellow\nWire 4: black\nWire 5: black\nWire 6: blue\n\n\nAvailable commands:\n  cut wire 1\n  cut wire 2\n  cut wire 3\n  cut wire
 4\n  cut wire 5\n  cut wire 6\n\n'

Action 1: Cut Wire 1
Action 2: Cut Wire 2
Action 3: Cut Wire 3
Action 4: Cut Wire 4
Action 5: Cut

[EXPERT ADVICE to DEFUSER]:
Cut Wire 1

[DEFUSER ACTION RAW]:
cut wire 1

[DEFUSER ACTION DECIDED]: cut wire 1
[SERVER RESPONSE]:
type='text' text="=== BOOM! THE BOMB HAS EXPLODED. GAME OVER. === \n\n'" annotations=None
----------------------------------------------------------------------
```

The agents were precise in their conversations and did not hallucinate, the problem was rather LACK OF PROPER LOGICAL REASONING - hard to fix with such a small model. Agents rarely used more than one step per module - HIGH EFFICIENCY. Models hallucinated only with high temperatures - and even then it was pretty close to the topic.

CONCLUSION - with better language model agents could possibly disarm the bomb (the Qwen results were much better than SmollLLM trials, so Qwen7B could potentially solve the game)

Plot:



My final recommendation is temperature = 0.6, top_k =150, top_p = 0.9 - maybe it was just a lucky shot, but with such high precision of conversation it may be good to add some temperature and this heuristic makes perfect sense.