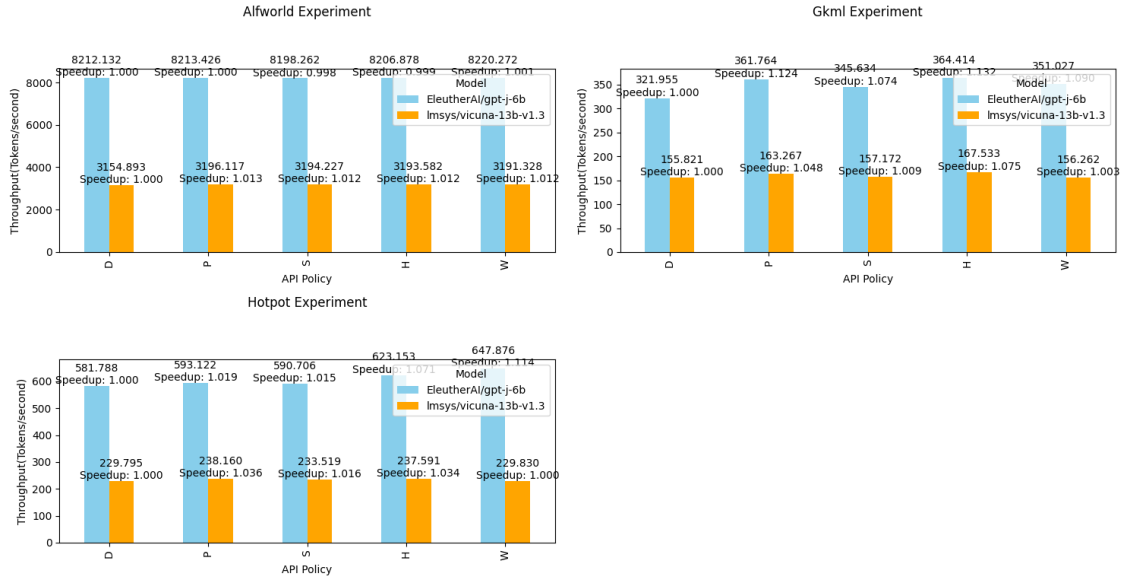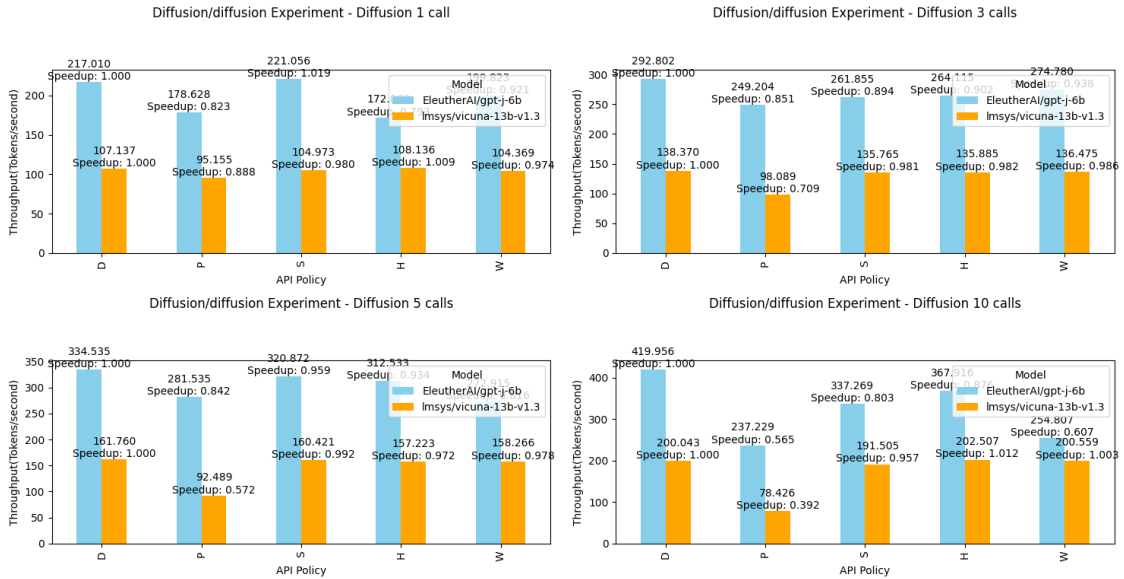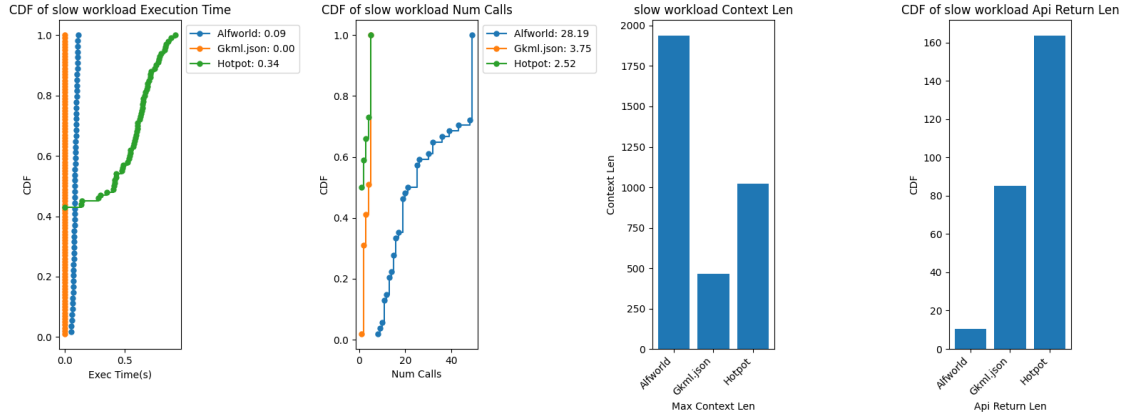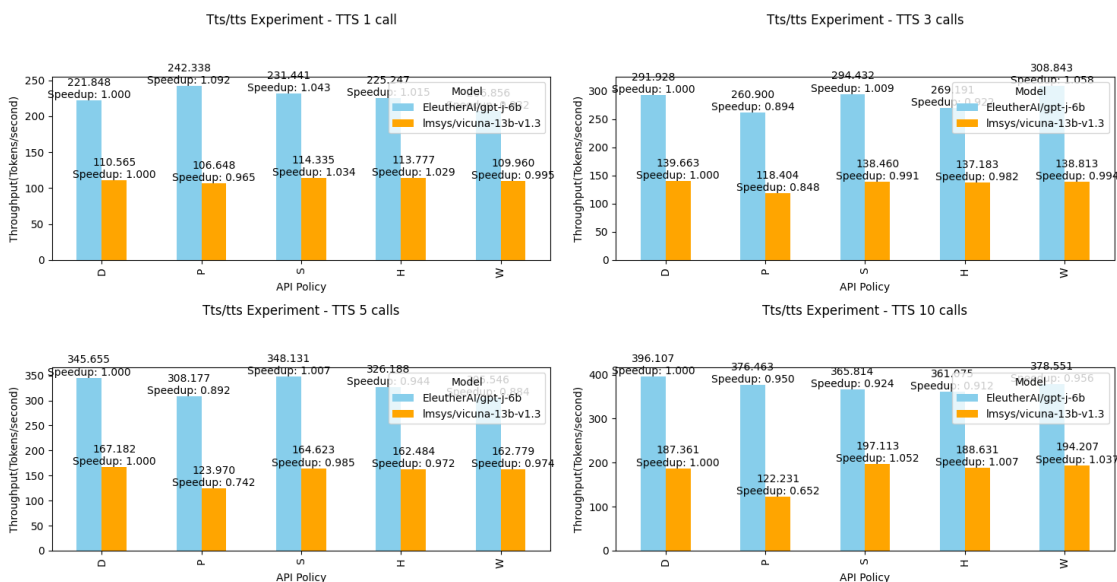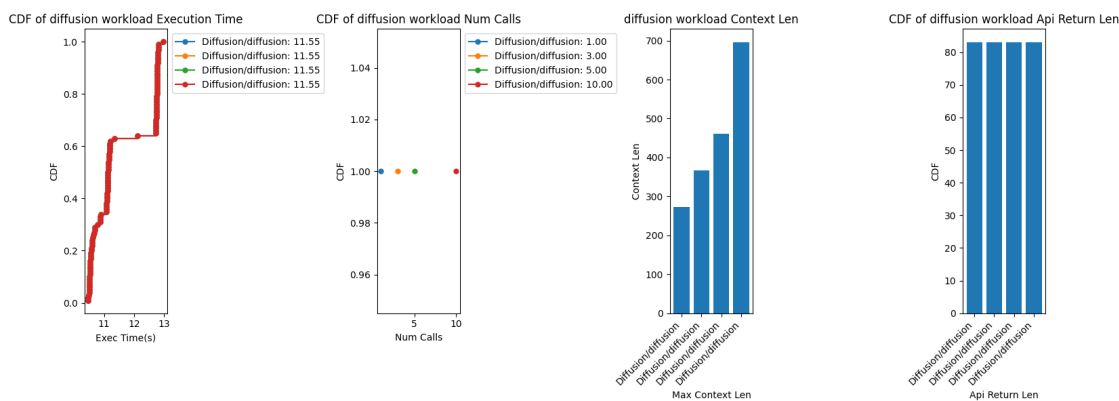# plot_over_heuristic

December 4, 2023


```
alfworld_trace.json Mean exec time 0.08521986754214392
alfworld_trace.json Mean num calls 28.185185185185187
alfworld_trace.json Mean max context len 1937.0
alfworld_trace.json Mean api return len 9.51758342719694
gkml.json Mean exec time 5.8569597342284395e-05
gkml.json Mean num calls 3.75
gkml.json Mean max context len 463.1
gkml.json Mean api return len 119.60816666666666
hotpot_qa.json Mean exec time 0.34395439928036653
hotpot_qa.json Mean num calls 2.52
hotpot_qa.json Mean max context len 1020.4
hotpot_qa.json Mean api return len 242.50683333333333
```
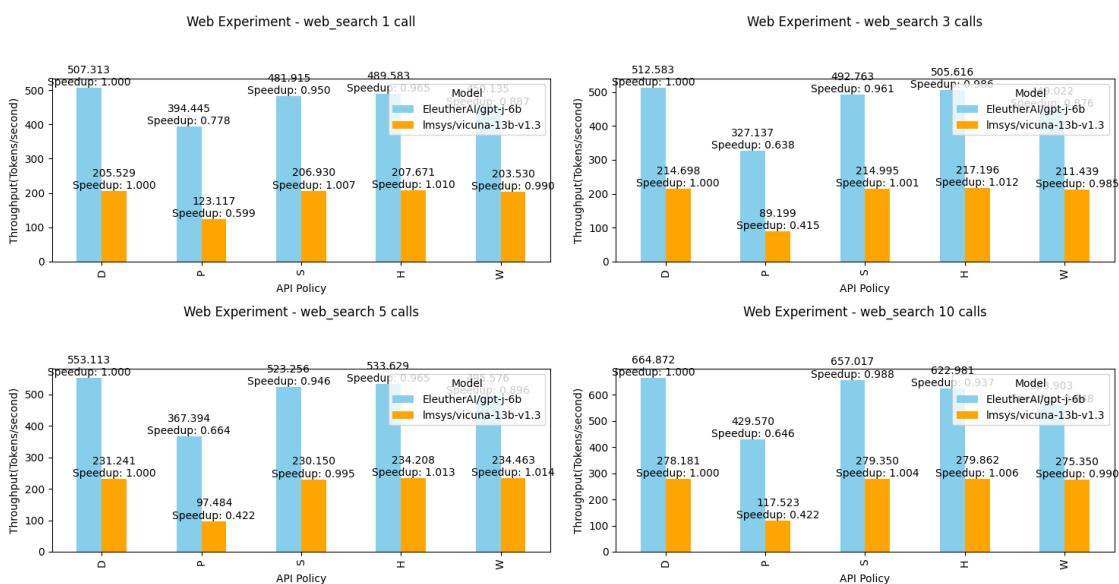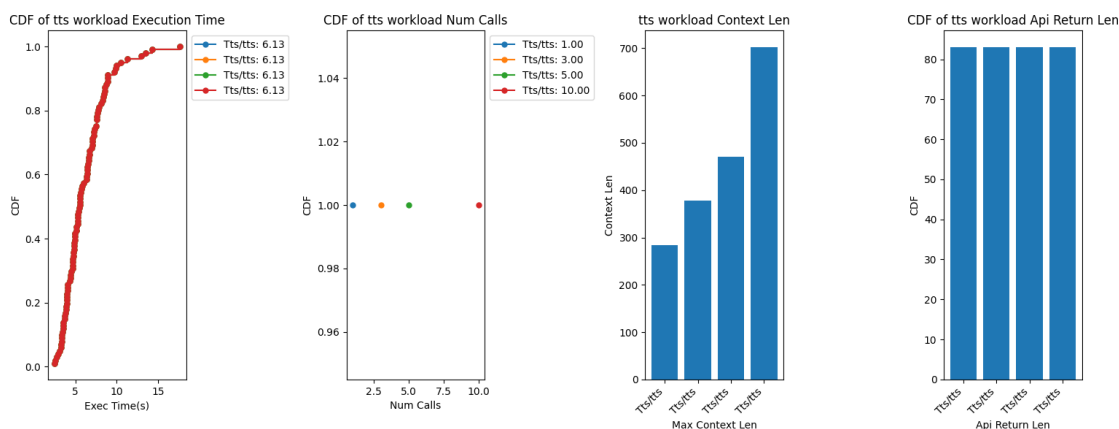
CDF of slow workload Execution Time | CDF of slow workload Num Calls | slow workload Context Len | CDF of slow workload Api Return Len

Diffusion/diffusion Experiment - Diffusion 1 call | Diffusion/diffusion Experiment - Diffusion 3 calls

Diffusion/diffusion Experiment - Diffusion 5 calls | Diffusion/diffusion Experiment - Diffusion 10 calls

```
diffusion/diffusion_1.json Mean exec time 11.552928400039672
diffusion/diffusion_1.json Mean num calls 1.0
diffusion/diffusion_1.json Mean max context len 272.03
diffusion/diffusion_1.json Mean api return len 82.35
diffusion/diffusion_3.json Mean exec time 11.552928400039672
diffusion/diffusion_3.json Mean num calls 3.0
diffusion/diffusion_3.json Mean max context len 366.09
diffusion/diffusion_3.json Mean api return len 82.35
diffusion/diffusion_5.json Mean exec time 11.552928400039672
diffusion/diffusion_5.json Mean num calls 5.0
diffusion/diffusion_5.json Mean max context len 460.15
diffusion/diffusion_5.json Mean api return len 82.35
diffusion/diffusion_10.json Mean exec time 11.552928400039672
```

```
diffusion/diffusion_10.json Mean num calls 10.0
diffusion/diffusion_10.json Mean max context len 695.3
diffusion/diffusion_10.json Mean api return len 82.35
```
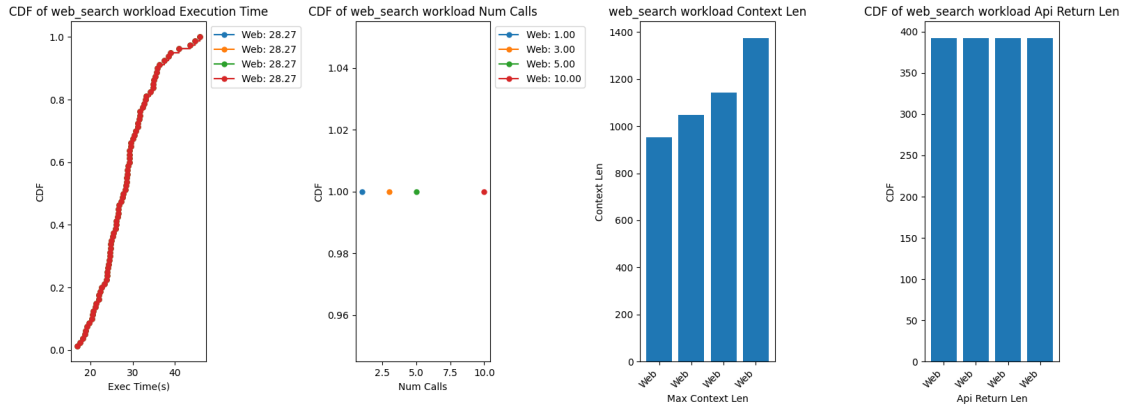


```
tts/tts_1.json Mean exec time 6.130424914973797
tts/tts_1.json Mean num calls 1.0
tts/tts_1.json Mean max context len 284.41584158415844
tts/tts_1.json Mean api return len 83.02970297029702
tts/tts_3.json Mean exec time 6.130424914973797
tts/tts_3.json Mean num calls 3.0
tts/tts_3.json Mean max context len 377.2475247524753
tts/tts_3.json Mean api return len 83.02970297029702
tts/tts_5.json Mean exec time 6.130424914973797
tts/tts_5.json Mean num calls 5.0
```
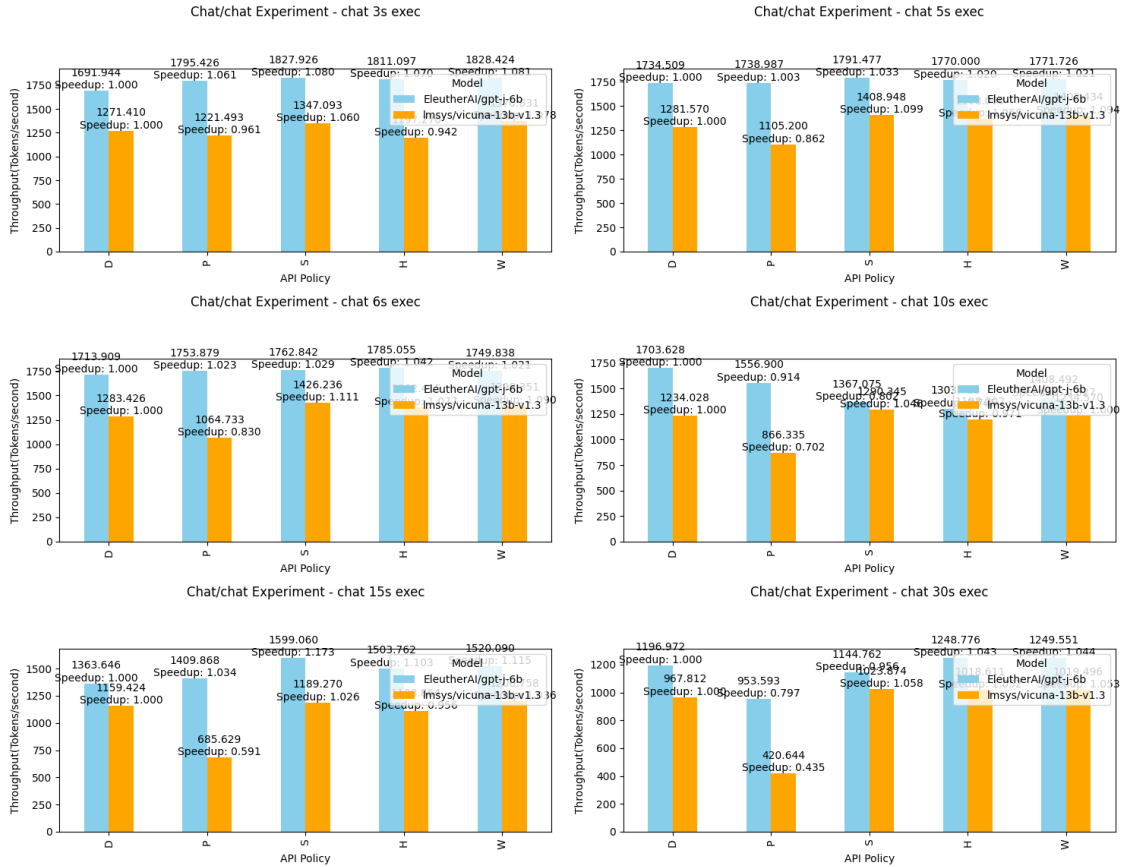
```
tts/tts_5.json Mean max context len 470.0792079207921
tts/tts_5.json Mean api return len 83.02970297029702
tts/tts_10.json Mean exec time 6.130424914973797
tts/tts_10.json Mean num calls 10.0
tts/tts_10.json Mean max context len 702.1584158415842
tts/tts_10.json Mean api return len 83.02970297029702
```



```
web_search/web_search_1.json Mean exec time 28.266159227490427
web_search/web_search_1.json Mean num calls 1.0
web_search/web_search_1.json Mean max context len 953.825
web_search/web_search_1.json Mean api return len 401.075
web_search/web_search_3.json Mean exec time 28.266159227490427
web_search/web_search_3.json Mean num calls 3.0
web_search/web_search_3.json Mean max context len 1047.475
```

```
web_search/web_search_3.json Mean api return len 401.075
web_search/web_search_5.json Mean exec time 28.266159227490427
web_search/web_search_5.json Mean num calls 5.0
web_search/web_search_5.json Mean max context len 1141.125
web_search/web_search_5.json Mean api return len 401.075
web_search/web_search_10.json Mean exec time 28.266159227490427
web_search/web_search_10.json Mean num calls 10.0
web_search/web_search_10.json Mean max context len 1375.25
web_search/web_search_10.json Mean api return len 401.075
```

Chat/chat Experiment - chat 3s exec

Chat/chat Experiment - chat 5s exec

Chat/chat Experiment - chat 6s exec

Chat/chat Experiment - chat 10s exec

Chat/chat Experiment - chat 15s exec

Chat/chat Experiment - chat 30s exec

```
chat/chat_bot_3.json_filtered.json Mean exec time 3.0200154795730856
chat/chat_bot_3.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_3.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_3.json_filtered.json Mean api return len 286.48806613756614
chat/chat_bot_5.json_filtered.json Mean exec time 4.9226500621959755
chat/chat_bot_5.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_5.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_5.json_filtered.json Mean api return len 286.48806613756614
chat/chat_bot_6.json_filtered.json Mean exec time 5.907420159634362
chat/chat_bot_6.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_6.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_6.json_filtered.json Mean api return len 286.48806613756614
chat/chat_bot_10.json_filtered.json Mean exec time 10.019736349577151
chat/chat_bot_10.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_10.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_10.json_filtered.json Mean api return len 286.48806613756614
chat/chat_bot_15.json_filtered.json Mean exec time 15.008369356333418
chat/chat_bot_15.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_15.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_15.json_filtered.json Mean api return len 286.48806613756614
```

```
chat/chat_bot_30.json_filtered.json Mean exec time 29.988225714774046
chat/chat_bot_30.json_filtered.json Mean num calls 3.2266666666666666
chat/chat_bot_30.json_filtered.json Mean max context len 417.62666666666667
chat/chat_bot_30.json_filtered.json Mean api return len 286.48806613756614
```