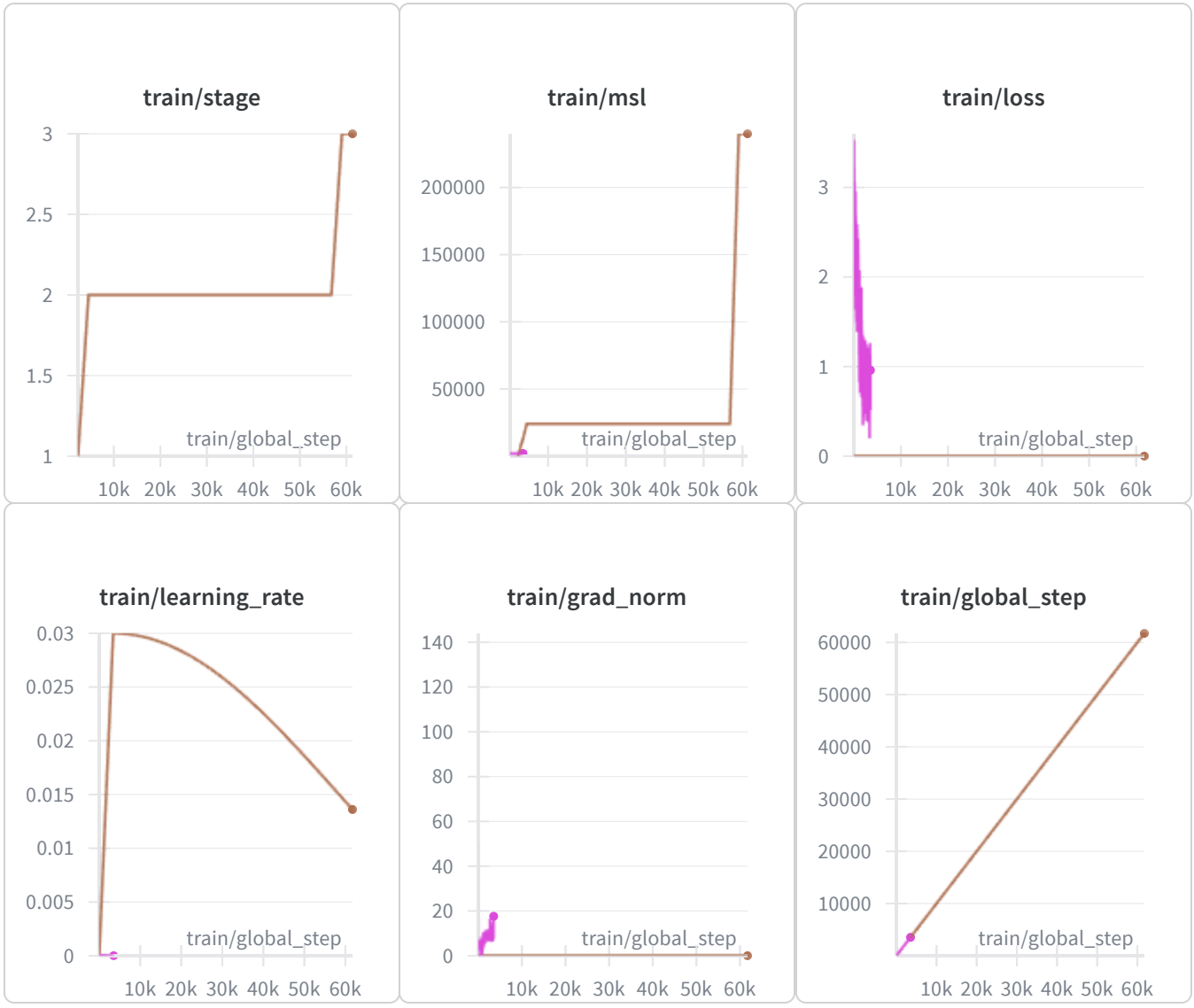
Hermes 4 14B Full Fine-Tuning

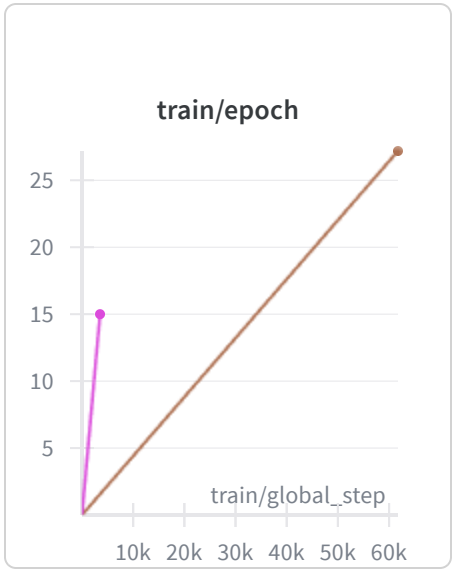
These are the training metrics from full fine-tuning of Nous Research's Hermes 14B (runs v4 and v7 noted in Research Paper)

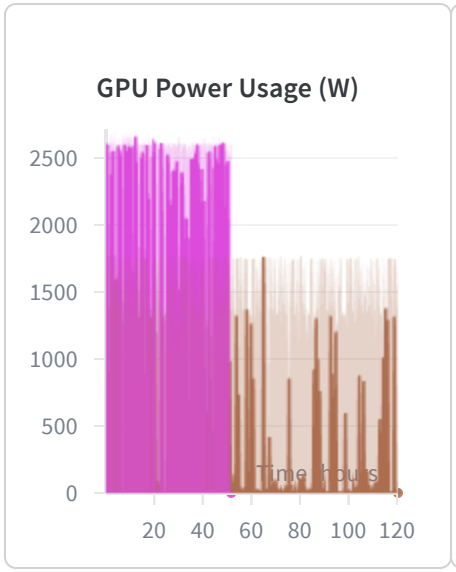
kurtis cobb

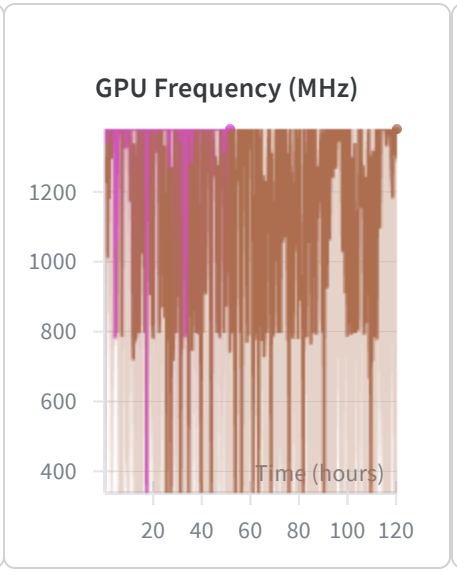
Created on January 1 | Last edited on January 1

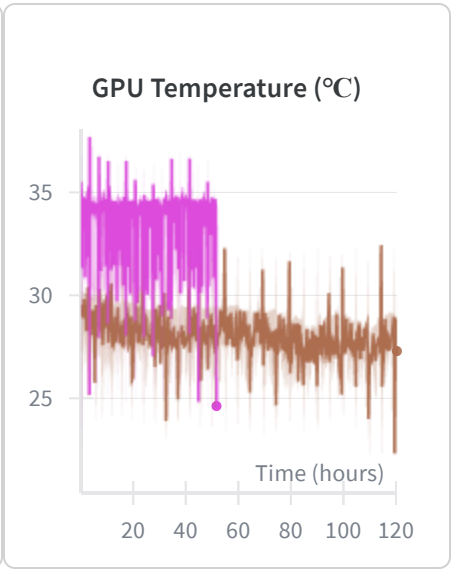
ADL Journal/Training/System Metrics


train/stage
3
2.5
2
1.5
1
train/global_step
10k 20k 30k 40k 50k 60k
train/msl
200000
150000
100000
50000
train/global_step
10k 20k 30k 40k 50k 60k
train/loss
3
2
1
0
train/global_step
10k 20k 30k 40k 50k 60k
train/learning_rate
0.03
0.025
0.02
0.015
0.01
0.005
0
train/global_step
10k 20k 30k 40k 50k 60k
train/grad_norm
140
120
100
80
60
40
20
0
train/global_step
10k 20k 30k 40k 50k 60k
train/global_step
60000
50000
40000
30000
20000
10000
train/global_step
10k 20k 30k 40k 50k 60k


train/epoch
25
20
15
10
5
train/global_step
10k 20k 30k 40k 50k 60k


GPU Power Usage (W)
2500
2000
1500
1000
500
0
Time (hours)
20 40 60 80 100 120

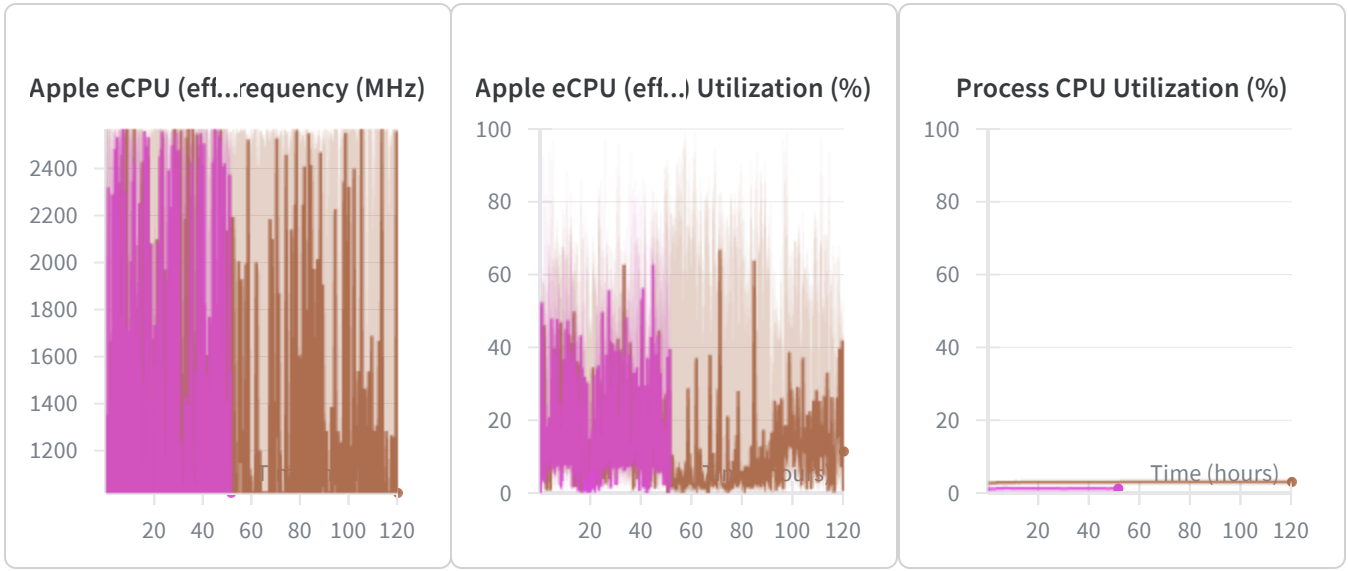

GPU Frequency (MHz)
1200
1000
800
600
400
Time (hours)
20 40 60 80 100 120


GPU Temperature (°C)
35
30
25
Time (hours)
20 40 60 80 100 120

GPU Utilization (%)
100
80
60
40
20
0
20 40 60 80 100 120
System Power Usage (W)
2500
2000
1500
1000
500
20 40 60 80 100 120
Network Traffic (Bytes)
4e+11
3e+11
2e+11
1e+11
Time (hours)
20 40 60 80 100 120
Disk I/O Utilization (MB)
2500000
2000000
1500000
1000000
500000
Time (hours)
20 40 60 80 100 120
Disk Utilization (GB)
2600
2400
2200
2000
Time (hours)
20 40 60 80 100 120
Disk Utilization (%)
100
80
60
40
20
0
Time (hours)
20 40 60 80 100 120

Process CPU Threads In Use
18
16
14
12
10
Time (hours)
20 40 60 80 100 120
Process Memory Available (MB)
400000
350000
300000
250000
200000
Time (hours)
20 40 60 80 100 120
Process Memory In Use (%)
100
80
60
40
20
0
Time (hours)
20 40 60 80 100 120
Process Memory In Use (MB)
60000
50000
40000
30000
20000
10000
0
Time (hours)
20 40 60 80 100 120
System Memory Utilization (%)
100
80
60
40
20
0
Time (hours)
20 40 60 80 100 120
CPU Power (W)
2
1
0
-1
-2
Time (hours)
20 40 60 80 100 120

Apple Neural Engine Power (W)
2
1
0
-1
-2
Time (hours)
20 40 60 80 100 120
RAM Used (%)
100
80
60
40
20
0
Time (hours)
20 40 60 80 100 120
RAM Used (Bytes)
5e+11
4e+11
3e+11
2e+11
Time (hours)
20 40 60 80 100 120
Average CPU Temperature (°C)
70
65
60
55
50
45
Time (hours)
20 40 60 80 100 120
Apple pCPU (pe...requency (MHz)
4000
3500
3000
2500
2000
1500
20 40 60 80 100 120
Apple pCPU (pe...) Utilization (%)
100
80
60
40
20
0
Time (hours)
20 40 60 80 100 120



Created with ❤️ on Weights & Biases.

https://wandb.ai/harmony-kc/hermes-mac-training/reports/Hermes-4-14B-Full-Fine-Tuning--VmlldzoxNTUxMjI2OA