

# BATTLE-TESTING SWARM

SETTING FOUNDATIONS

MIKA  
SENGHAAS

EPFL

## Next Steps

① Verify 1-GPU Baseline

NanoGPT setup  
→ (GPT-2 (140m) on  
Fine webEdv on  
AOC)

② Implement traditional PP → Minimal (20xx)  
PP Gist

③ Implement SWARM → Yandex SWARM +  
Pi Zero Band

Commits on Oct 11, 2024

Minor change	<a href="#">Verified</a>	c15d4cf		
mikasenghaas committed 4 days ago				
Implemented AMP (again, note: can only run on CUDA for simplicity for now)	<a href="#">Verified</a>	4570e8e		
mikasenghaas committed 4 days ago				
Toy examples of torch.amp	<a href="#">Verified</a>	87b684f		
mikasenghaas committed 4 days ago				
Fixed another W&B checkpointing bug + increase micro batch size for 4090	<a href="#">Verified</a>	c17e074		
mikasenghaas committed 4 days ago				
Install tmux and vim on setup	<a href="#">Verified</a>	c81bcd8		
mikasenghaas committed 4 days ago				
Added data configs (accidentally gitignored)	<a href="#">Verified</a>	da57cb1		
mikasenghaas committed 4 days ago				
Fix bug in checkpointing	<a href="#">Verified</a>	39ffba6		
mikasenghaas committed 4 days ago				
Processed + Uploaded Fineweb Edu 1BT + 100MT and wrote configs	<a href="#">Verified</a>	64af241		
mikasenghaas committed 4 days ago				
Prepare FinewebEdu	<a href="#">Verified</a>	45f4dd4		
mikasenghaas committed 4 days ago				
Modularised configs	<a href="#">Verified</a>	3ca2180		
mikasenghaas committed 5 days ago				
Initialised and pushed fresh Llama 2 14M, GPT-2 124M, and Llama 3.2 1B to HF Hub	<a href="#">Verified</a>	afcc357		
mikasenghaas committed 5 days ago				
Nicer plots	<a href="#">Verified</a>	22b19c8		
mikasenghaas committed 5 days ago				

Commits on Oct 10, 2024

Added performance benchmark results	<a href="#">Verified</a>	d08bbaa8		
mikasenghaas committed 5 days ago				
Removed redundant logging	<a href="#">Verified</a>	74866d7		
mikasenghaas committed 5 days ago				
Make port argument optional in setup (A100/H100)	<a href="#">Verified</a>	719cf33		
mikasenghaas committed 5 days ago				
Increase maximal micro batch size in perf. benchmark	<a href="#">Verified</a>	cfe7517		
mikasenghaas committed 5 days ago				
Added performance benchmarking config and updated config	<a href="#">Verified</a>	64e8364		
mikasenghaas committed 5 days ago				
Show per step loss and perplexity in train progress bar	<a href="#">Verified</a>	44ccb53d		
mikasenghaas committed 5 days ago				
Fixed bug in loss accumulation (not scaled by grad accum steps)	<a href="#">Verified</a>	bccb7a7		
mikasenghaas committed 5 days ago				
Added GPU benchmark experiment script	<a href="#">Verified</a>	bd587cf		
mikasenghaas committed 5 days ago				

Commits on Oct 9, 2024

Added mixed precision training (removed autocasting from codebase for now)	<a href="#">Verified</a>	534b162		
mikasenghaas committed last week				
Updated requirements to support toml config	<a href="#">Verified</a>	93db4c2		
mikasenghaas committed last week				
Added cosine scheduler experiment and fixed small bug	<a href="#">Verified</a>	8f78b1c		
mikasenghaas committed last week				
Added grad accumulation verification experiment	<a href="#">Verified</a>	55255b2		
mikasenghaas committed last week				
Separated requirements	<a href="#">Verified</a>	9634c3d		
mikasenghaas committed last week				
Added file-based config	<a href="#">Verified</a>	409549d		
mikasenghaas committed last week				
Updated setup scripts to build in persistent dir on remote server	<a href="#">Verified</a>	1c91993		
mikasenghaas committed last week				
Use max_norm from config in gradient clipping + simplify step function	<a href="#">Verified</a>	cff1145d		
mikasenghaas committed last week				
Moved baseline config into training script (+fix pickling error upon tokenize)	<a href="#">Verified</a>	1c6fd8c		
mikasenghaas committed last week				
Updated .env.example	<a href="#">Verified</a>	e7e0e39		
mikasenghaas committed last week				
Improved auto-setup/cleanup for Prime instance	<a href="#">Verified</a>	fcb54bc		
mikasenghaas committed last week				

## - Verification Experiments

## - Scaled model

## - Scaled data

## - Misc DX improvements (reduced LOC)

# Scaled Models

The screenshot shows a dark-themed user interface for a model repository. At the top left, there is a icon of a 3D cube labeled "Models 3". To its right is a magnifying glass search icon. On the far right, a button reads "Sort: Recently updated" with a downward arrow icon. The main content area displays three model cards, each with a circular profile picture of a person, the model name, a brief description, and a "Text Generation" button.

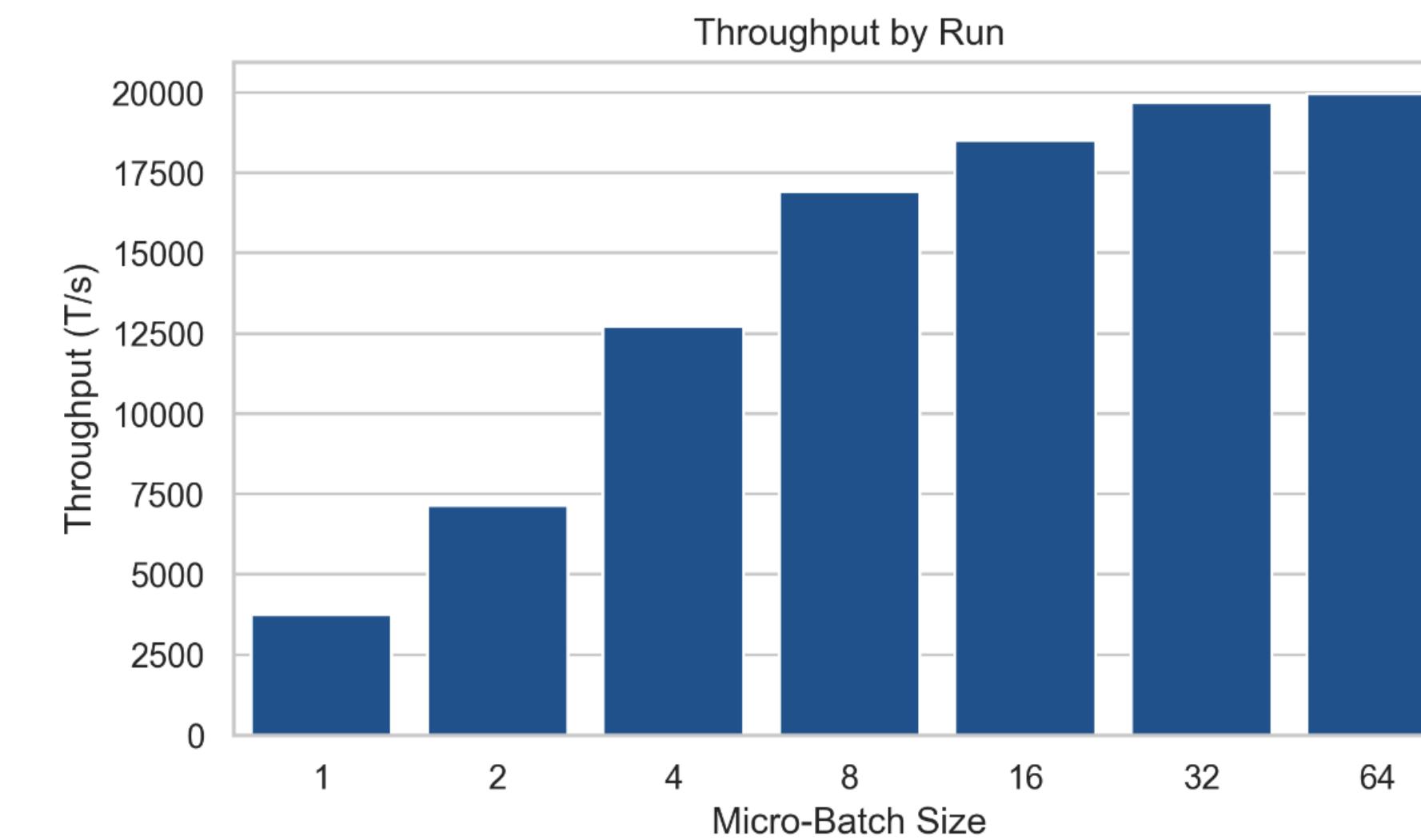
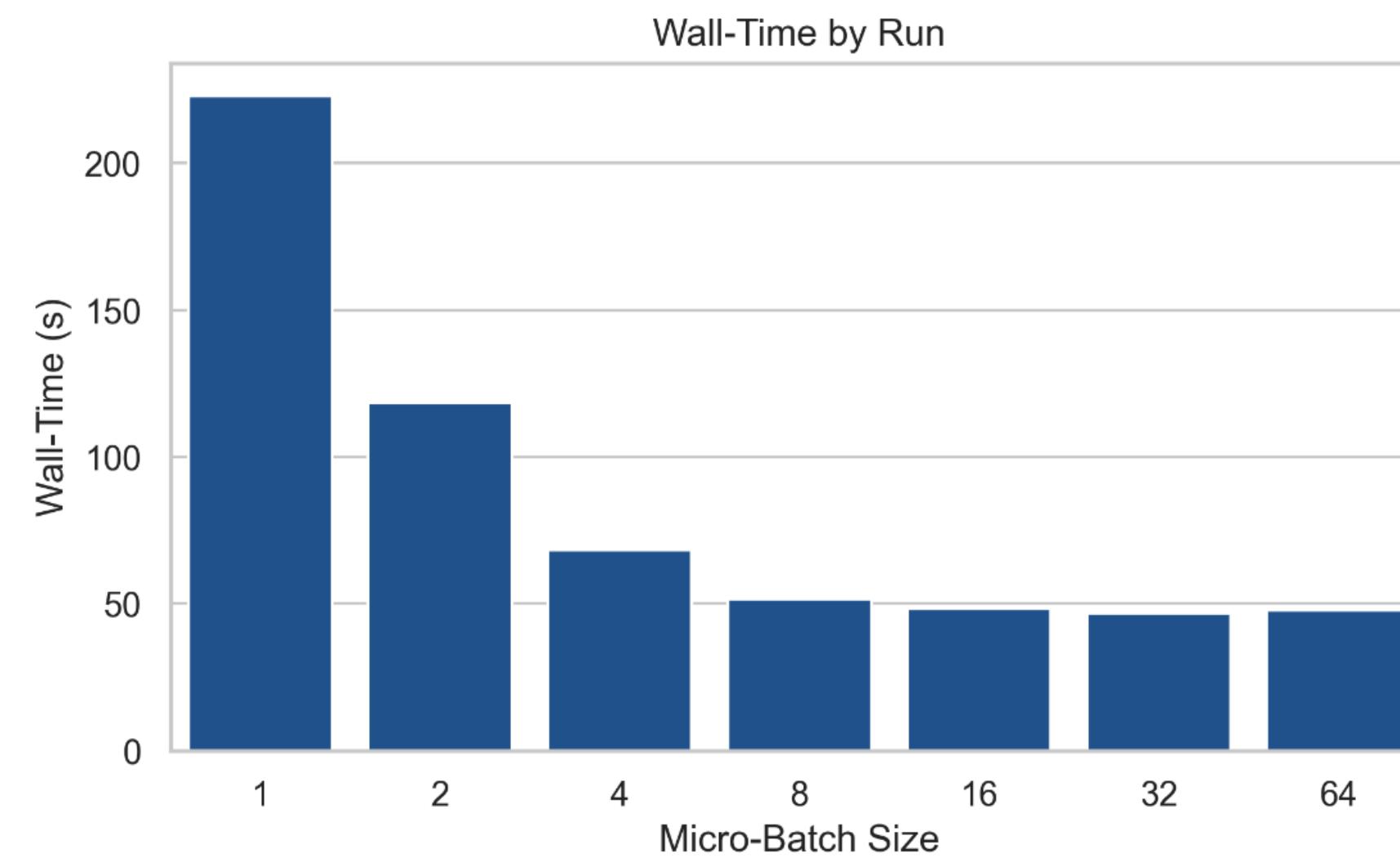
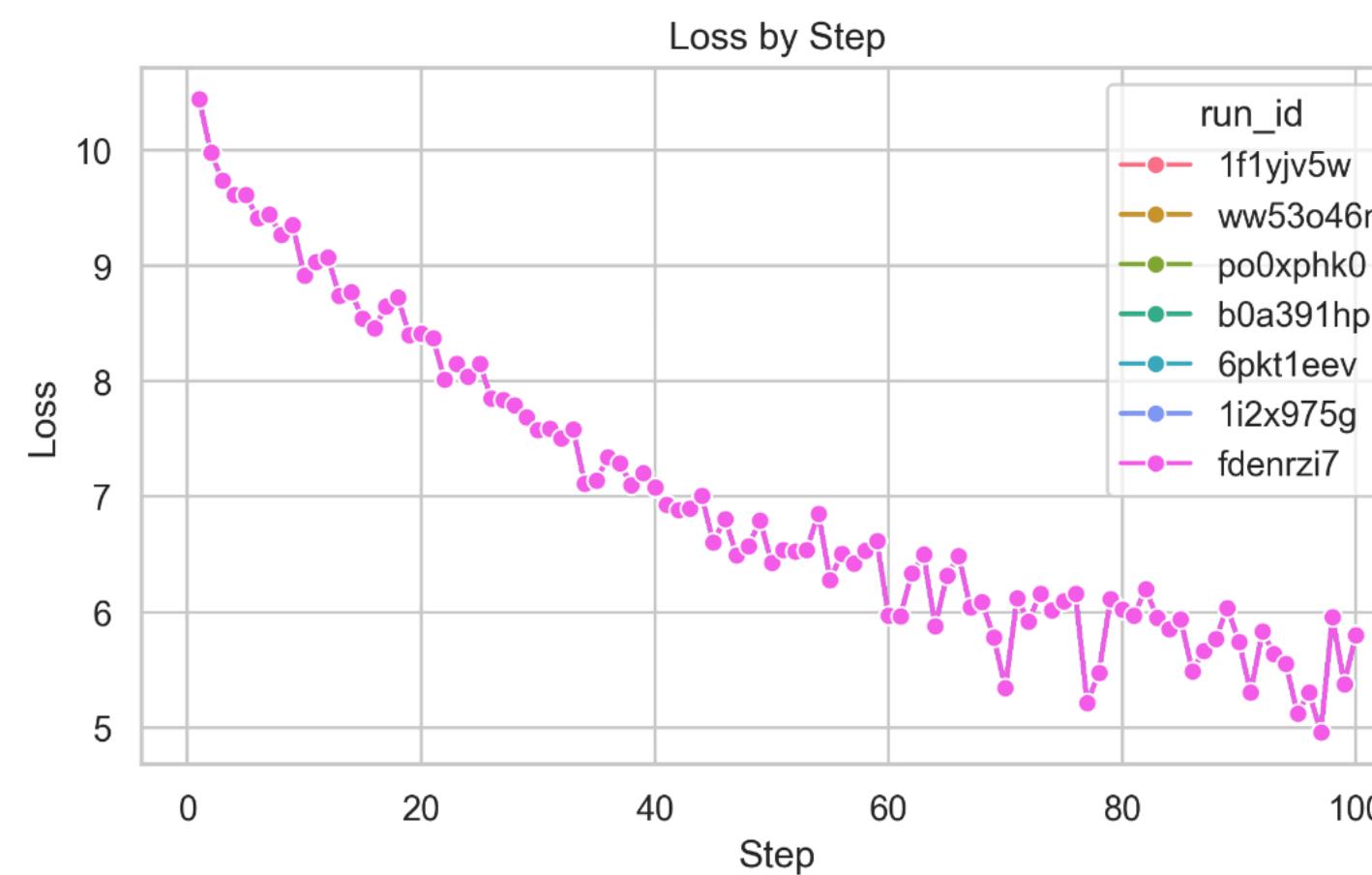
- mikasenghaas/llama32-1b-fresh**  
Text Generation • Updated 5 days ago • ↓ 23
- mikasenghaas/gpt2-124m-fresh**  
Text Generation • Updated 5 days ago • ↓ 125
- mikasenghaas/llama2-9m-fresh**  
Text Generation • Updated 5 days ago • ↓ 24

# Scaled Dataset

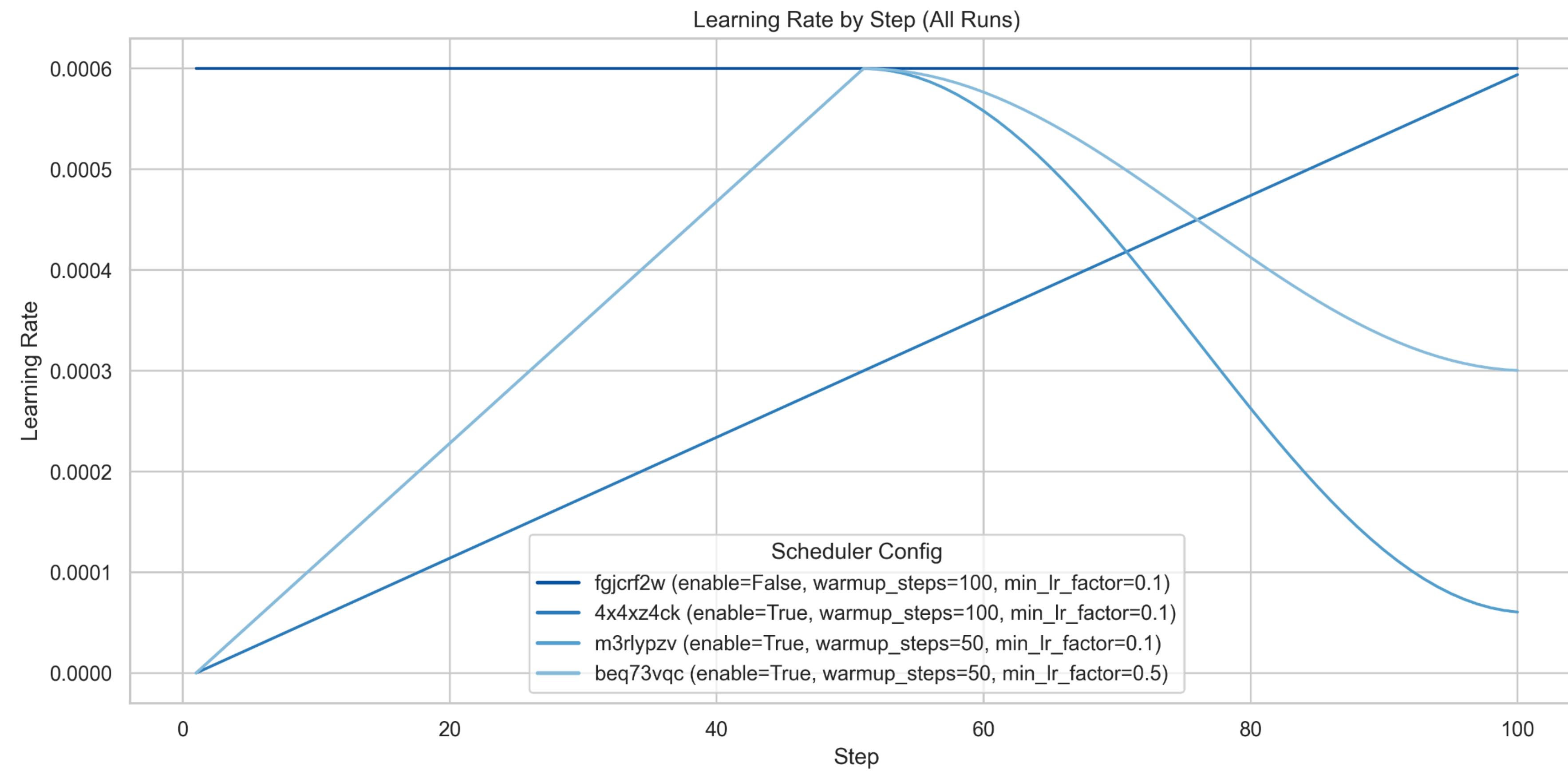
The screenshot shows a dark-themed interface for viewing datasets. At the top left, there is a circular icon with a white 'D' and the text "Datasets 3". To its right is a magnifying glass icon inside a circle. On the far right, a button says "Sort: Recently updated" with a downward arrow. The main area displays three dataset cards:

- mikasenghaas/fineweb-edu-1bt**  
Viewer • Updated 4 days ago • 1.21M
- mikasenghaas/fineweb-edu-100mt**  
Viewer • Updated 4 days ago • 121k • 53
- mikasenghaas/wikitext-2**  
Viewer • Updated 4 days ago • 21.6k • 130

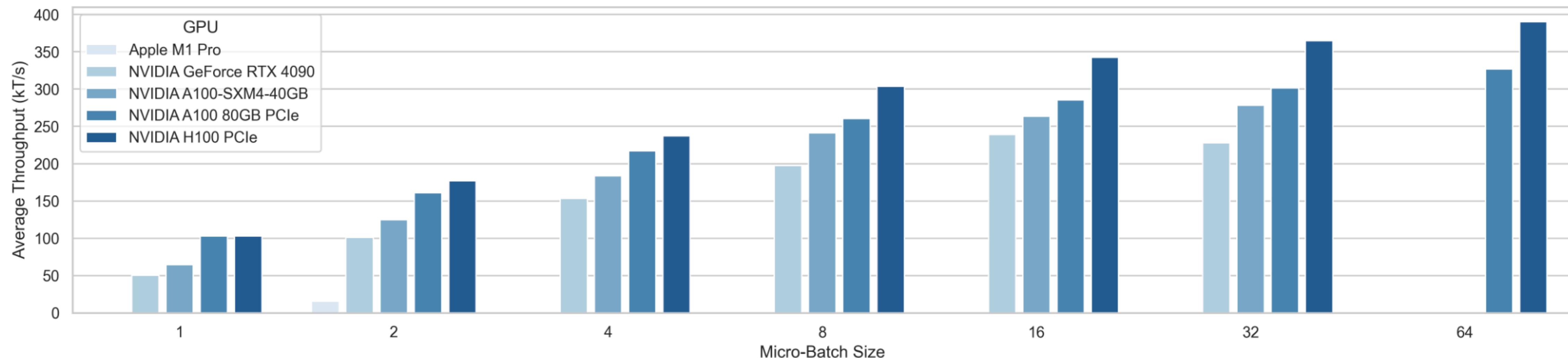
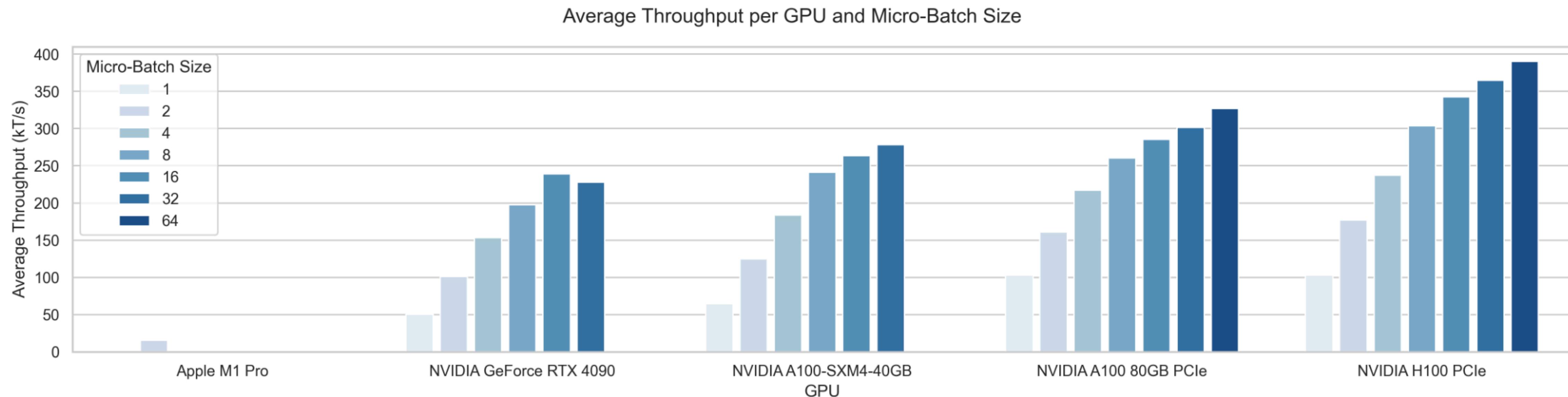
# Experiment 1: Verify Gradient Accumulation



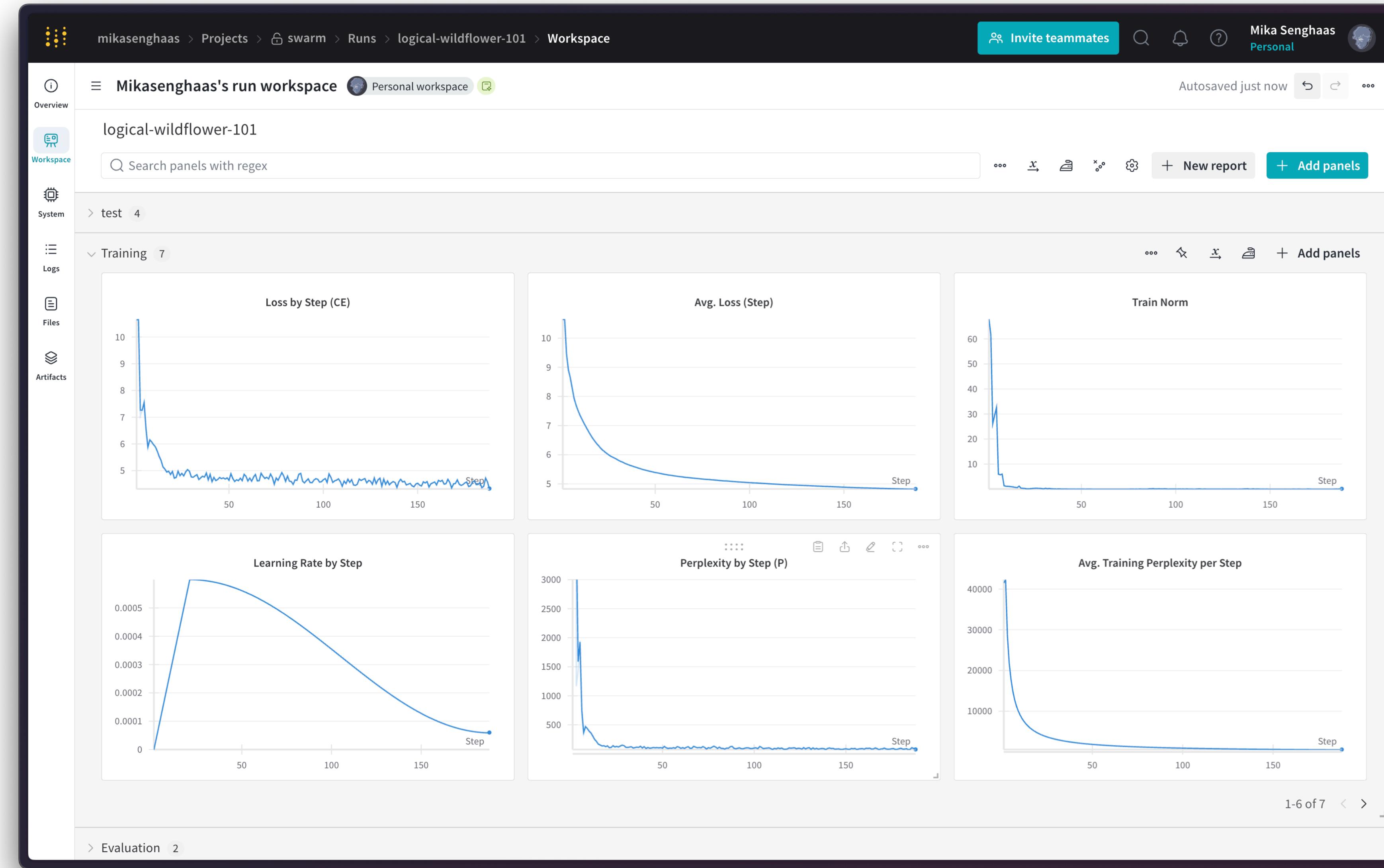
# Experiment 2: Verify LR Scheduling



# Experiment 3: Benchmarking GPUs (L=1024, B=512, Llama 2 9M)



# Experiment 4: GPT-2 Reproduce (GPT-2 (124M) + FinewebEdu 1bt)



## Next Steps

① Verify 1-GPU Baseline

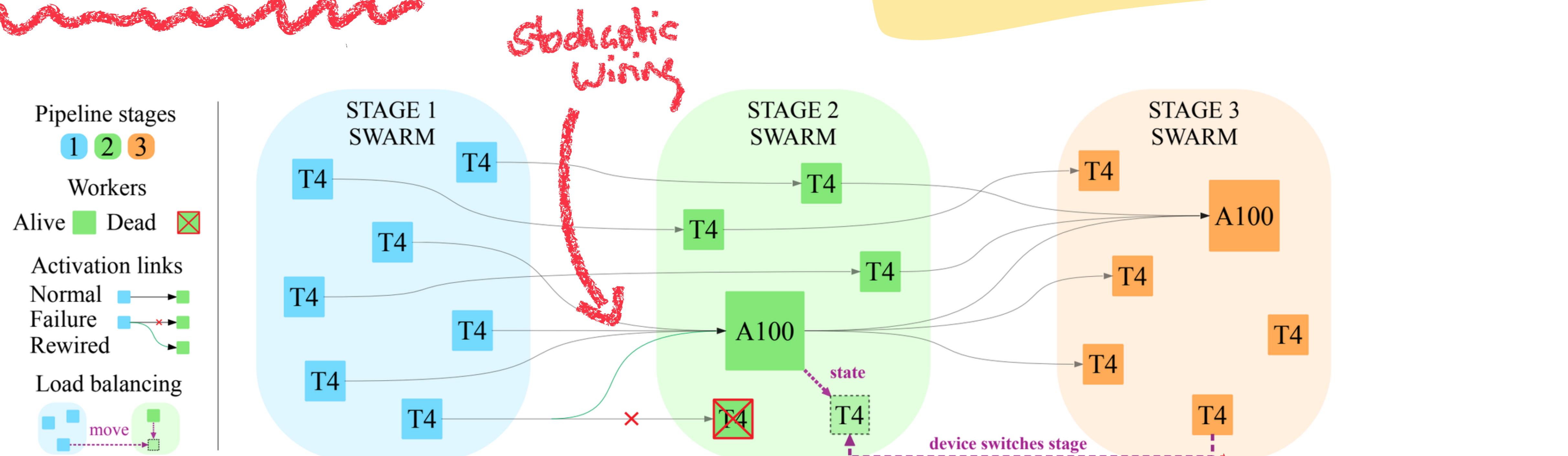
NanoGPT setup  
→ (GPT-2 (140m) on  
Fine webEdv on  
AOC)

② Implement traditional PP → Minimal (20xx)  
PP Gist

③ Implement SWARM → Yandex SWARM +  
Pi Zero Band

References

# SWARM



DYNAMIC + REDUNDANT PP

Remember?