CRUCIBLE LABS | RESEARCH

A Dive into Bittensor's Decentralized Training Subnets

Sami Kassab and Seth Bloomberg

December 9, 2024

Validator Strategy Overview

- We are conducting deep-dive analyses of different sector within the Bittensor ecosystem to:
  - Develop comprehensive expertise in each domain
  - Make more informed emission allocation decisions
  - Guide network development toward highest-impact areas
- Our first sector analysis focuses on decentralized training:
  - Foundational infrastructure layer for decentralized AI development
  - Heavy research interest and innovation pace from major AI labs and decentralized AI participants

CRUCIBLE LABS

Decentralized Training Session Agenda

- Decentralizing Training Landscape
  - The Field + Bittensor
- Categorizing Training Networks
  - Model Competitions and Distributed Training Networks
    - Overviews
    - Analysis & Thoughts
  - Emergence of a Third Category
- Takeaways

CRUCIBLE LABS

Decentralized Training Landscape

While numerous decentralized AI projects outside of the Bittensor ecosystem focus on training, only a few are operational today.
This analysis will primarily focus on pre-training networks.

Bittensor

Fine-tuning Focused:

- Dippy (SN 11)
- OMEGA Any-to-Any (SN 21)
- Fine-tuning (SN 37)
- EdgeMaxxing (SN 39)
- Gradients (SN 56)
- LogicNet (SN 35)

Pre-training Focused:

- Templar (SN 3)
- Pre-training (SN 9)
- Coldint (SN 29)
- Distributed Training (SN 38)

External

Pre-training Focused:

- Prime Intellect
- Nous Research

CRUCIBLE LABS

Two Main Types of Decentralized Training Networks

Model Competition Networks

- **Objective**: SOTA model development
- **Overview**: Miners each train and submit models for evaluation, forcing and incentivizing them to develop the best model by any means necessary
- **Projects**:
  - Bittensor Subnets: 9, 11, 21, 29, 36, 37, 39
  - External: None

Distributed Training Networks

- **Objective**: Global supercomputer capable of outperforming centralized data centers
- **Overview**: Miners communicate and work together to train a model collaboratively within a unifying training framework
- **Projects**:
  - Bittensor Subnets: 3, 38
  - External: Nous Research, Prime Intellect

CRUCIBLE LABS

Model Competition (MC) Networks: Overview

- Decentralized Kaggle competitions with continuous economic incentives
  - Serves as a novel funding mechanism for open-source model development
- Miners independently train and submit complete models
  - Open-sourcing models enables collaboration leveraging previous work
- Skill-based subnet requiring deep ML/AI expertise, enables miners to leverage proprietary knowledge and optimizations
- Acts as a permissionless R&D environment, allowing for unexpected solutions to emerge through trial and error at scale across many miners

CRUCIBLE LABS

MC Networks: Highlighting Subnet 9

- **Subnet Team: Macrocosmos**
  - Led by Will Squires (CEO), Steffan Cruz (CTO) - Both ex-OpenTensor Foundation
  - Macrocosmos is a team of 25 comprised of numerous physics and ML PhDs
  - Registered Subnet 9 on October 14, 2023
- **Subnet 9 Competitions and Current State**
  - Based on model sizes (700M, 3B, 7B, 14B is an active competition)
    - Competitions have primarily benchmarked models against a single dataset
  - Winner-take-nearly-all incentive structure
    - Very rewarding for winners, but creates high barrier to entry and high upfront capital cost with uncertain reward (6 miners on the subnet)
  - Achieved SOTA perplexity level on FineWeb-EDU2 dataset
  - Beginning to leverage additional datasets within pre-training runs
    - Should improve benchmark performance, but introduces new subnet complexities

CRUCIBLE LABS

MC Networks: Highlighting Subnet 29

- **Subnet Team: Coldint**
  - Led by RWH and μ
  - RWH has PhD in experimental quantum physics, both RWH and μ have been involved with Bittensor mining since Q1 2024
  - Registered Subnet 29 on July 13, 2024
- **Subnet 29 Competitions and Current State**
  - Codebase is a fork of Subnet 9
  - Competitions based on model types and sizes
    - Model Types = Llama, Phi; Model Sizes = 10.5B, 20.1B is active
  - More distributed reward structure than Subnet 9
    - Incentivizing broader participation and competition (16 miners on the subnet)
  - Massive innovation on validation mechanism, drastically reducing validator computational overhead
    - Validating 10.5B parameter model on RTX4090

CRUCIBLE LABS

MC Networks: Analysis & Thoughts

- **Current State**
  - Relatively mature subnet designs with well-established incentive mechanisms
  - Primary focus now on optimization, exploit prevention, expanding model size
  - Significant gap remains between subnet outputs and SOTA models against popular benchmarks
    - This is in large part due to lack of focus on fine-tuning, post-training, and training data curation
- **Key Challenges**
  - Talent bottleneck: Success heavily dependent on ML expertise in Bittensor ecosystem
  - Economic barriers: High upfront costs with uncertain returns; Individual miners bear substantial financial risk for large training runs
  - Knowledge retention: Valuable insights often remain with winning miners (SN 9)
- **Strategic Questions**
  - Should emissions be tied to absolute performance benchmarks rather than relative performance?
  - Is the current reward structure sustainable as model sizes and training costs increase?

CRUCIBLE LABS

Distributed Training (DT) Networks: Overview

- Enforces a specific distributed training framework across all miners
- Harnesses and aggregates the collective miner compute capacity
  - Participation is more accessible, as miners are simply pooling together compute resources
- Unlike MC Networks, miner coordination is paramount to success
- Removes some creative freedom and skill-based competition from miners in favor of aggregated compute capacity

CRUCIBLE LABS

DT Networks: Highlighting Subnet 3

- **Subnet Team: Templar**
  - Samuel Dare: Previously blockchain lead at OpenTensor Foundation
  - Previously co-founder of Khalani (Intents Protocol) and CTO of Akropolis (DeFi Protocols)
  - Registered Subnet 3 on October 31, 2024
- **Subnet 3 Overview and Current State**
  - Miners locally train a model with an assigned subset of the training data
  - Miners constantly upload compressed, random model slices to R2 buckets (Cloudflare servers)
    - Continuous streaming of model slices creates high communication overhead, differing from most distributed training frameworks
  - Miners pull slices as needed to remain in-sync across the network
  - Validators compute gradients locally and compare with miner outputs, rank accordingly
  - Currently training a 1.2B parameter model

CRUCIBLE LABS

DT Networks: Subnet 38

- **Subnet Team: DistributedTraining**
  - Karim Foda has 10 years of ML experience, contributor to multiple open-source
  - Mikkel Loose has 6+ experience as an AI researcher and developer, focused on LLMs and computer vision
  - Both Karim and Mikkel and have extensive Bittensor experience, with Karim joining the Bittensor community in November 2022 and contributing to multiple subnets
  - Registered Subnet 38 on September 4, 2024
- **Subnet 38 Overview and Current State**
  - Miners locally train a model, accumulating gradient changes, then perform Butterfly all reduce step
    - The Butterfly technique reduces miner communication overhead
    - Each miner only sends/receives O(d*log(n)) data instead of the typical O(d*n), where d is the size of the gradient vector and n is the number of miners on the network
  - Miners communicate P2P with other miners and validators
    - Avoids centralization concerns around Subnet 3's R2 bucket usage
  - Validators score both miner bandwidth (via proxy requests) and gradient calculations
    - Strong miner collaboration and loss minimization are both explicitly incentivized
  - Currently training a 1B parameter model

CRUCIBLE LABS

DT Networks: Outside of Bittensor

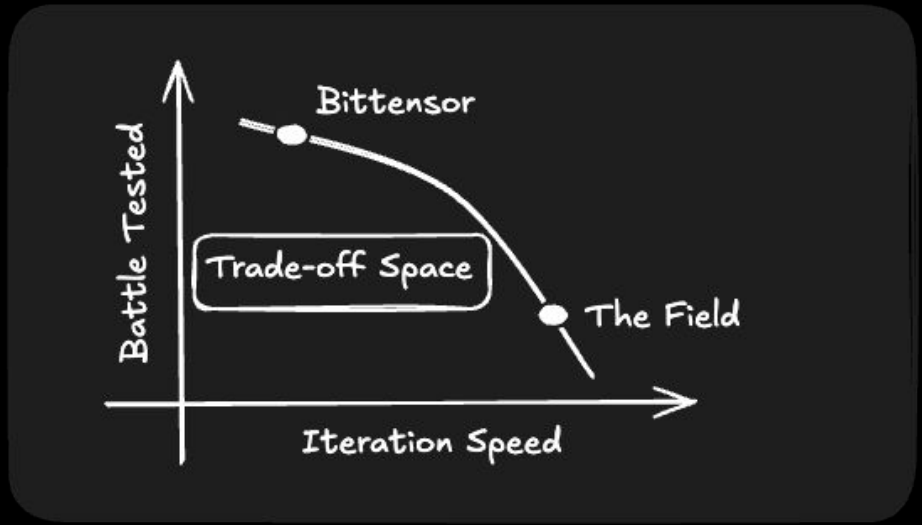
Nous Research and Prime Intellect are the only decentralized training networks in operation today

- Nous Research
  - Recently completed training of a 15B parameter model
  - DisTrO = DT Framework, DeMo = New Optimizer, Psyche = “a decentralized network that builds on Nous DisTrO to autonomously coordinate compute for model training and more”
- Prime Intellect
  - Recently completed training of a 10B parameter model
  - Not at SOTA benchmark levels, but more for large-scale PoC of their new training framework, PRIME, which is based the the Open DiLoCo framework
  - Not much public discussion yet on their network/chain

CRUCIBLE LABS

DT Networks: Bittensor vs The Field

- Bittensor Subnets
  - Incentivized
  - Permissionless Access
  - Adversarial Environment
- The Field
  - No Incentives
  - Permissioned Access
  - Controlled Environments
- Trade-off Space
  - The Field can experiment and iterate quickly on designs with little-to-no risk
  - Subnets must move more cautiously, but are immediately battle-tested and can incentivize optimizations via miner competition



CRUCIBLE LABS

DT Networks: Analysis & Thoughts

- **Current State**
  - The Field has trained larger models than DT subnets and published academic papers, which has also given them brand recognition
  - DT subnets are nascent, registered in September and late October this year
  - DT subnets architectures are likely more solid than The Field due to adversarial environment, miner competition, and validator actions
- **Key Challenges**
  - Incentive balance: Can you incentivize miner innovation while keeping a broad miner base?
  - R&D pace: Experimenting with new methods could introduce breaking changes to the subnet
  - Emissions: Similar to MCs, DTs need to scale emissions with model size
- **Strategic Questions**
  - How long should MCs and/or DTs be funded before we expect external monetization?

CRUCIBLE LABS

The Emergence of a Third Category

- A third category is emerging: on-demand training networks
  - Different than MC subnets
    - Miners are completing specific tasks, rather than participating in a competition to train a specific model
    - Product-focused. Represents a shift from R&D-centric to consumer-oriented subnets in the Bittensor ecosystem
- **Gradients** (SN 56) pioneers this approach
  - Offers a user-friendly interface for model fine-tuning, abstracting away technical complexity
  - User simply selects a model and uploads their dataset; enables anyone to fine-tune a model
- Current State & Evolution:
  - Currently operates with single-miner execution
  - Expected to evolve toward distributed training for improved performance and speed

CRUCIBLE LABS

Takeaways - Which Type of Training Network is Better?

- It's too early to determine whether MC Networks or DT Networks will ultimately produce the best open-source models. However, we lean toward Distributed Training Networks because:
  - Scalability Advantages
    - Can coordinate larger pools of compute resources
    - Better positioned to train massive models by aggregating and orchestrating distributed compute
    - Lower barrier to entry means potentially more total compute contribution
  - Economic Sustainability
    - More predictable returns for participants since rewards are based on compute contribution
    - Lower financial risk compared to MC networks where miners risk large training runs without guaranteed rewards
    - More sustainable incentive structure encourages long-term participation
- With Model Competition Networks
  - More likely to produce research breakthroughs due to permissionless R&D environment, but the outcomes and timelines are less predictable
  - Success depends on quality of participating miners

CRUCIBLE LABS

Takeaways - Investment Case

Today, we think Distributed Training Networks offer a clearer path to monetization; essentially an infrastructure play

- Distributed Training Networks
  - Could function as a rentable distributed computing cluster
  - Organizations can access large-scale compute power for model training that could be cheaper and/or faster than centralized alternatives
- Model Competition Networks
  - Entities can commision competitions for specific model development
  - Essentially crowdsourcing R&D
    - Entities pay for miners to develop models for them

CRUCIBLE LABS

Takeaways - 2025 Predictions

- Proliferation of DT networks
  - New and diverse techniques and frameworks will be explored
  - Top-tier ML teams likely to launch these subnets to crowdsource compute and experiment with novel training frameworks
- MC networks incentivizing dataset innovation
  - Data curation has proven to be instrumental in model performance, MC networks so far have been benchmarking loss minimization against single datasets
- MC and DT networks leveraging fine-tuning subnets
  - Collaborating with fine-tuning subnets will improve model performance and production-readiness

CRUCIBLE LABS

Final Thoughts

- The decentralized training sector within Bittensor has attracted exceptional talent across all levels - from subnet developers with deep ML expertise, to skilled miners pushing training boundaries
- While adversarial conditions and economic incentives initially slow subnet development compared to permissioned networks, this hostile environment ultimately produces more robust and secure training frameworks
- Looking ahead, we expect DT Networks to establish leadership in framework and incentive design across deAI. While Bittensor's adversarial, incentivized environment creates initial challenges, it ultimately produces more robust training frameworks.
- Though MCNs remain a valuable experiment, we think DTNs show greater potential for achieving SOTA results through their focus on scalable infrastructure and collaborative training.

CRUCIBLE LABS