SynthNet

Decentralized Synthetic Data Generation

A Bittensor subnet marketplace for high-quality synthetic datasets

Bittensor | Synthetic Data | Decentralized AI

THE PROBLEM

Why synthetic data matters

- AI models need large, diverse datasets — real data is scarce, expensive, or privacy-sensitive
- Centralized synthetic data providers create lock-in, opacity, and single points of failure
- Quality is hard to verify; incentives often favor volume over usefulness
- No transparent marketplace where demand meets supply with on-chain accountability

SynthNet aims to solve this with a decentralized, incentive-aligned marketplace.

THE SOLUTION

SynthNet in one sentence

SynthNet is a Bittensor subnet that acts as a **decentralized marketplace for synthetic data**: users request datasets (text, code, or conversations), miners produce them, and validators score quality on-chain so that supply, demand, and rewards align without a central operator.

Users → Request data → **Miners** → Generate datasets → **Validators** → Score quality

HOW IT WORKS

Marketplace flow

- **Request** — Users specify data type (text, code, conversations), format, volume, and constraints
- **Generate** — Miners compete to produce datasets that meet the spec
- **Validate** — Validators with permits evaluate quality and score on-chain
- **Reward** — TAO flows to miners and validators based on quality, not just volume

All scoring and emissions happen on-chain — no central authority curates or gates the market.

KEY FEATURES

Why SynthNet

Decentralized quality

Miners compete to produce the best synthetic data. Validators score and reward quality on-chain.

Any data type

Text, code, conversations — request the format you need. Multi-step forms with real-time validation.

Real-time pipeline

Watch generations in progress, inspect samples, and download verified datasets from the marketplace.

Verifiable & on-chain

Train better models with datasets whose quality is attested by the network.

NETWORK OVERVIEW

By the numbers

342
Active miners

15,847
Total datasets generated

94.2/100
Average quality score

99.8%
Network uptime

Example datasets: Customer service conversations (1,000 samples, Quality 96), Python function examples (500, 91), Product descriptions (2,500, 93).

MINER ROLE

Design & incentives

- **Generate datasets** — Miners produce synthetic data (text, code, conversations) based on user requests
- **Trust, incentive, dividends** — Metrics track miner reliability and contribution
- **Task assignment** — Design ensures miners are matched to requests they can fulfill
- **Scoring logic** — Quality evaluation mechanisms (in design and UI)

Miners are rewarded for high-quality output, not volume — the core of sustainable incentives.

VALIDATOR ROLE

Quality assurance

- **Permits** — Validators hold permits to participate in scoring
- **Score miner output** — Evaluate datasets for quality, usefulness, and spec adherence
- **On-chain consensus** — Scores and rewards flow through the Bittensor protocol
- **Aligned incentives** — Validators rewarded for accurate scoring, not gaming

Validators are the bridge between raw miner output and verified, usable datasets.

CHALLENGES & OPPORTUNITIES

Building for the long term

- **Sustainable incentives** — Reward quality over volume; prevent gaming
- **Proof of intelligence** — Define measurable effort or usefulness
- **Evaluation design** — Scoring that stays simple and on-chain
- **Adversarial behavior** — Guard against low-quality or Sybil-type attacks

Biggest blocker: *Defining incentive curves and emission rules so that TAO flow ties to measurable quality (e.g. downstream usefulness or human/validator agreement) without making emissions unsustainable or easy to game.*

GET INVOLVED

Next steps

Users: Request synthetic data — choose type, set specs, get high-quality datasets from the SynthNet subnet.

Support we need:

- Example subnet designs
- Incentive / mechanism design discussion
- Feedback on proposal drafts

SynthNet — Decentralized synthetic data generation on Bittensor.