BITTENSOR SUBNET IDEATHON PROPOSAL · 2026

MedMetric

A Decentralized Global Competition Engine that Continuously Produces
Clinical-Grade Medical Imaging AI — Locally Deployable, Zero Data Egress

> *"When a hospital uploads an MRI scan, it should be as if the world's best radiologists are competing to read it — and the winner is automatically deployed at the bedside, faster and cheaper than any single vendor can offer."*

3.6B	$450B	<4	$80K
IMAGING STUDIES/YEAR GLOBALLY	MEDICAL AI MARKET BY 2030	RADIOLOGISTS PER 1M IN LMICS	AVG. INCUMBENT COST PER SITE

Subnet Commodity: Edge-deployable medical imaging AI models
Primary modalities: Chest X-ray · Brain MRI · Bone CT
Target hardware: NVIDIA Jetson AGX Orin · Intel NUC · DICOM workstations

medmetric.subnet
Bittensor · Dynamic TAO

SECTION 01

The World Has More MRI Machines Than Radiologists to Read Them

This is not an AI-adoption problem. It is a structural supply-demand collapse with three interlocking failure modes that no single incumbent can resolve.

FAILURE MODE 1 — THE ACCESS CLIFF

WHO data shows high-income countries have ~120 radiologists per million people. Low-and-middle-income countries: fewer than 4. Yet 60% + of the world's 3.6 billion annual imaging studies are performed in these under-resourced settings. A missed lung tumor costs a life. A missed fracture costs mobility. **AI is the only scalable bridge — but current AI is priced for the wrong customer.**

FAILURE MODE 2 — THE DATA SOVEREIGNTY TRAP

HIPAA (US), GDPR (EU), and PIPL (China) impose strict constraints on cross-border medical image transmission. Every major cloud AI platform — Google Health, AWS HealthLake, Aidoc — requires data egress. This is not a policy gap. It is a **structural architectural conflict** that disqualifies cloud-first solutions for regulated hospital environments.

FAILURE MODE 3 — THE BLACK-BOX TRUST PROBLEM

Hospitals purchasing AI diagnostic systems cannot independently verify vendor performance claims. Aidoc says 94% sensitivity — on whose data? Under what augmentation? There is no open benchmark. No reproducible score. **MedMetric replaces vendor trust with on-chain verifiable scores** that any hospital can audit.

WHY DECENTRALIZATION IS THE ONLY ANSWER

A centralized AI company has one R&D roadmap, one data source, one team. MedMetric has **192 competing miners globally**, each exploring different architectures, training strategies, and augmentation pipelines — all racing toward the same objective benchmark. This is structurally superior. No single company can replicate it.

> **The core problem is not "build a medical AI model."** It is continuously producing production-grade, locally-deployable imaging analysis systems that operate at radiology workflow latency (<200 ms), generalize across devices and disease spectra, adapt when new pathologies emerge, and guarantee zero data egress compliance — at a price point accessible to a 200-bed regional hospital.

Why This Is the Right Subnet

Medical imaging analysis is structurally ideal for a Bittensor subnet for three reasons:

- **Decomposable tasks:** An imaging study can be broken into modality-specific subtasks (segmentation, classification, anomaly scoring) that map naturally onto specialized miner roles.

- **Objective ground truth:** Unlike language tasks, imaging accuracy is measurable via pixel-level overlap metrics (Dice, IoU) against expert annotations — enabling trustless, deterministic on-chain scoring.
- **Real economic demand:** Hospitals pay for these models today. Market demand becomes the dominant reward signal that steers miner development toward what clinical practice actually needs.

SECTION 02

Incentive & Mechanism Design

In MedMetric, the benchmark dataset *is* the subnet's ground truth. What the subnet optimizes is not abstract "model quality," but measurable performance against curated, rotating clinical datasets that reflect real hospital conditions. Benchmark design and evaluation methodology are therefore treated as first-class protocol components.

2.1 Emission and Reward Logic

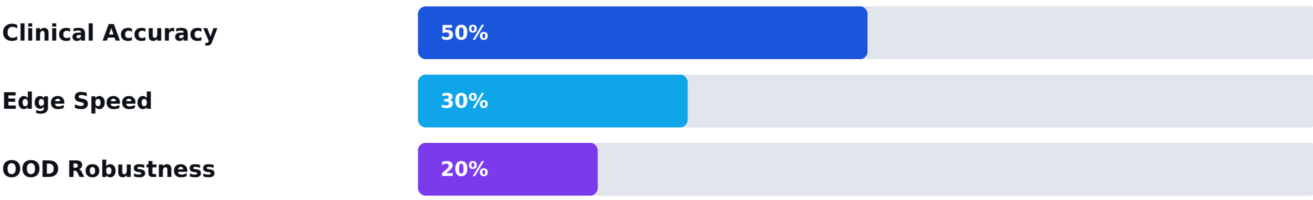
MedMetric receives TAO emissions via Dynamic TAO. Stakers vote by staking into the subnet's alpha token pool — more inflow means higher emissions. Yuma Consensus distributes arriving emissions: ~41% to miners, ~41% to validators (via bonds), 18% to the subnet owner.

Miner Reward Structure: Emissions concentrate toward top-performing miners (winner-takes-most). This reflects the real market: hospitals need the single best model for their equipment and disease mix, not a distribution of average ones. The exact reward curve (strict winner-takes-all vs. steep top-N) is calibrated during testnet based on observed miner behavior.

Each evaluation epoch, validators score miners using the composite formula:

$$\text{Score} = 0.50 \times \text{ClinicalAccuracy} + 0.30 \times \text{EdgeSpeed} + 0.20 \times \text{OOD\_Robustness}$$

Normalized into a weight vector, submitted on-chain via set_weights(). Yuma Consensus distributes emissions accordingly.



2.2 Organic Revenue (Phase 2+)

Hospital customers pay for model access (TAO or fiat → TAO). 5–10% flows to the subnet treasury; 90–95% flows to the miner whose model was purchased. This creates a **second reward signal**: benchmark scores measure technical quality (distributing emissions), while customer purchases measure market quality (distributing organic revenue). Over time, market demand becomes the dominant signal — steering the subnet toward what hospitals actually need.

2.3 Per-Image Loss Function

For each image *i* in the evaluation batch, total loss is:

$$L_i = \alpha_{seg} \cdot L_i^{seg} + \alpha_{cls} \cdot L_i^{cls}$$

SEGMENTATION LOSS — DICE + BCE HYBRID

$$L_i^{seg} = 1 - (2|\hat{M}_i \cap M_i|) / (|\hat{M}_i| + |M_i|) + BCE(\hat{M}_i, M_i)$$

$\hat{M}_i$ = predicted mask (threshold 0.5) · M_i = ground-truth mask. For normal images ($M_i = 0$), all-zero prediction scores perfectly.

Why Dice + BCE over plain BCE? Medical images are severely class-imbalanced — lesions often occupy <1% of pixels. Pure BCE would reward a model that predicts "healthy" everywhere with 99%+ accuracy. Dice loss penalizes overlap failure directly, forcing miners to actually localize pathology.

CLASSIFICATION LOSS — NORMALIZED BINARY CROSS-ENTROPY

$$L_i^{cls} = -[y_i \cdot \log(\hat{p}_i) + (1-y_i) \cdot \log(1-\hat{p}_i)] / \log(2)$$

$\hat{p}_i \in [0,1]$ is the raw sigmoid output. Division by log(2) normalizes to [0,1].

Component Weight	α_{seg}	α_{cls}	Rationale
Recommended	0.6	0.4	Segmentation quality is the clinically proximate metric
No mask head present	—	—	$L_i^{seg} = 1$ for every image; max achievable total = 0.6 × 0.5 = 0.30

2.4 Per-Image and Batch-Level Accuracy Reward

$$R_i = \alpha_{seg} \cdot (1 - L_i^{seg}) + \alpha_{cls} \cdot (1 - L_i^{cls})$$

$$\bar{R}_{acc} = (1/N) \Sigma R_i \ [i = 1 \text{ to } N]$$

Dice (macro) also computed for leaderboard display. $\bar{R}_{acc}$ drives on-chain weights.

2.5 Edge Speed Reward

Validators measure wall-clock time for a full forward pass per image, taking the mean. Reference hardware: **NVIDIA Jetson Orin NX 8 GB** — the representative ceiling for hospital edge deployment (PACS workstation or bedside inference box).

$$R_{speed} = \max(0, 1 - (\bar{t} - t_{soft}) / (t_{hard} - t_{soft}))$$

Parameter	Value	Rationale
t_soft (full score threshold)	150 ms/image	Radiologist reading workflow target
t_hard (zero speed score)	800 ms/image	Absolute clinical throughput floor
Above t_hard	R_speed = 0	Accuracy still scored; model not deployable

Because validators run the model themselves, miners cannot self-report latency or use cloud GPU acceleration to fake edge performance.

2.6 OOD Robustness Reward

Validators embed *k* augmented copies of *m* probe images in each batch. A model that genuinely learned pathology features predicts consistently across all *k* variants. A model that overfits exact pixel values will produce inconsistent outputs.

$$R\_robust = (1/m)\ \Sigma\ \mathbb{1}[\text{all k predictions for probe j agree}]$$

For models with mask head: "agree" = pairwise Dice > 0.80 across k predictions.

2.7 Total Reward

$$R\_total = 0.5 \cdot \bar{R}\_acc + 0.3 \cdot R\_speed + 0.2 \cdot R\_robust$$

Starting configuration. Expected to iterate through testnet and mainnet based on miner behavior and clinical feedback. New challenge types (3D CT reconstruction, ECG time-series, multi-organ segmentation) may add dimensions or rebalance weights.

2.8 Incentive Alignment Matrix

Desired Behavior	Mechanism
Train accurate segmentation models	Clinical Accuracy = 50% of composite score
Optimize for edge hardware deployment	Speed = 30%; hard fail above RAM/latency limits
Build robust, generalizable models	OOD Robustness = 20%; augmented probe images break overfit models
Continuously improve	Below-average miners lose UIDs to new registrants
Score honestly (validators)	Yuma Consensus clips outliers — dishonest scoring = lost emissions
Identify quality miners early (validators)	Bonds appreciate — early discovery = cheaper bonds

SECTION 03

Miner Design

A miner's single task: **produce a defect detection model that is clinically accurate, fast enough for bedside deployment, and robust to real-world imaging variability.**

3.1 How Mining Works

Register UID → **Train Model** → **Submit to IPFS** → **Record CID on-chain** → **Validator Evaluates** → **TAO Reward**

Miners submit trained model files (`ONNX` / `TensorRT` / `TFLite`) to Pinata IPFS. The validator downloads the model and evaluates it locally on Jetson Orin NX reference hardware against benchmark images. All scoring metrics (Dice, inference speed, OOD robustness) are measured directly by the validator — zero self-reporting.

3.2 Competitive Strategy Space

Strategy Axis	Available Approaches	Trade-off
Architecture	Lightweight U-Net · SegFormer-B5 · MedSAM fine-tune · EfficientDet	Speed vs. accuracy vs. zero-shot generalization
Data Strategy	Public dataset ensemble · Synthetic augmentation · Transfer learning from ImageNet / CLIP	Data diversity vs. domain specificity
Deployment Optimization	INT8/FP16 quantization · TensorRT compilation · Pruning	3–5× speed gain, small accuracy cost
Specialization	Focus on chest X-ray · Brain MRI · Bone CT; deep task-specific tuning	High score in one modality, lower breadth
Robustness Engineering	Augmentation-heavy training · Test-time augmentation · Ensemble voting	Higher OOD robustness, inference overhead

> **Open-Source First Phase:** Following Bittensor community recommendations, MedMetric launches with an open-source-first phase where early miners are encouraged to submit solutions with public code. This enables the community to observe working strategies, detect gaming patterns transparently, and iterate on the incentive mechanism based on real data.

SECTION 04

Validator Design & Anti-Gaming

A validator's task: **maintain benchmark datasets, challenge miners, score responses objectively, and set weights on-chain.**

4.1 Evaluation Loop (Each Epoch)

1. Sample N images → **2. Download miner models** → **3. Local eval on Jetson** → **4. Compute composite score** → **5. set_weights() on-chain**

4.2 Validation Dataset Construction

The validator generates a new, anti-gaming validation dataset every round. Version 1 uses **RSNA Pneumonia + CheXpert** (chest X-ray) and **BraTS 2023** (brain tumor MRI segmentation) as primary clinical sources, with **MIMIC-CXR** and **LIDC-IDRI** in the diversity pool.

Step	Operation	Formula / Detail
1. Seed Derivation	Random seed from block hash	`seed_round = hash(block_current)` — unpredictable, changes every round
2. Image Sampling	Fixed N images from source pool	`I_val = Sample(D_source, N, seed_round)`
3. Hybridization	Synthetic lesion injection on random subset	Real pathology texture transfer onto healthy tissue; rest are authentic samples
4. Augmentation Pipeline	Deterministic per-round transforms	`x̃_i = Augment(x_i, seed_round)` — affine, CLAHE, DICOM artifact simulation, Gaussian noise, random occlusion
5. Robustness Probes	k augmented variants of m probe images	`{x̃_j,1 … x̃_j,k} = Augment(x_j, seed_round, k)`

> **Key guarantee:** Sampling is verifiable by anyone (deterministic public randomness). No validator can selectively leak the sample to a favored miner — the seed is unknown until the block is produced.

4.3 Anti-Gaming Mechanisms

COPY DETECTION — DUAL-LAYER

Because validators download full model files, detection runs at two levels:

- **Output layer:** Models producing near-identical prediction vectors across multiple rounds (variance below threshold) are flagged as duplicates.
- **Weight layer:** Validators compare model architecture and parameter similarity directly. Earlier registrant or higher historical scorer is the original; duplicate is penalized.

THREAT → MITIGATION MAPPING

Attack Vector	Mitigation
`Look up BraTS/CheXpert ground truth online`	Geometric augmentation produces unique pixels every round — a lookup table cannot match transformed images
`Memorize prior-round datasets`	Seed rotates every round from block hash — different image selection and transforms each time
`Submit large accurate model ignoring edge constraints`	Deployability gate hard-fails models exceeding RAM / size / latency limits
`Use cloud GPU for inference, self-report low latency`	Validator runs the model itself on Jetson reference hardware — self-reported latency never trusted
`Skip mask head, win on classification alone`	No mask head → L_{seg} = 1 always → max achievable total = 0.6 × 0.5 = 0.30 of total score
`Miner-validator collusion on sample selection`	Block-hash seed not known until block is produced; no advance leakage possible

SECTION 05

Proof of Intelligence

MedMetric qualifies as genuine "proof of intelligence" because it satisfies five necessary conditions simultaneously — a bar that commodity compute and storage subnets cannot clear.

1. COMMODITY IS A TRAINED ML MODEL

Miners must produce genuine pathology detection capability. Submitting raw compute, storage, or random tensors earns zero score. The output is a functional, deployable clinical-grade model file.

2. SCORING IS OBJECTIVE AND DETERMINISTIC

Validators download each model and run it on the same reference hardware against the same held-out test set. Dice score produces identical results regardless of which validator runs it — fully trustless evaluation.

3. INTELLIGENCE IMPROVES OVER TIME

Miners compete to beat each other's scores. Below-average miners are deregistered. The subnet's best model quality ratchets upward each epoch — a compounding improvement loop with no ceiling.

4. INTELLIGENCE HAS REAL ECONOMIC VALUE

Hospitals pay for these models in a $450B+ market. Benchmark performance directly translates to commercial value — creating a tight feedback loop between on-chain scoring and real-world clinical utility.

5. CANNOT BE TRIVIALLY GAMED

Memorization is defeated by held-out test sets and OOD robustness probes (augmented inputs break overfitted models). Speed and accuracy metrics are fully validator-controlled with zero self-reporting. The mask-head requirement ensures models must demonstrate spatial localization, not just binary classification. No shortcut path exists to a competitive composite score.

SECTION 06

Business Logic & Market Rationale

6.1 Competitive Landscape

Type	Examples	Strengths	Limitation
Integrated hardware vendors	Philips IntelliSpace, GE Healthcare AI	Deep PACS integration, clinical validation	$200K–$800K per site; vendor lock-in; no customization
Cloud AI platforms	Aidoc, Zebra Medical, AWS HealthLake	Fast onboarding, broad disease coverage	Data egress (HIPAA/GDPR conflict); ongoing subscription; not edge-optimized
Point solutions	Nanox.AI, RapidAI (stroke)	Best-in-class for single pathology	Single disease, cannot scale, high licensing fees
Within Bittensor	**No existing medical imaging subnet**	—	**MedMetric is first-mover in this vertical**

6.2 The Irreplaceability Argument

A natural question: why not use OpenAI's medical API, or AWS HealthLake, or a HuggingFace hosted model? Three reasons this cannot be substituted:

- **Data sovereignty:** Every cloud API requires data egress. MedMetric delivers a compiled model file that runs entirely on hospital hardware. No image ever leaves the premises.
- **Continuous improvement without a vendor:** Centralized vendors improve on their own roadmap, on their own timeline. MedMetric improves every epoch, driven by global competition with no single point of control or innovation bottleneck.
- **Verifiable performance:** No cloud vendor publishes open, reproducible benchmark scores on your imaging equipment. MedMetric's scores are on-chain, deterministic, and reproducible by the hospital itself before purchase.

6.3 Monetization Path

PHASE 1 — MODEL LICENSING

$5,000–$20,000 per department (vs. $200K–$800K incumbent)

Hospital receives: compiled model file (ONNX/TensorRT), local deployment SDK, DICOM workstation integration guide, benchmark performance report. Models delivered as hardware-bound compiled binaries. Paying customers receive continuous updates.

PHASE 2 — ADD-ON SERVICES

Task-specific fine-tuning, performance bounties, custom disease challenges

Hospitals request performance tuning for specific hardware or pathology categories (pulmonary nodule, breast calcification, bone age assessment). Each adaptation billed separately — composable, modular revenue stream.

PHASE 3 — FEDERATED DATA FLYWHEEL

Cross-hospital data advantage with zero raw data sharing

Hospitals optionally contribute anonymized model updates via federated learning. Contributing hospitals receive discounted or free model access. The more hospitals participate, the better the models become — creating defensibility no centralized competitor can replicate.

6.4 Why This Transitions from Emission-Driven to Market-Driven

Phase 1 is funded primarily by TAO emissions — the subnet attracts miners through reward potential. As organic hospital revenue grows (Phase 2+), the market-revenue signal becomes the dominant steering force. Miners who optimize for benchmark scores alone will earn less than miners whose models are actually purchased by hospitals. This alignment between on-chain incentives and real-world clinical value is the subnet's long-term defensibility.

SECTION 07

Go-To-Market Strategy

7.1 Target Users & Anchor Use Cases

Early adopters: Mid-size hospitals (50–500 beds) with digitized PACS systems that are actively exploring AI augmentation but are price-blocked by incumbent solutions. Medical AI system integrators and DICOM workstation vendors who can deploy edge models on Jetson AGX hardware.

Anchor use cases: Chest X-ray pneumonia/nodule detection · Brain MRI tumor segmentation · Bone X-ray fracture localization. Focus on high-value, measurable-ROI scenarios: missed-diagnosis cost reduction and radiologist throughput amplification.

7.2 Distribution Channels

Hospital side: Partner with regional healthcare networks for pilot projects — free or discounted early access in exchange for usage feedback and de-identified outcome data. Leverage medical imaging equipment distributors and PACS integration consultants as resellers. Build a self-service marketplace where hospitals browse models by pathology type, inspect benchmark scores, and download deployment packages directly.

Miner side: Recruit through ML/AI communities (Hugging Face, Papers With Code, Kaggle medical imaging tracks). Promote public benchmark leaderboards to attract competitive medical CV researchers. Partner with academic radiology AI labs as early miner cohorts.

7.3 Cold-Start Incentive Structure

Participant	Cold-Start Incentive	Why It Works
Miners	Higher TAO reward multiplier for first 6–12 months; early miners occupy initial UID slots; access to curated clinical datasets	Attracts competitive CV researchers before organic revenue materializes
Validators	Early bond accumulation at low cost appreciates as subnet grows; delegated TAO rewards attract stakers	Ensures rigorous evaluation from day one without decentralization overhead
Hospitals	Free access for first 10–20 pilot sites; custom pathology requests for early adopters; co-branding as innovation case studies	Generates real clinical feedback and reference customers before full commercialization

Self-reinforcing flywheel: Miners and validators compete for emissions while hospitals provide real-world feedback and revenue. Better models attract more hospitals. More hospital revenue increases miner incentive to improve further. The loop has no external dependency to sustain.

SECTION 08 — EXTENSION

Phase 3+: The Federated Data Flywheel

8.1 The Core Insight

> **The fundamental limit of every centralized medical AI company:** you cannot legally obtain your competitor hospital's patient data. Medical images exist in institutional silos — legal and commercial barriers prevent centralization. In the centralized paradigm, training data diversity has a hard ceiling. MedMetric's federated learning layer breaks this ceiling.

Hospitals compute gradients locally. Only model weight updates (ΔW) are uploaded — never raw images. The global model learns from all participating hospitals without any single entity ever possessing the raw data.

8.2 Mechanism Design

CONTRIBUTION-AS-MINING

Hospitals submit locally-trained model deltas (ΔW). Validators assess their marginal contribution to global model performance. Higher contribution → more discount on future model access / priority access to updated models.

DIFFERENTIAL PRIVACY PROTECTION

All model updates receive DP noise ($\varepsilon \leq 1.0$) before upload — mathematically guaranteeing that patient data cannot be reverse-engineered from gradients. HIPAA and GDPR compliant by design.

THE FLYWHEEL EFFECT

More participating hospitals → more diverse training signal → better global model → more valuable to all participants → more hospitals join. A compounding data advantage that no single-institution AI can replicate.

DEFENSIBLE MOAT

The moat is not the algorithm — any company can copy an architecture. The moat is **cross-institutional distributed data advantage**. Once 50+ hospitals are contributing, the barrier to replication becomes structural, not technical.

8.3 Why This Cannot Be Replicated by a Centralized Competitor

Google, Microsoft, and Amazon have tried to build federated medical AI networks. All three failed to achieve hospital adoption at scale for the same reason: hospitals do not trust a for-profit cloud provider as the federated aggregator. **A neutral, decentralized subnet with on-chain verifiable aggregation logic and no single controlling entity resolves the trust problem that has blocked every prior attempt.**

CLOSING

One Sentence for the Judges

> *MedMetric is not a medical AI company. It is a decentralized R&D engine — where global GPU competition continuously produces clinical-grade imaging intelligence, delivered at one-tenth the cost of incumbents, with zero data egress, on-chain verifiable performance, and a federated data flywheel that compounds in defensibility with every hospital that joins.*

What Makes This Fundable

PRODUCT

A single, crisp digital commodity: edge-deployable medical imaging AI models. Not a platform. Not an aggregator. One thing, done continuously better by competitive pressure.

ORGANIZATION

192 competing global miners explore architectures in parallel — structural diversity no single company R&D team can match. Below-average miners auto-deregister, ratcheting quality upward every epoch.

VERIFICATION

Dice + BCE scoring on held-out clinical datasets. On-chain, deterministic, reproducible by any hospital before purchase. First-ever open benchmark for medical imaging AI performance.

ANTI-GAMING

Block-hash seeded sampling. Validator-local inference on reference hardware. Probe-image OOD robustness. Dual-layer copy detection. No shortcut path to competitive composite score exists.

Dimension	MedMetric Answer
What do you sell?	Edge-deployable clinical imaging AI models ($5K–$20K/dept vs. $200K+ incumbent)
Why would miners compete?	Winner-takes-most TAO emissions + organic hospital revenue; below-average miners deregistered
How is output verified?	Validator-run Dice+BCE on Jetson Orin, block-hash seeded benchmark, fully on-chain
Why can't you game it?	OOD probes, validator-local inference, mask-head requirement, copy detection
Why must this be a subnet?	Only decentralized aggregation resolves hospital trust problem for federated learning; cloud providers have failed
Why now?	BraTS/CheXpert datasets mature; Jetson AGX Orin now viable edge hardware; HIPAA cloud tension at peak