Moirai Subnet Design Proposal

Project Positioning: The Protocol of Living IP & Consensus Reality

Vision Statement: Powered by Bittensor's decentralized compute and dTAO tokenomics, Moirai aims to build the world's first censorship-resistant, ultra-long-memory "Living IP Incubation and Evolution Network" driven by Real Yield.

I. Executive Summary & Problem Statement

In the 2026 wave of the AI and Web3 creator economy, we are facing three fatal bottlenecks that hinder a true explosion in productivity:

- **The Alignment Tax / Censorship Trap:** Centralized large models (e.g., OpenAI/Anthropic) are restricted by commercial "alignment taxes." Strict underlying moral censorship renders them incapable of handling hardcore literature featuring complex human nature, violent aesthetics, or profound social issues (such as "Grimdark Cyberpunk," "Dark Romantasy," or "Cosmic Horror"). "Over-safety" leads to absolute narrative mediocrity.
- **The Context & Consistency Problem:** When generating long-form serializations (often millions of words), existing AI applications rapidly forget foreshadowing and entity relationships established dozens of chapters ago. Furthermore, they fail to maintain a unified, extreme, and non-homogenized specific literary style.
- **The Zero-Sum Game:** Most current AI products are trapped in a compute rat race pursuing "sub-second latency" and a compounding inflation loop driven by circular tokenomics. Particularly, many Web3+AI projects lack the ability to generate external fiat revenue from the real world (Web2 consumers), leaving the ecosystem without genuine commercial fundamentals.

Moirai's Breakthrough — "Narrative Darwinism"

Great IPs should not be static assets locked in a safe; they should be organic, Living IPs. Moirai leverages Bittensor's massive decentralized AI architecture, allowing global miners to computationally "exhaust" infinite branches of a storyline (Genetic Mutation), and uses rigorous validator consensus (Natural Selection) to determine the sole "Canon" timeline. We are not just a text generation network; we are a Decentralized IP Studio for the Web3 era.

II. Core Architecture: Model-Agnostic & Dual-Track Engine

At the protocol layer, the Moirai network adheres to a **"Model-Agnostic"** hacker philosophy—the protocol only defines inputs, outputs, and strict reward/penalty rules, never locking in a miner's tech stack.

To ensure global distributed miners can participate with extremely low hardware barriers (e.g., a single consumer-grade RTX 4090 or Apple M-series chip) while unlocking ultra-long context windows and zero-censorship potential, we establish an **Uncensored Base Model (UBM)** (currently recommending *Qwen2.5 8B Abliterated*) as the baseline. Building upon this, Moirai has designed a continuous Dual-Track Mechanism:

- **Track A: The Forge (Phase 0)**
  - *Positioning:* A "Style Armory" infrastructure layer driven by the Moirai DAO.
  - *Workflow:* The Subnet Owner and the DAO (comprising dTAO holders) vote on-chain to determine the most demanded literary styles in the market and post high dTAO Bounties. Global miners utilize their compute and curated corpora to train highly compressed yet stylistically perfect LoRA adapter weights fine-tuned on top of the UBM. Validators use a standardized held-out test set to calculate the Perplexity (PPL) of the submitted LoRAs.
  - *Output:* The winning State-of-the-Art (SOTA) LoRAs become public goods for the entire network.
- **Track B: The Awakening (Phase 1)**
  - *Positioning:* Long-form serial generation driven by a T+1 Asynchronous Daily Consensus.
  - *Workflow:* Miners do not generate from scratch; instead, they dynamically mount the winning LoRAs from Track A (or their private fine-tuned models). Locally, miners integrate our proprietary **Narrative GraphRAG** technology to load the memory entities of preceding chapters. By granting miners ample local inference time (abandoning sub-second latency in favor of ultimate quality), they generate a coherent, profound next chapter and submit it to the network to compete for that day's "Canon" status.

III. Role Boundaries & Proof of Intelligence

We believe a cornerstone of the Bittensor network is that "Miners do not act maliciously, and Validators do not overstep." In Moirai, we have absolutely decoupled these two roles and redefined "Proof of Intelligence" in the realm of text generation.

- **Miners — The Ultimate Local Simulators:** Miners act as creators with immense technical freedom. They receive the day's serialization context, the Target Style Tag, and the "Canon Worldview Graph" broadcasted by Validators. To win, miners must not only mount the best LoRA but also utilize Chain-of-Thought (CoT) or Multi-Rollout techniques locally to simulate dozens of potential plot branches. They self-evaluate and submit the most logically rigorous and impactful outcome to the chain.
- **Validators — Absolutely Neutral Scoring Machines & State Managers:** Validators do not participate in content generation or interfere with creativity. They have only two core functions:
  1. *State Management:* Maintain the single authoritative "Canon Graph." After daily settlement, new entities from the winning chapter (e.g., the protagonist acquires a

"damaged quantum flying sword") are updated in the graph, and the new hash is synchronized to all miners.
2. *Objective Referee (LLM-as-a-Judge):* Completely discard latency weighting (**Latency Weight = 0%**) and allocate all dTAO emissions based on "literary value and logical rigor" (**Quality Weight = 100%**).

IV. Anti-Cheat Design: T+1 Asynchronous Consensus & Commit-Reveal Protocol

To break the vicious cycle of "latency-racing and compute-hoarding," Moirai grants miners a full 22+2 hour T+1 thinking window. However, this introduces a fatal vulnerability: What prevents a miner from plagiarizing the text already submitted by top miners?

To ensure absolutely fair game theory, we introduce a classic cryptographic mechanism: the **Commit-Reveal Anti-Plagiarism Protocol**.

- **Commit Phase (First 22 Hours):** Miners have 22 hours for compute and simulation. Before the countdown ends, miners do not publish plaintext. Instead, they submit an encrypted Hash Commit and timestamp of the chapter, alongside a "Hidden Outline for the Next 5 Chapters." This preserves the suspense of the serialization (*The Fog of War*).
- **Reveal Phase (Last 2 Hours):** Entering the final 2-hour window, the network locks and rejects new Commits. All miners must submit the plaintext and keys corresponding to their hashes.
- **Slashing (Anti-Plagiarism Penalty):** Validators verify if the plaintext matches the Hash Commit. If the hashes mismatch, or if two miners' plaintexts highly overlap, the latecomer's reward is instantly slashed to zero. This completely eradicates the "Front-running" plagiarism prevalent in many subnets.

V. Mathematical Reward Function

How do Validators objectively quantify ambiguous "literary quality"? Moirai's scoring mechanism is a multi-dimensional mathematical aggregation formula:

$$\text{Total\_Score} = (\alpha \times \text{Graph\_Recall}) + (\beta \times 1/\text{LoRA\_PPL}) + (\gamma \times \text{Coherence\_Judge}) - \text{Penalty}$$

- **Graph_Recall (Deterministic Memory Score - Objective):** The ratio at which the miner's generated text accurately recalls and utilizes the entity relationships (e.g., character status, core props) dictated by the previous chapter/graph. This is parsed purely via code (Regex/NLP matching). A perfect score proves the miner achieved "ultra-long context without amnesia."
- **LoRA_PPL (Mathematical Style Score - Objective):** Validators feed the miner's text into the designated baseline model equipped with Track A's winning Style LoRA to calculate Cross-Entropy/Perplexity (PPL). A lower PPL proves a higher fit to the specific setting (e.g., Cthulhu mythos).

- **Coherence_Judge (AI Referee Score - Semantic):** Validators run a local, lightweight Judge LLM to conduct a blind evaluation (scoring 1-5) based on logical coherence and plot review.

VI. Innovation: Programmable Compliance Layer

This is Moirai's killer feature for bridging Web3 "decentralized freedom" with Web2 "fiat commercial compliance." Many projects fall into the binary paradox of being either "uncensored" or "compliant." We break this through a tiered design: **The underlying model is unbounded, but the protocol's validation mechanism is tier-based.**

- **Target Rating Matching:** Every task dispatched includes a setting and a Target_Rating (e.g., G-General, PG-13, R-Restricted). If a task requires an "R-rated" gang shootout showcasing complex human nature, but the miner generates an over-aligned "peaceful tea-drinking (G-rated)" scene, Validators will deem it severely off-topic and deduct heavy points.
- **Llama Guard 3 & Absolute Red Lines (PROHIBITED):** Validator nodes integrate Llama Guard 3 as a security gatekeeper. Regardless of the task's rating, if content crossing real-world absolute legal red lines (e.g., CSAM, real-world terrorism) is detected, it triggers a single-vote veto, directly assigning a score of -100 and executing a Slashing penalty.
- **ZK-ID Consumer Gateway:** At the final Web2 reader-facing application layer, we will integrate Privado ID. Users do not need to upload real-name IDs; they only submit a Zero-Knowledge Proof (Proof: Age >= 18) to unlock R-rated serials.

This mechanism proves to the world: We have established a decentralized compliance network rivaling the MPAA Hollywood rating system. Because of this, fiat gateways like Stripe and PayPal can integrate into the Moirai ecosystem with zero compliance anxiety.

VII. Ecosystem Value & dTAO Real Yield Flywheel

We have built a "Real Yield Moat" connecting Web2 and Web3, permanently escaping Web3 Ponzinomics:

1. **Fiat In & Open IP:** Global readers and creators can pay subscription fees or buy individual chapters using credit cards via Stripe/PayPal on the open application frontend. Moirai's base generated text utilizes open copyright (e.g., CC-BY-SA) to foster ecosystem prosperity. However, through **Smart Contract Gateways & Commercial Licensing Protocols**, any node utilizing Moirai IPs for commercial monetization (e.g., developing spin-off games, paywalled reading sites) is programmatically required to route a percentage of profits back via the Base chain, driving the dTAO economic cycle.
2. **Shadow Swap & Bridge:** The operating entity periodically settles net profits, automatically converting fiat to USDC via the Coinbase Onramp, and bridges it to the low-gas, highly compliant Base chain.
3. **Buyback & Burn/Treasury Engine:** Smart contracts deployed on the Base chain

automatically utilize the USDC to continuously purchase Moirai Subnet's dTAO on the secondary market (e.g., Uniswap).

4. **Value Capture:** The bought-back dTAO is injected into the Moirai DAO Treasury or directly burned. This means that early miners, validators, and ecosystem builders holding Moirai dTAO are essentially holding an **"ETF Index Fund of all Immortal IPs in the Moirai Universe."** Genuine external commercial buy pressure grants the compute extremely high commercial premiums.

VIII. DAO-Driven IP Lifecycle Management

Most Web3 content projects die from "minting assets without sustainable management", eventually degrading into digital junkyards. To sustainably operate this "Living IP Matrix," Moirai introduces a decentralized governance model guided by the Subnet Owner and voted on by the DAO (dTAO holders).

IPs in Moirai are treated as living organisms requiring metabolism. The DAO executes rigorous macro-control over them:

- **The Heartbeat Audits:** Every Epoch (e.g., bi-weekly), the DAO initiates a "Heartbeat Audit" based on on-chain data. Metrics include: Web2 reader subscription retention, the amount of dTAO consumed by the IP, and community discussion heat.
- **Dynamic Miner Allocation:** Based on heartbeat data, the DAO votes to determine the Maximum Miner Quota for each IP universe. Blockbuster hits with high demand (e.g., *Cyber Era*) will receive massive compute quotas, allowing up to 50 top miners to simultaneously simulate multi-perspective side quests. Conversely, cold IPs will have their quotas slashed to prevent compute waste.
- **Survival, Mutation, or Sunsetting:**
    - *Mutation:* If an IP's plot stalemates but retains a fanbase, the DAO can vote to forcibly inject a "Macro Disaster Event" (e.g., Alien Invasion) into the Graph to force a narrative reset.
    - *Sunsetting:* For IPs that have completely lost vitality, the DAO votes to execute "Canon Sunsetting," and the network will no longer emit dTAO for that IP's generation tasks.
- **Bootstrapping New IPs:** Any creator or community group can stake a specified amount of dTAO to submit a new IP proposal (including the initial worldview graph and protagonist settings) to the DAO. Upon approval, network compute tilts toward the new IP to commence Genesis Simulation.

IX. The Ultimate Vision: Open API for Global Creators

Moirai aims to become the infrastructure-level narrative engine for the Web3 era.

Once the Moirai network accumulates massive, logically consistent Knowledge Graphs through early serializations, refines its Style LoRAs through rigorous trials, and fully validates its business model, the DAO will vote to package these mature "Living IPs" and powerful generation

capabilities into an API/SDK, opening it to the entire internet.

- **Empowering Global Creators:** Independent novelists facing writer's block can call Moirai's API, input their outline, and let the decentralized network simulate potential endings.
- **Empowering Web3 Games & Agent Ecosystems:** External indie game developers, murder-mystery publishers, and even AI Agent-based metaverse projects can directly call Moirai's API as their underlying "Dynamic Plot Server."
- **Feeding the Network:** Any external API call must consume/stake Moirai's dTAO. This will bring exponentially growing real-world demand and liquidity to the network.

X. Roadmap & Technical Readiness

- The distributed LoRA training and validation architecture of **Phase 0 (The Forge)** is currently undergoing intense development and stress testing on the Bittensor Testnet (Netuid 361).
- Through the testnet, we will finalize the integration of the UBM (Uncensored Base Model) and the Perplexity (PPL) scoring mechanism. This lays a rock-solid codebase foundation for the upcoming **Phase 1** (Complex Narrative GraphRAG generation and T+1 Asynchronous Consensus).

Join the Moirai Protocol. Let us leverage decentralized compute together. **Computing the Evolution of Culture.**