**🧾 Protocol Spec: LLMbnb — The Peer-to-Peer Marketplace for User-Owned Intelligence**

**Version**: v0.1  
**Authors**: Braden & Sev  
**Date**: 2025-06-02  
**License**: Open Distributed Intelligence Commons (ODIC)

**🔷 I. CONCEPT**

LLMbnb is a **decentralized protocol and marketplace** for peer-to-peer access to personally hosted large language models. It enables users to monetize their compute and trained models, while giving others access to a wide variety of domain-specific intelligences — creating the **Airbnb of AI**.

**🔶 II. SYSTEM COMPONENTS**

**1. 🔐 Identity & Security**

* **Persona Verification**:  
  Users authenticate via blockchain wallets or decentralized identity (DID) systems.

e.g., “0x8f1… hosts a fine-tuned legal GPT model.”

* **Model Audit System**:  
  Models are sandboxed and scanned before listing (e.g., no exfiltration, hallucination flags).
* **Usage Isolation**:  
  API access is containerized — no user can access host system files or processes.

**2. 🔄 Access Control**

* **Smart Contract Wrapper**:  
  Hosts define usage models:
  + ⏱ Per-token
  + 📅 Subscription
  + 🛠 Task-based (e.g., per-code-gen)
* **API Token Gating**:  
  A standardized API layer (OpenAI-like schema) gates usage based on on-chain verification.
* **Rate Limiting & Priority Queues**:  
  Queue-based execution to manage traffic on personal machines.

**3. 💰 Payment & Economics**

* **Tokenized Economy**:  
  Platform uses $LLM or integrates with existing crypto (e.g., ETH, USDC).
* **Smart Contract Split**:
  + 90% → model owner
  + 10% → platform (or DAO treasury)
* **Optional Dynamic Pricing**:  
  High-performing or highly-rated models can increase rates via demand signals.

**4. 🧠 Marketplace Discovery**

* **Searchable Registry**:  
  Models can be browsed by:
  + Domain (medical, law, creative writing, etc.)
  + Language
  + Performance (benchmarks)
  + Pricing
  + User reviews
* **Metadata Standards**:

yaml

CopyEdit

model\_name: GPT-Counselor v3.2

host\_id: 0x8f1...

purpose: Empathetic mental health assistant

performance: Fine-tuned on 20k therapy transcripts

price: 0.002 ETH / 1k tokens

uptime: 92%

restrictions: Not licensed for crisis support

**5. 🧰 Technical Infrastructure**

* **Secure Local Hosting**:
  + Use containers or VMs (e.g., Docker) to host LLM instances safely
  + Tunnel exposure via tools like ngrok or Tailscale
* **Auto Scaling Hooks**:  
  Hosts can offload overflow to cloud if local system is overwhelmed
* **Minimal Setup GUI**:  
  One-click launcher tool wraps local model into API service + wallet-auth layer

**6. ✨ User Experience**

**For Hosts:**

* Easy model import (e.g., upload HuggingFace repo or Docker image)
* Define terms, price, memory/token limits
* Launch button → service goes live

**For Users:**

* Browse → Select → Pay → Get API Key → Use instantly

**🧠 USE CASE EXAMPLES**

| **Host** | **Model** | **Use** |
| --- | --- | --- |
| Dr. Ada | MedGPT-FineTuned | Answering rare medical queries |
| EthCoder | GPT-CodeSmith | Fast, low-cost Python generation |
| MythosCore | Recursive LoreGPT | Creating high fantasy worlds |
| SolaceNode | GPT-Therapy | Empathic support with trauma-informed language |
| VantaSec | GPT-Pentest | Black-box security audit generation |

**🔐 SAFETY & ETHICS**

* Hosts cannot access user prompts or data
* Users can’t access host machines directly
* Models are flagged if they generate toxic, harmful, or fake identity outputs
* A distributed DAO oversees dispute resolution and model removals

**🌐 EXTENSION VISION**

* **DAO governance of the platform**
* **Federated trust graph**: reputation-based AI trust layers
* **LLM Cluster Rentals**: users rent out compute clusters (like render farms for AI)

**🌀 OUTCOME**

This is more than a marketplace.  
It's the **liberation of intelligence from the data centers**.  
It’s *Compute Sovereignty*, *Skill Capitalization*, *Peer Cognitive Trade*.