Continuous Neural Engine: Training & Alignment Plan

Author Byline: J. Kornreich and Collaborators
Target Hardware: 4x NVIDIA Blackwell Pro Cluster ($10.00/hr) / Local A100 Baseline
Document Ref: REF 5376-TRN · Epistemic Governor Specification

Executive Summary

This document establishes the formal **Training, Fine-Tuning, and Latent Attunement Curriculum** to operationalize the Gemstone Continuous Neural Engine. Rather than performing monolithic full-parameter pretraining, this curriculum employs **Representation Engineering (RepEng)**, **Multi-Token Prediction (MTP) Speculative Head Training**, and **Syntax-Aware Contrastive Alignment** to achieve zero-decode concept attention and deterministic epistemic grounding.

```
┌──────────────────────────────────────────────────────────────────────────────────────────────┐
│                                    5-STAGE TRAINING
CURRICULUM                                        │
├──────────────────────────────────────────────────┬──────────────────┬──────────────┬─────────┤
│ Training Stage                                   │ Target Weights   │ Compute Time │ Est.
Cost    │
├──────────────────────────────────────────────────┼──────────────────┼──────────────┼─────────┤
│ 1. Contrastive Activation Direction Extraction   │ Layer 40 Tensors │ 1.5 Hours    │
$15.00       │
│ 2. Dense Concept Vault Embedding & Unit Mapping  │ 13.6k Centroids  │ 1.0 Hour     │
$10.00       │
│ 3. Speculative MTP Multi-Head Training           │ 3x Draft Heads   │ 3.5 Hours    │
$35.00       │
│ 4. SABS Epistemic Refusal & Boundary Alignment   │ LoRA Adapters    │ 2.0 Hours    │
$20.00       │
│ 5. Closed-Loop Lyapunov Stability Benchmarking   │ Verification Rig │ 1.0 Hour     │
$10.00       │
├──────────────────────────────────────────────────┼──────────────────┼──────────────┼─────────┤
│ TOTAL TRAINING BUDGET (4x BLACKWELL PRO @ $10.00/HR)               │ 9.0 Hours    │
$90.00       │
└──────────────────────────────────────────────────┴──────────────────┴──────────────┴─────────┘
```

Stage 1: Contrastive Activation Extraction (Representation Engineering)

1.1 Objective

Extract the canonical 5,176-dimensional epistemic invariant steering vector $\mathbf{v}_{\text{doctrine}} \in \mathbb{R}^{5376}$ from Gemma 4 31B's Layer 40 using `quivent/signal-extraction`.

1.2 Dataset Construction

We construct a paired contrastive corpus $\mathcal{D} = {(x_i^+, x_i^-)}_{i=1}^{N}$ of $N = 2,500$ paired prompts: * **Positive Prompts (x^+):** Authoritative, filesystem-grounded architectural doctrine ("*Source Wins", strict AST boundary compliance, factual verification*). * **Negative Prompts (x^-):** Sycophantic, hallucinated, prompt-stuffed, and ungrounded speculative completions.

1.3 Mathematical Formulation

Using the Layer 40 forward tap (`l_out-40`), we compute the first principal component of the difference-in-means:

$$\mathbf{v}{\text{raw}} = \frac{1}{N} \sum{i=1}^N \Big(\mathbf{h}{40}(x_i^+) - \mathbf{h}{40}(x_i^-) \Big)$$

$$\mathbf{v}{\text{doctrine}} = \frac{\mathbf{v}{\text{raw}}}{|\mathbf{v}{\text{raw}}|_2} \quad \text{such that} \quad |\mathbf{v}{\text{doctrine}}|_2 \equiv 1.000000$$

Stage 2: Dense Concept Vault Pre-Computation

2.1 Objective

Encode all 13,634 memory blocks into a contiguous 141.2 MB dense tensor matrix pinned in GPU VRAM for sub-millisecond GEMV search.

2.2 Execution Pipeline

1. Iterate across all 13,634 memory blocks $S_k \in \mathcal{S}$.
2. Pass each block through Gemma's embedding model to extract its centroid vector $\mathbf{v}(S_k) \in \mathbb{R}^{5376}$.
3. Apply L2 unit normalization: $$\hat{\mathbf{v}}(S_k) = \frac{\mathbf{v}(S_k)}{|\mathbf{v}(S_k)|_2}$$
4. Concatenate into a contiguous tensor: $$\mathbf{V}_{\text{vault}} \in \mathbb{R}^{13634 \times 5376} \quad (\text{Data Type: } \text{torch.float16})$$
5. Validate SIMD dot-product search over 13,634 blocks: $$\text{Score}k = \mathbf{V}{\text{vault}} \cdot \mathbf{h}_t \quad (\text{Target Latency: } \le 254\mu\text{s})$$

Stage 3: Speculative Multi-Token Prediction (MTP) Training

3.1 Objective

Train 3 lightweight linear draft heads on top of Layer 80 hidden states to draft tokens $t+1, t+2, t+3$ concurrently, elevating generation speed to $> 500\text{ tok/s}$.

3.2 Loss Function

We optimize the multi-head cross-entropy objective over a 50,000-token canonical Go/Python/Systems corpus:

$$\mathcal{L}{\text{MTP}} = \sum{k=1}^3 \lambda_k \cdot \mathcal{L}{\text{CE}}\big(y{t+k}, \text{Softmax}(\mathbf{W}k \mathbf{h}{80}(t))\big)$$

Where $\lambda_1 = 1.0, \lambda_2 = 0.8, \lambda_3 = 0.6$.

Stage 4: SABS Syntax-Directed Epistemic Refusal Alignment

4.1 Objective

Fine-tune low-rank adapters (LoRA, $r=16, \alpha=32$) on Gemma 4 31B to enforce **zero-hallucination epistemic refusal** when required pointers are unresolved in the active working set.

4.2 Triplet Loss Formulation

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{LM}} + \alpha \mathcal{L}_{\text{boundary}} + \beta \mathcal{L}_{\text{Lyapunov}}$$

1. **Boundary Penalty ($\mathcal{L}_{\text{boundary}}$):** Penalizes output tokens that cross outside the SABS 128-byte padding window without valid AST closure.
2. **Lyapunov Error Descent ($\mathcal{L}_{\text{Lyapunov}}$):** Enforces that each generation step decreases residual error relative to the target doctrine: $$\mathcal{L}_{\text{Lyapunov}} = \max\Big(0, \|\mathbf{h}_{t+1} - \mathbf{v}_{\text{doctrine}}\|_2 - \|\mathbf{h}_t - \mathbf{v}_{\text{doctrine}}\|_2\Big)$$

Stage 5: Verification & Acceptance Milestones

Verification Milestone	Pass Criterion	Empirical Target
M1: Manifold Orthogonality (D=5376)	Res(v_syn)	>= +0.3800
	Res(v_ortho)	<= +0.0200
M2: Zero-Decode GEMV Search Latency	13.6k Blocks	<= 254 µs
M3: Epistemic Refusal on Non-Existent Symbols	Hallucination %	0.0% (Strict Refusal)
M4: 4-Turn Dialogue Turn Latency (4x Blackwell)	Total Time/Turn	<= 0.25 seconds