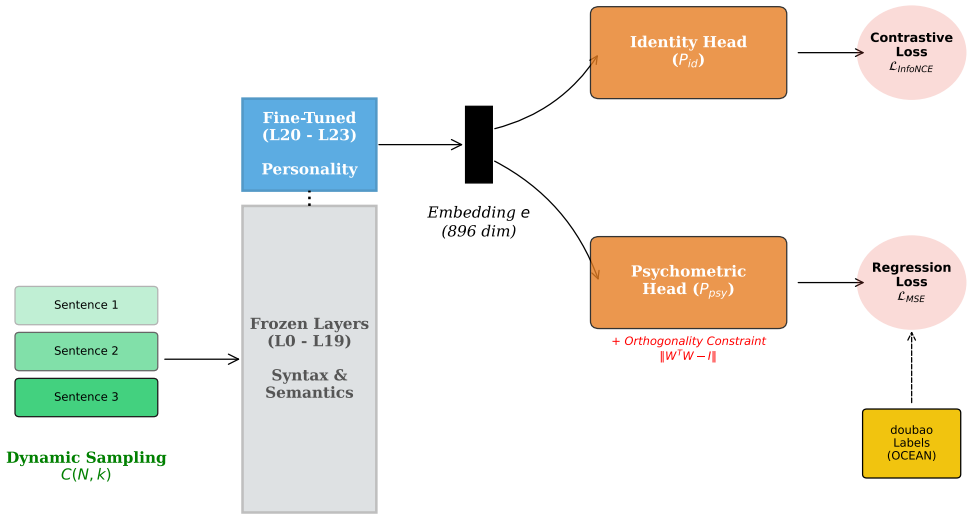

Figure 1: **The Soul Engine Architecture.** The lower layers (Grey) are frozen to preserve general intelligence. The upper layers (Blue) are fine-tuned. The embedding is projected into orthogonal Identity and Psychometric spaces.

2.3 Optimization Objective

We propose a hybrid loss function designed to simultaneously maximize discrimination and measurement accuracy, while enforcing geometric disentanglement.

$$\mathcal{L}_{total} = \mathcal{L}_{InfoNCE} + \lambda_1 \cdot \mathcal{L}_{MSE} + \lambda_2 \cdot \mathcal{L}_{Orth} \tag{3}$$

1. Stylistic Contrastive Loss ($\mathcal{L}_{InfoNCE}$). To learn a robust identity representation, we employ a contrastive objective with in-batch negatives. For an anchor chunk A_i and a positive chunk P_i (sampled from the same character but different texts), the loss is:

$$\mathcal{L}_{InfoNCE} = -\log \frac{\exp(\text{sim}(P_{id}(A_i), P_{id}(P_i))/\tau)}{\sum_{j=1}^{N} \exp(\text{sim}(P_{id}(A_i), P_{id}(A_j))/\tau)} \tag{4}$$

This forces the model to ignore semantic content and focus on the invariant "voice" of the character.

2. Psychometric Regression ($\mathcal{L}_{MSE}$). We minimize the divergence between the predicted traits and the ground truth OCEAN scores $\mathbf{y}$:

$$\mathcal{L}_{MSE} = \|P_{psy}(e) - \mathbf{y}_{truth}\|_2^2 \tag{5}$$

3. Orthogonality Regularization ($\mathcal{L}_{Orth}$). To strictly enforce the hypothesis that personality vectors should be independent of each other (e.g., Neuroticism should not correlate with Openness in the vector space), we impose an orthogonality constraint on the projection matrix W_{psy}:

$$\mathcal{L}_{Orth} = \|W_{psy}^T W_{psy} - I\|_F^2 \tag{6}$$