Total Loss Function (L_total)

$$L_{\text{total}} = (1-\alpha)\, L_{\text{CE}} + \alpha\, L_{\text{contrastive}} + \lambda_{\text{aux}}\, L_{\text{masked}} + \lambda_{\text{ortho}}\, L_{\text{orthogonality}}$$

1. Cross-Entropy Loss with Label Smoothing (L_CE)

$$\tilde{y}_{b,i} = (1-\varepsilon)\cdot \mathbb{1}[i = y_b] + \varepsilon/N$$
$$L_{\text{CE}} = -\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{N} \tilde{y}_{b,i}\cdot \log p_{b,i}$$

2. Contrastive InfoNCE Loss (L_contrastive)

$$L_{\text{contrastive}} = -\frac{1}{B}\sum_{b=1}^{B} \log\left(\frac{1}{K}\sum_{k=1}^{K} \exp((z_b \cdot z_{b,k}^{-})/\tau)\right)$$

3. Masked Item Prediction Loss (L_masked)

$$L_{\text{masked}} = -\frac{1}{|\mathcal{M}|}\sum_{(b,\ell)\in\mathcal{M}} \log p_{b,\ell}(y_{b,\ell})$$

4. Orthogonality Regularization Loss (L_orthogonality)

$$L_{\text{orthogonality}} = \frac{2}{M(M-1)}\sum_{i=1}^{M}\sum_{j=i+1}^{M} \mathbb{E}[v_i^{\top} v_j]$$

5. Relevance Score Computation

The predicted relevance score $s_{b,i}$ between user sequence $\mathbf{h}_b$ and item i is computed as:
$s_{b,i} = \mathbf{h}_b^{\top}\mathbf{e}_i$, where $\mathbf{h}_b$ is the sequence representation and $\mathbf{e}_i$ is the embedding of item i. These scores are used to rank candidates or passed through a softmax layer for classification.