========== ==========

Algorithm 1 Token Embedding (Phuong & Hutter, 2022)

Require: $v \in \mathcal{V}$ (token from vocabulary)
$W_e \in \mathbb{R}^{d_e \times |\mathcal{V}|}$ (embedding matrix)
Ensure: $e \in \mathbb{R}^{d_e}$ (embedding vector)

```
return e = W_e[:, v]
```

$$\textbf{return } e = W_e[:, v]$$

========== ==========

Algorithm 2 Multi-Head Attention (Phuong & Hutter, Alg 5)

Require: $X \in \mathbb{R}^{d \times n}$ (sequence of n tokens)
H attention heads; W_q^h, W_k^h, W_v^h for each head $h \in [H]$; W_o (output projection)
Ensure: $Y \in \mathbb{R}^{d \times n}$ (transformed sequence)

for each head $h \in \{1, \ldots, H\}$ **do**
  $Q^h \leftarrow W_q^h X$ ▷ Query projections
  $K^h \leftarrow W_k^h X$ ▷ Key projections
  $V^h \leftarrow W_v^h X$ ▷ Value projections

  $Y^h \leftarrow \text{Attention}(Q^h, K^h, V^h)$
  $\quad = \text{softmax}\left(\frac{Q^h {K^h}^T}{\sqrt{d_k}}\right) V^h$
end for

$Y \leftarrow W_o[Y^1; Y^2; \ldots; Y^H]$ ▷ Concatenate & project

return Y

========== ==========

Algorithm 3 Layer Normalization (Phuong & Hutter, Alg 6)

Require: $x \in \mathbb{R}^d$ (activations)
$\gamma, \beta \in \mathbb{R}^d$ (learned scale & shift)
Ensure: $\hat{x} \in \mathbb{R}^d$ (normalized activations)

$\mu \leftarrow \frac{1}{d}\sum_{i=1}^{d} x[i]$ ▷ Mean
$\sigma^2 \leftarrow \frac{1}{d}\sum_{i=1}^{d}(x[i]-\mu)^2$ ▷ Variance
$\tilde{x} \leftarrow \frac{x-\mu}{\sqrt{\sigma^2+\epsilon}}$ ▷ Normalize
return $\gamma \odot \tilde{x} + \beta$ ▷ Scale & shift

========== ==========

Algorithm 4 Semantic Search via Cosine Similarity

Require: Query embedding $q \in \mathbb{R}^d$; Database embeddings $D \in \mathbb{R}^{N \times d}$
k (number of top matches to return)
Ensure: Top-k most similar incidents

for each incident $i \in \{1, \ldots, N\}$ **do**
  similarity$[i] \leftarrow \frac{q \cdot D[i]}{\|q\| \times \|D[i]\|}$ ▷ Cosine similarity
end for
indices $\leftarrow$ argsort(similarity) ▷ Sort descending
return Top-k incidents from indices

========== Diagnosis ==========

Algorithm 5 Similarity-Weighted Diagnosis

Require: Top-n similar incidents with similarity scores $\{s_1, \ldots, s_n\}$ and root causes $\{c_1, \ldots, c_n\}$

Ensure: Probability distribution over all unique causes

causes $\leftarrow \{c_1, c_2, \ldots, c_n\}$ ▷ All unique causes
total_weight $\leftarrow \sum_{i=1}^{n} s_i$ ▷ Sum of all similarities

for each unique cause $j \in$ causes **do**
    weight$_j \leftarrow 0$
    for each incident $i \in \{1, \ldots, n\}$ **do**
        if $c_i = j$ **then**
            weight$_j \leftarrow$ weight$_j + s_i$ ▷ Accumulate similarity
        end if
    end for

    $P(j) \leftarrow \frac{\text{weight}_j}{\text{total\_weight}}$ ▷ Weighted probability
end for

return Probability distribution P over all causes