Sparse Activation: Only 2 of 4 experts

Embedding

Token Input

Router Gate

Expert 1 Mathematics

Expert 2 Language

Expert 3 Code

Expert 4 Vision

0.7

0.3

0.0

0.0

Weighted Sum

Representation

Output Final

$$y = \sum_{i=1}^{k} G(x)_i \cdot E_i(x) \text{ where } G(x) = \text{routing weights}, E_i = \text{expert } i$$