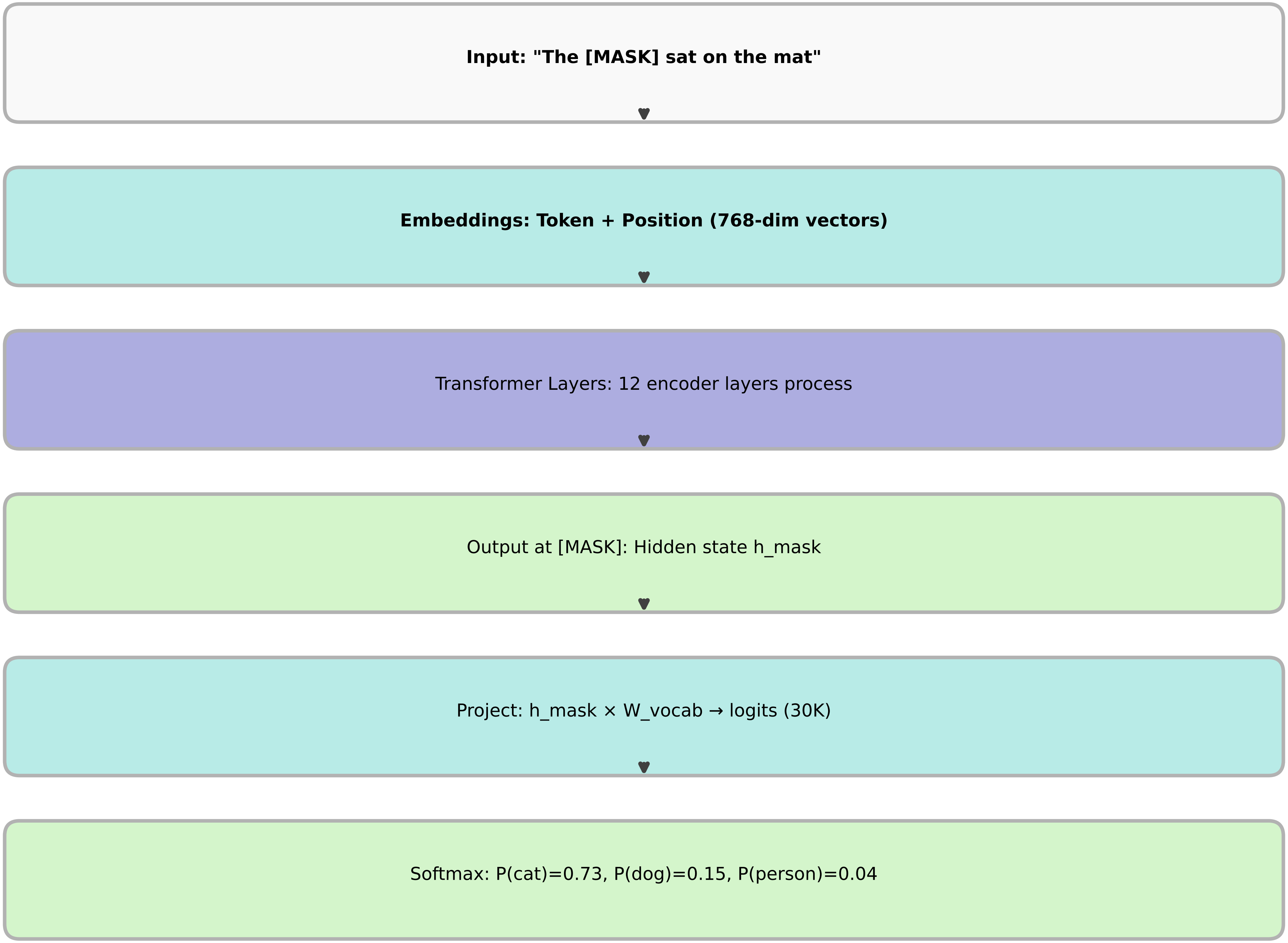
Masked Token Prediction: Step-by-Step Process



Input: "The [MASK] sat on the mat"

Embeddings: Token + Position (768-dim vectors)

Transformer Layers: 12 encoder layers process

Output at [MASK]: Hidden state h_mask

Project: h_mask × W_vocab → logits (30K)

Softmax: P(cat)=0.73, P(dog)=0.15, P(person)=0.04