Positional Encoding & BPE — Report
Reference: Vaswani et al., 2017 — 'Attention Is All You Need'

Input sentence: I love india
d_model: 64
Token count: 4
Tokens: ['■I', '■love', '■india', '■']
Token IDs: [18, 23, 22, 13]

Shapes:
 - token_embeddings: (4, 64)
 - positional_enc: (4, 64)
 - summed_matrix: (4, 64)

We compute: X = sqrt(d_model) * E + PE, where E are token embeddings and PE are sinusoidal enco
dings.

Preview: token embeddings (E)
cols 0..7 (rounded to 4 dp)
row 0: [-0.0128, -0.0265, 0.0328, 0.0202, -0.0138, 0.0450, 0.0196, -0.0065]
row 1: [-0.0027, 0.0003, -0.0157, 0.0130, -0.0024, 0.0084, -0.0177, -0.0087]
row 2: [-0.0065, -0.0050, 0.0266, 0.0111, 0.0091, 0.0433, -0.0129, 0.0186]
row 3: [-0.0201, -0.0154, -0.0007, 0.0047, 0.0310, -0.0200, 0.0197, -0.0043]
Preview: positional encodings (PE)
cols 0..7 (rounded to 4 dp)
row 0: [0.0000, 1.0000, 0.0000, 1.0000, 0.0000, 1.0000, 0.0000, 1.0000]
row 1: [0.8415, 0.5403, 0.6816, 0.7318, 0.5332, 0.8460, 0.4093, 0.9124]
row 2: [0.9093, -0.4161, 0.9975, 0.0709, 0.9021, 0.4315, 0.7469, 0.6649]
row 3: [0.1411, -0.9900, 0.7783, -0.6279, 0.9933, -0.1160, 0.9536, 0.3010]
Preview: X = sqrt(d_model) * E + PE
cols 0..7 (rounded to 4 dp)
row 0: [-0.1022, 0.7883, 0.2627, 1.1616, -0.1101, 1.3604, 0.1571, 0.9480]
row 1: [0.8200, 0.5427, 0.5560, 0.8355, 0.5138, 0.9131, 0.2673, 0.8424]
row 2: [0.8577, -0.4563, 1.2100, 0.1599, 0.9751, 0.7779, 0.6439, 0.8134]
row 3: [-0.0195, -1.1128, 0.7727, -0.5905, 1.2413, -0.2757, 1.1111, 0.2667]