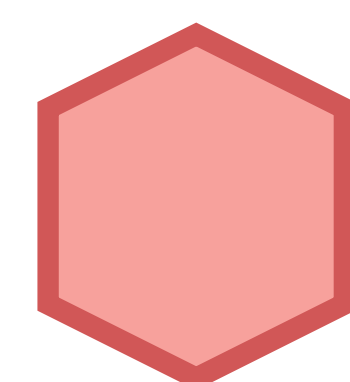
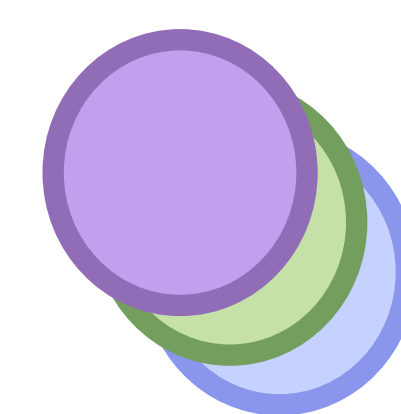


Token Vector



Feed Forward Network



Attention Head



Adding Residuals

Where

Layer 1

Layer 2

Layer 3

Layer 4

Layer 5

Layer 6

Layer 1

Layer 2

Layer 3

Layer 4

Layer 5

Layer 6

Layer 1

Layer 2

Layer 3

Layer 4

Layer 5

Layer 6

Layer 1

Layer 2

Layer 3

Layer 4

Layer 5

Layer 6

Shallow

Middle

Deep

is

Paris

?