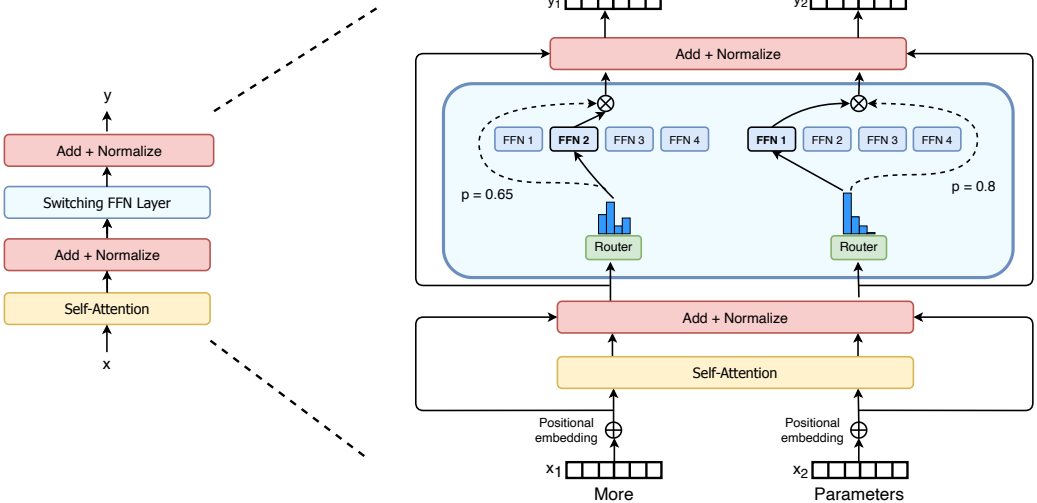

y
Add + Normalize
Switching FFN Layer
Add + Normalize
Self-Attention
x
y1
y2
Add + Normalize
FFN 1
FFN 2
FFN 3
FFN 4
p = 0.65
Router
FFN 1
FFN 2
FFN 3
FFN 4
p = 0.8
Router
Add + Normalize
Self-Attention
Positional embedding
Positional embedding
x1
More
x2
Parameters