**Transformer**

# Language Modeling (MLE)

# Attention

Soft Attention: , Hard Attention:

# Complexity

Time: . Space:

Time: All:

Space:

## MHA

: Time: . Space:

Time:

Projection + Attention + Output Projection

Space:

# Layer Norm

# Point-wise FFN (PFFN)

A screenshot of a computer

AI-generated content may be incorrect.

假设输入

FFN