

Input

“Hello ladies and ____”

Tokens

Hello ladies and

Transformer LM

Embed.
Layer +
Position.
Encodings

1

...

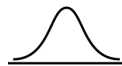
L

Pred.
Head

Low-level
Information

Transformer
Block

Contextual
Information



Probabilities
over Vocab.
Tokens

0.2 %

aardvark

...

1 %

badgers

...

90 %

gentlemen

...