From Prediction to Generation: The Decoding Challenge

Input
(Text)

DEC
IT

Output
Text

P(sat) = 0.45

WeBic = 0.8 Beam

RNN P(jump Temp=Top-k

→ BERT Nucleus,

50,0 Contrastive

6 Strategies:

Gre →

"The cat sat
on the mat"

Key Question: Model gives 50,000 probabilities - which word do we choose?