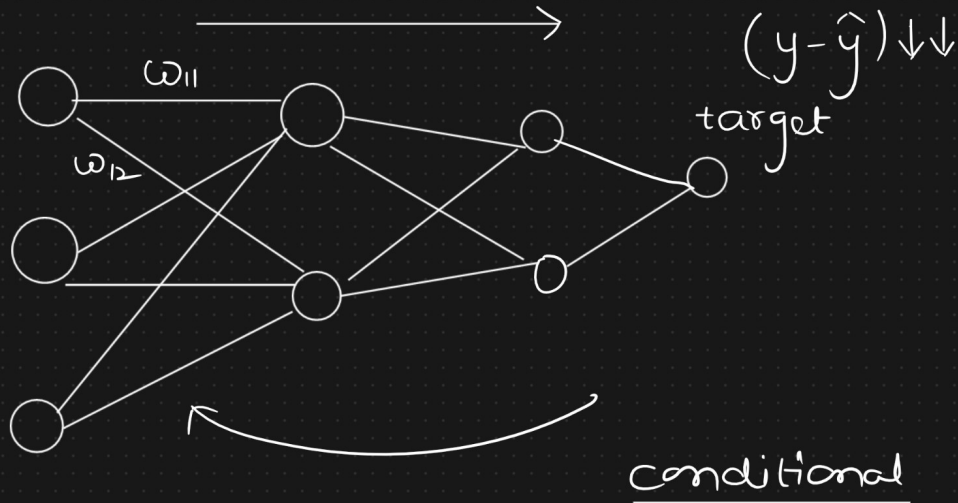


GPT 2 (Generative Pretrained Transformers)



↳ 1.5 billion Parameters

↳ 8 Million web pages dataset

conditional

Probability

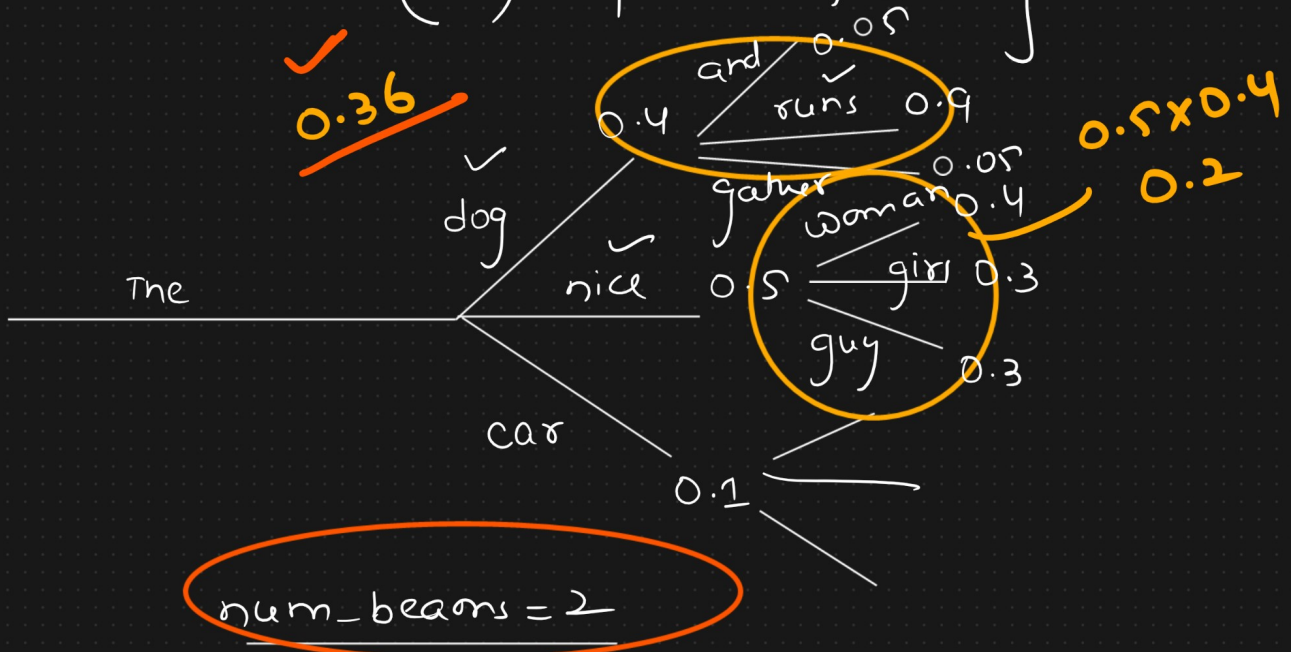
$$P(y | x_1, x_2, x_3)$$

x_1	x_2	x_3	<u>target(y)</u>
I	love	reading	novel
(((
92	104	156	

numeric

id

- 1) unsupervised pretraining
- 2) supervised finetuning



num_beams = 2