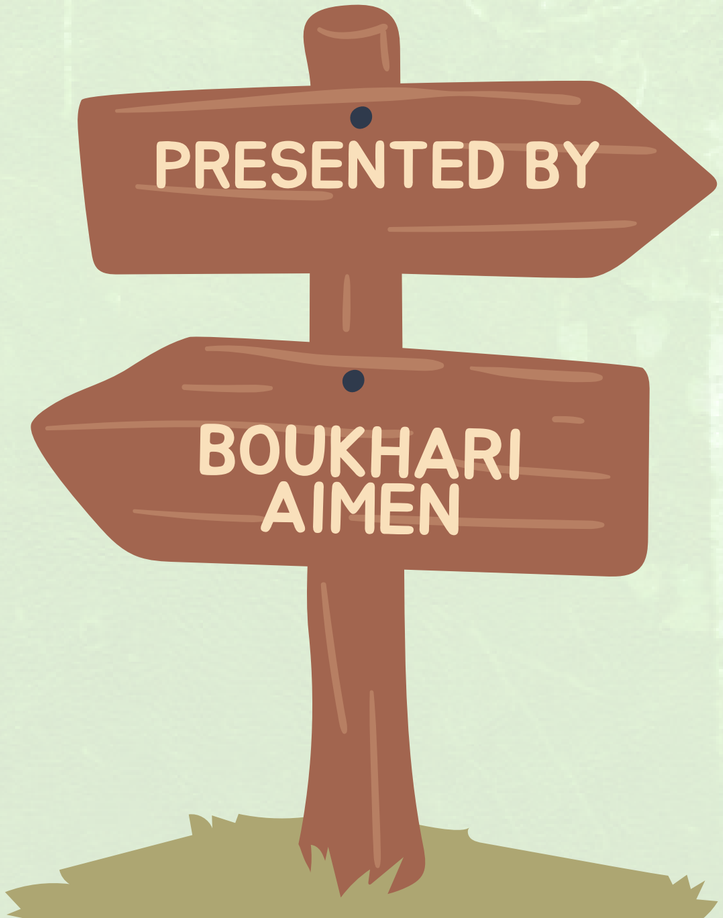
AI Camp

ADVANCED NLP TECHNIQUES

LLMs



school of ai Algiers

AI Camp

INTRODUCTION TO GENERATIVE MODELS:

GPT

T5

school of ai Algiers

AI Camp

TEXT GENERATION

USE CASE GPT 2

$$P(y_1, \ldots, y_t | \mathbf{x}) = \prod_{t=1}^{N} P(y_t | y_{<t}, \mathbf{x})$$

school of ai Algiers

AI Camp

TEXT GENERATION

school of ai Algiers

AI Camp

TEXT GENERATION

Methods of sampling

Greedy Search

Beam Search

Top-k and Nucleus Sampling

school of ai Algiers

AI Camp

TEXT GENERATION

Greedy search

- we can get the probability distribution over the next possible token w_i by taking the softmax :

$$P(y_t = w_i \mid y_{<t}, \mathbf{x}) = \text{softmax}(z_{t,i})$$

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\text{argmax}}\ P(\mathbf{y} \mid \mathbf{x})$$

school of ai Algiers

AI Camp

TEXT GENERATION

Greedy search

- we can get the probability distribution over the next possible token w_i by taking the softmax :

$$P(y_t = w_i \mid y_{<t}, \mathbf{x}) = \text{softmax}(z_{t,i})$$

$$\hat{\mathbf{y}} = \underset{\mathbf{y}}{\text{argmax}}\ P(\mathbf{y} \mid \mathbf{x})$$

school of ai Algiers

AI Camp

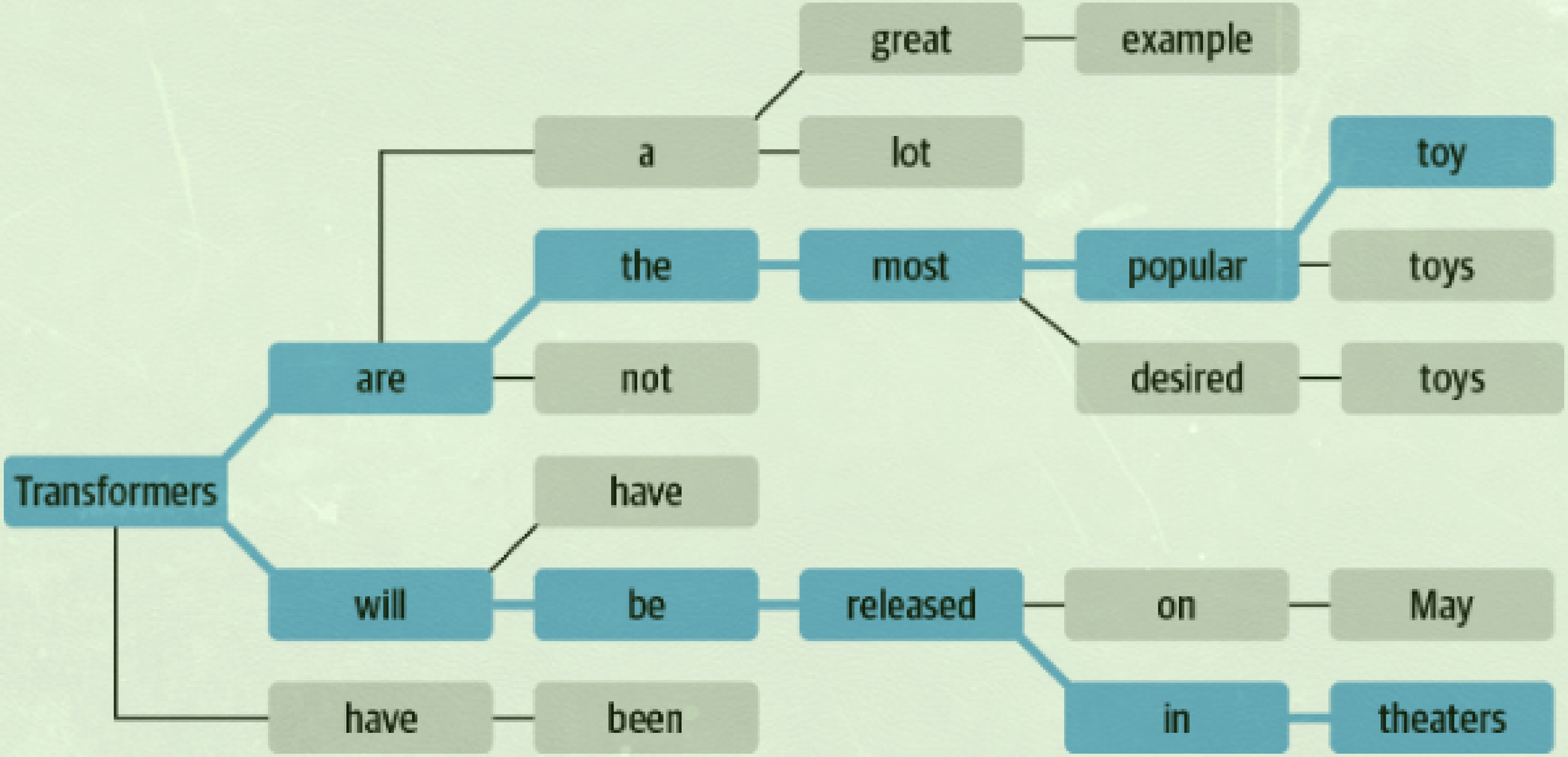
TEXT GENERATION

Beam Search

- beam search keeps track of the top-b most probable next tokens
- and the most likely sequence is selected by ranking the b beams according to their log probabilities

school of ai Algiers

AI Camp



school of ai Algiers

AI Camp

TEXT GENERATION

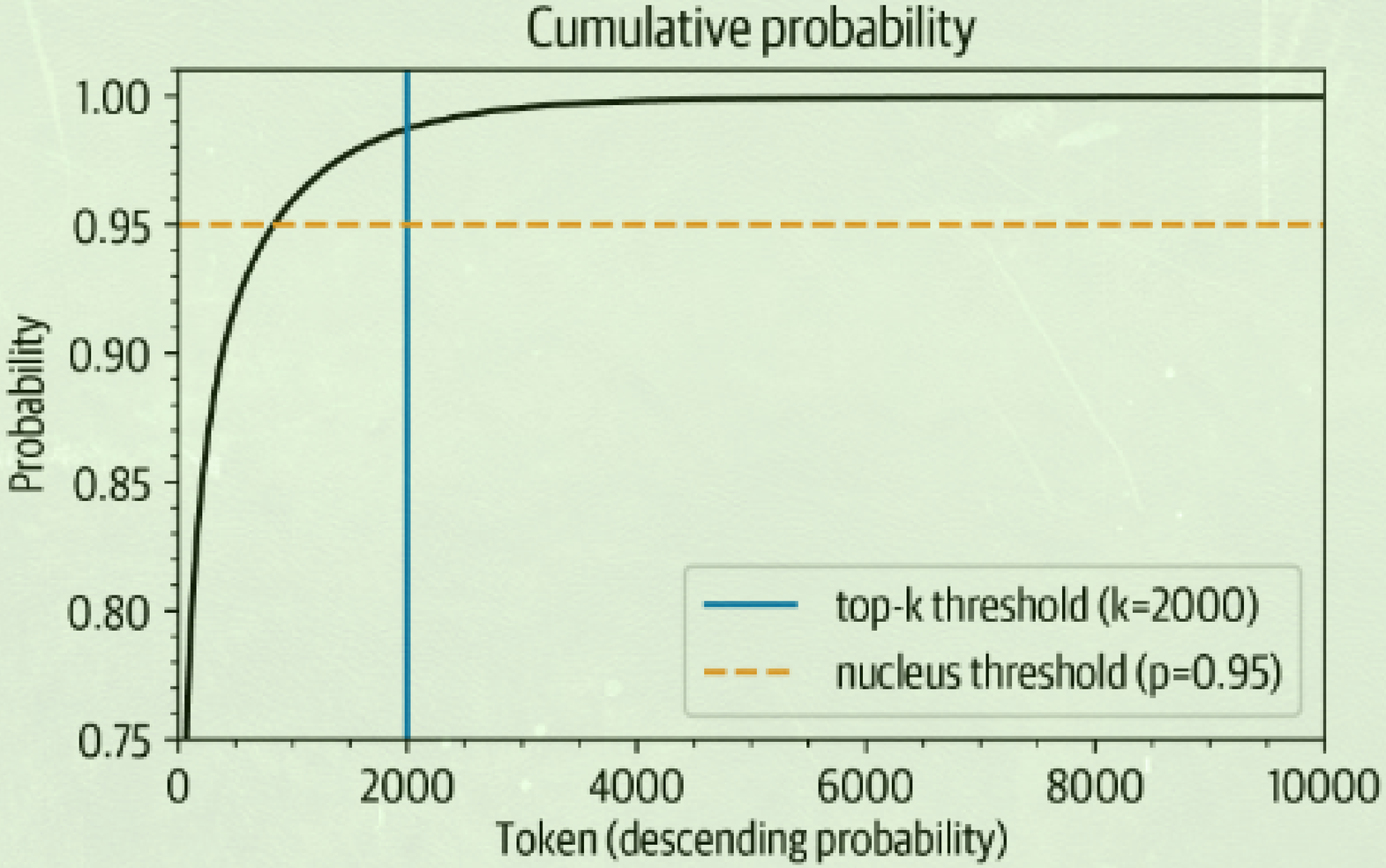
Top-p and Nucleus Sampling

- the basic idea is to restrict the number of possible tokens we can sample from at each timestep

school of ai Algiers

AI Camp

TEXT GENERATION



school of ai Algiers

AI Camp

BUT
HOW CAN WE EVALUATE OUR MODELS !!

school of ai Algiers

AI Camp

MEASURING THE QUALITY

Methods of sampling

BLEU

Perplexity

ROUGE.

school of ai Algiers

AI Camp

TEXT GENERATION

Bilingual Evaluation Understudy (BELU)

$$\mathrm{BP} = \begin{cases} 1 & \text{if } c > r \\ e^{(1-r/c)} & \text{if } c \leq r \end{cases}.$$

Then,

$$\mathrm{BLEU} = \mathrm{BP} \cdot \exp\left(\sum_{n=1}^{N} w_n \log p_n\right).$$

school of ai Algiers

AI Camp

TEXT GENERATION

ROUGE-1 Score

Model → "the hello a cat dog fox jumps"

reference text

"the fox jumps" → ['the', 'fox', 'jumps']

$$\frac{count_{match}(gram_n)}{count(gram_n)}$$

$$\frac{3}{3} = \textbf{100\%}\ \text{recall}$$

school of ai Algiers

AI Camp

TEXT GENERATION

ROUGE-1 Score

"the hello a cat dog fox jumps"

longest common subsequence length = **2**

"the fox jumps"

reference text

$$\frac{LCS(gram_n)}{count(gram_n)}$$

$$\frac{2}{7} = \textbf{29\%} \text{ precision}$$

school of ai Algiers

AI Camp

TEXT GENERATION

ROUGE-1 Score

"the hello a cat dog fox jumps"

['the', 'fox', 'jumps']

1.0 recall
0.43 precision

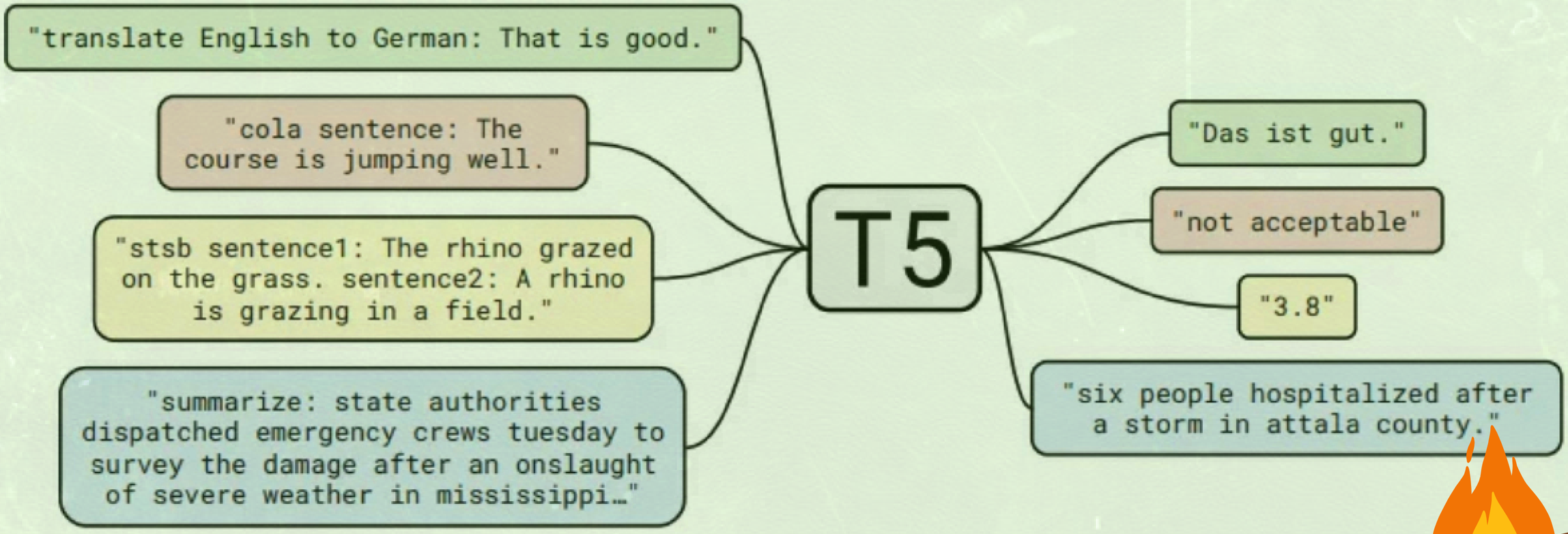
$$2 * \frac{0.43 * 1.0}{0.43 + 1.0} = 0.6$$

60% f1 score

school of ai Algiers

AI Camp

INTRODUCTION TO T5



school of ai Algiers

AI Camp

INTORDUCTION TO T5

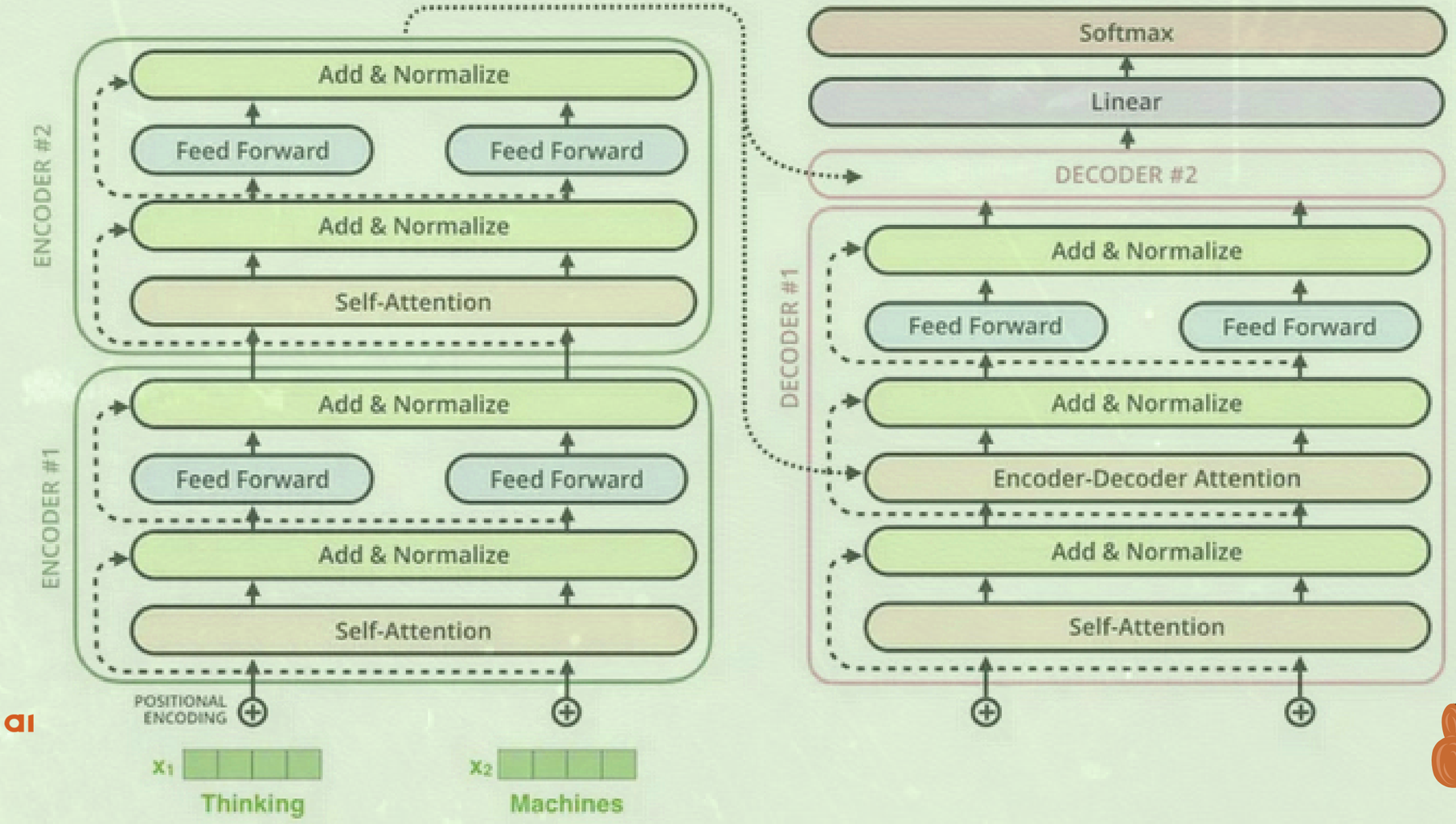
FACTS ABOUT T5

- The primary idea behind T5 is to convert all NLP tasks into a text-to-text format
- T5 reformulates existing transfer learning techniques into a unified format

school of ai Algiers

AI Camp

ARCHITECTURE OF THE T5 MODEL



school of ai Algiers

AI Camp

INTORDUCTION TO T5

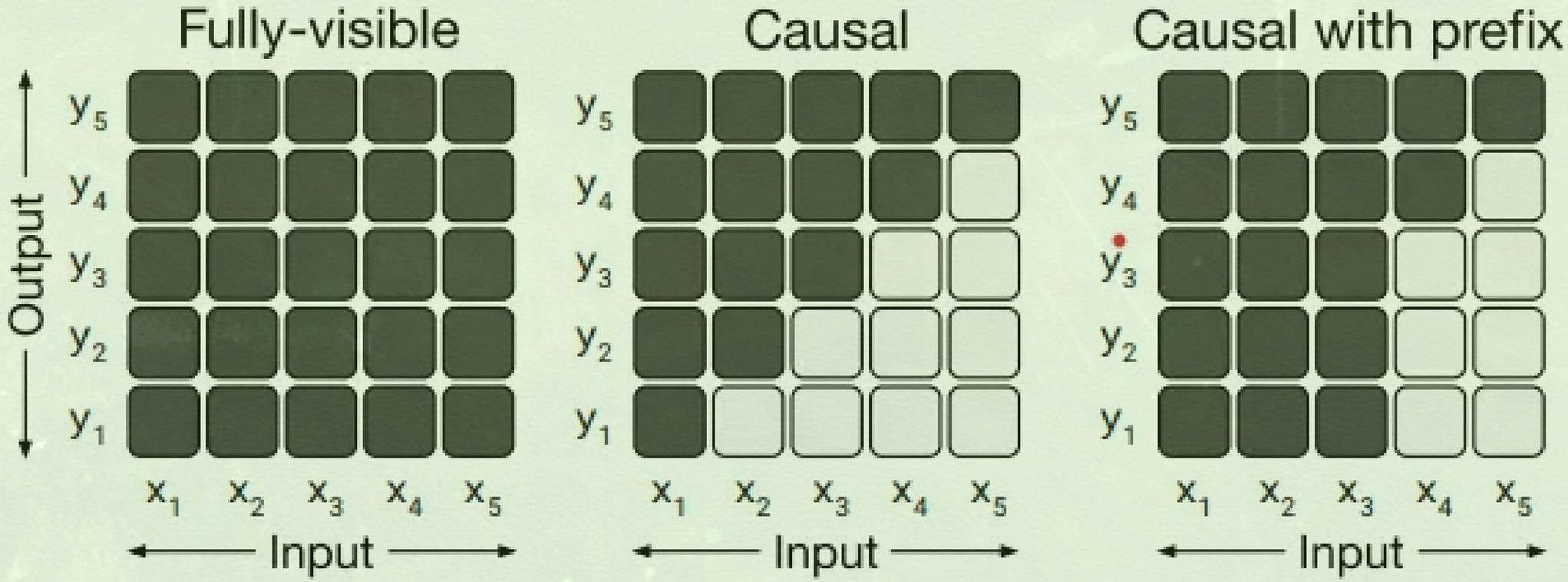
attention masks

- Variants of the transformer architecture have one major distinction: the type of masking used in their attention layers
- Certain applications (e.g., language modeling) require causal self-attention during training to prevent the transformer from “looking into the future”

school of ai Algiers

AI Camp

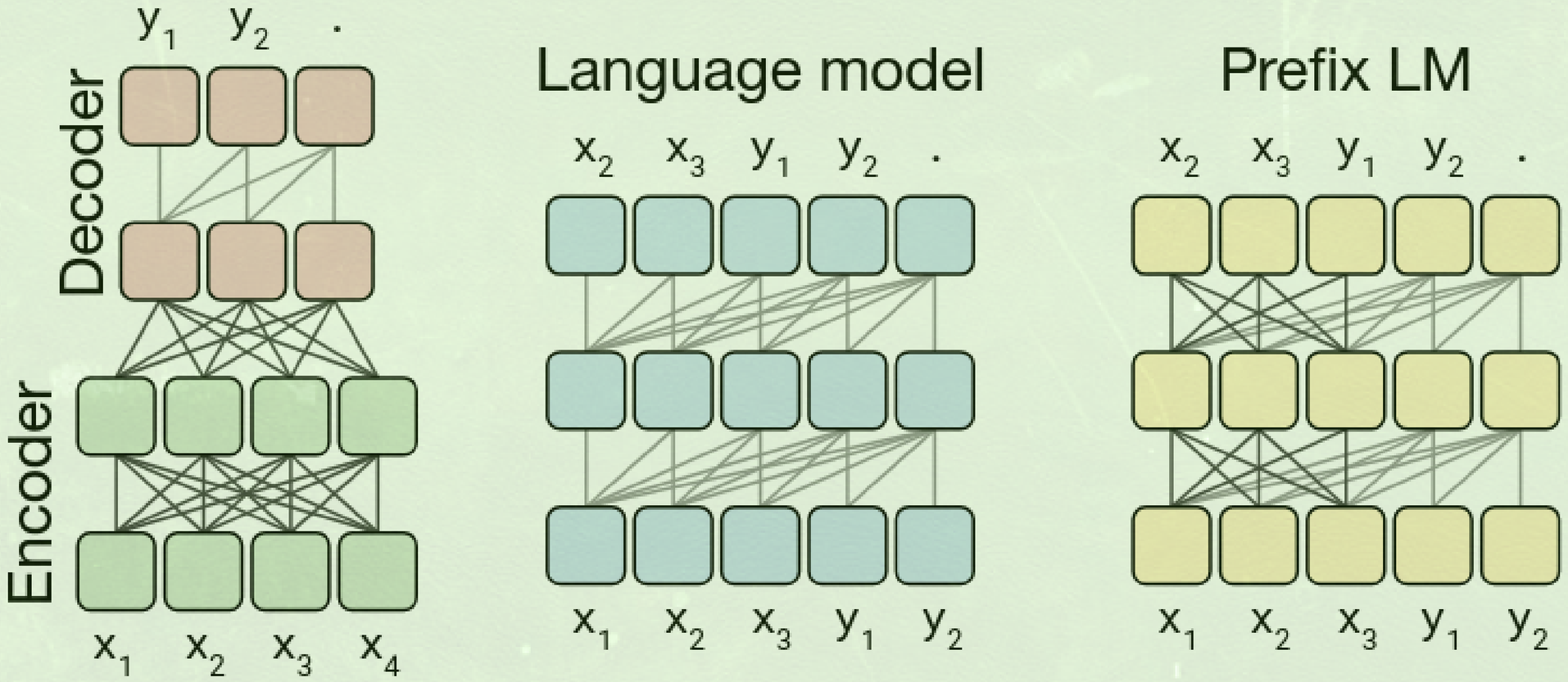
INTORDUCTION TO T5



school of ai Algiers

AI Camp

INTORDUCTION TO T5



school of ai Algiers

AI Camp

INTORDUCTION TO T5

	Architecture	Objective	Params	Cost	GLUE
★	Encoder-decoder	Denoising	$2P$	M	**83.28**
	Enc-dec, shared	Denoising	P	M	82.81
	Enc-dec, 6 layers	Denoising	P	$M/2$	80.88
	Language model	Denoising	P	M	74.70
	Prefix LM	Denoising	P	M	81.82
	Encoder-decoder	LM	$2P$	M	79.56
	Enc-dec, shared	LM	P	M	79.60
	Enc-dec, 6 layers	LM	P	$M/2$	78.67
	Language model	LM	P	M	73.78
	Prefix LM	LM	P	M	79.68

school of ai Algiers

AI Camp

INTORDUCTION TO T5

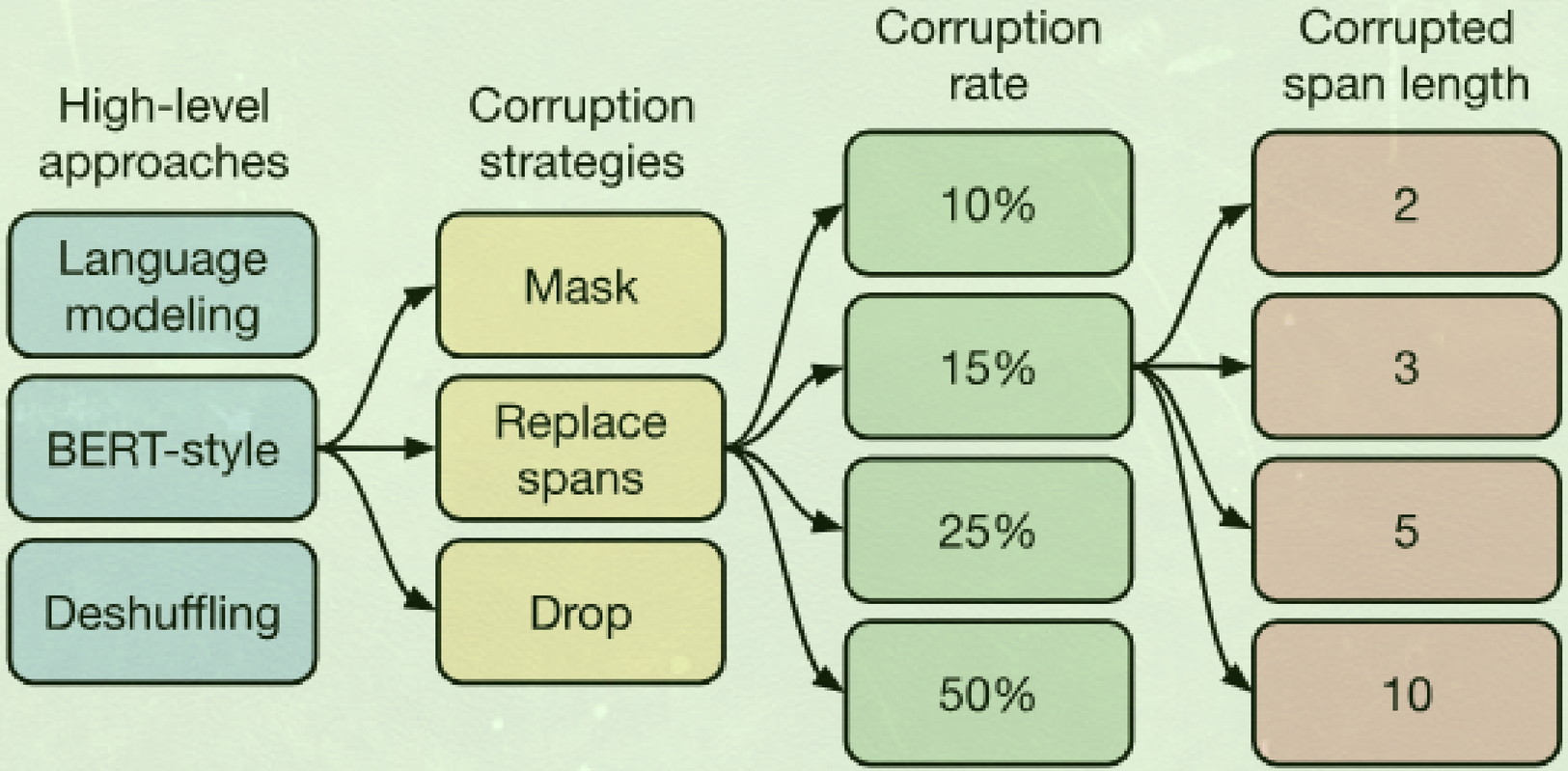
TRAINING T5

- The T5 model is pre-trained on a total of 34B tokens from the C4 corpus
- Randomly selects 15% of tokens in the input sequence and Replaces all consecutive spans of selected tokens with a single ”sentinel” token

school of ai Algiers

AI Camp

INTORDUCTION TO T5



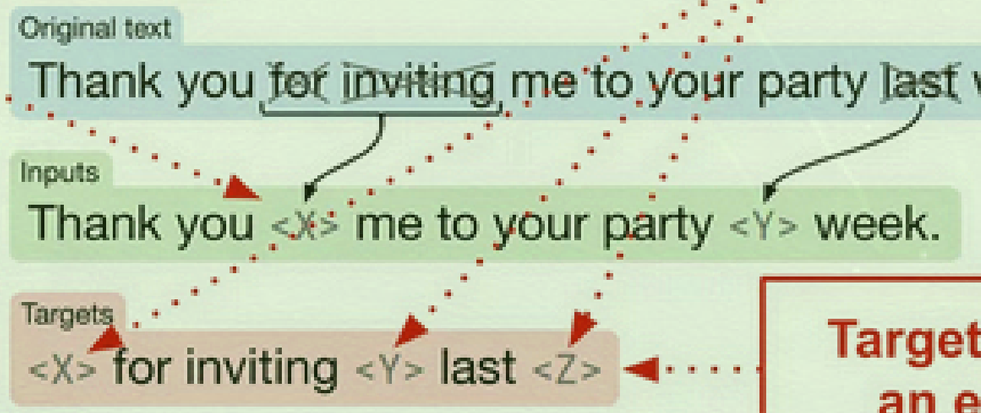
school of ai Algiers

AI Camp

INTORDUCTION TO T5

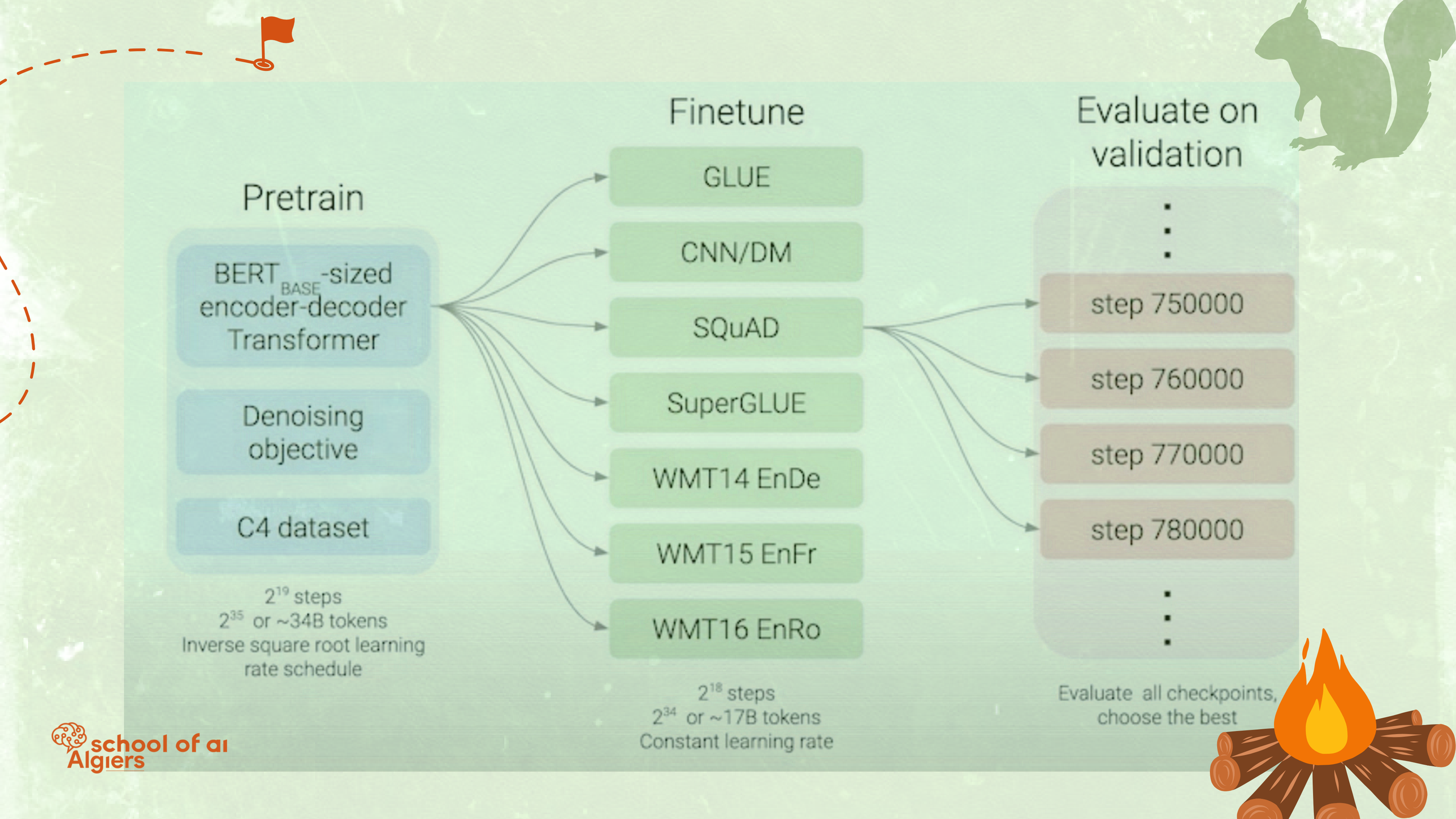
Sequences of tokens are assigned the same sentinel token

Each sentinel token is unique within the sequence



Target sequence ends with an extra sentinel token

school of ai Algiers


Pretrain
BERT BASE -sized encoder-decoder Transformer
Denoising objective
C4 dataset
2^19 steps
2^35 or ~34B tokens
Inverse square root learning rate schedule
Finetune
GLUE
CNN/DM
SQuAD
SuperGLUE
WMT14 EnDe
WMT15 EnFr
WMT16 EnRo
2^18 steps
2^34 or ~17B tokens
Constant learning rate
Evaluate on validation
step 750000
step 760000
step 770000
step 780000
Evaluate all checkpoints, choose the best
school of ai Algiers

AI Camp

school of ai Algiers

AI Camp

THANK YOU!

school of ai Algiers