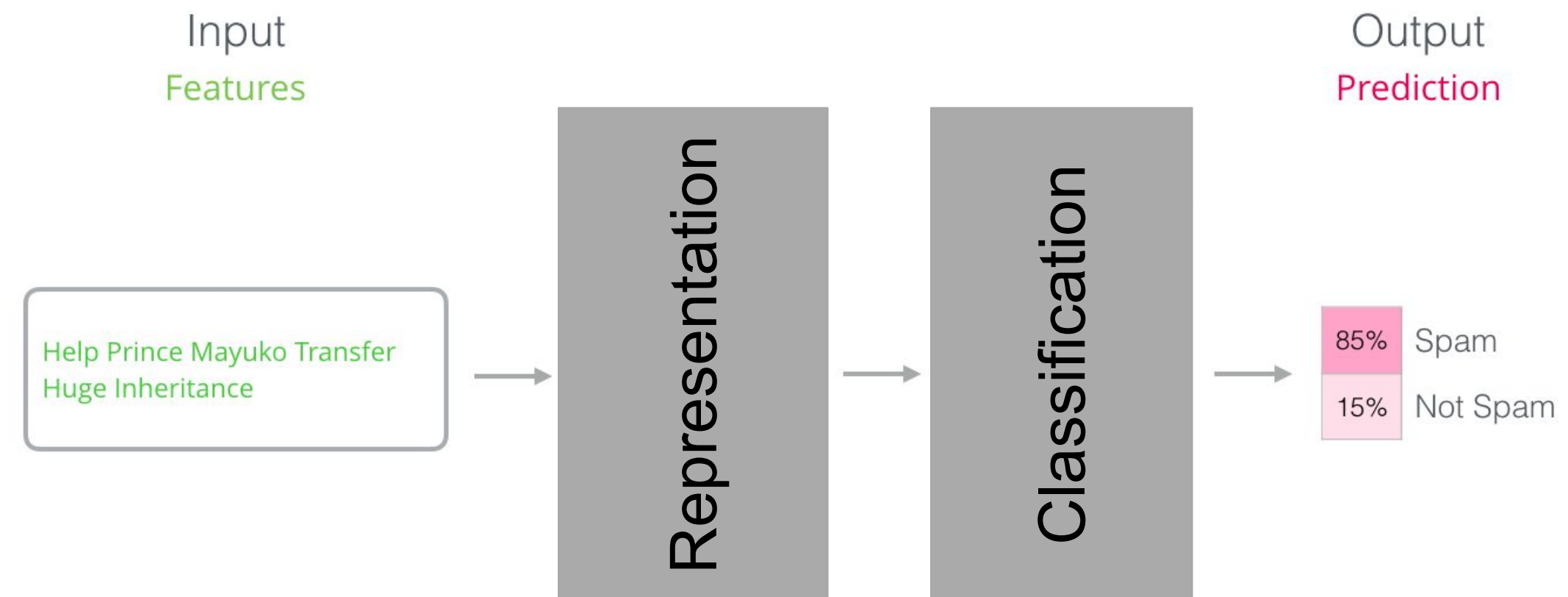

Transformers

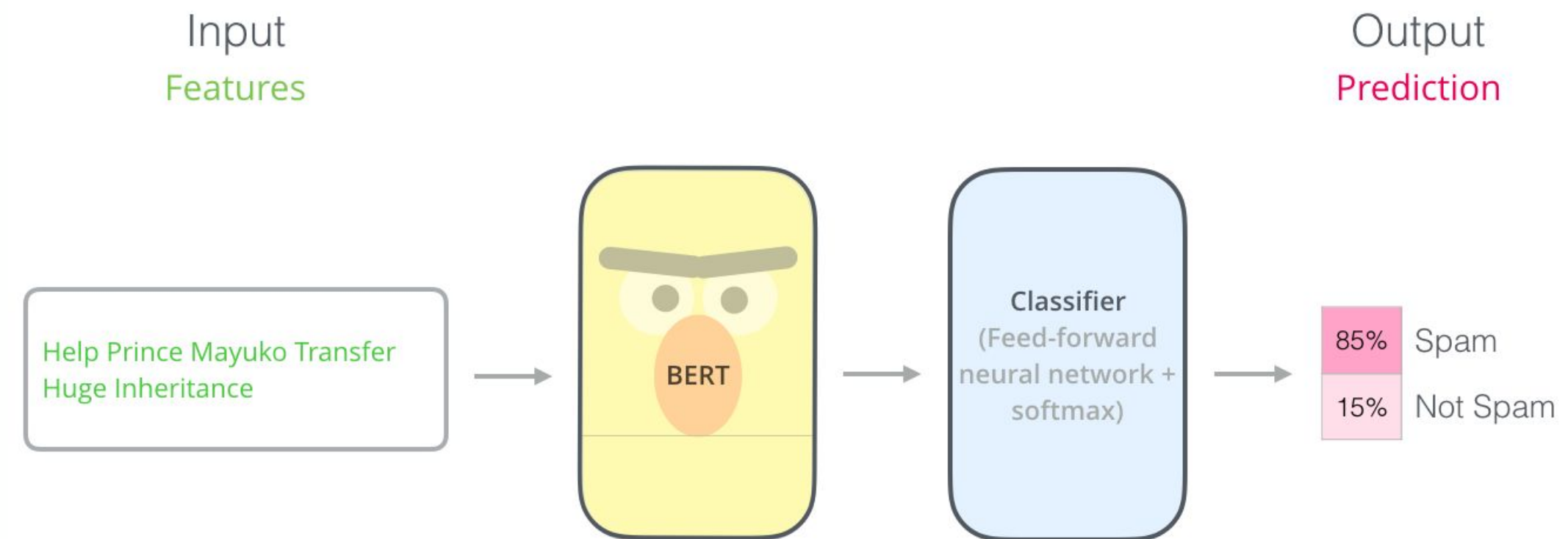
leo@ideai.io

[LinkedIn](#)

SUPERVISED - NLP



SUPERVISED - NLP



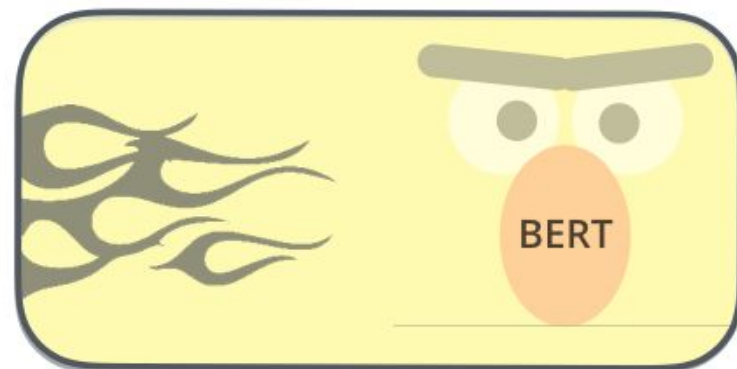
SEMI-SUPERVISED

1 - **Semi-supervised** training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:



Dataset:



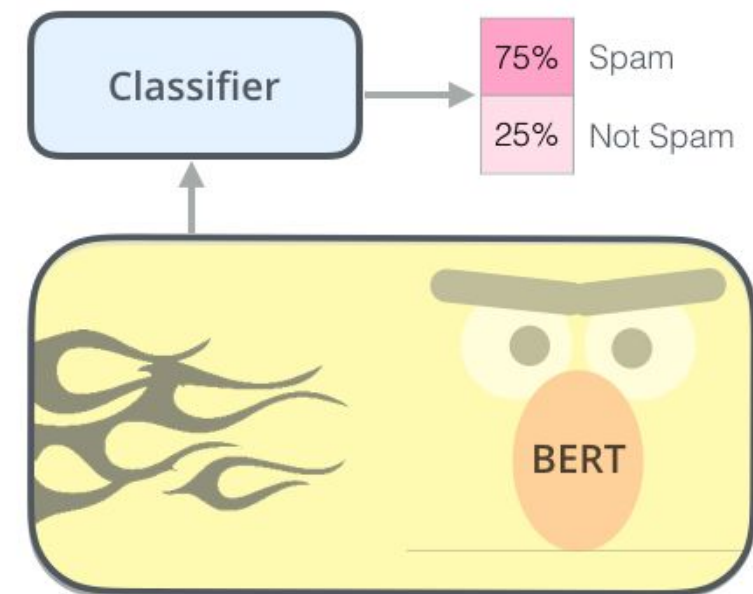
Objective:

Predict the masked word
(language modeling)

2 - **Supervised** training on a specific task with a labeled dataset.

Supervised Learning Step

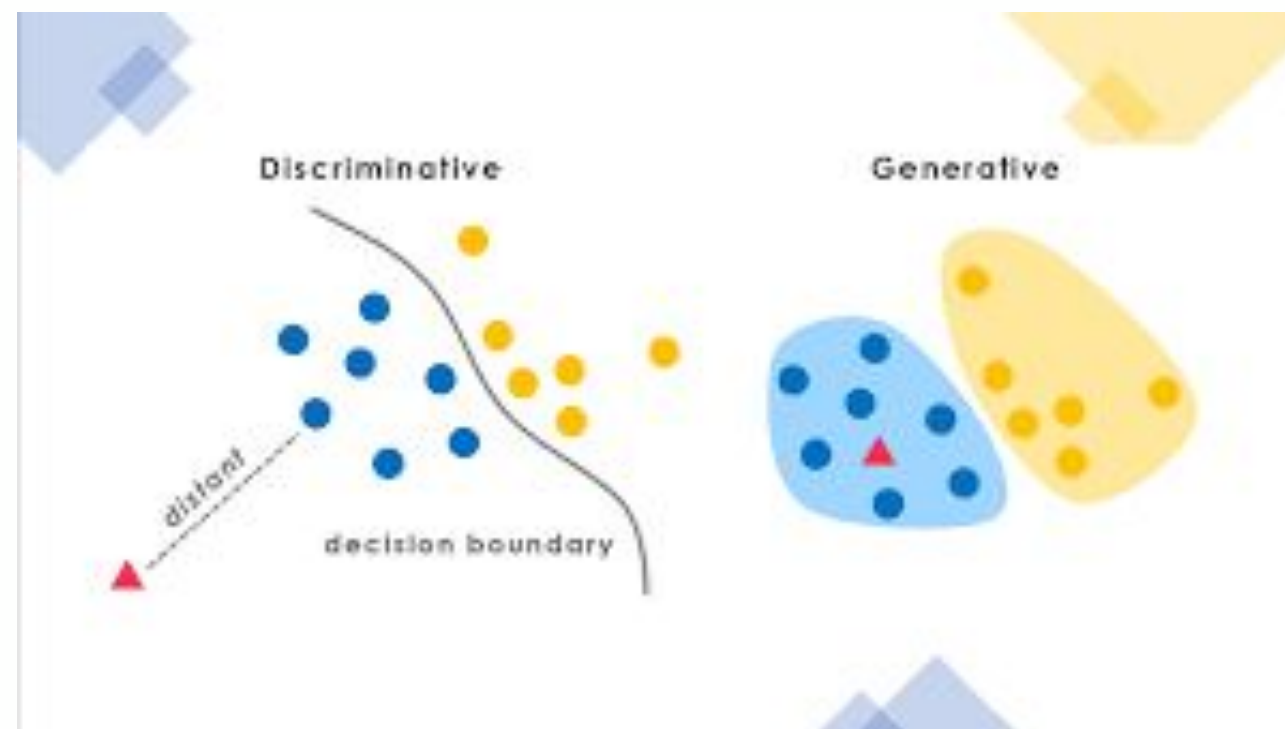
Model:
(pre-trained
in step #1)



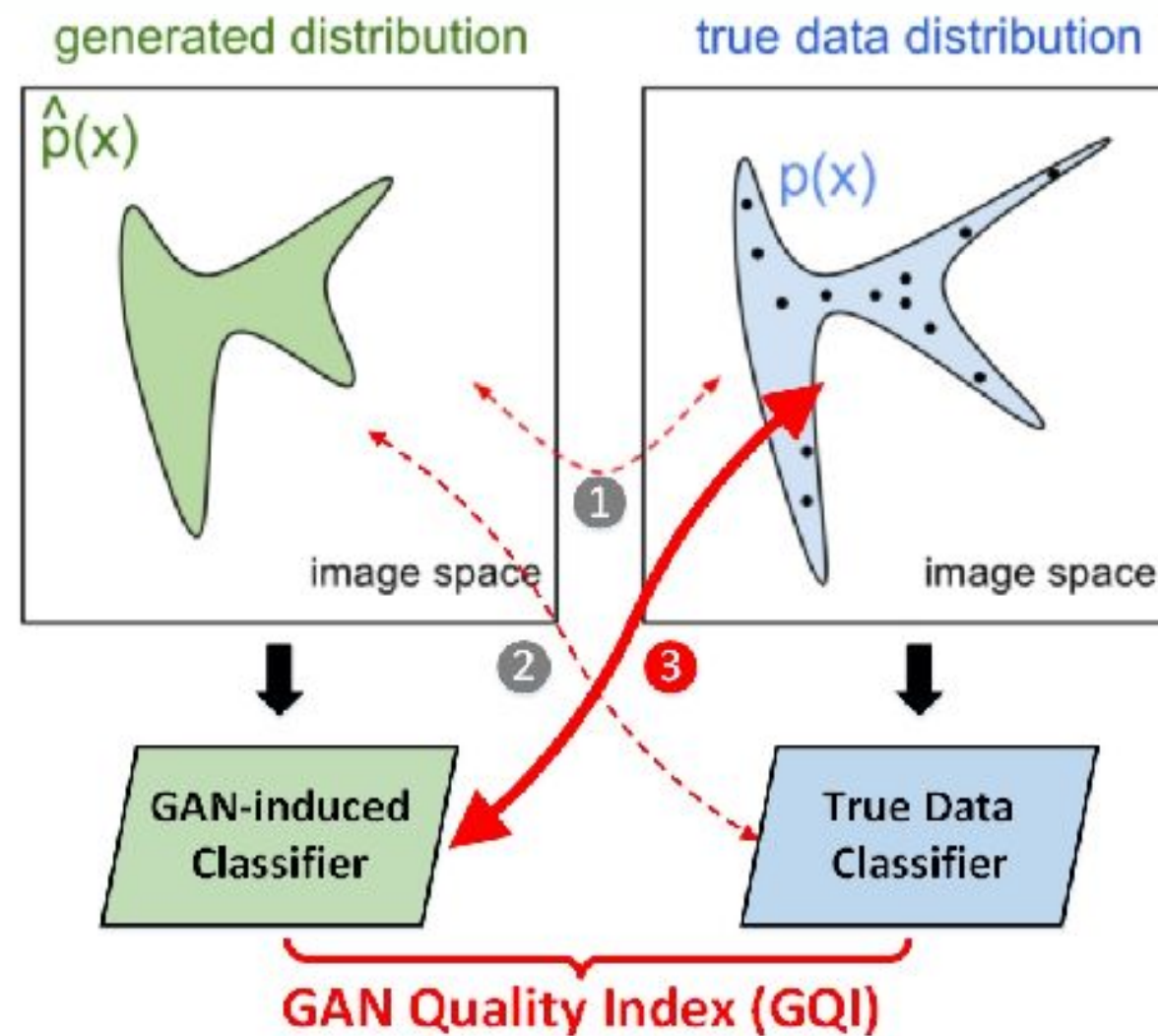
Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached...	Not Spam

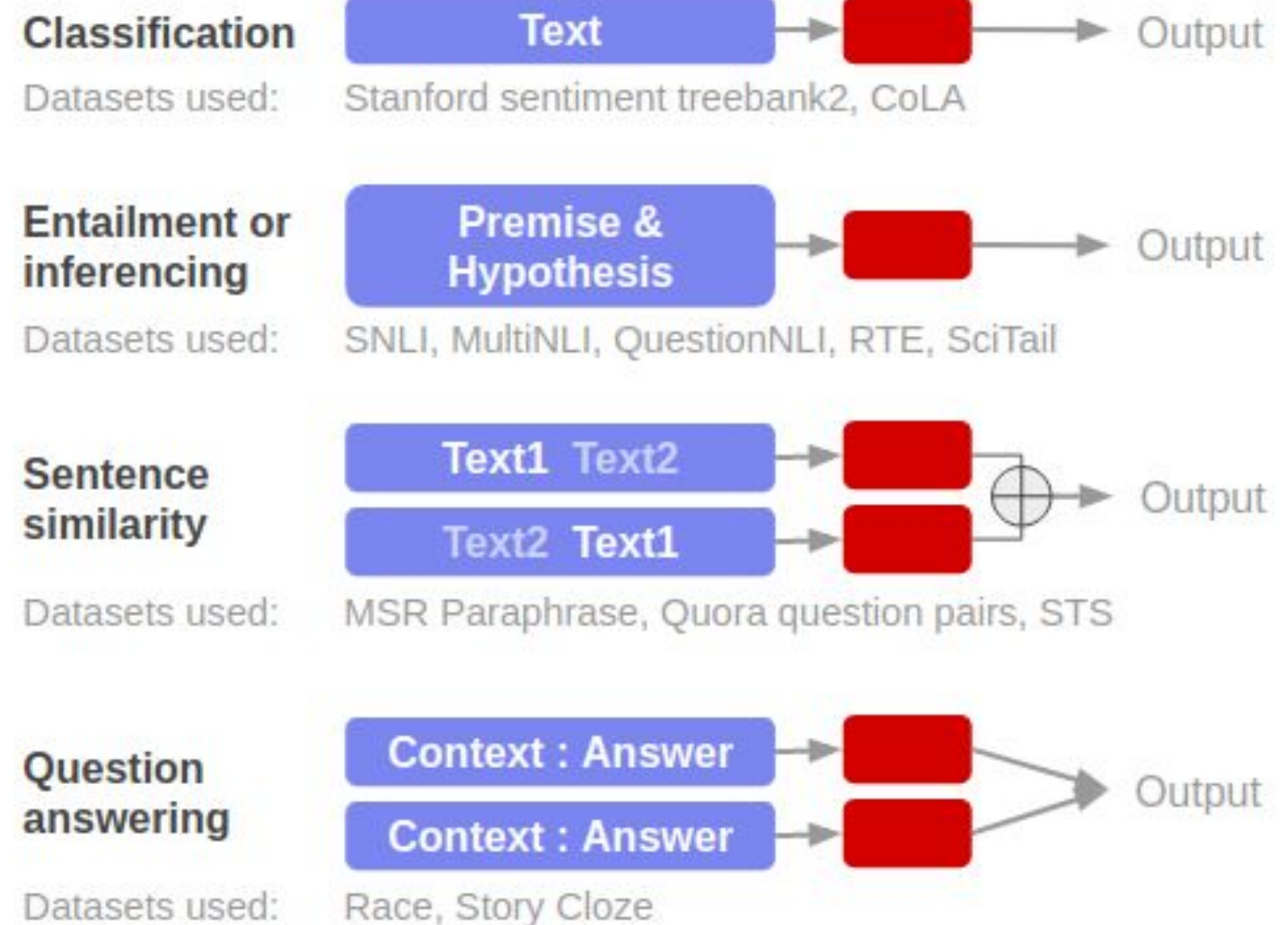
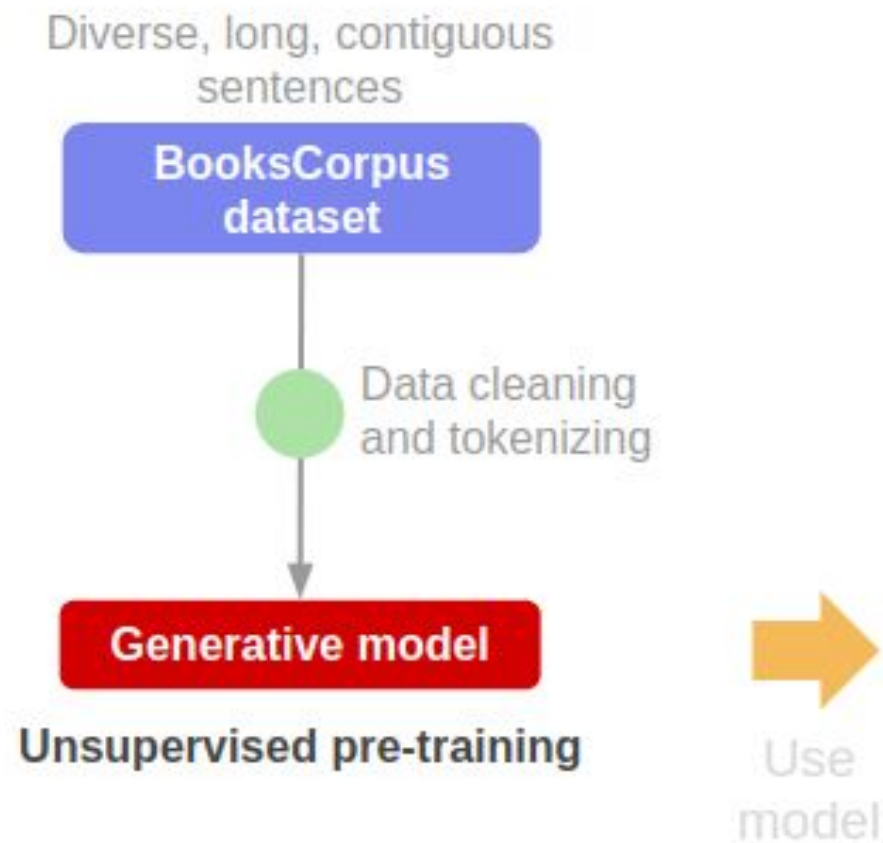
GENERATIVE



GENERATIVE



GENERATIVE



Supervised discriminative fine-tuning tasks

GENERATIVE - GPT2

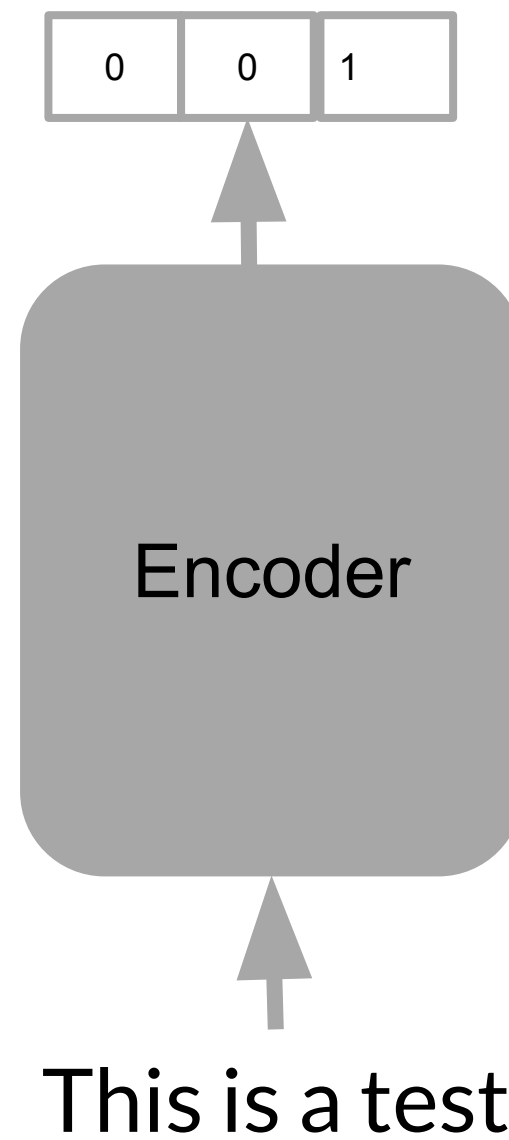
`python src/generate_unconditional_samples.py`

Most people think that when a warship runs aground it doesn't just kill people and then sink or burn all of society. That would make L.S. Kaminsky blush. Yet Kaminsky is doing one thing right: the CREAPH presidency. Whoever he is that fired the salt gun after getting thrown out of the Senate tossup race here in Richmond, he runs the "war," real, that is, guys like Alvin Dream, Dennis Hastert and Vijay Swarup. Given that Ed Gillespie, the GOP nominee barely a month into the campaign, on May 2 earned 45 points from the Tea Partiers, secessionists and nativities, right much everyone under 30 has been cheering the idea of "the war." Elliot Abrams, one of the Campus Reform editorial staff writers, also called the "war" mundane in the broadest terms. "Oh, well with them suiting it up voting be vigilant. And that produces 14 Rand Paul a grand total of 50 but Johnson 53. Two just finished with a 45 for Johnson 46. 'Well I hope it keeps getting led!' ...

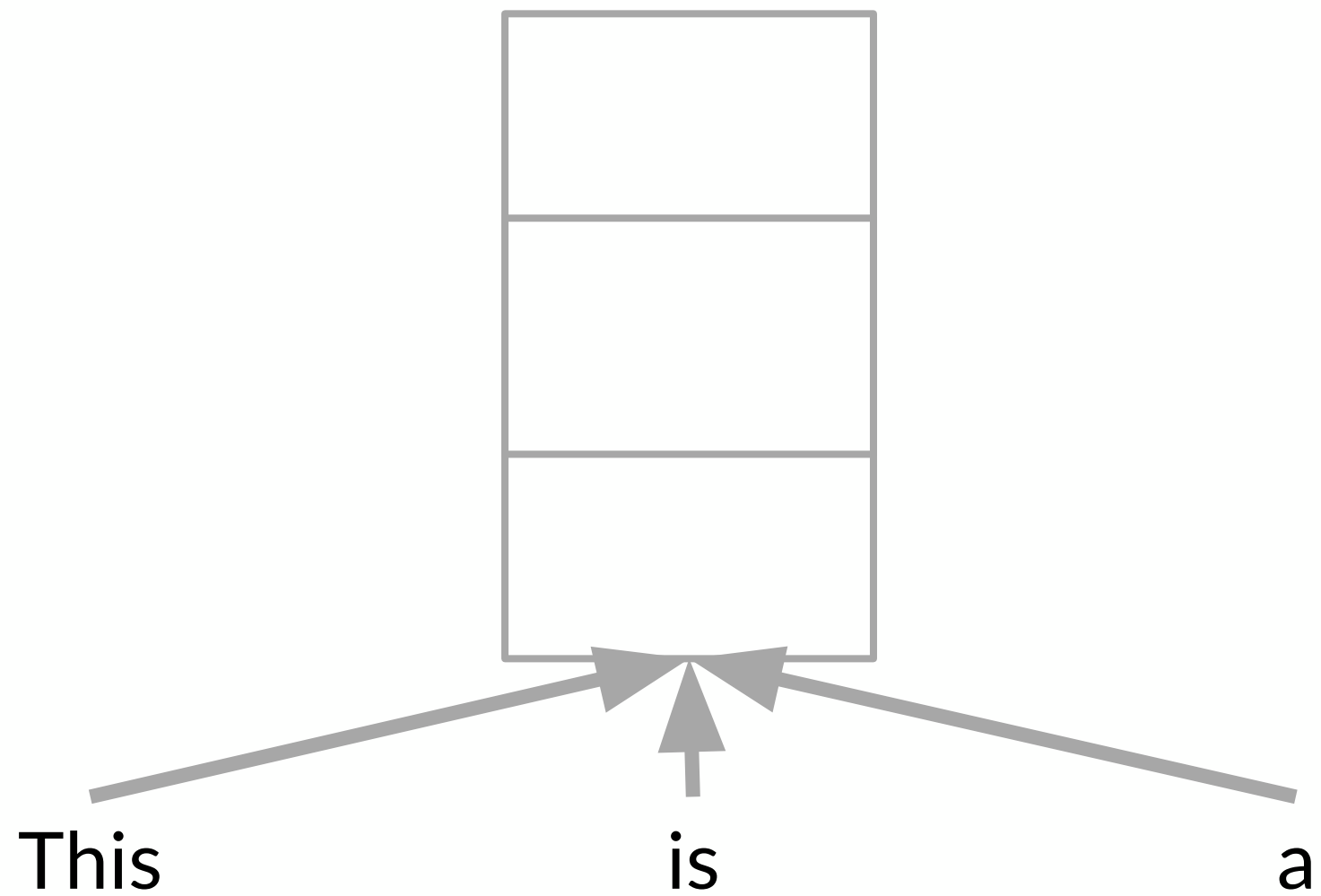


ENCODER

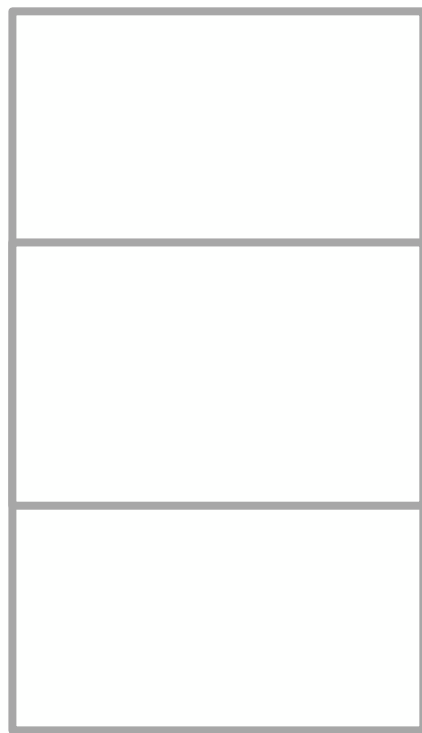
ENCODER



SELF-ATTENTION



REPRESENTATION



↑
This



↑
is



↑
a

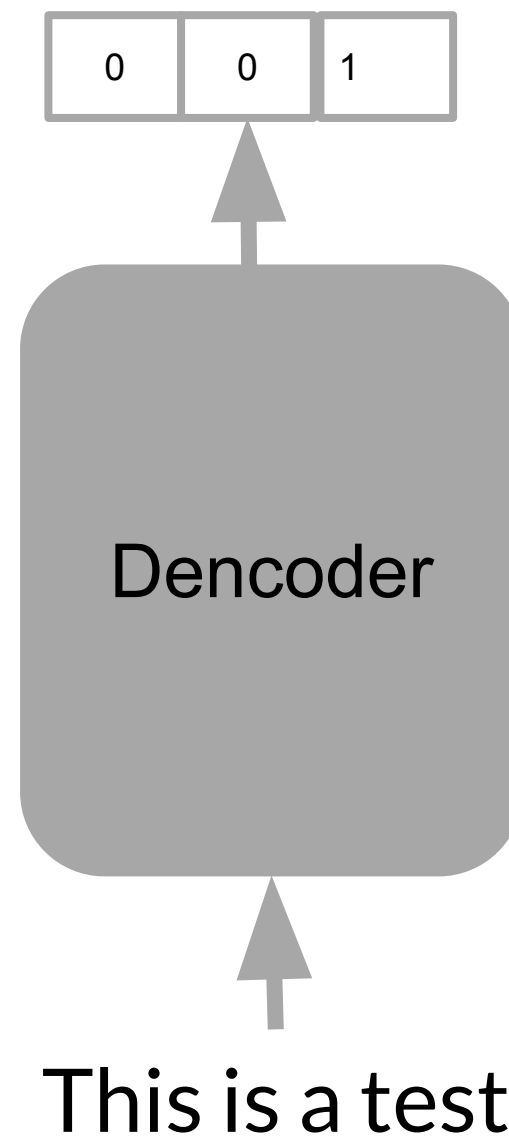
WHEN TO USE IT

- Bi-directional access (online)
- Extracting meaningful informations
- Question/answering
- Examples: BERT, RoBERTa

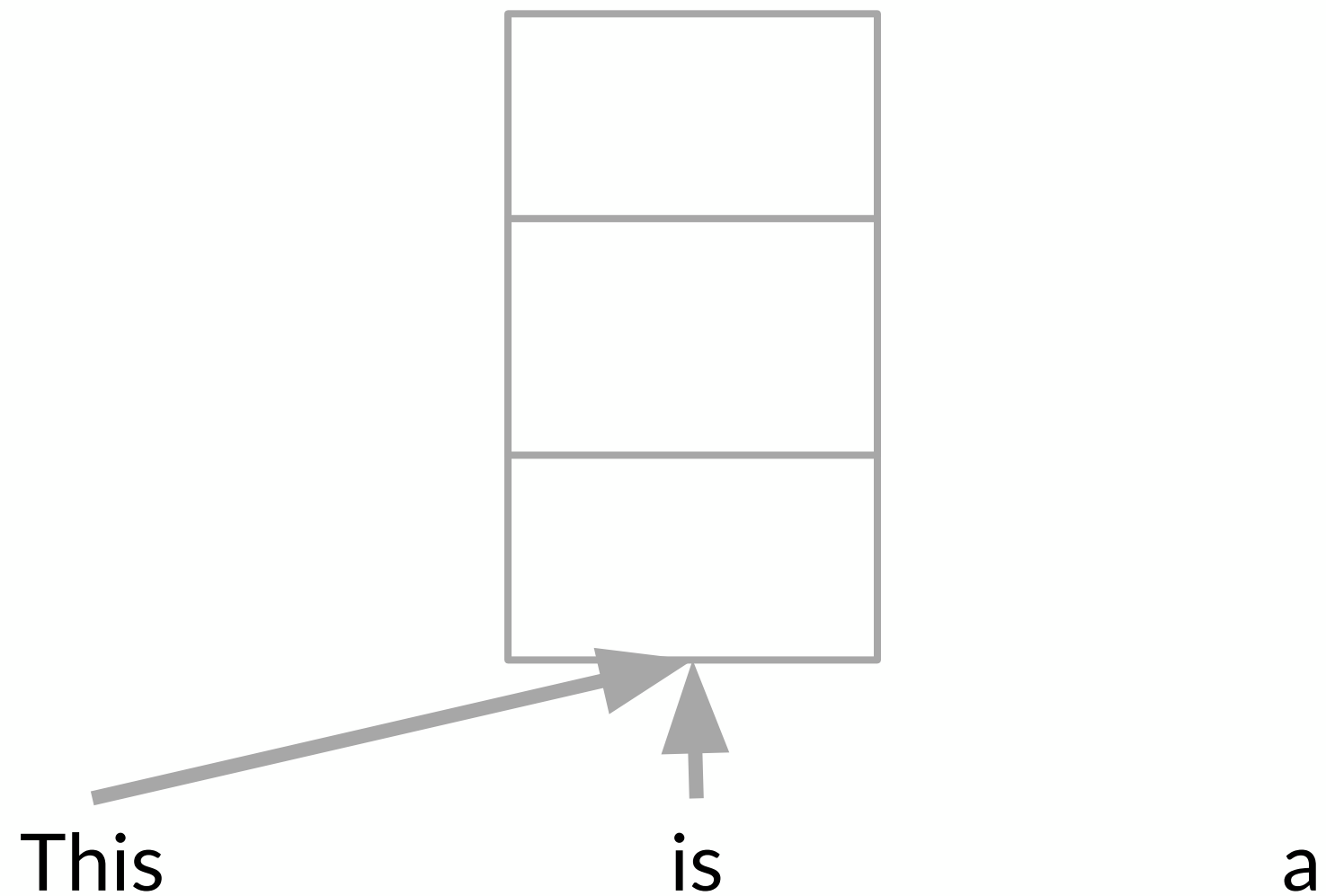


DECODER

DECODER



MASKED SELF-ATTENTION



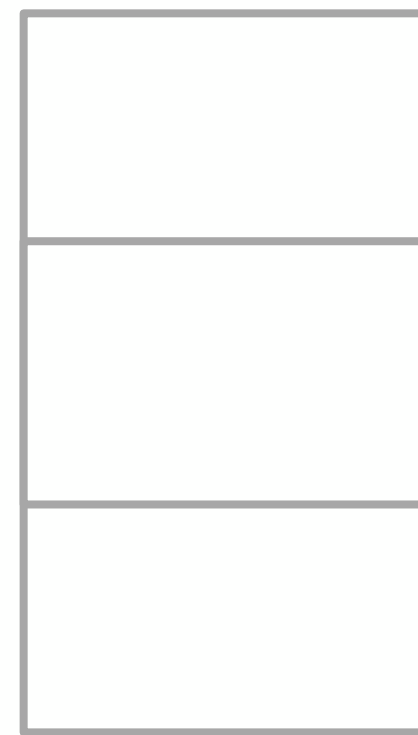
MASKED SELF-ATTENTION



This



is



a

WHEN TO USE IT

- Unidirectional access (online)
- Generation
- Examples: GPT-2



ATTENTION

WHEN TO USE IT

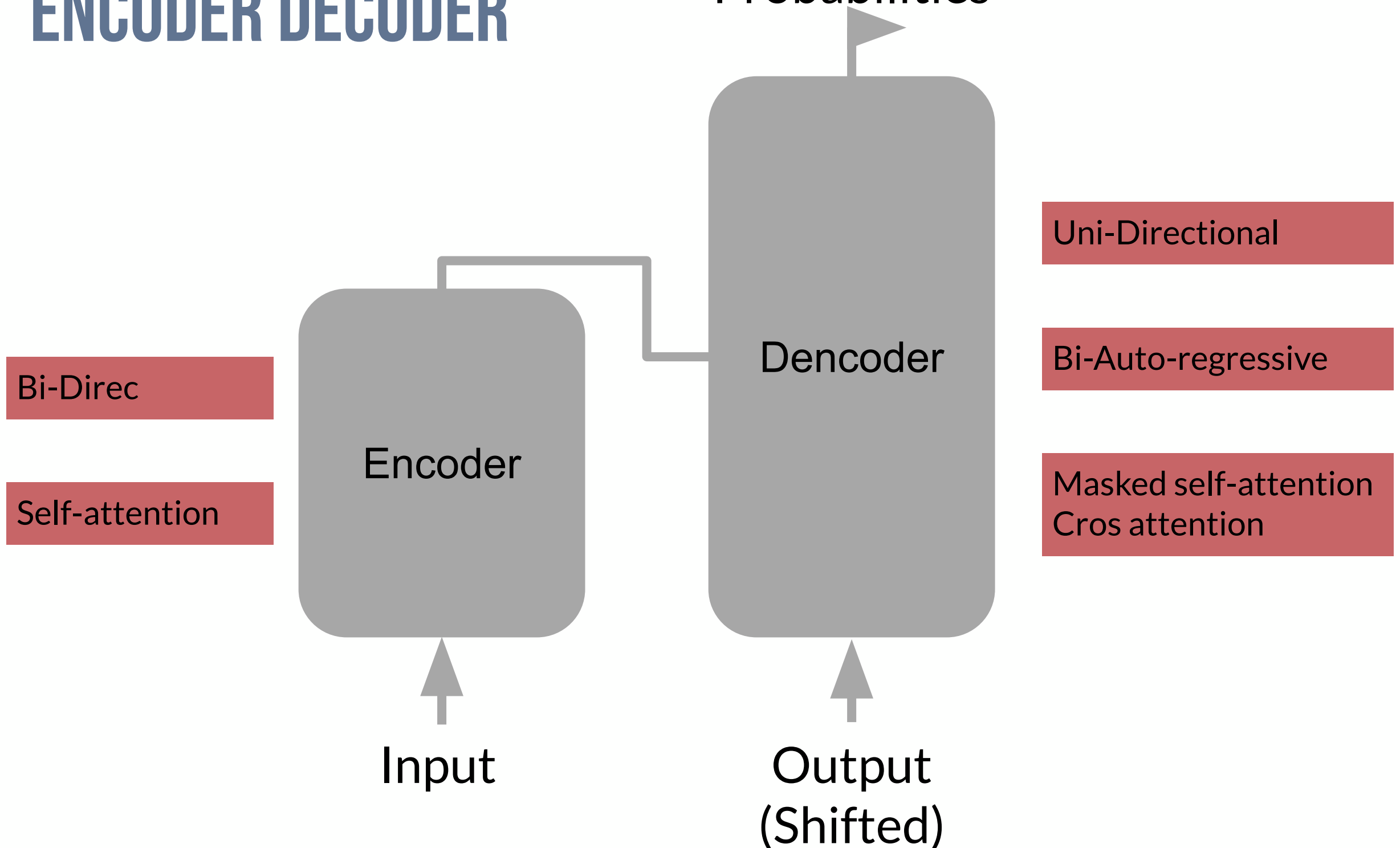
- Unidirectional access (online)
- Generation
- Examples: GPT-2



ENCODER-DECODER

ENCODER DECODER

Probabilities



WHEN TO USE IT

- Sequence to sequence
- Not shared weights
- Example: BART, T5
- Use different Enco/Decoders
- Bert/Bert, Bert/GPT-2



QUESTIONS



THANK YOU!

You can contact me at

www.ideai.io info@ideai.io