Large Language Models From ZERO To HERO

NLP

NLP is a subfield of AI that focuses on developing technologies that are capable of understanding, interpreting, and generating human language.

LANGUAGE MODEL

A Language Model (LM) is any model that learns to predict the next word (or token) in a sequence – based on what came before.

LANGUAGE MODEL

CORE OBJECTIVE

$$P(\text{next\_token} \mid \text{previous\_tokens})$$

TYPES OF LANGUAGE MODELS

STATS ERA

NEURAL ERA

TRANSFORMERS

TYPES OF LANGUAGE MODELS

STATS ERA

N-grams, Markov Chains

NEURAL ERA

RNN, LSTM

TRANSFORMERS

ENCODER-DECODER, DECODER-ONLY

LARGE LANGUAGE MODELS

A Large Language Model (LLM) is a Transformer-based Language Model that's trained at massive scale.
So while the core idea is still "predict the next token," the scale and architecture make all the difference.

TYPES OF LLMS

ENCODER-DECODER

T5

ENCODER ONLY

BERT

DECODER ONLY

GPT

TOKENIZATION & TOKEN ID

Tokenization is the process of breaking down text into smaller units called tokens. These tokens can be words, subwords, characters, or even sentences, depending on the level of granularity required for a specific application.

TOKENIZATION & TOKEN ID

There's one more step involved after breaking down the text into tokens, which is converting these tokens into numerical representations (like token IDs) that the model can process.

TOKENIZER

Tokenizer is a tool or algorithm that performs tokenization. Every language model has its own tokenizer that is specifically designed to work with that model. That's why when you use different models, you often need to use their respective tokenizers to ensure compatibility.

EMBEDDINGS

Embeddings are vector representations of words that attempt to capture its meaning.

EMBEDDINGS

TYPES OF EMBEDDINGS

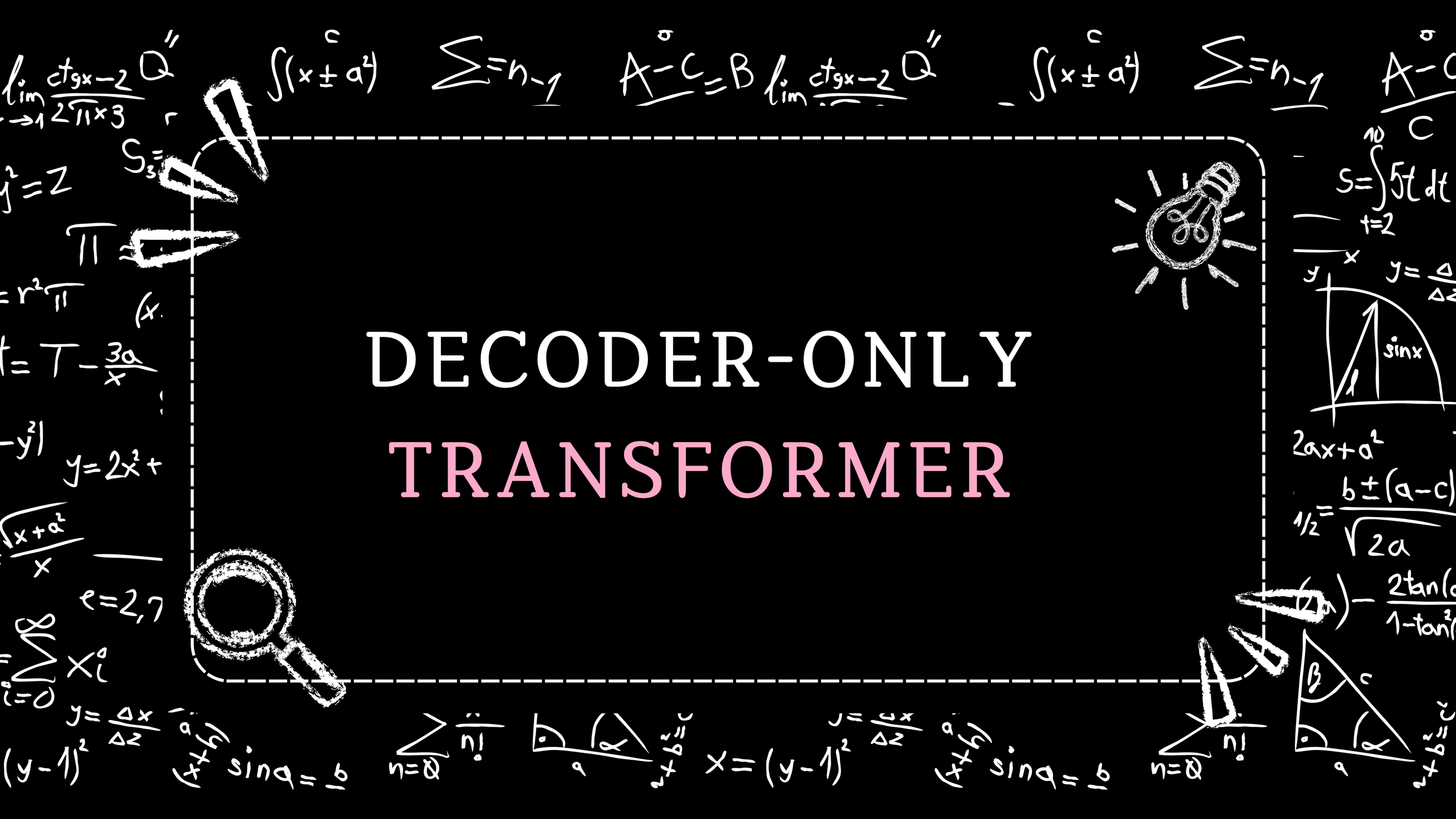
STATIC

Same and constant for all the processing.

CONTEXTUAL

It changes as per the sentence and context

DECODER-ONLY TRANSFORMER



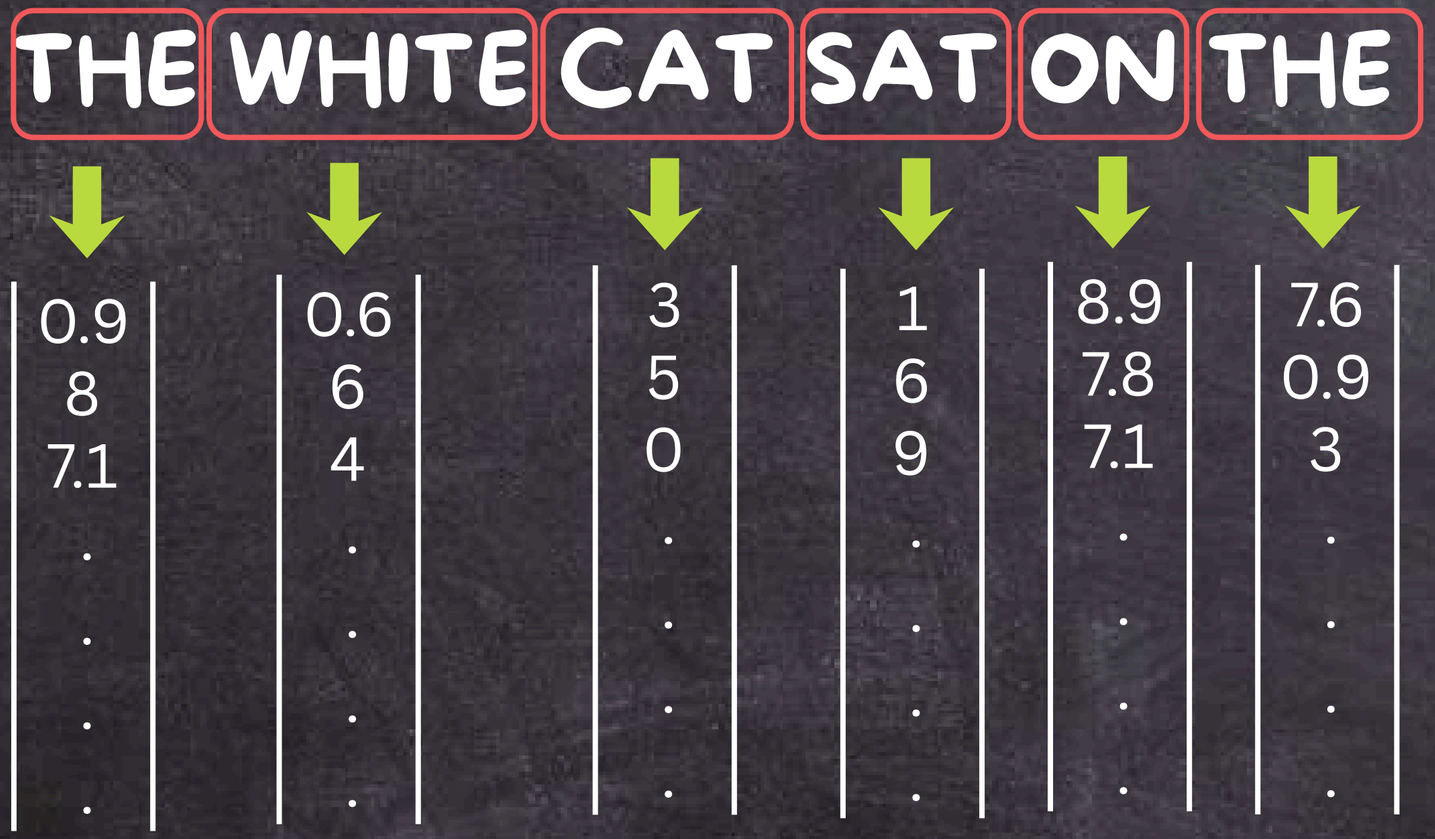
EXAMPLE

THE WHITE CAT SAT ON THE

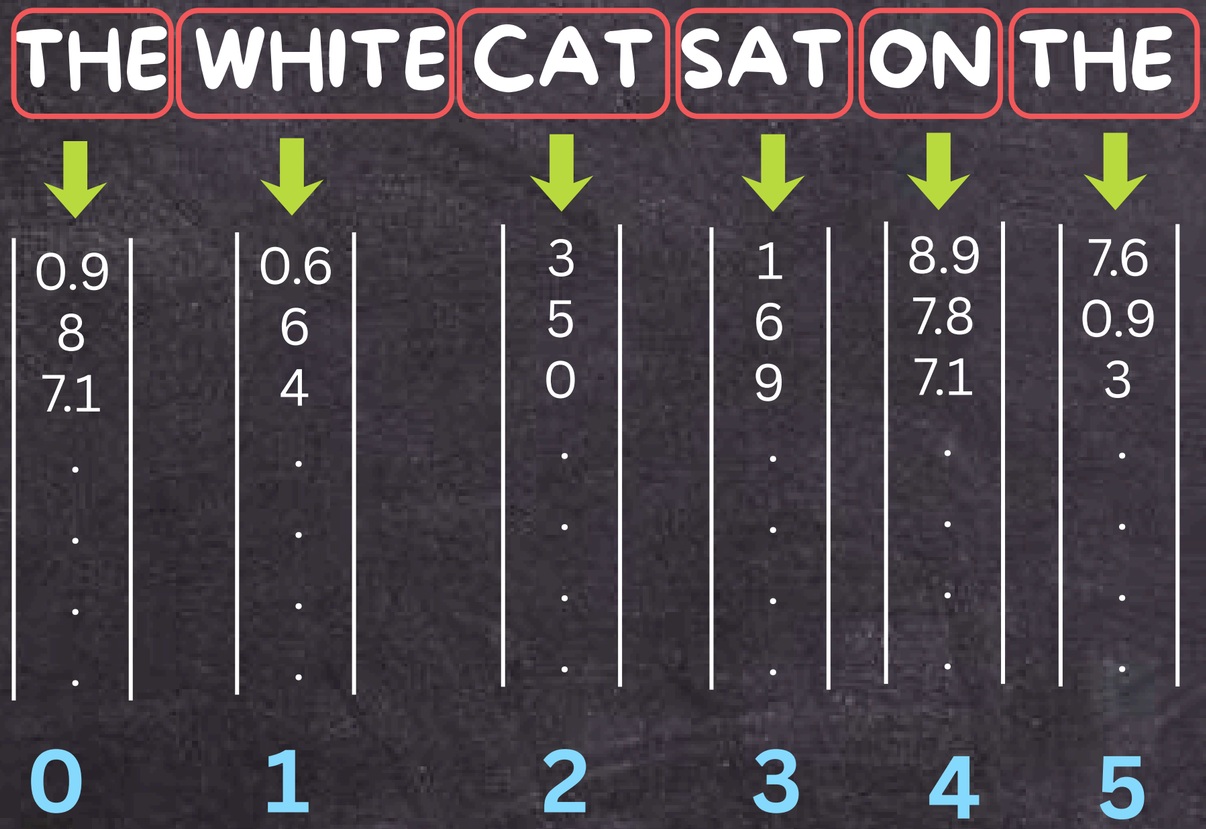
TOKENIZER

THE WHITE CAT SAT ON THE

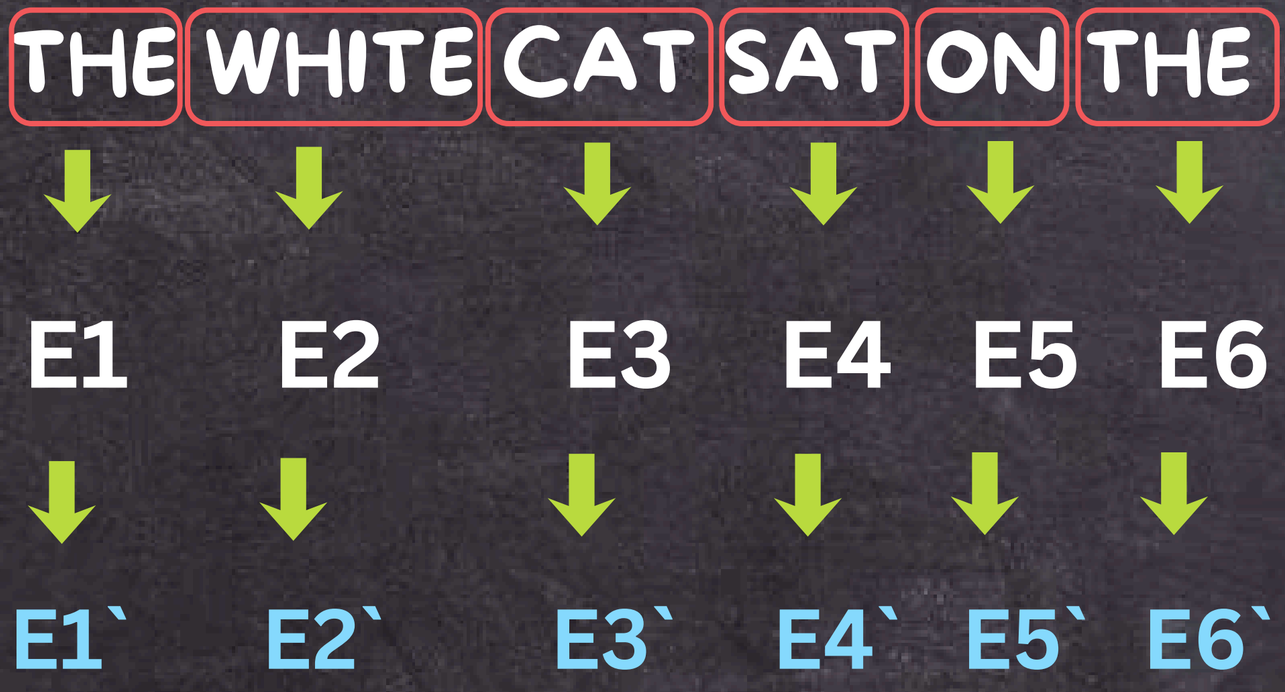
STATIC EMBEDDING

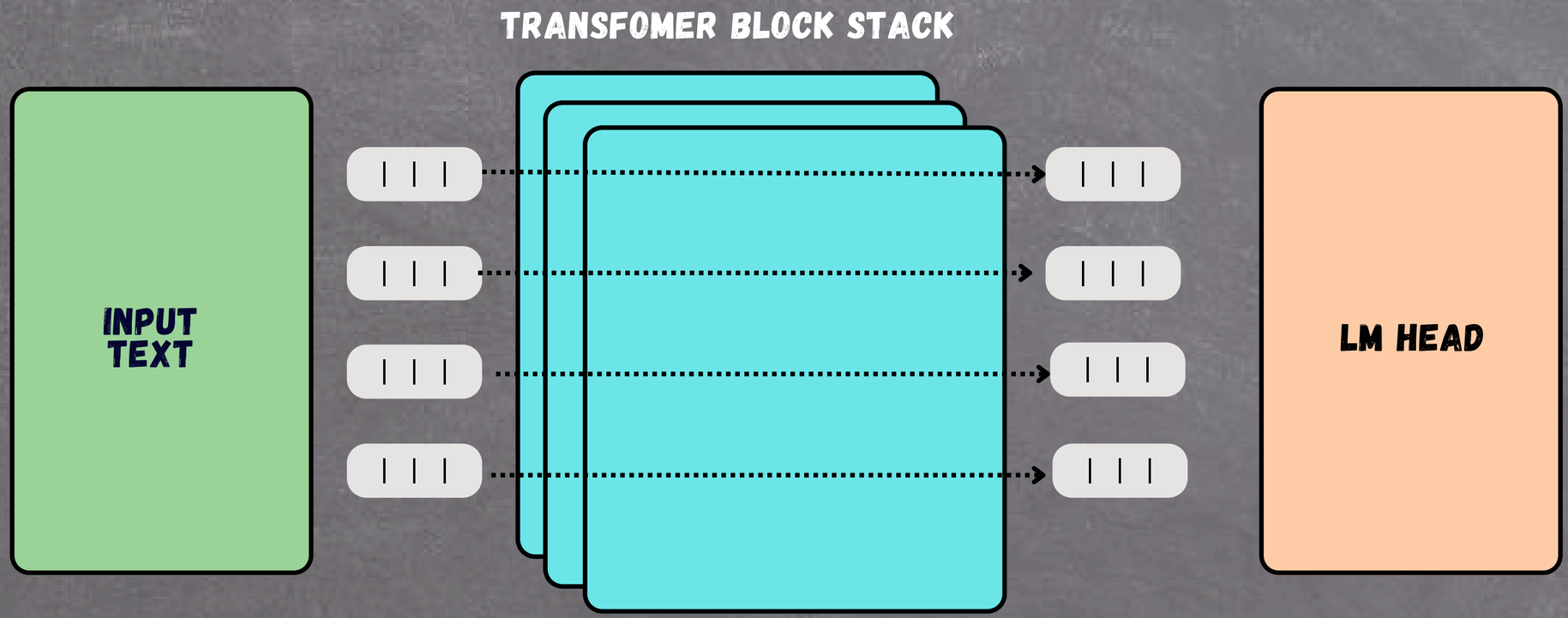


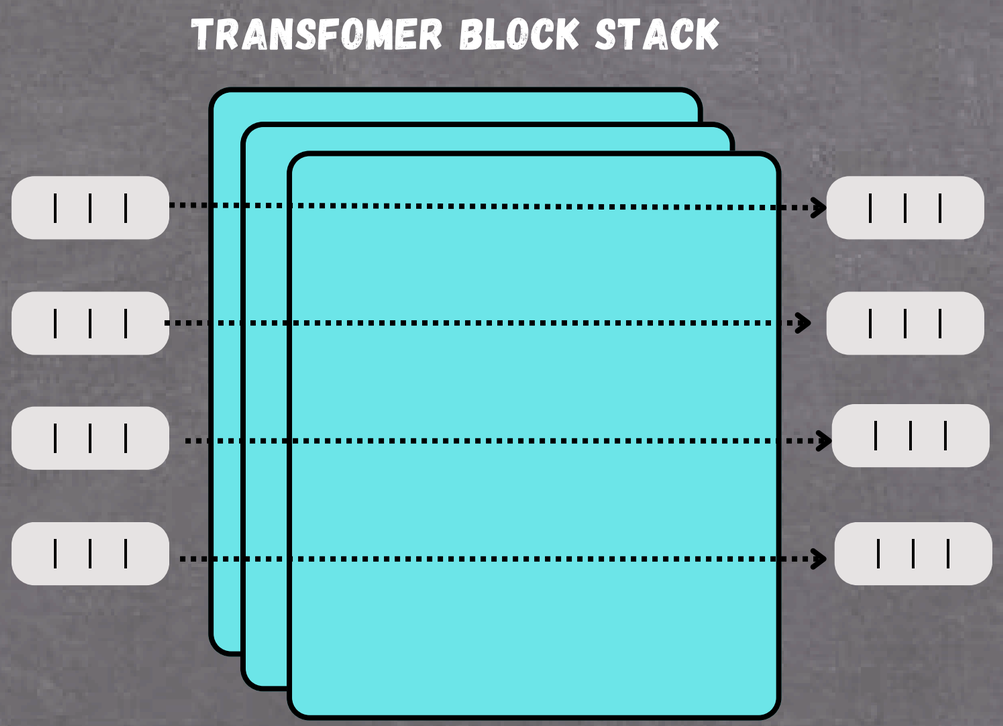
POSITIONAL ENCODING



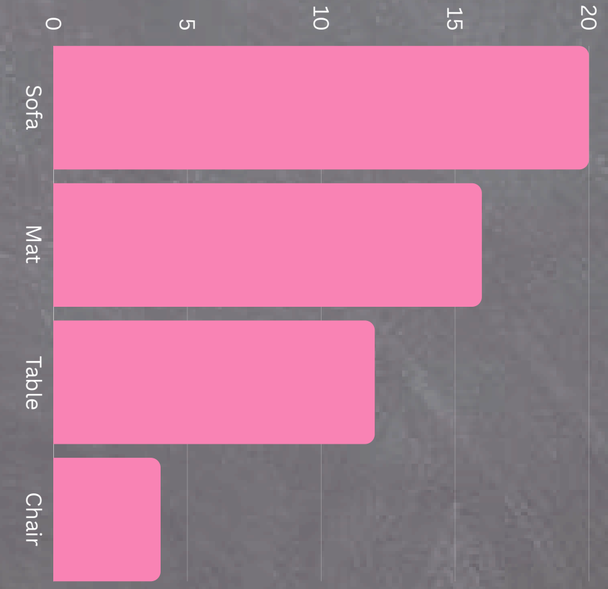
GOAL


THE WHITE CAT SAT ON THE
E1 E2 E3 E4 E5 E6
E1` E2` E3` E4` E5` E6`
CONTEXTUAL EMBEDDINGS

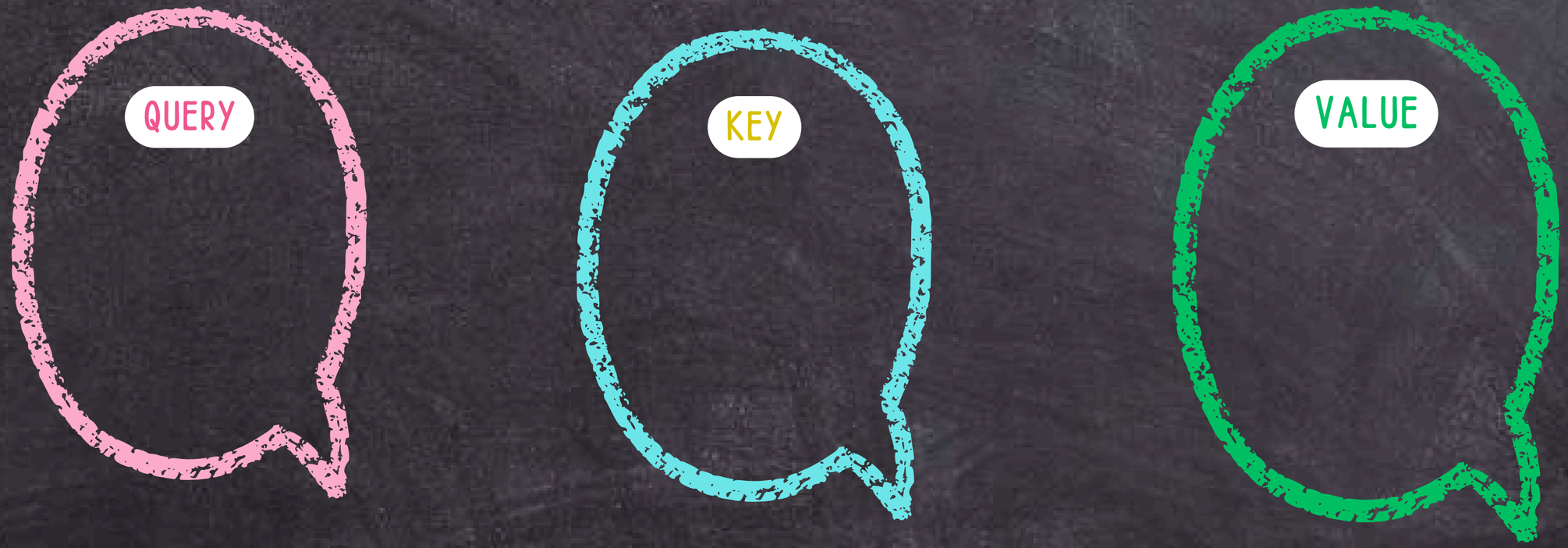

TRANSFOMER BLOCK STACK
INPUT TEXT
LM HEAD


TRANSFOMER BLOCK STACK

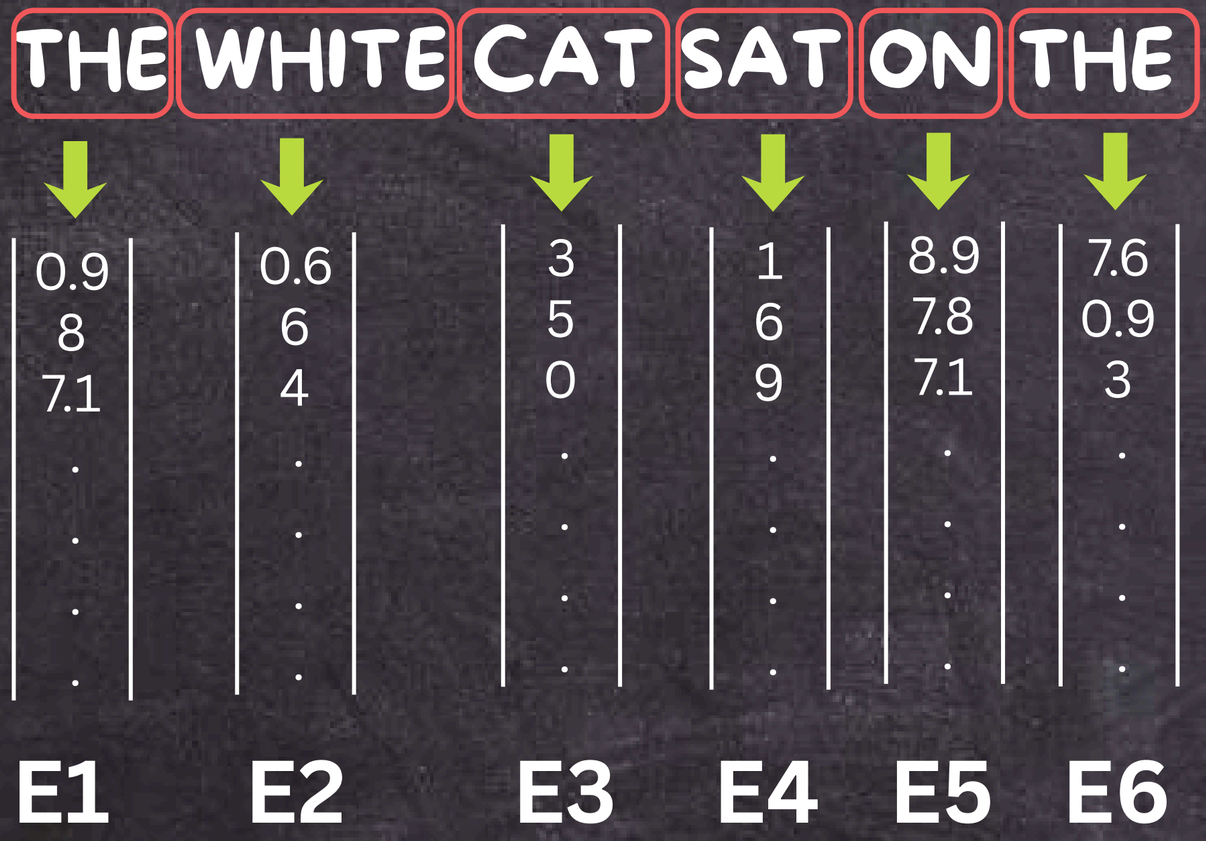

LM HEAD


0
5
10
15
20
Sofa
Mat
Table
Chair

SELF ATTENTION



SELF ATTENTION



QUERY (Q)

CAT

E3

W(Q)

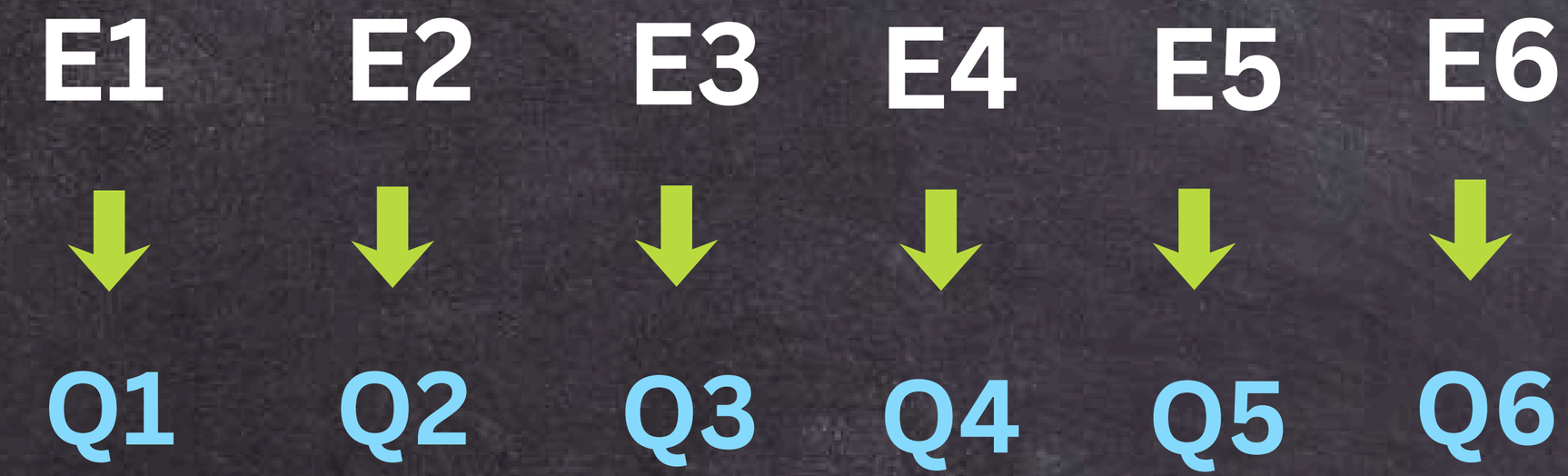
3.8	1.9	-08	0.6	-1
3.8	1.9	-08	0.6	-1
3.8	1.9	-08	0.6	-1
3.8	1.9	-08	0.6	-1
3.8	1.9	-08	0.6	-1

=

Q3

0.3
3.2
-1
.
.
.
.

QUERY (Q)



KEY (K)

WHITE

3
5
0
.
.
.
.

E3

W(K)

3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1

=

K3

0.3
3.2
-1
.
.
.
.

KEY (K)

RELEVANCY SCORE

	Q1	Q2	Q3	Q4	Q5	Q6
K1	0.7	12	91	1	8	0.7
K2	0.6	23	98	23	0.6	0.6
K3	0.5	0.9	21	0.8	8	0.5
K4	-21	11	-21	-21	13	2
K5	-42	21	0.9	5	7	-42
K6	0.1	10	0	1	1	2

MASKING

	Q1	Q2	Q3	Q4	Q5	Q6
K1	0.7	12	91	1	8	0.7
K2	-∞	23	98	23	0.6	0.6
K3	-∞	-∞	-∞	0.8	8	0.5
K4	-∞	-∞	-∞	-∞	13	2
K5	-∞	-∞	-∞	-∞	-∞	-42
K6	-∞	-∞	-∞	-∞	-∞	2

SOFTMAX

	Q1	Q2	Q3	Q4	Q5	Q6
K1	0.7	0.2	0.1	0.01	0.1	0.1
K2	0	0.8	0.8	0.4	0.4	0.3
K3	0	0	0.1	0.5	0.2	0.3
K4	0	0	0	0.1	0.2	0.1
K5	0	0	0	0	0.1	0.1
K6	0	0	0	0	0	0.1

VALUE (V)

CAT

3
5
0
.
.
.
.

E3

W(V)

3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1
3.8, 1.9, -08, 0.6, -1

=

V3

0.3
3.2
-1
.
.
.
.

DELTA CONTEXT

CAT

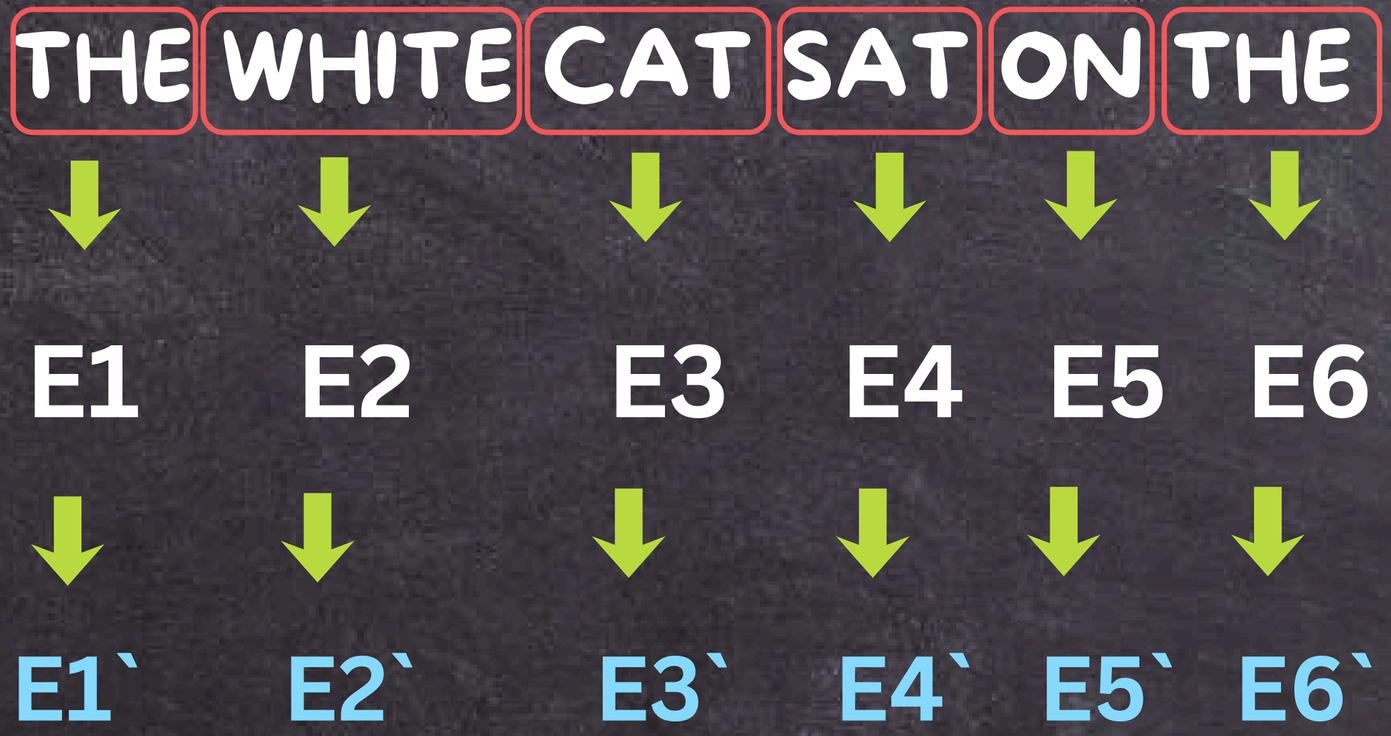
$$\Delta E3 = V1*0.1 + V2*0.9 + ...$$

In short, 10% of “The” and 90% of “White” we are having a DELTA CONTEXT

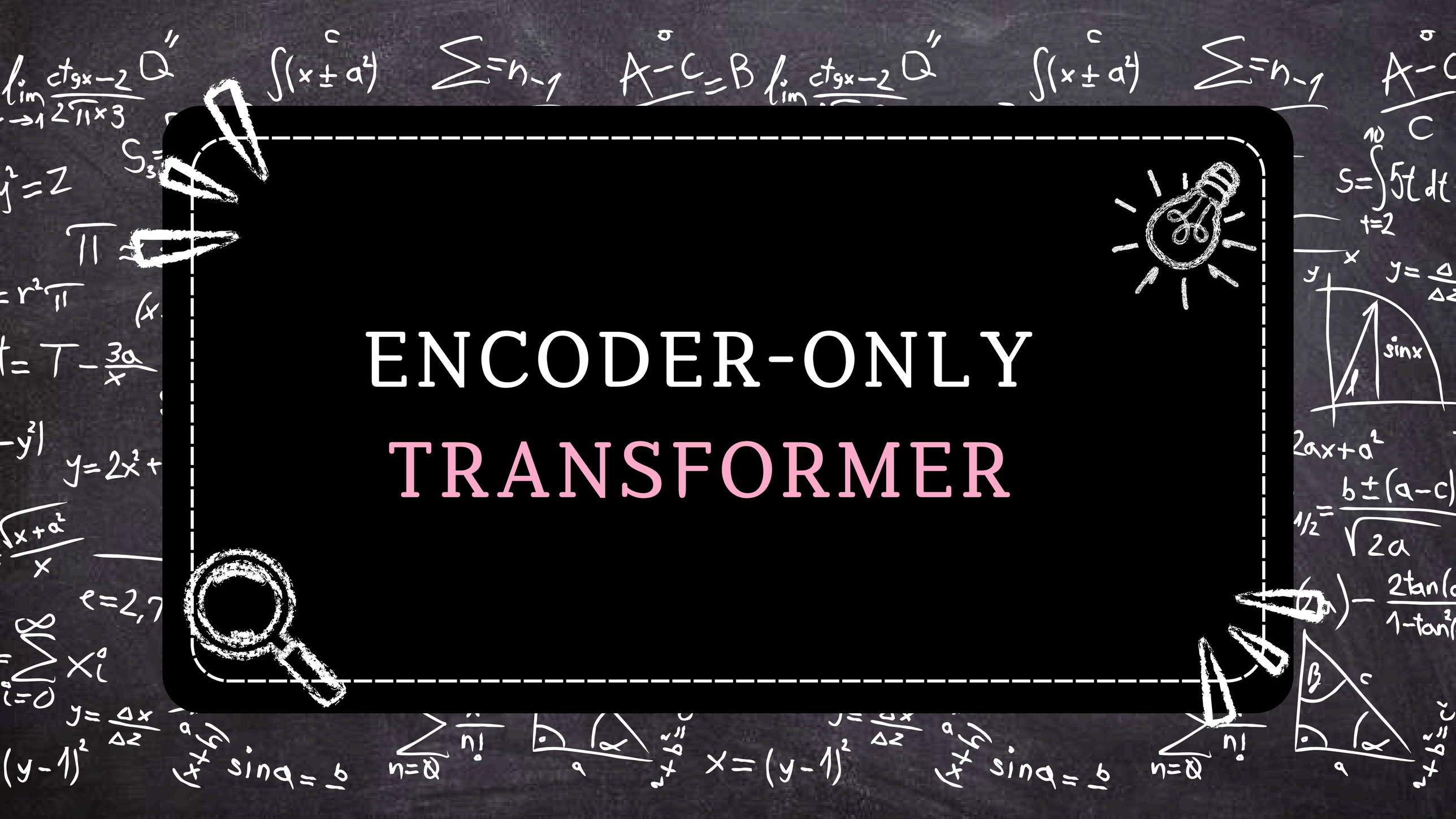
FINAL CONTEXTUAL EMBEDDING

$$E3` = E3 + \Delta E3$$

FINAL CONTEXTUAL EMBEDDING



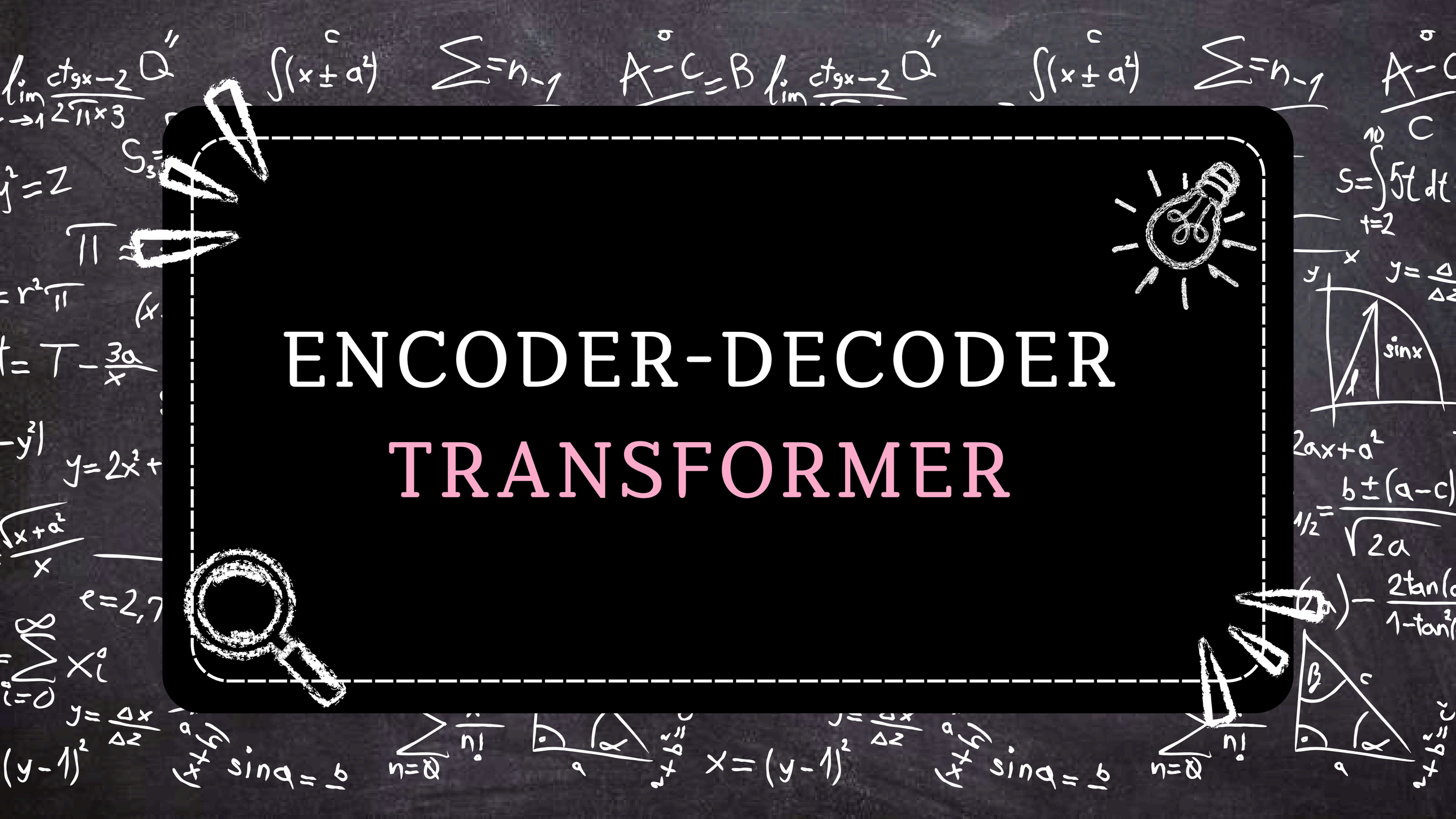
ENCODER-ONLY TRANSFORMER

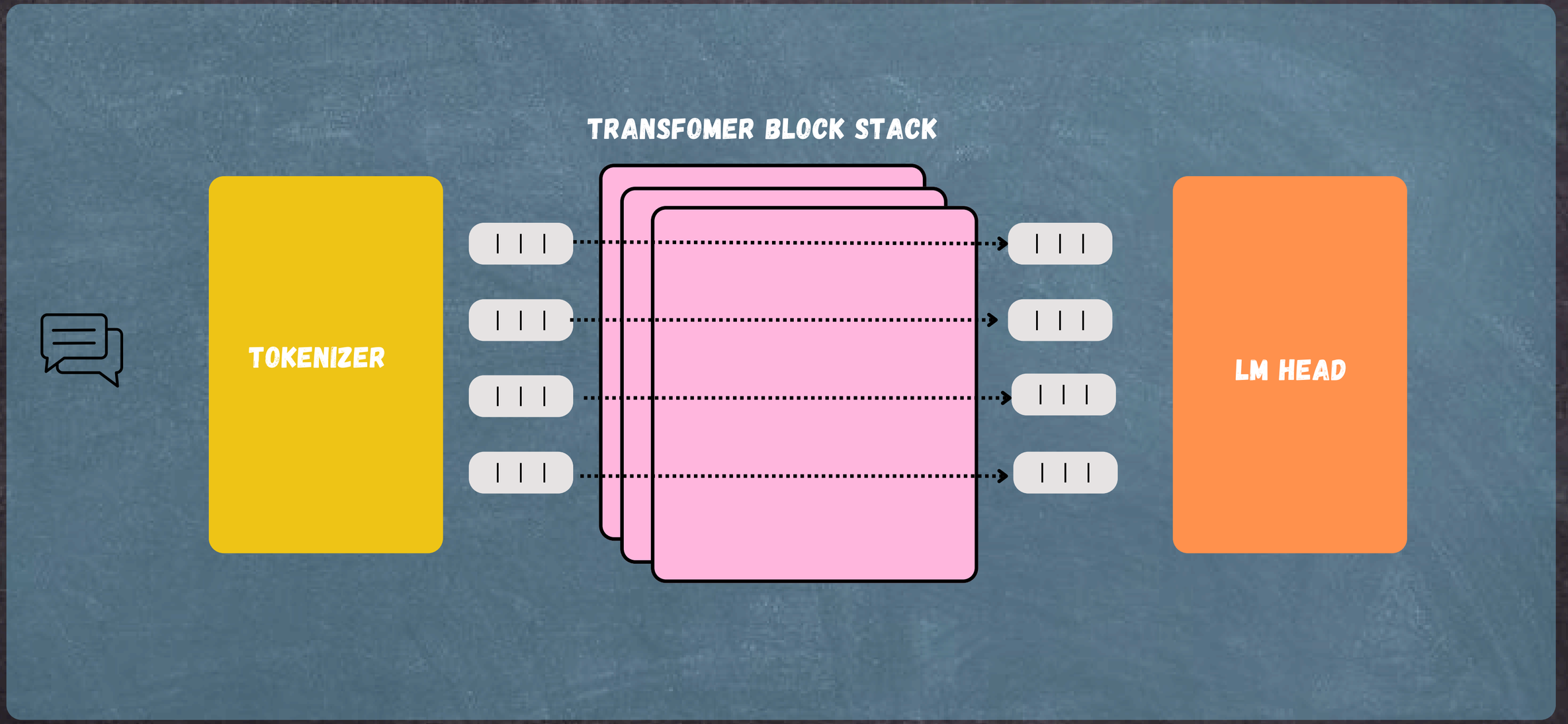


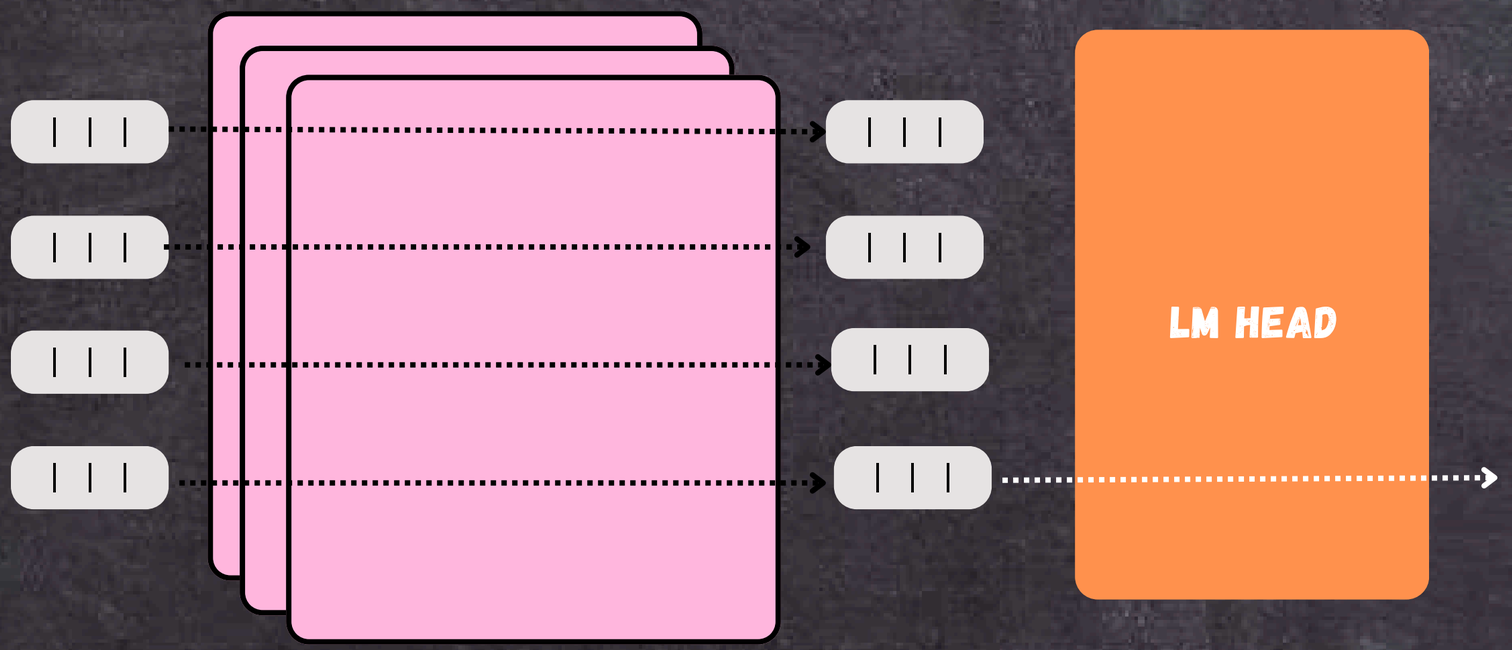
Encoder-Only transformers are designed primarily for understanding and processing input sequences, making them ideal for tasks such as text classification, sentiment analysis, and named entity recognition.

By architecture, it is exactly same as the Decoder-Only transformer we discussed earlier, except that it does not have the mechanism to generate output sequences. Instead, it focuses on encoding the input data into rich contextual embeddings that can be used for various downstream tasks.

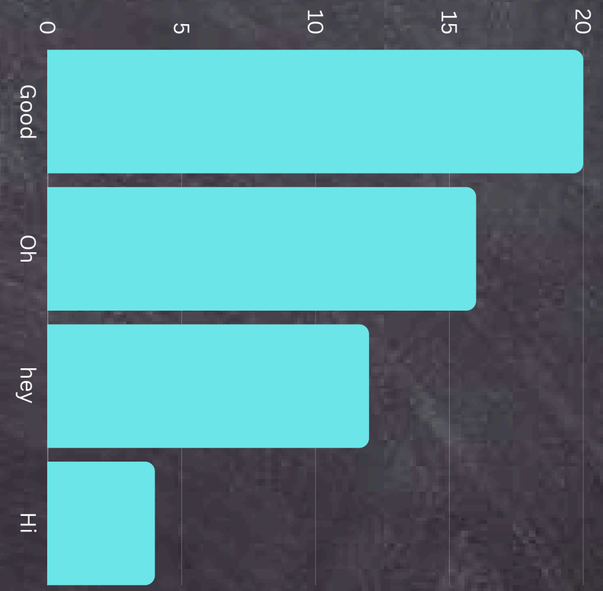
ENCODER-DECODER
TRANSFORMER

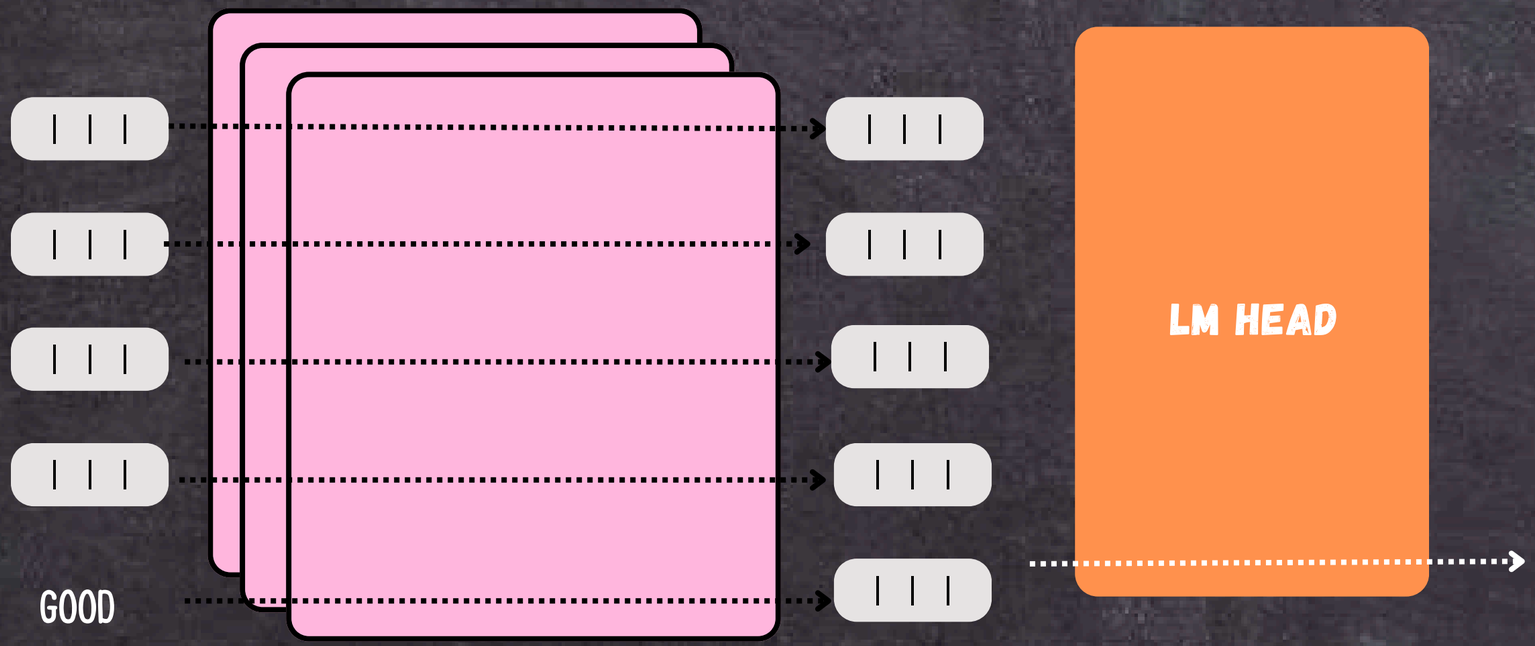



TRANSFOMER BLOCK STACK
TOKENIZER
LM HEAD



LM HEAD




GOOD
LM HEAD


0
5
10
15
20
Choice
boy
bro
bad

@anaghak.s8750 • 15 hours ago

You're amazing! Your content has been a huge help in my Azure data journey. I'm exc
offer and joined a new organization yesterday. Thank you so much for your support!

Reply 0 replies 1

@AnshLambaJSR • 0 seconds ago

Wow, congratulations!

Reply

@pavankumar12121 • 8 hours ago

I just wanted to let you know that I got the job — and it honestly wouldn't have been
projects you helped me with. Thank you so much from the bottom of my heart. I'm really, really

Reply 2 replies 1

a

@amoltare2843 • 3 hours ago

Same words ❤️ ..He is a saviour

Reply

S

@ShafiqAhmed-h7t • 10 hours ago

I got placed in two companies waiting for more all kudos to your help and videos anc
us thanks a ton 🎉🎉🎉 am grateful to you am learning a lot from you I have 10 year experienc
the knowledge I got from past two months from I precious you are jem n jewel bro.

Reply 4 replies 2

@prateekjha5162 • 15 hours ago

Bro , I do have 7 years of exp in DE already.. about to receive 2 offers because of you (more incoming h
even service notice period which is 90 days for me !!! thank you so much! If you're in India sometime, pl
the community know so that we can sit together atleast once

Reply 0 replies 1

@AnshLambaJSR • 0 seconds ago

Proud of my DATA FAM :)

@mohithkariappak6148 • 7 hours ago

Hi Ansh,

Just wanted to take moment to thank you. I have cracked 2 offer's watching you're videos. You're vide
a great impact during my preparations for the interviews and learnt a lot of new things at the same ti

Kudos to your hard work. Will keep following you throughout my data engineer career.

Reply 0 replies 1

D

@DEEPAKPARMAR-u5j2c • 7 hours ago

Hey Bro, Thanks for all practical knowledge. I am able to crack 2 offers from your contents.
are providing all this contents are high in quality

Reply 1 reply 2

M

@miky7595 • 7 hours ago

I got 5 offers after preparing from your videos .A big thanks to you bro from the bottom of my
under confident of cracking even a single round of any company 3-4 months back..Happy birt
shining

HANDLES

ANSH LAMBA

ANSH LAMBA JSR