Transformers

Cornelia Ilin, Ph.D.
UC Berkeley
School of Information
May 10, 2023

“ChatGPT could threaten 300 million jobs”

“ChatGPT is the hottest in-demand job skill”

“ChatGPT is the hottest in-demand job skill”

Powered by the Transformer architecture

Learning Objectives: Transformers

- History
- Architecture
- Application (with a deep-dive into electronic medical records)
- Conclusions

RNN/LSTM

"model of choice"

2017

Transformers

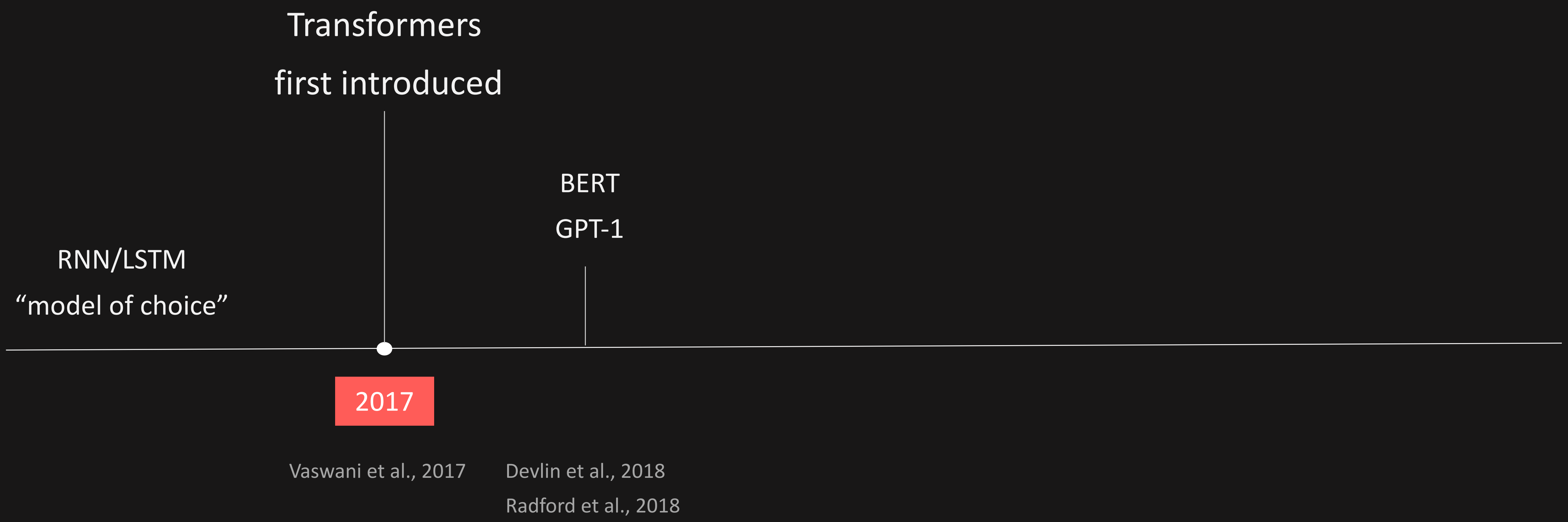
first introduced

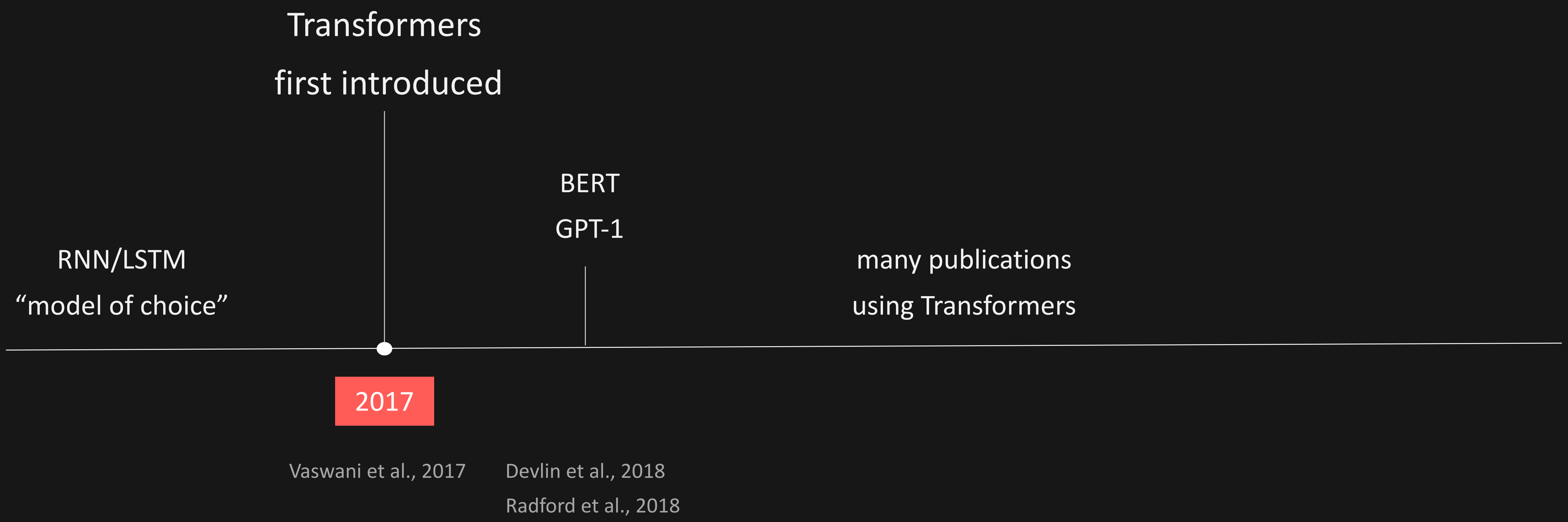
RNN/LSTM

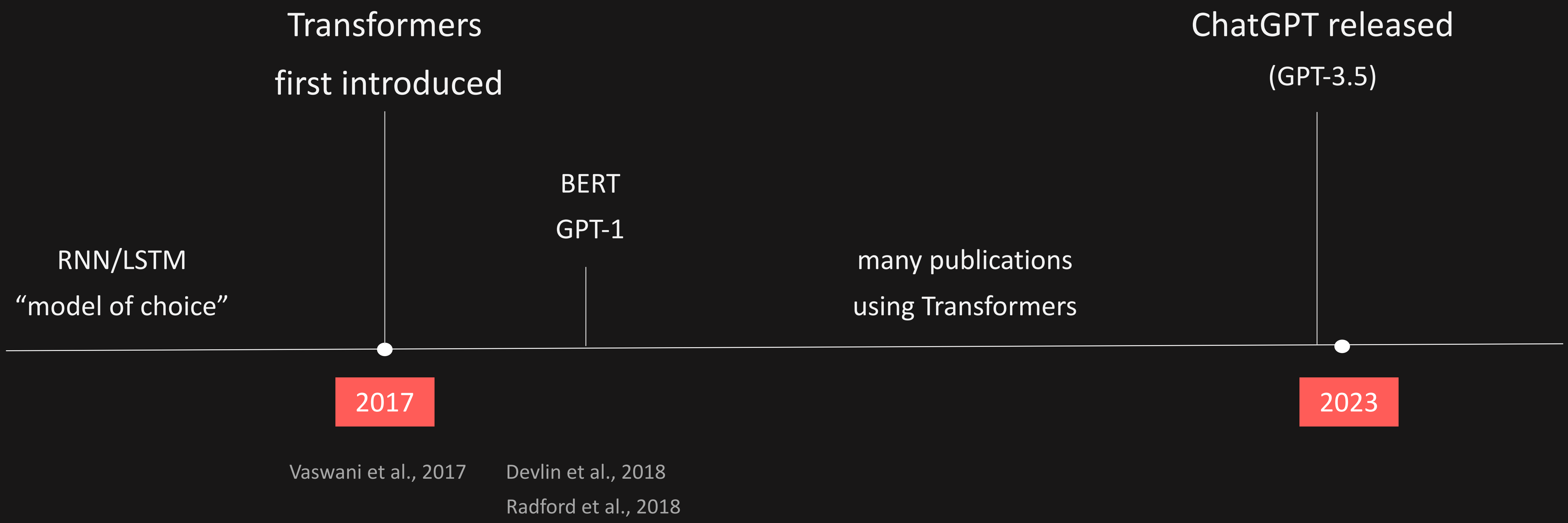
“model of choice”

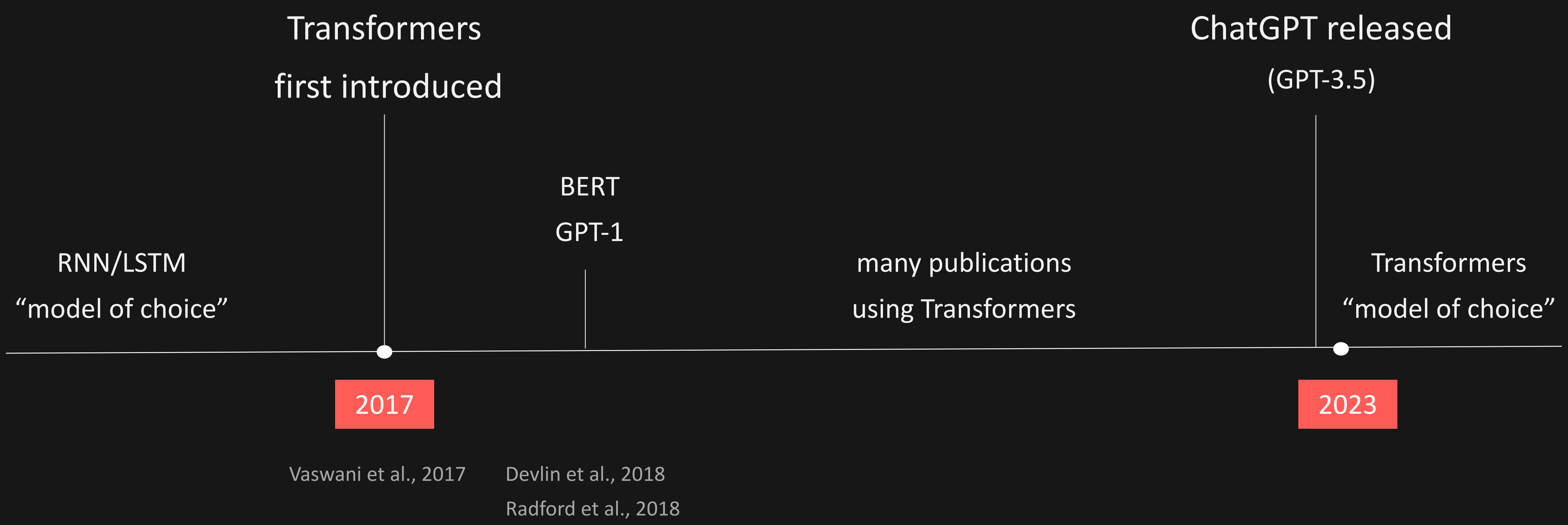
2017

Vaswani et al., 2017


Transformers
first introduced
BERT
GPT-1
RNN/LSTM
“model of choice”
2017
Vaswani et al., 2017
Devlin et al., 2018
Radford et al., 2018


Transformers
first introduced
BERT
GPT-1
RNN/LSTM
“model of choice”
many publications
using Transformers
2017
Vaswani et al., 2017
Devlin et al., 2018
Radford et al., 2018


Transformers
first introduced
ChatGPT released
(GPT-3.5)
BERT
GPT-1
RNN/LSTM
“model of choice”
many publications
using Transformers
2017
2023
Vaswani et al., 2017
Devlin et al., 2018
Radford et al., 2018


Transformers
first introduced
ChatGPT released
(GPT-3.5)
BERT
GPT-1
RNN/LSTM
“model of choice”
many publications
using Transformers
Transformers
“model of choice”
2017
2023
Vaswani et al., 2017
Devlin et al., 2018
Radford et al., 2018

Attention is All You Need

Vaswani et al., 2017

Issue with RNN/LSTMs

Cannot parallelize due to recurrency, i.e., very long training times

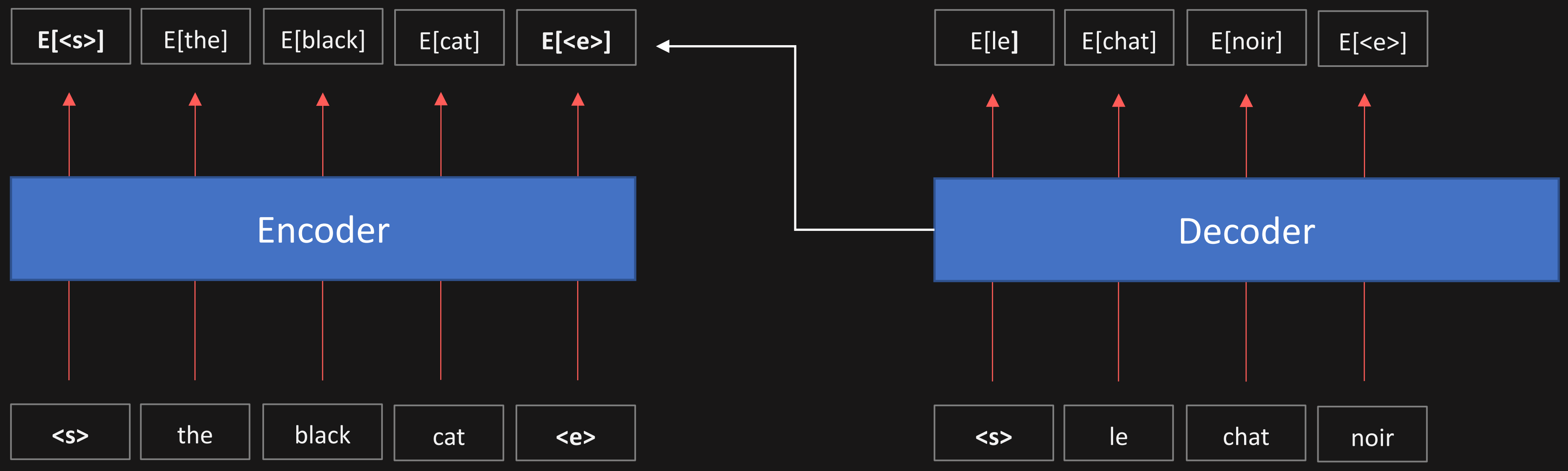
Proposed solution

Remove recurrency and rely on positional encodings and self-attention only

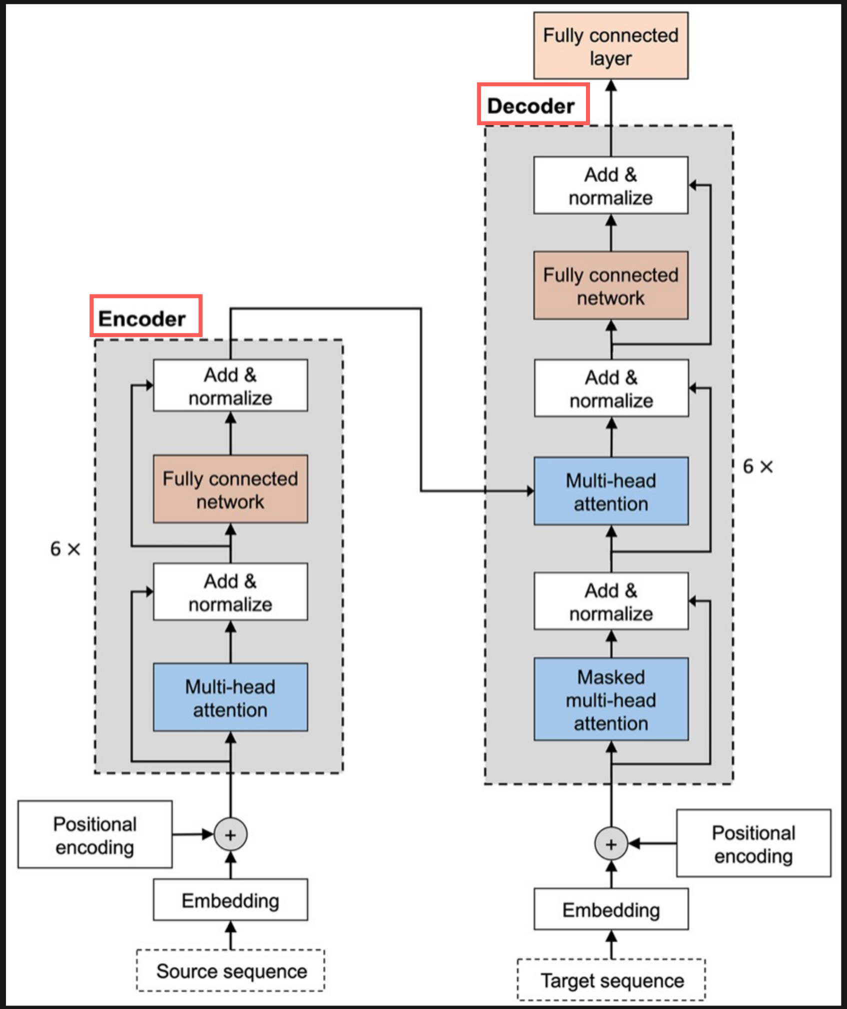
- Transformer architecture

Transformer (High Level View)

- Vaswani et al., 2017 considered the task of machine translation
- Encoder encodes source sentence (source token embeddings)
- Decoder predicts target sentence, one token at a time

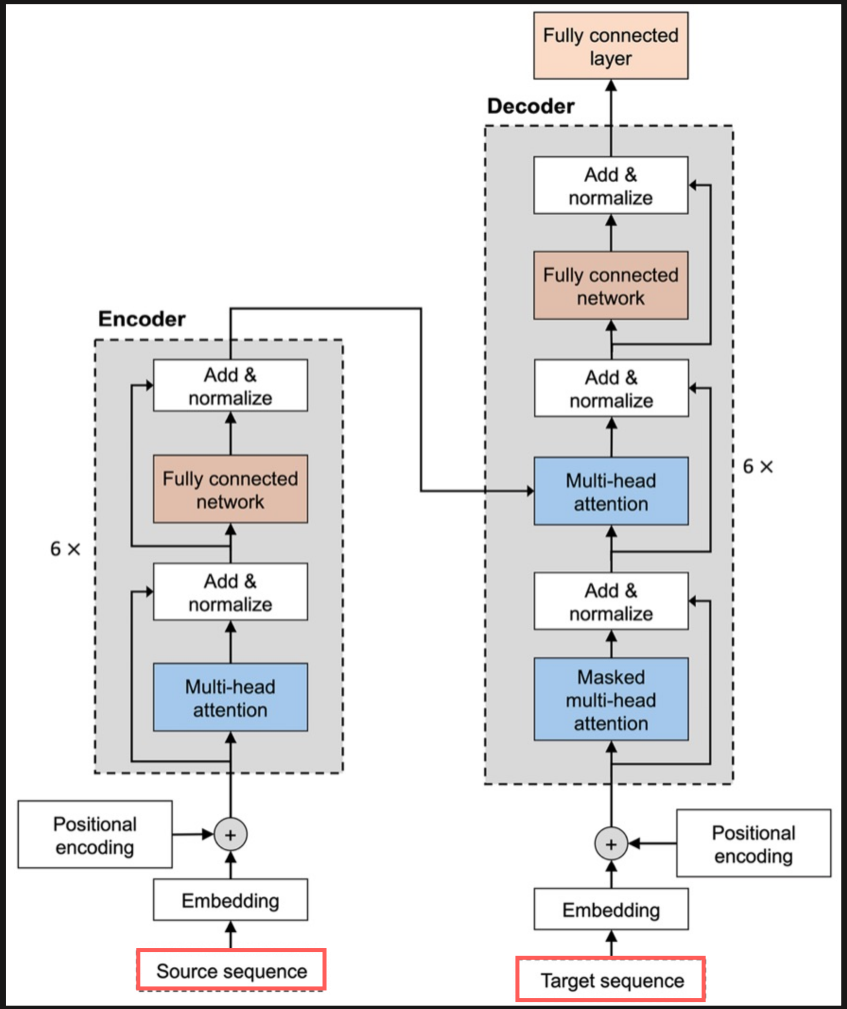


Transformer



Transformer

Architecture walkthrough (model training)

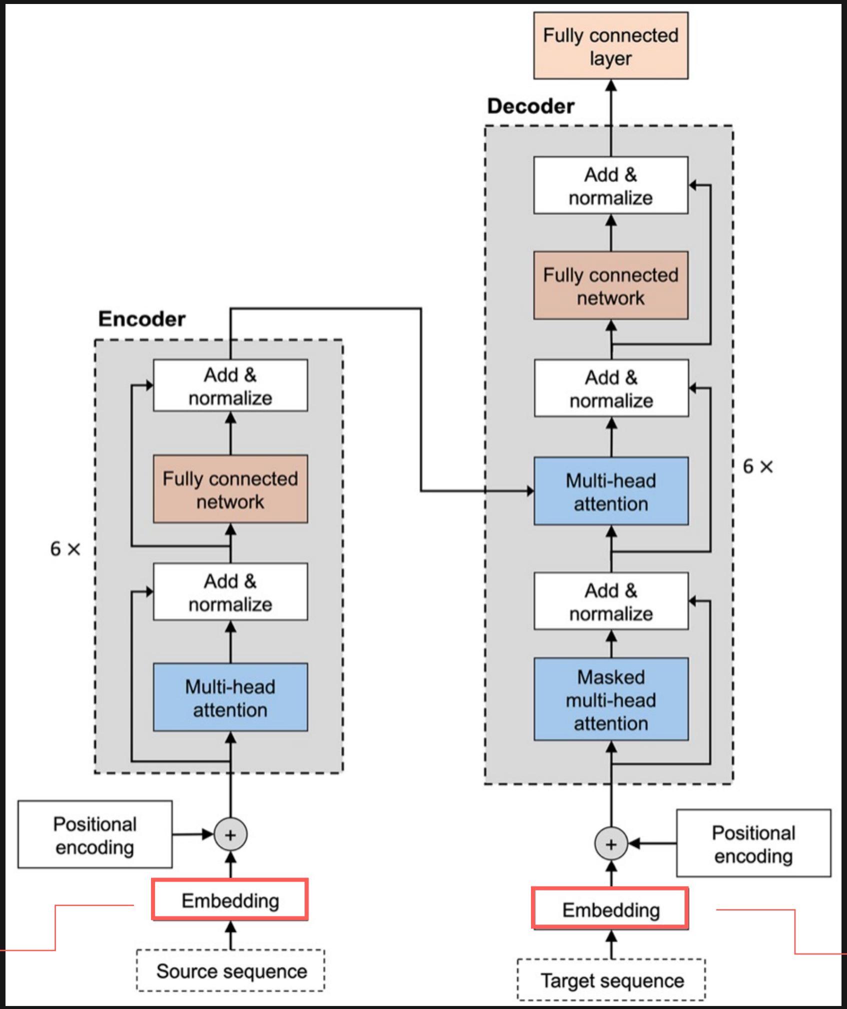


the black cat

le chat noir

Transformer

Architecture walkthrough (model training)



Reminder: embeddings are vector representations

E[<s>] | E[the] | E[black] | E[cat] | **E[<e>]**

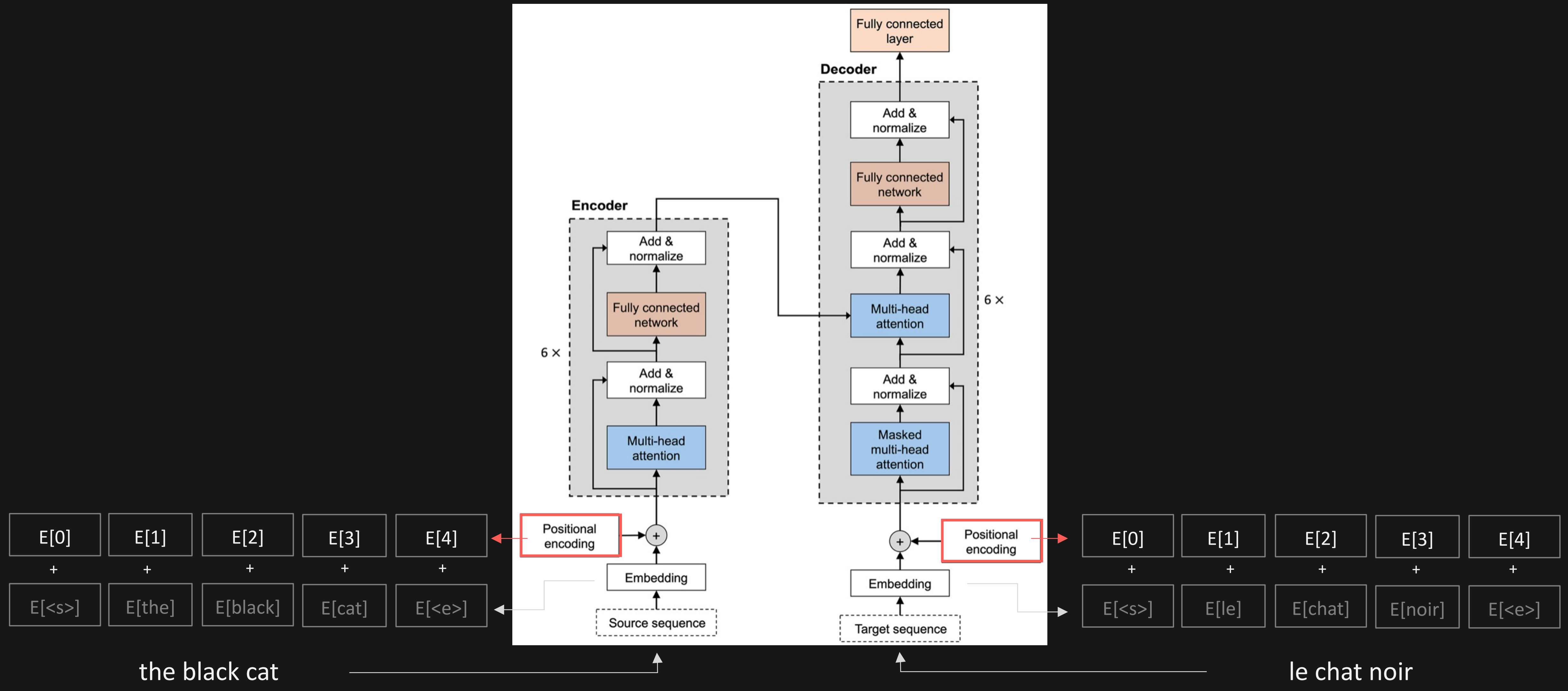
E[<s>] | E[le] | E[chat] | E[noir] | **E[<e>]**

the black cat

le chat noir

Transformer

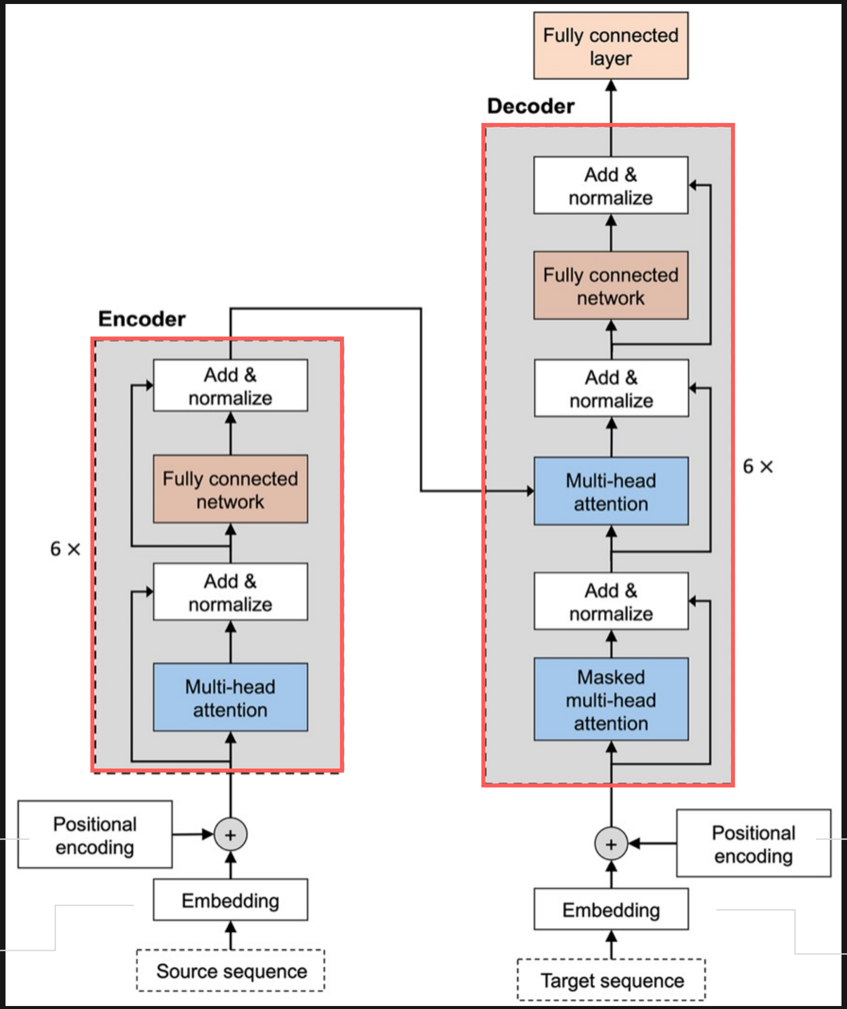
Architecture walkthrough (model training)



Transformer

Architecture walkthrough (model training)

E[0] + E[<s>] | E[1] + E[the] | E[2] + E[black] | E[3] + E[cat] | E[4] + E[<e>]



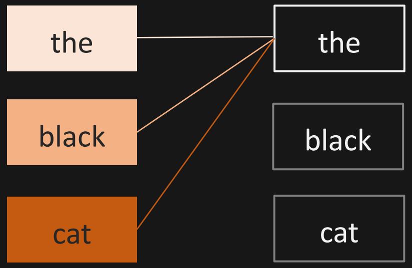
E[0] + E[<s>] | E[1] + E[le] | E[2] + E[chat] | E[3] + E[noir] | E[4] + E[<e>]

the black cat

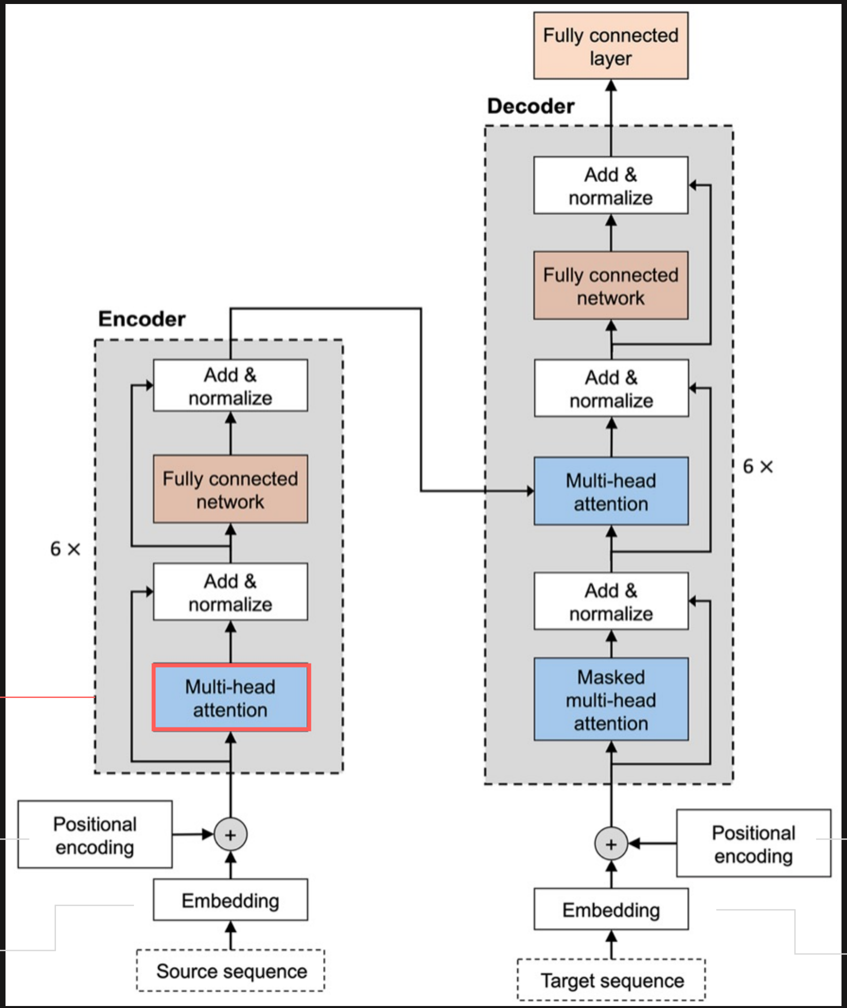
le chat noir

Transformer

Architecture walkthrough (model training)



Find how each token is related to all the tokens in the sequence



E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

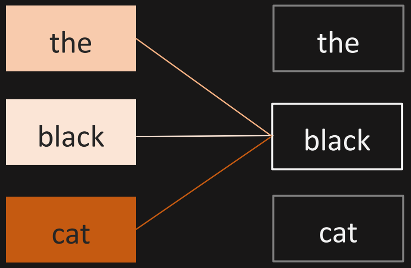
E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

the black cat

le chat noir

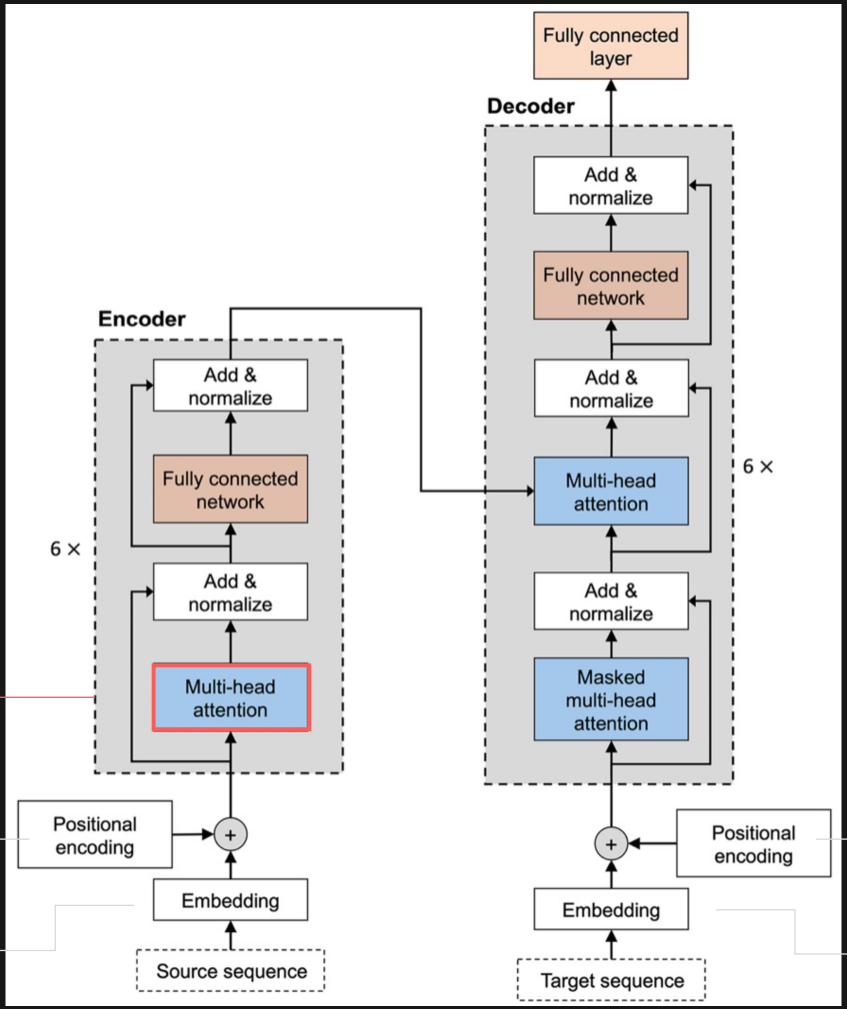
Transformer

Architecture walkthrough (model training)



Find how each token is related to all the tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]



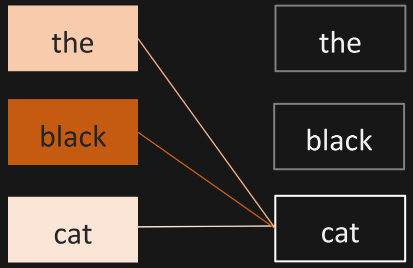
E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

the black cat

le chat noir

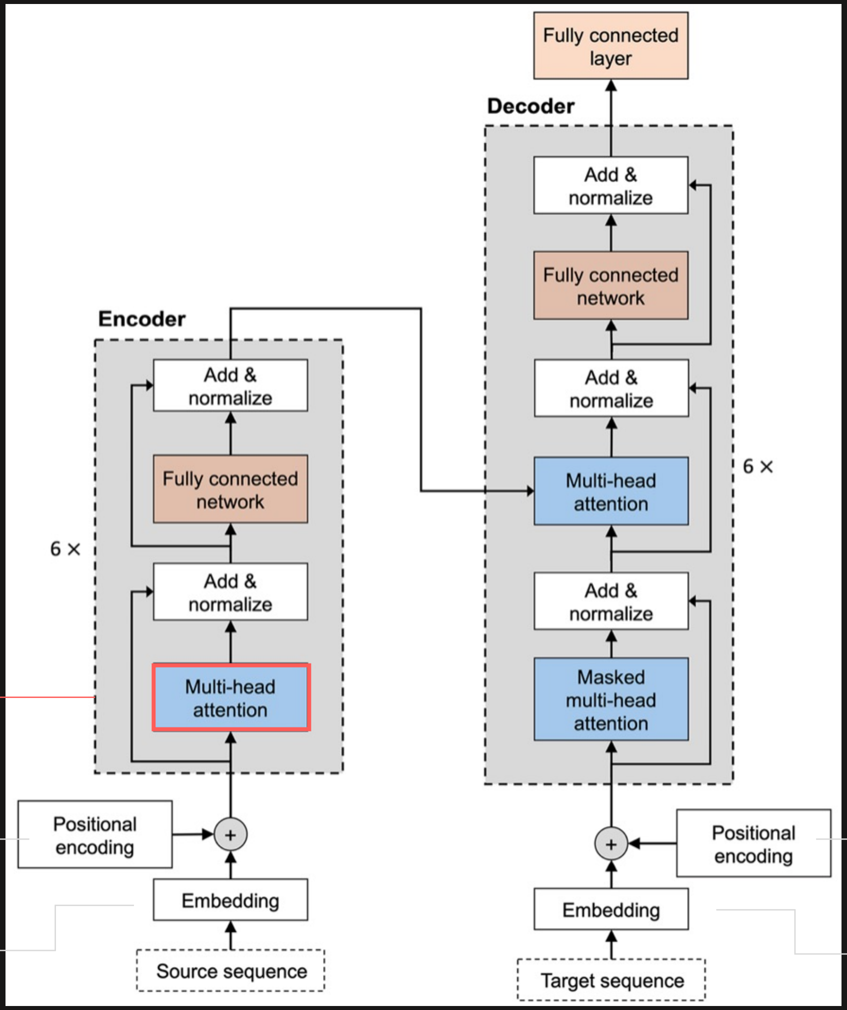
Transformer

Architecture walkthrough (model training)



Find how each token is related to all the tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]



E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

the black cat

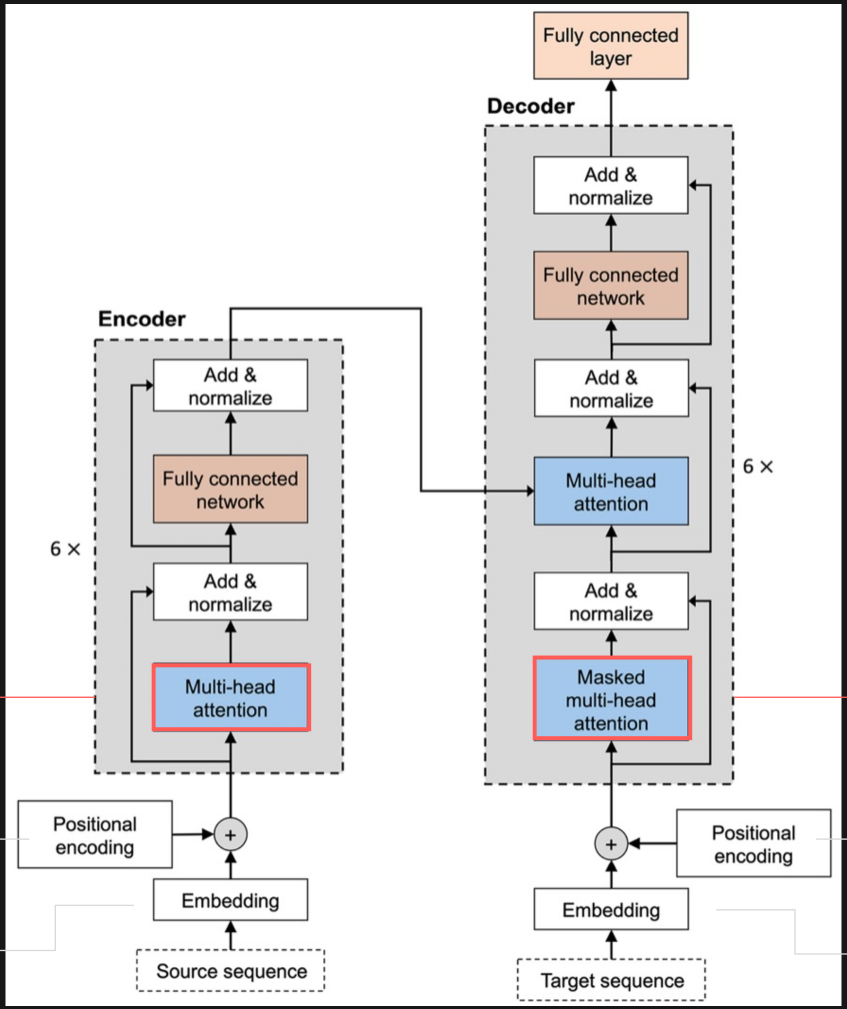
le chat noir

Transformer

Architecture walkthrough (model training)

the
black
cat

the
black
cat



le

le

Find how each token is related to all the tokens in the sequence

Find how each token is related to all unmasked tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

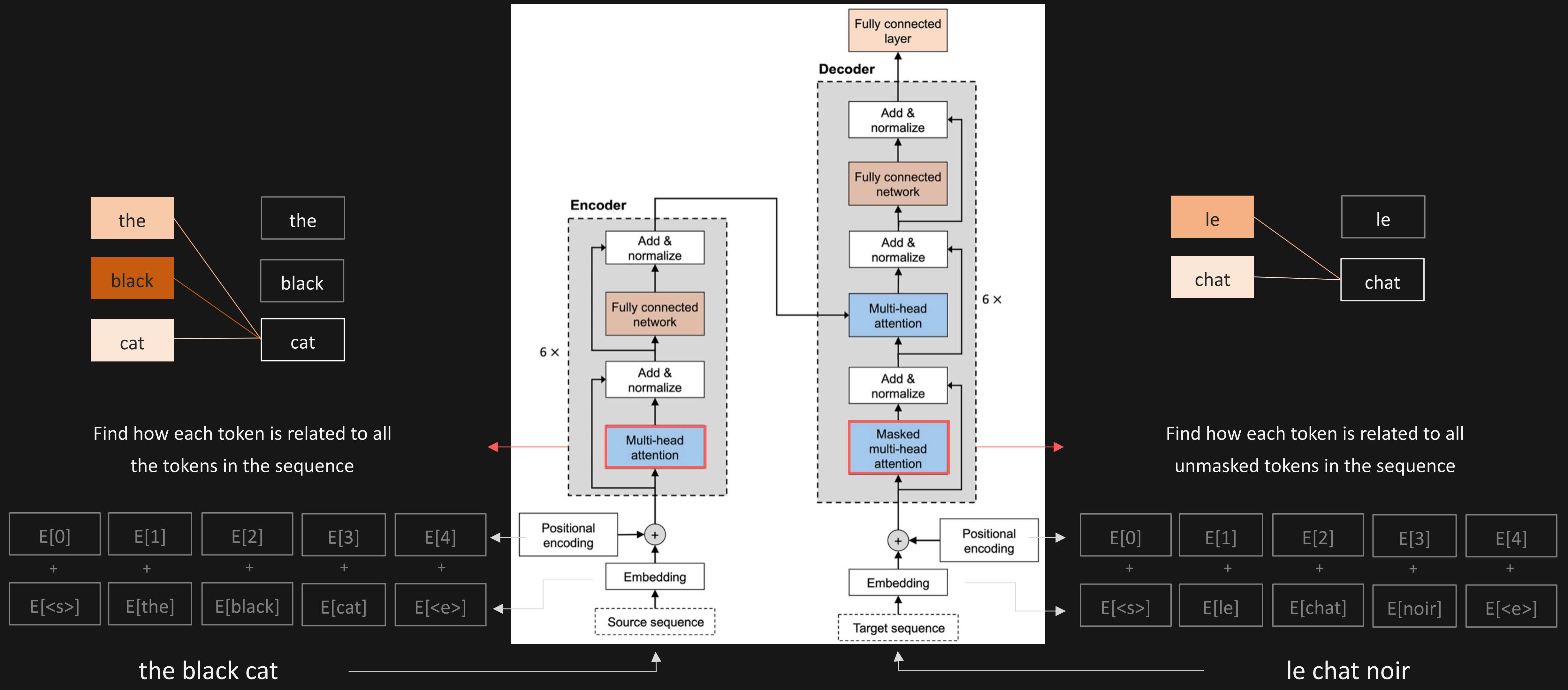
E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

the black cat

le chat noir

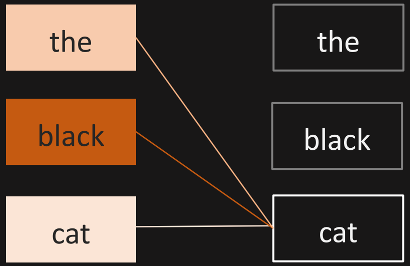
Transformer

Architecture walkthrough (model training)



Transformer

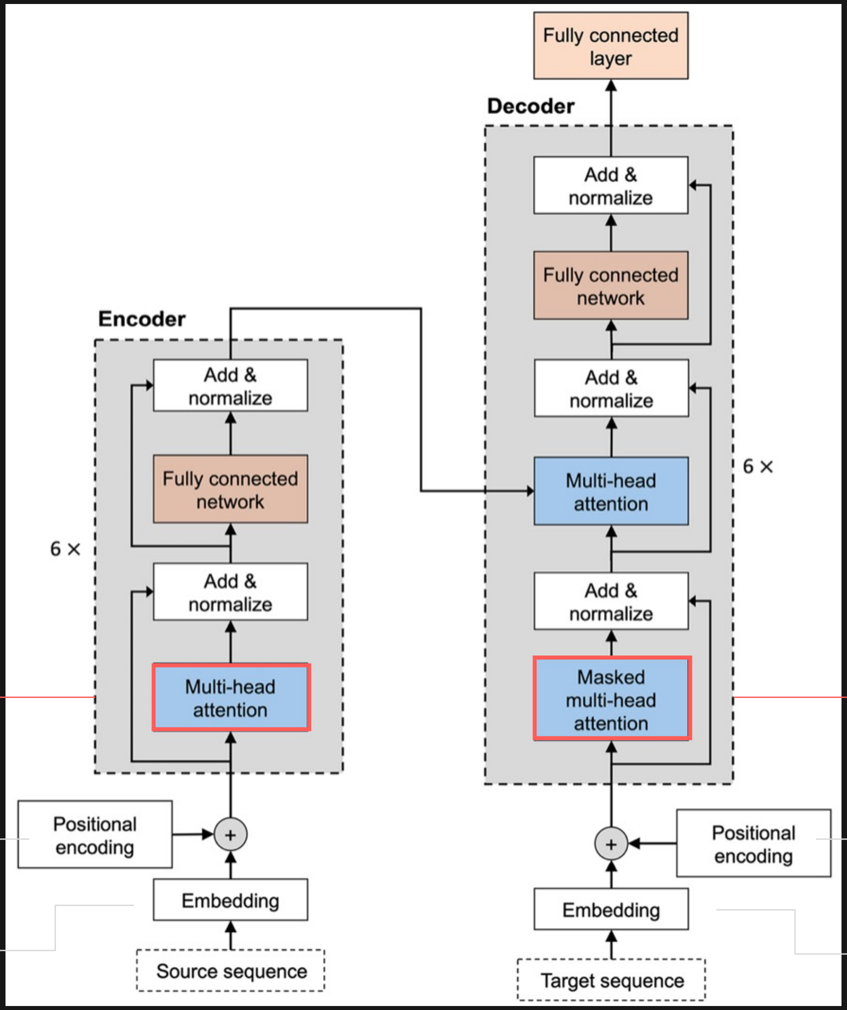
Architecture walkthrough (model training)

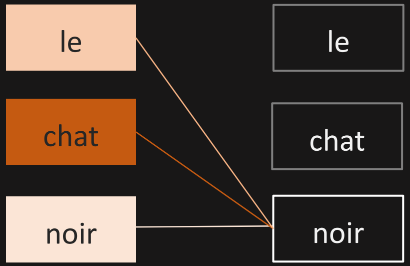


Find how each token is related to all the tokens in the sequence

E[0] + E[<s>]
E[1] + E[the]
E[2] + E[black]
E[3] + E[cat]
E[4] + E[<e>]

the black cat





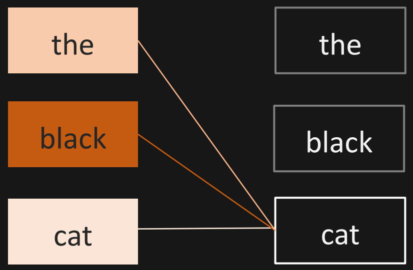
Find how each token is related to all unmasked tokens in the sequence

E[0] + E[<s>]
E[1] + E[le]
E[2] + E[chat]
E[3] + E[noir]
E[4] + E[<e>]

le chat noir

Transformer

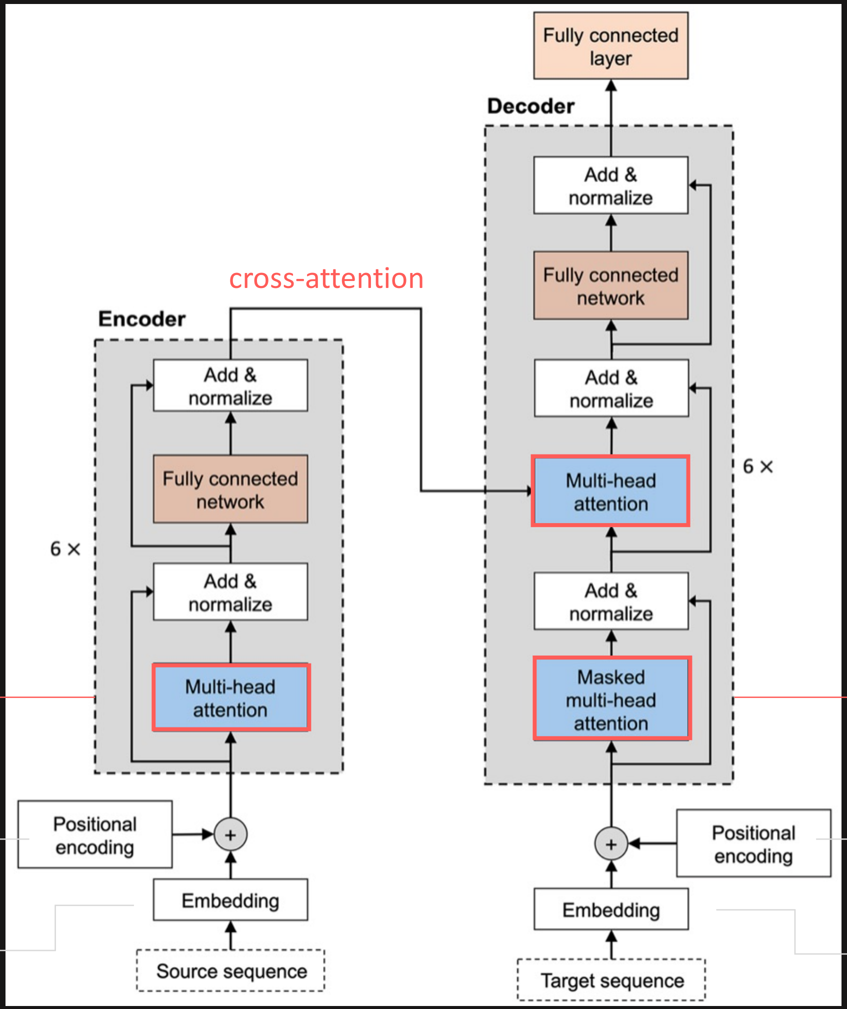
Architecture walkthrough (model training)

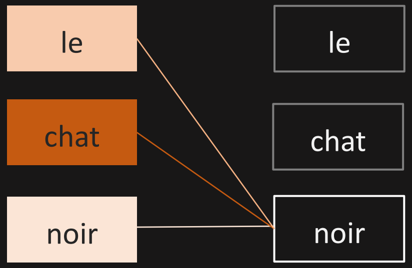


Find how each token is related to all the tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

the black cat





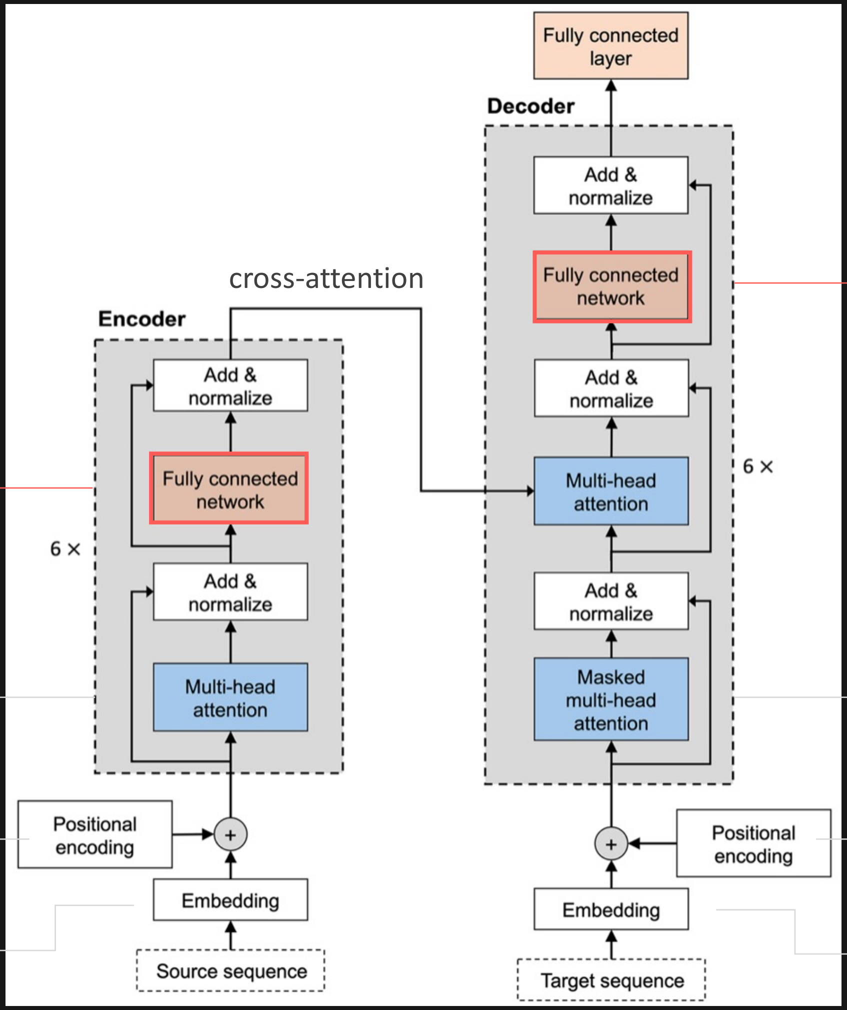
Find how each token is related to all unmasked tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

le chat noir

Transformer

Architecture walkthrough (model training)



FFNN

FFNN

Find how each token is related to all the tokens in the sequence

Find how each token is related to all unmasked tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

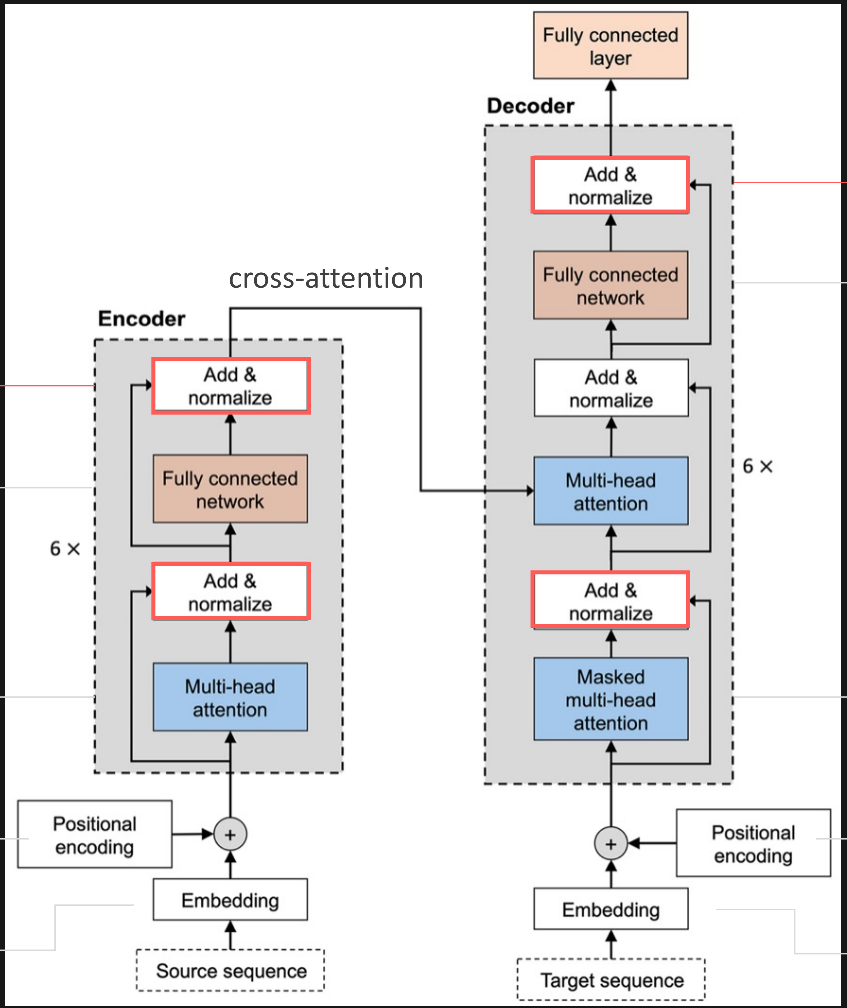
the black cat

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

le chat noir

Transformer

Architecture walkthrough (model training)



cross-attention

improves training/reduces overfitting

FFNN

Find how each token is related to all the tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

the black cat

improves training/reduces overfitting

FFNN

Find how each token is related to all unmasked tokens in the sequence

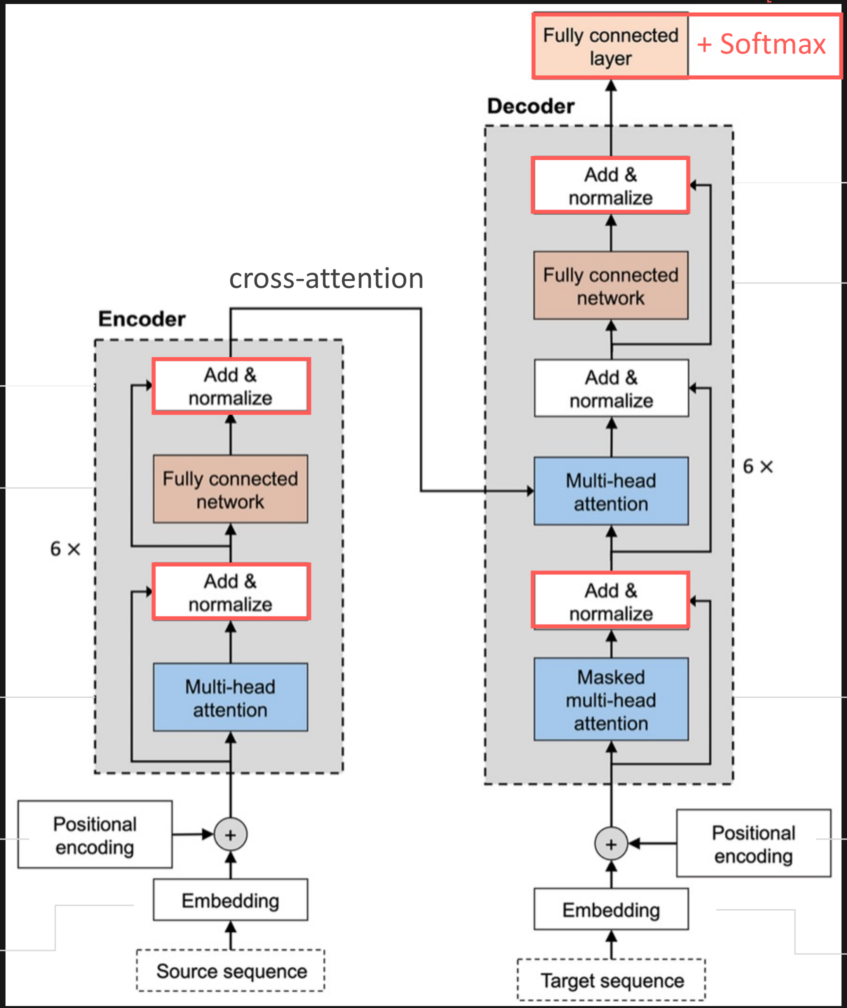
E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

le chat noir

Transformer

Architecture walkthrough (model training)

le | chat | noir | <e>



improves training/reduces overfitting

FFNN

improves training/reduces overfitting

FFNN

Find how each token is related to all the tokens in the sequence

Find how each token is related to all unmasked tokens in the sequence

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[black]	E[cat]	E[<e>]

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[le]	E[chat]	E[noir]	E[<e>]

the black cat

le chat noir

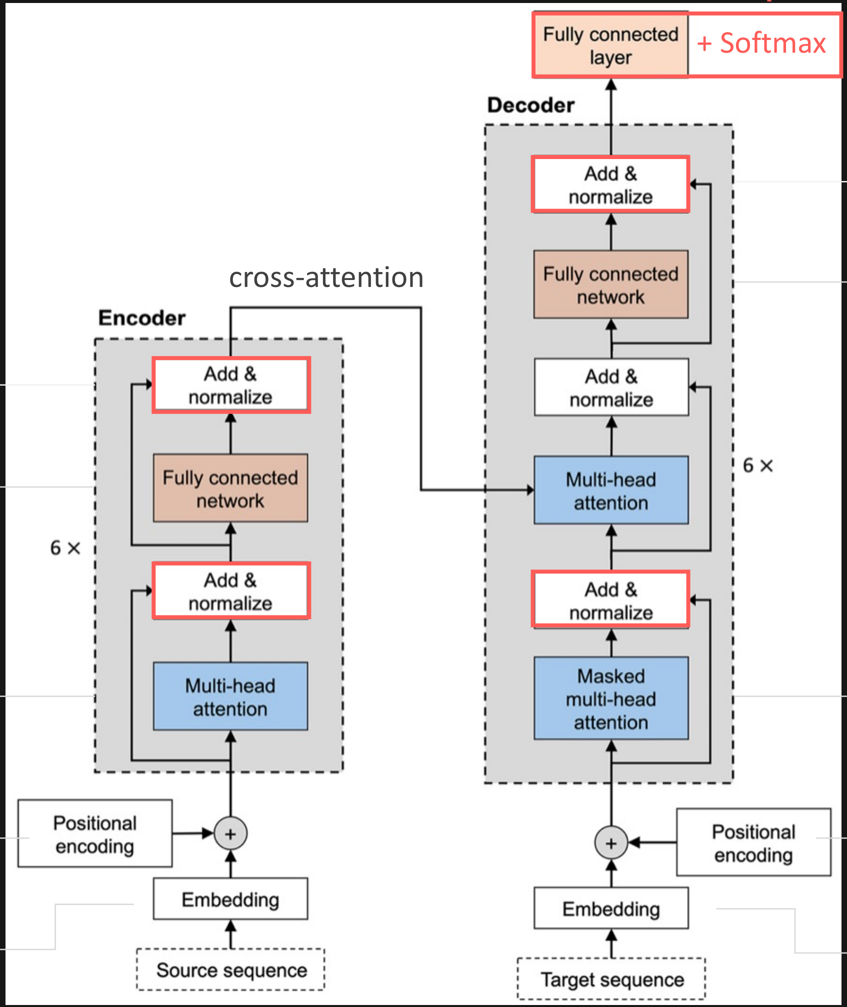
Transformer

Architecture walkthrough (model training)

The error backpropagates;
The labels the decoder needs are always the **next token** in the sequence

chien

le | ~~chat~~ | noir | <e>



improves training/reduces overfitting

FFNN

Find how each token is related to all the tokens in the sequence

E[0] | E[1] | E[2] | E[3] | E[4]
+ | + | + | + | +
E[<s>] | E[the] | E[black] | E[cat] | E[<e>]

the black cat

improves training/reduces overfitting

FFNN

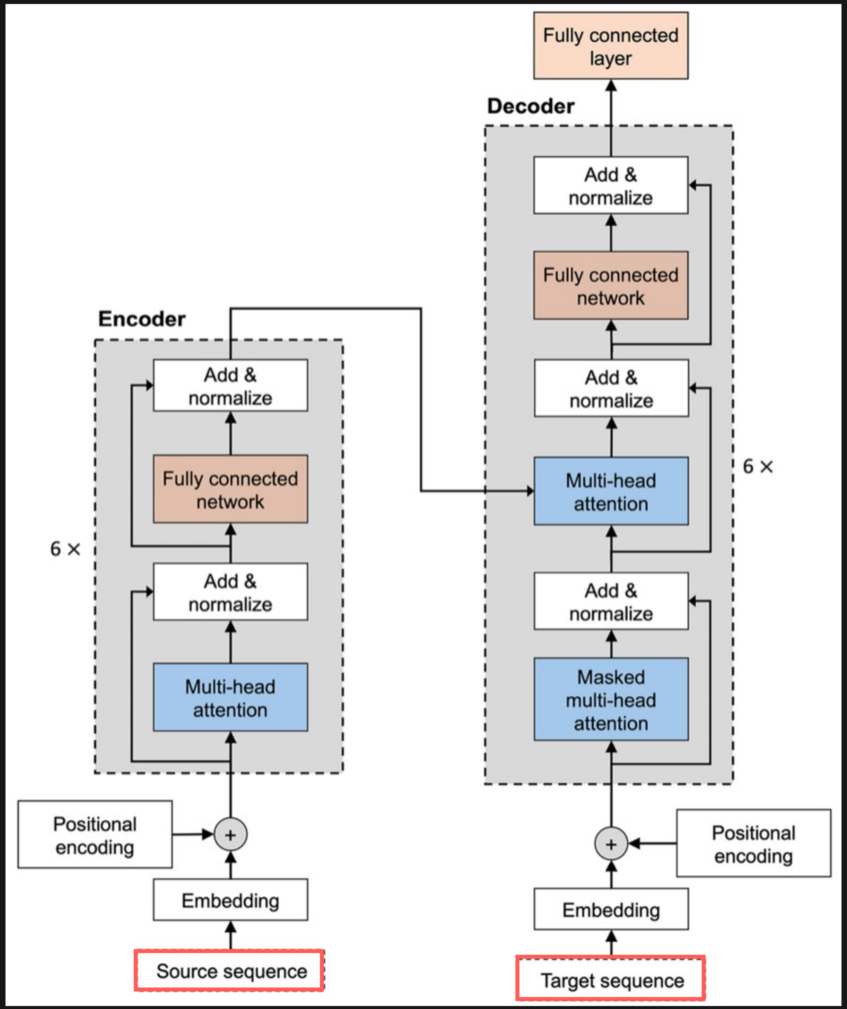
Find how each token is related to all unmasked tokens in the sequence

E[0] | E[1] | E[2] | E[3] | E[4]
+ | + | + | + | +
E[<s>] | E[le] | E[chat] | E[noir] | E[<e>]

le chat noir

Transformer

Architecture walkthrough (inference)

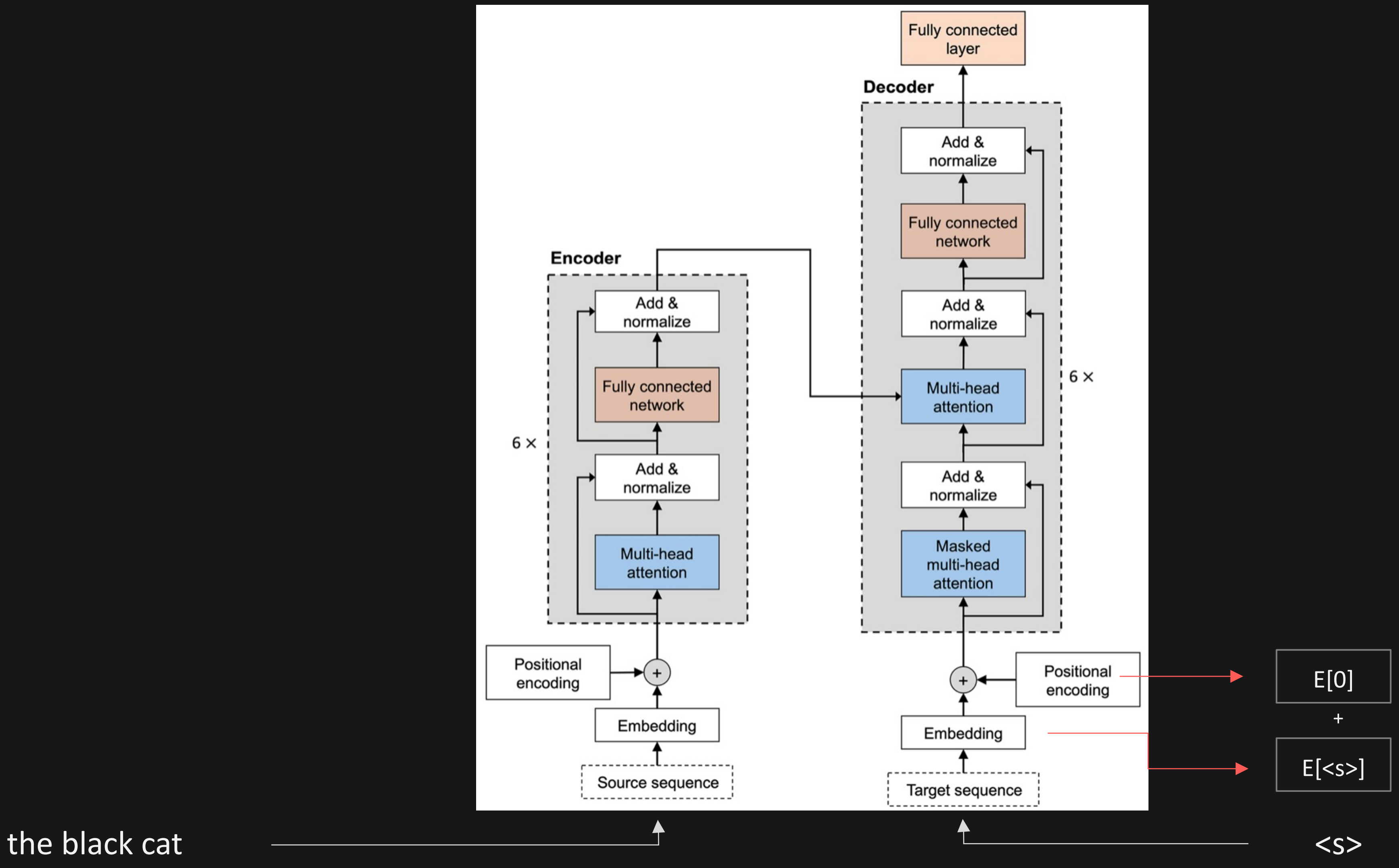


the black cat

<s>

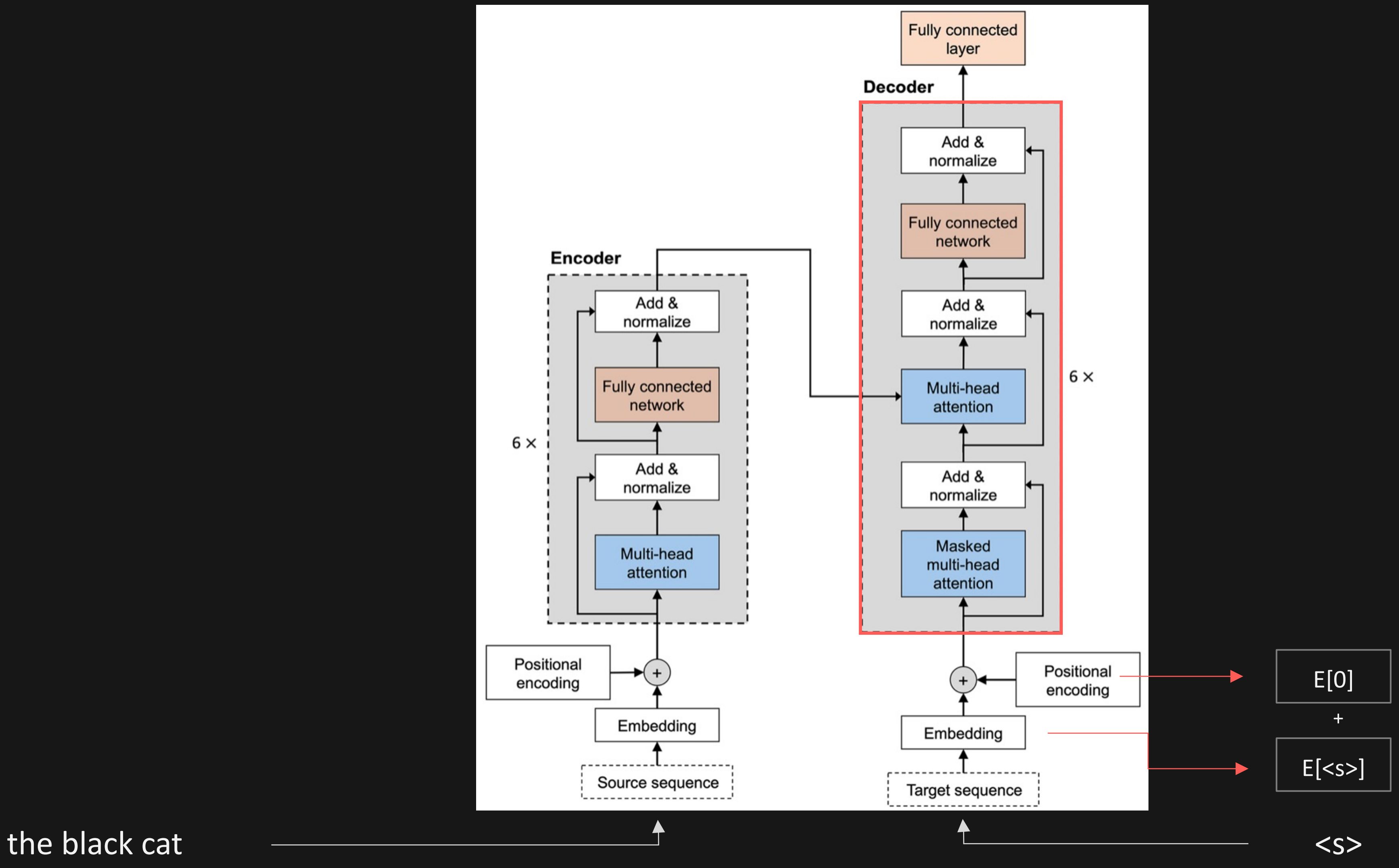
Transformer

Architecture walkthrough (inference)



Transformer

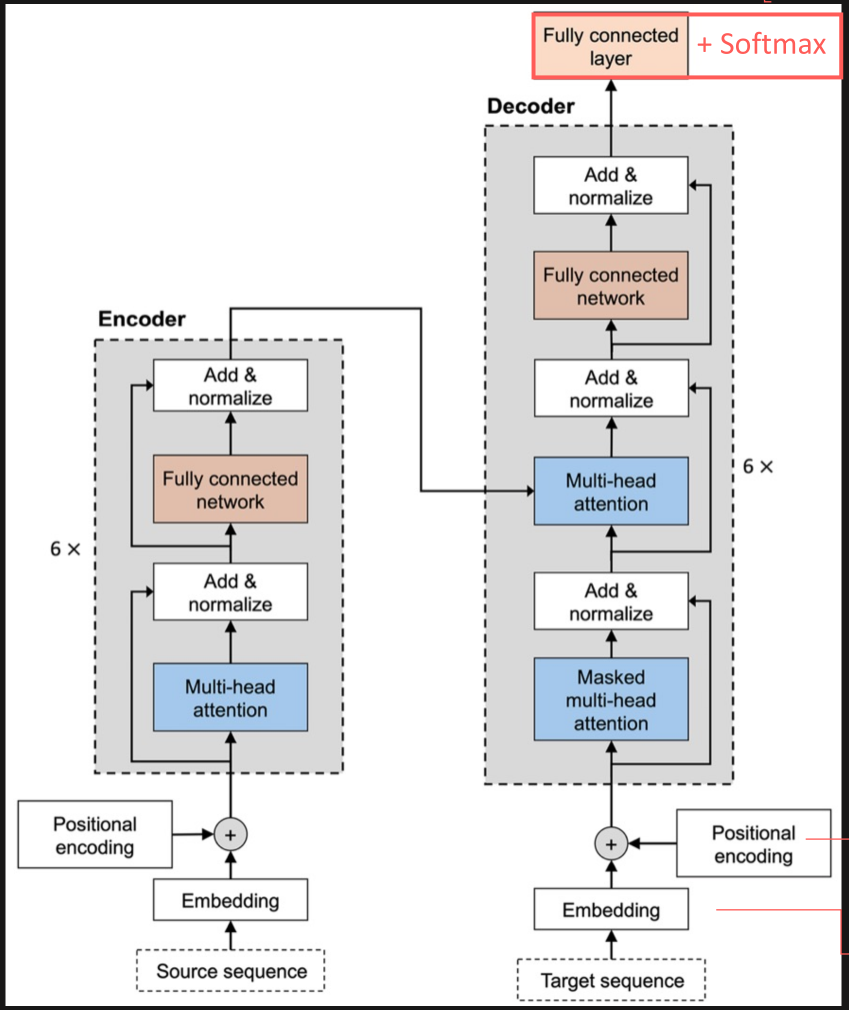
Architecture walkthrough (inference)



Transformer

Architecture walkthrough (inference)

le



E[0]

+

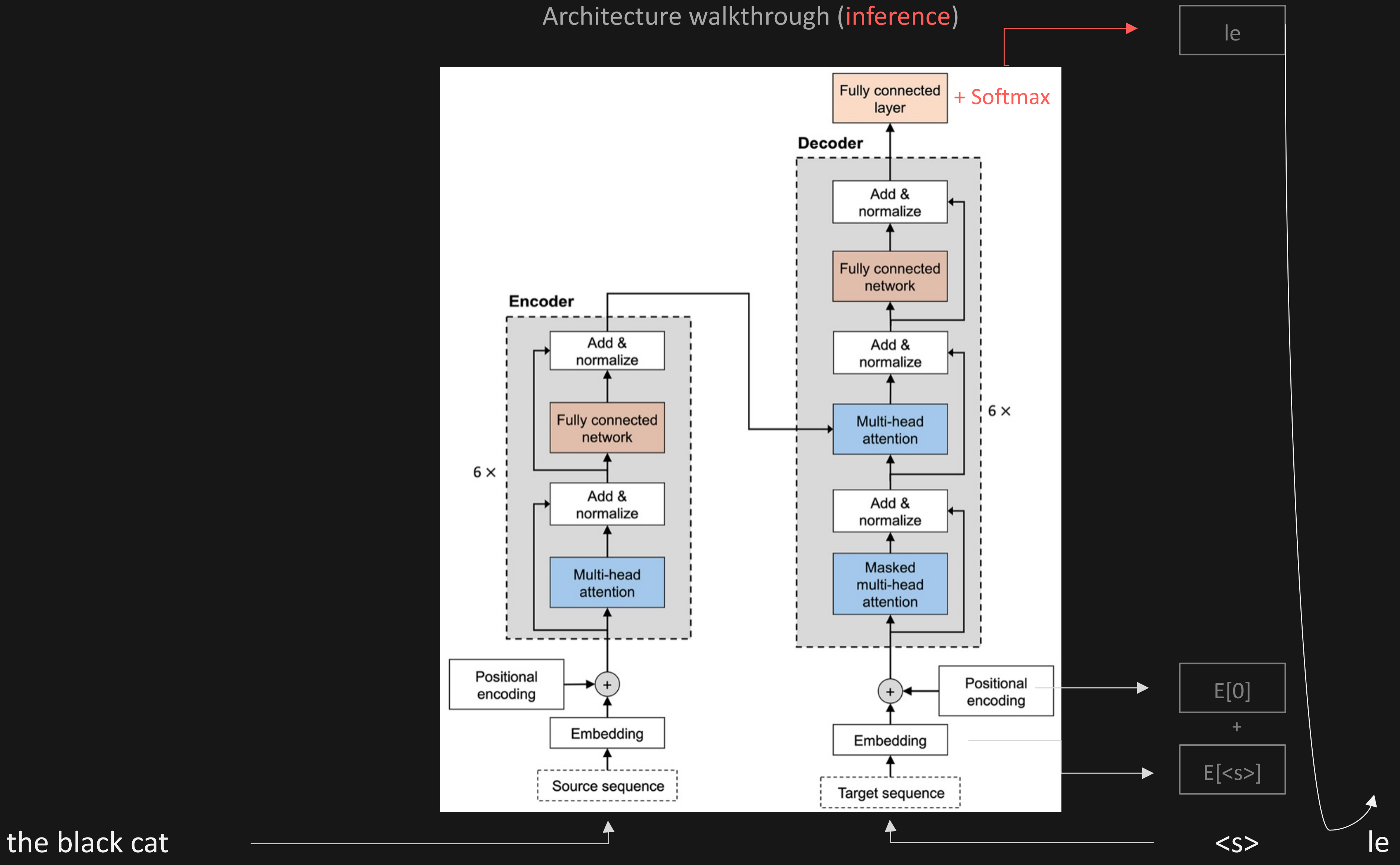
E[<s>]

the black cat

<s>

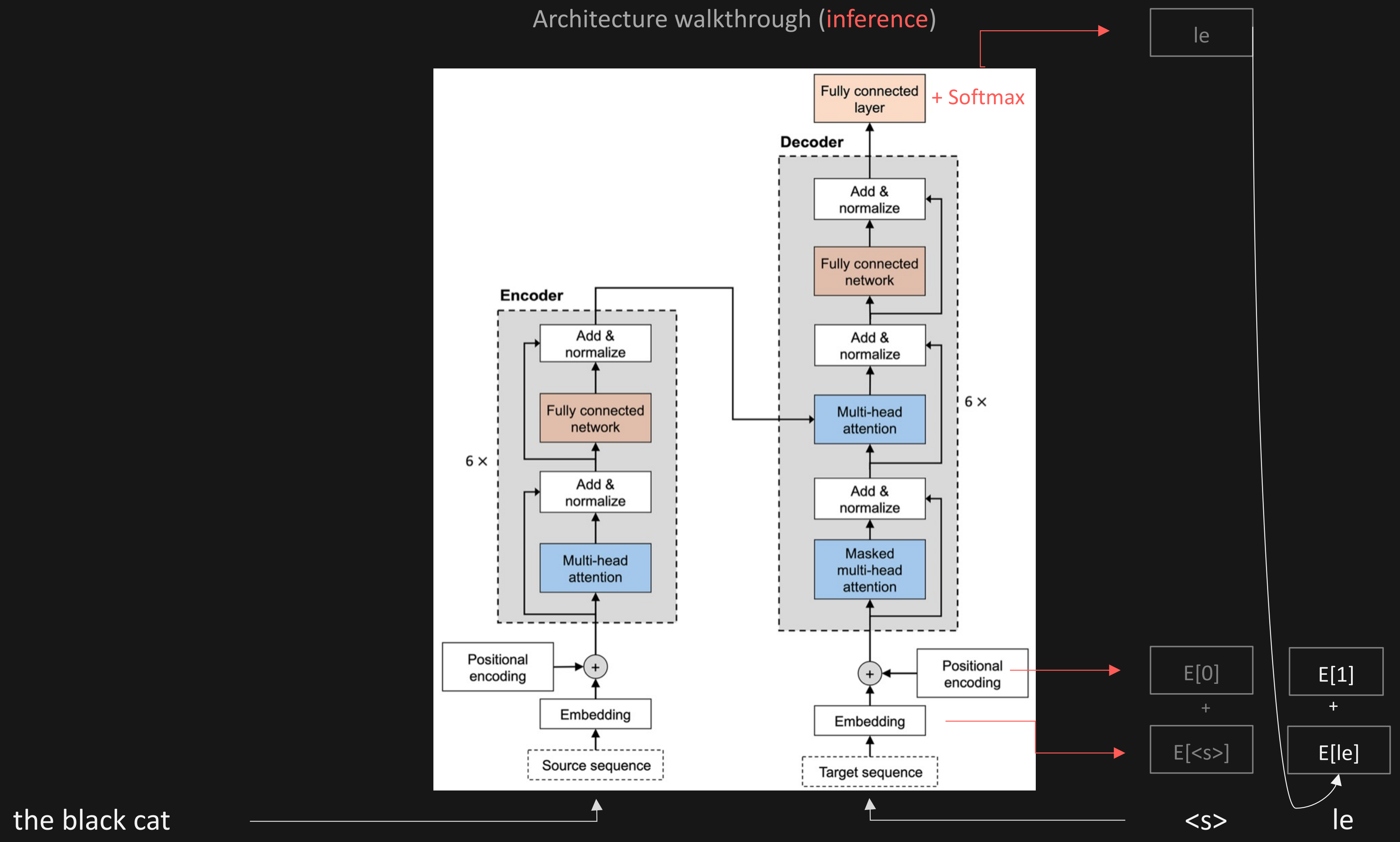
Transformer

Architecture walkthrough (inference)



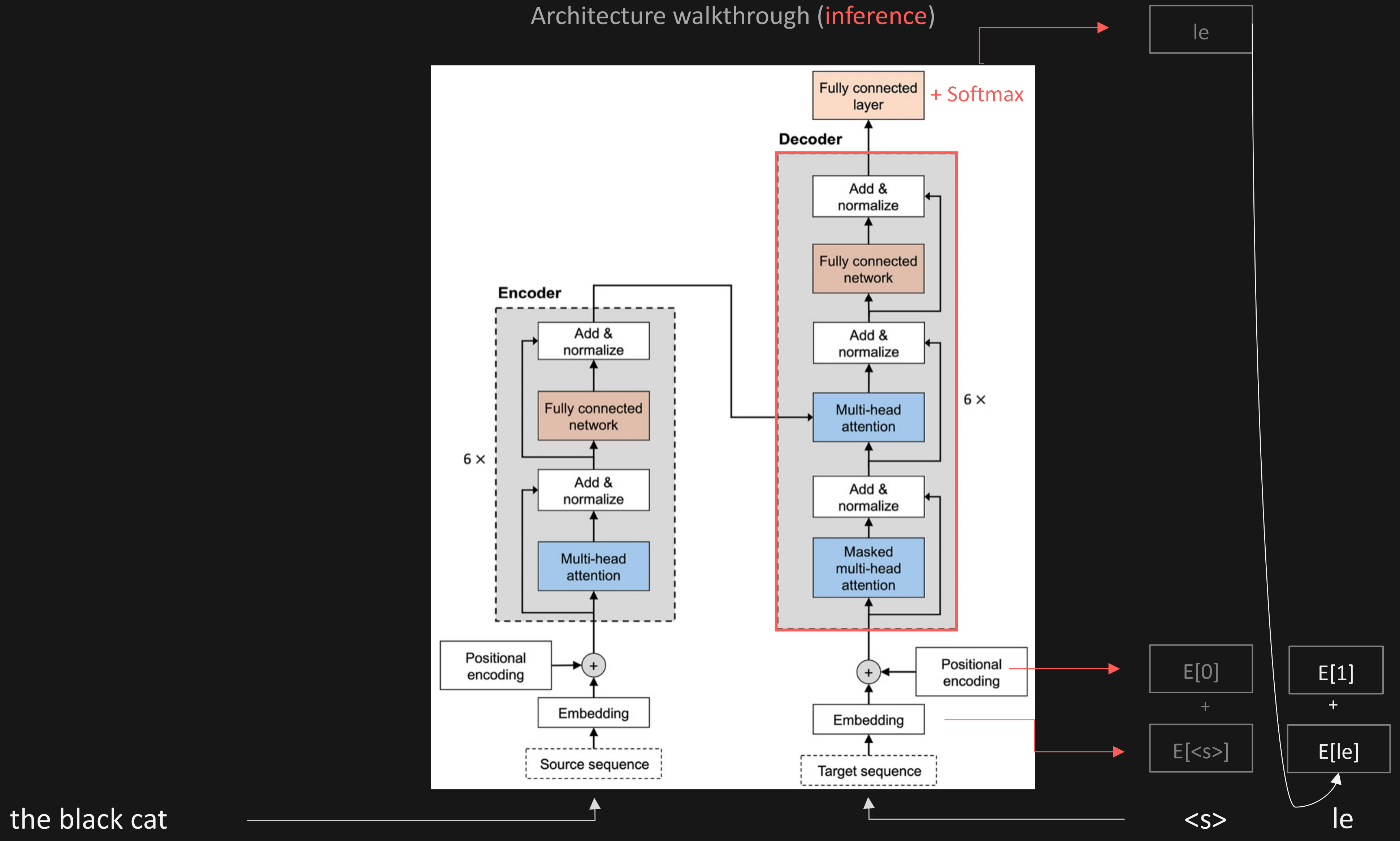
Transformer

Architecture walkthrough (inference)



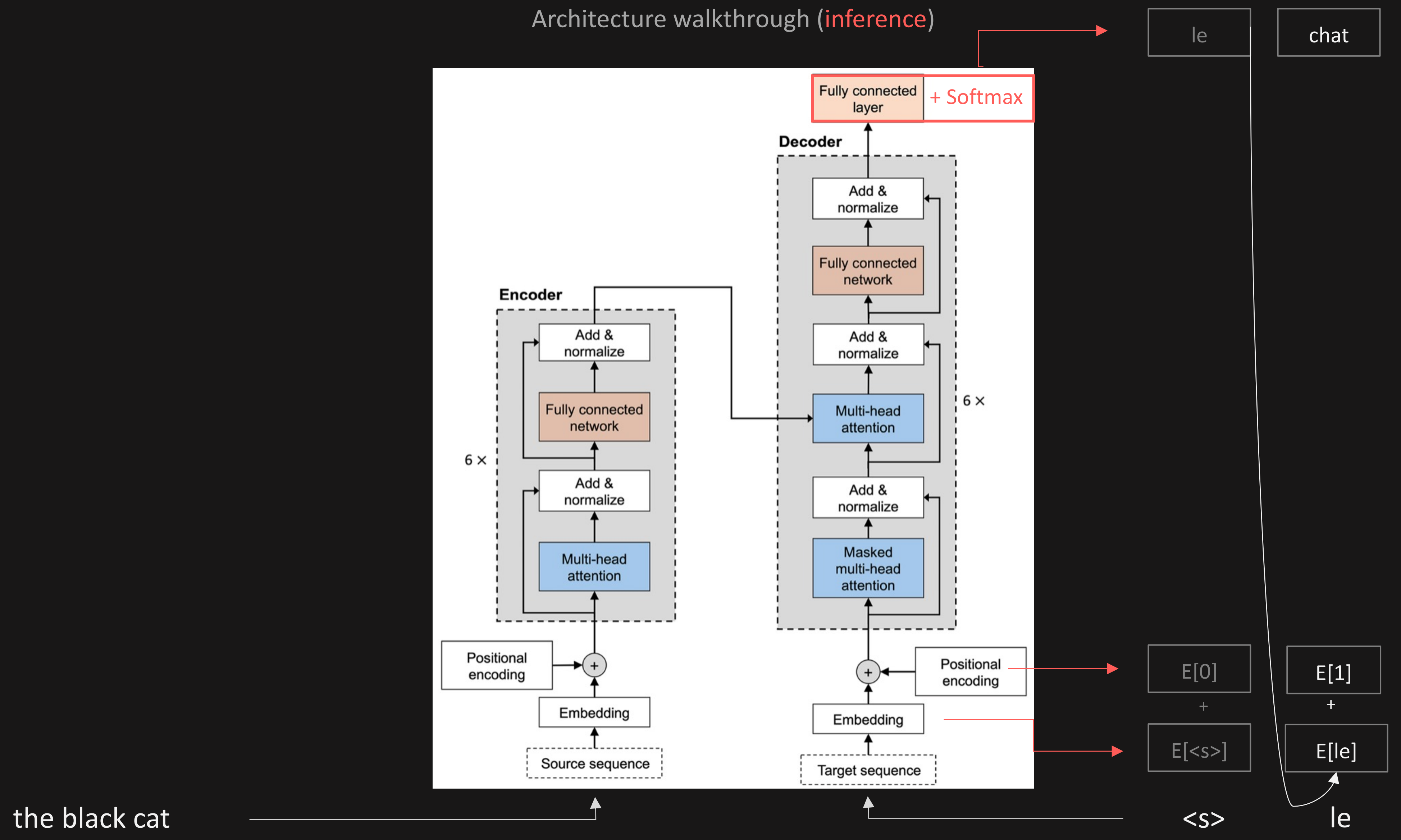
Transformer

Architecture walkthrough (inference)



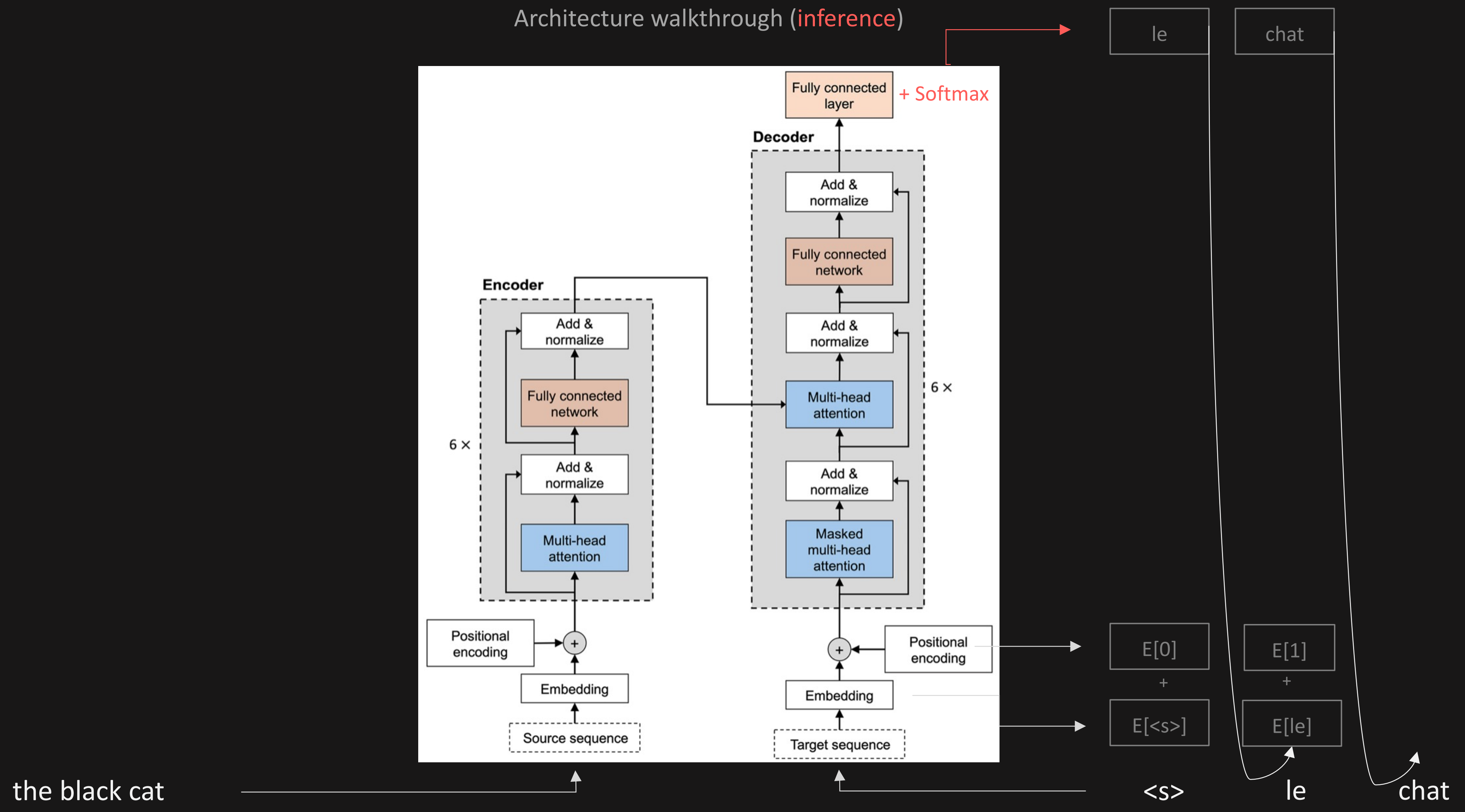
Transformer

Architecture walkthrough (inference)



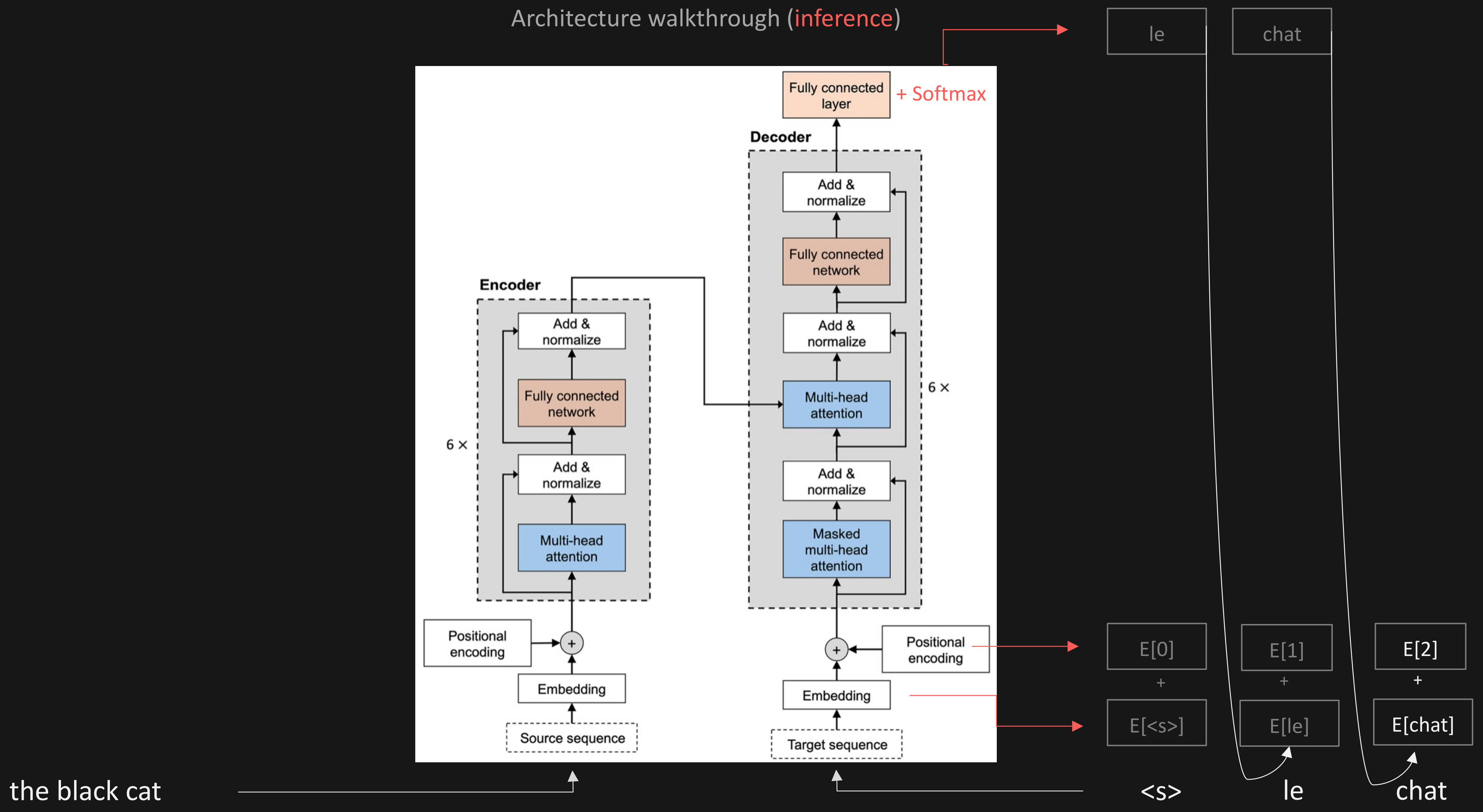
Transformer

Architecture walkthrough (inference)



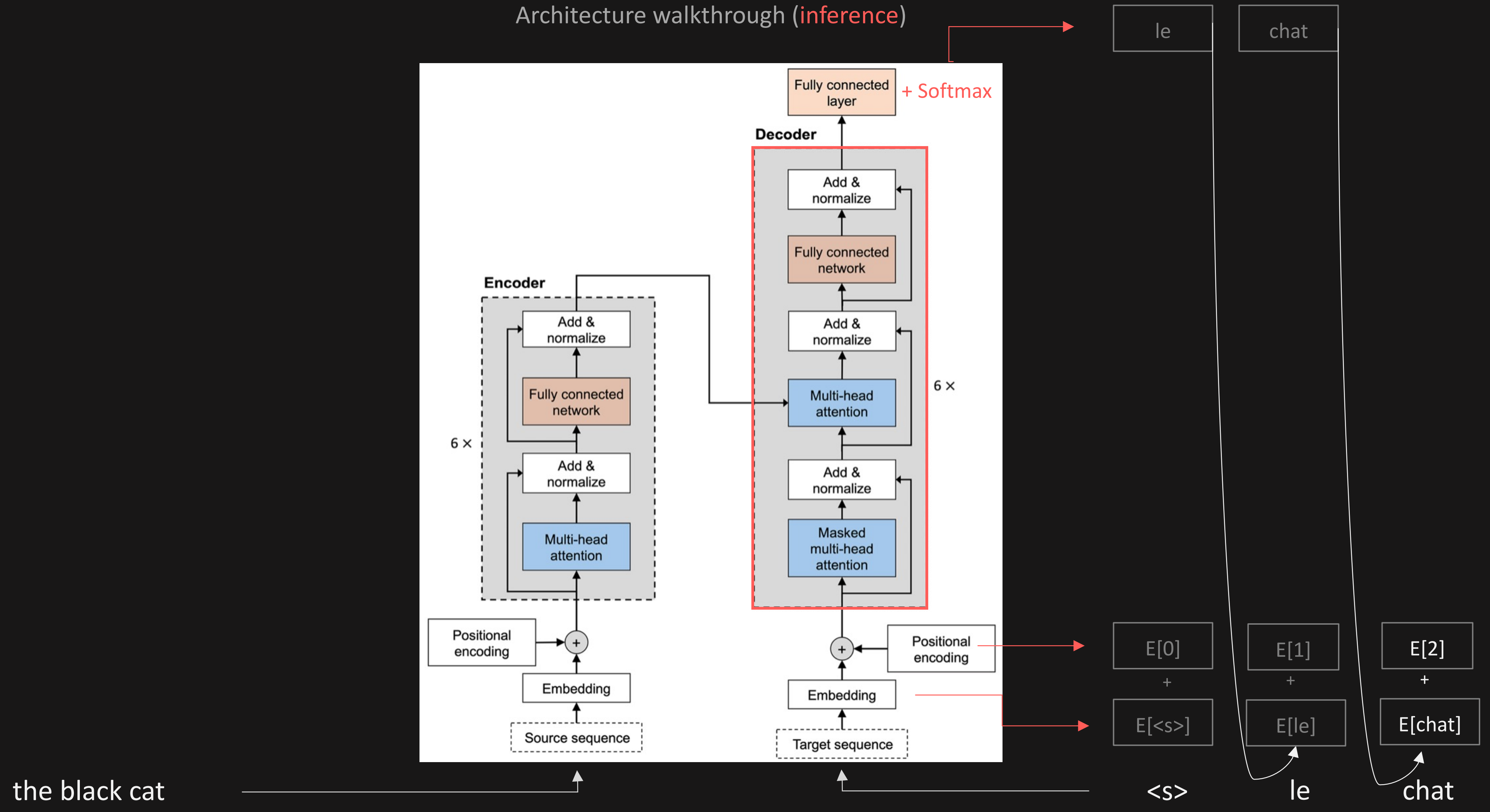
Transformer

Architecture walkthrough (inference)



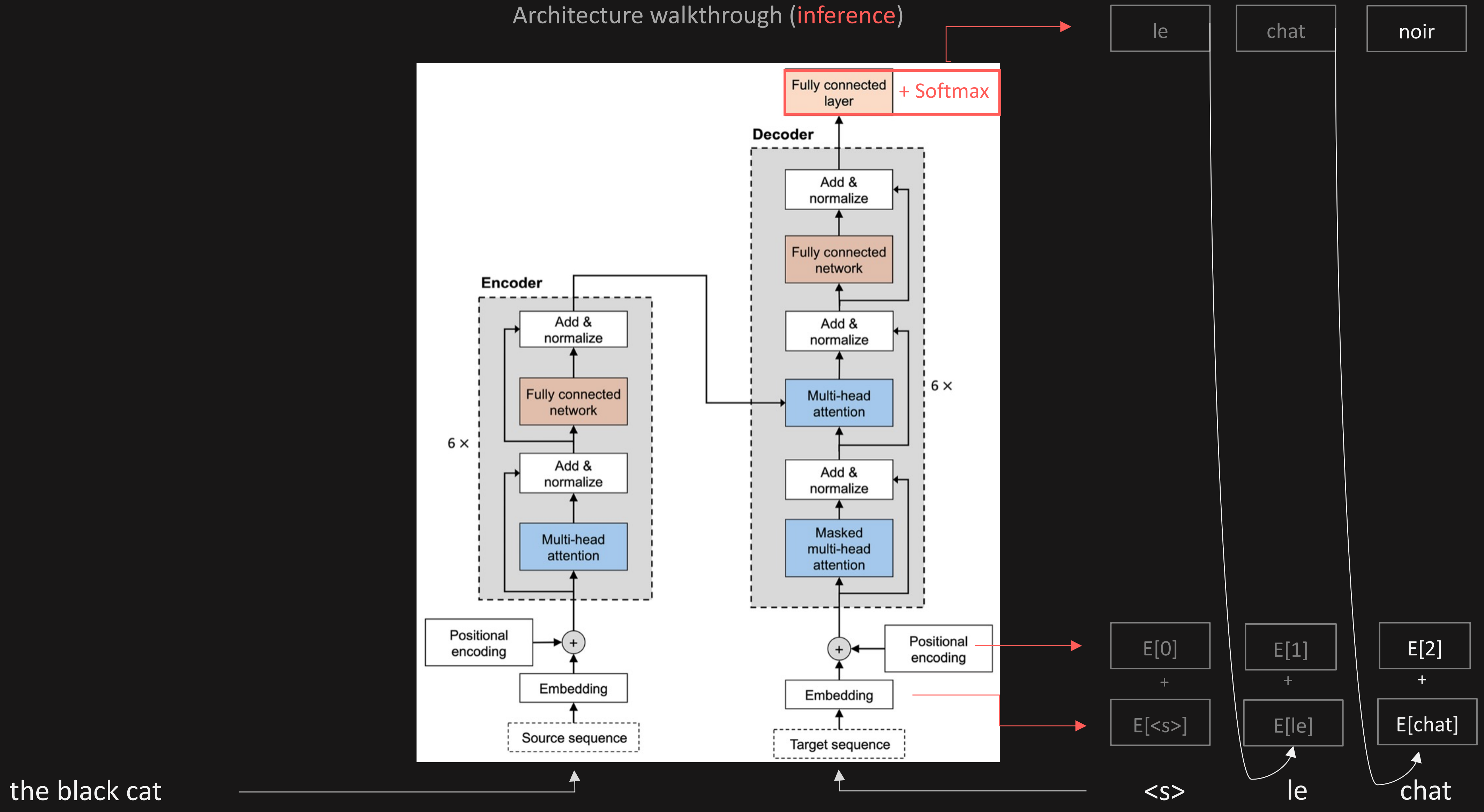
Transformer

Architecture walkthrough (inference)



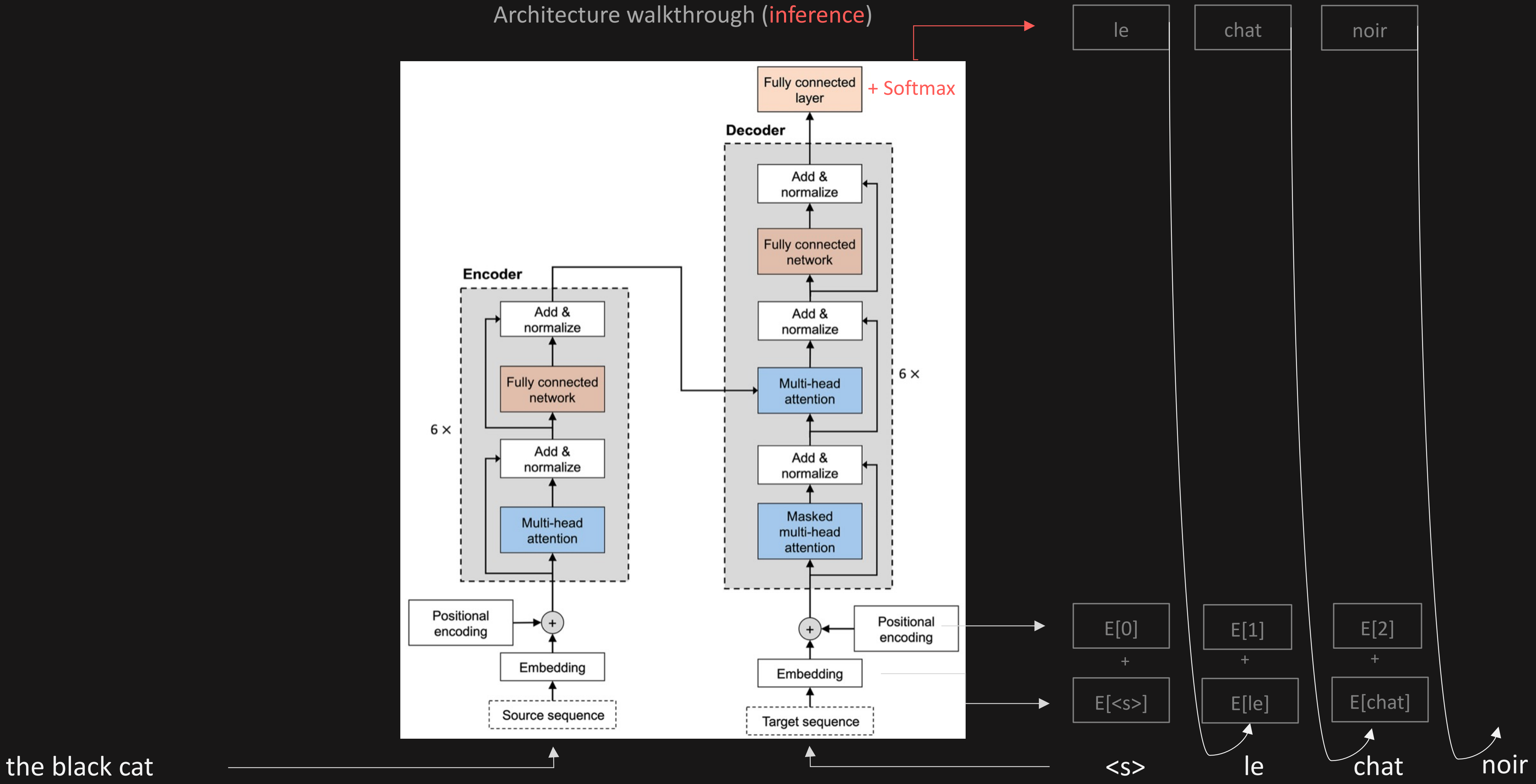
Transformer

Architecture walkthrough (inference)



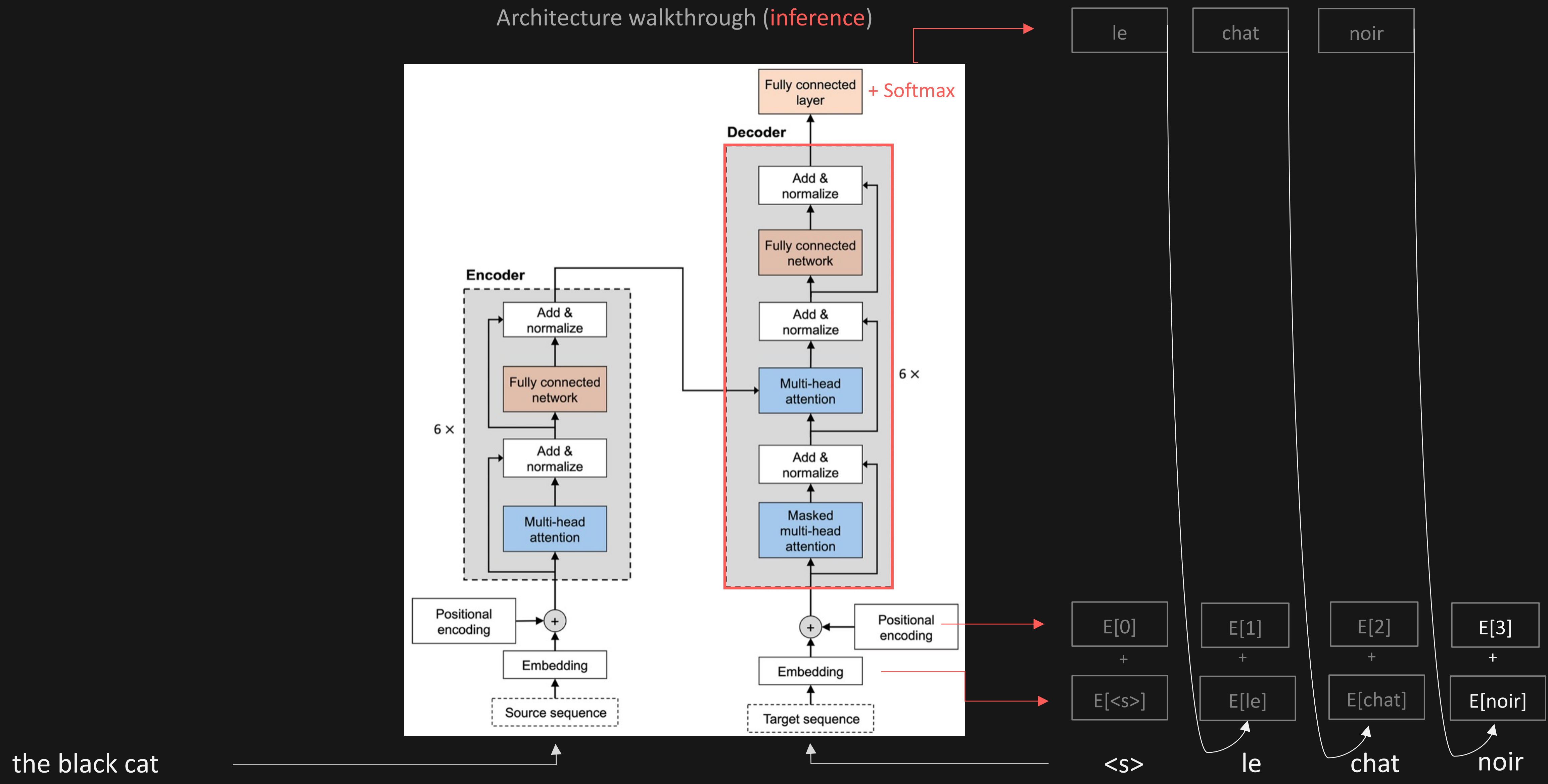
Transformer

Architecture walkthrough (inference)



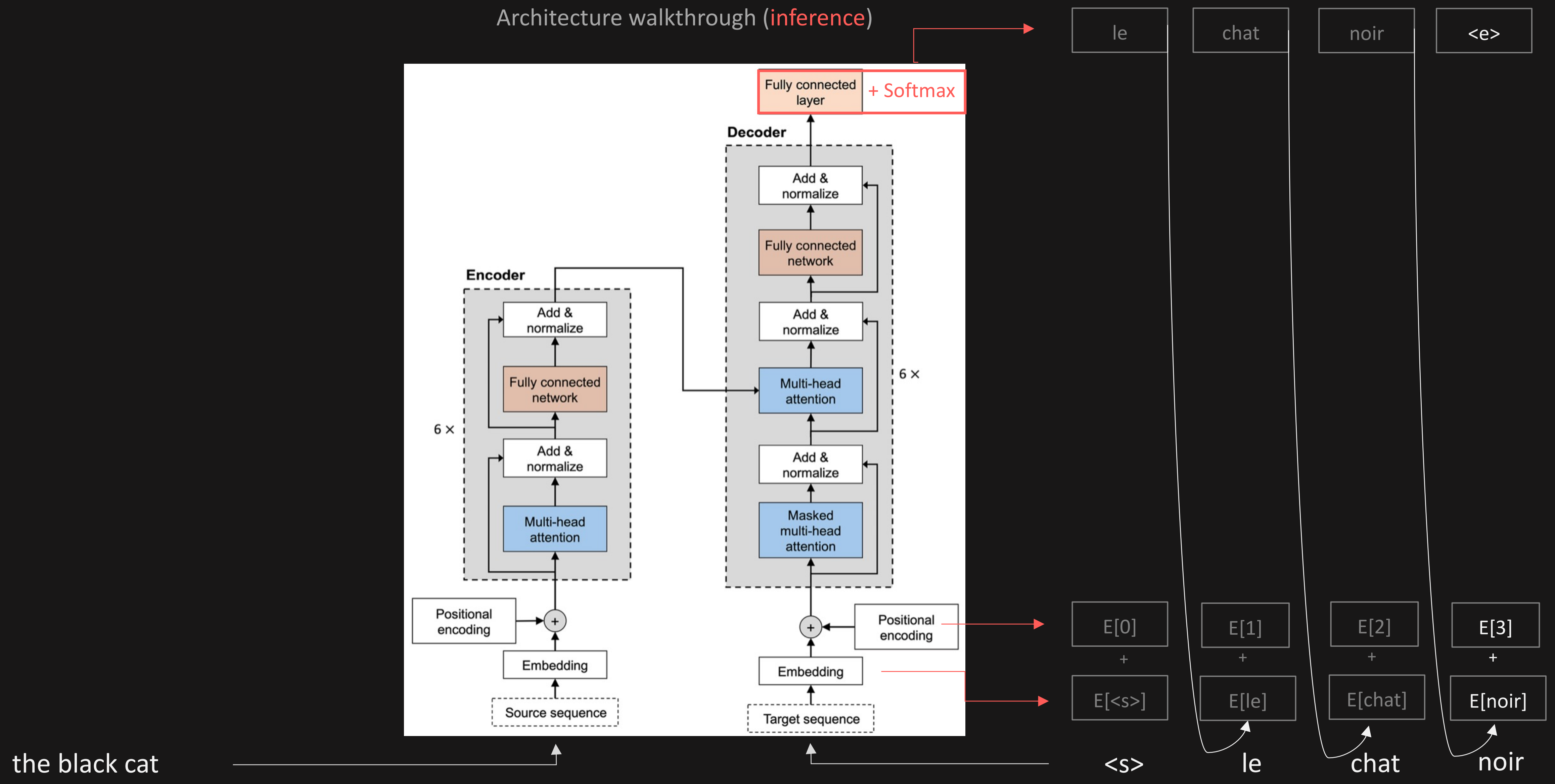
Transformer

Architecture walkthrough (inference)



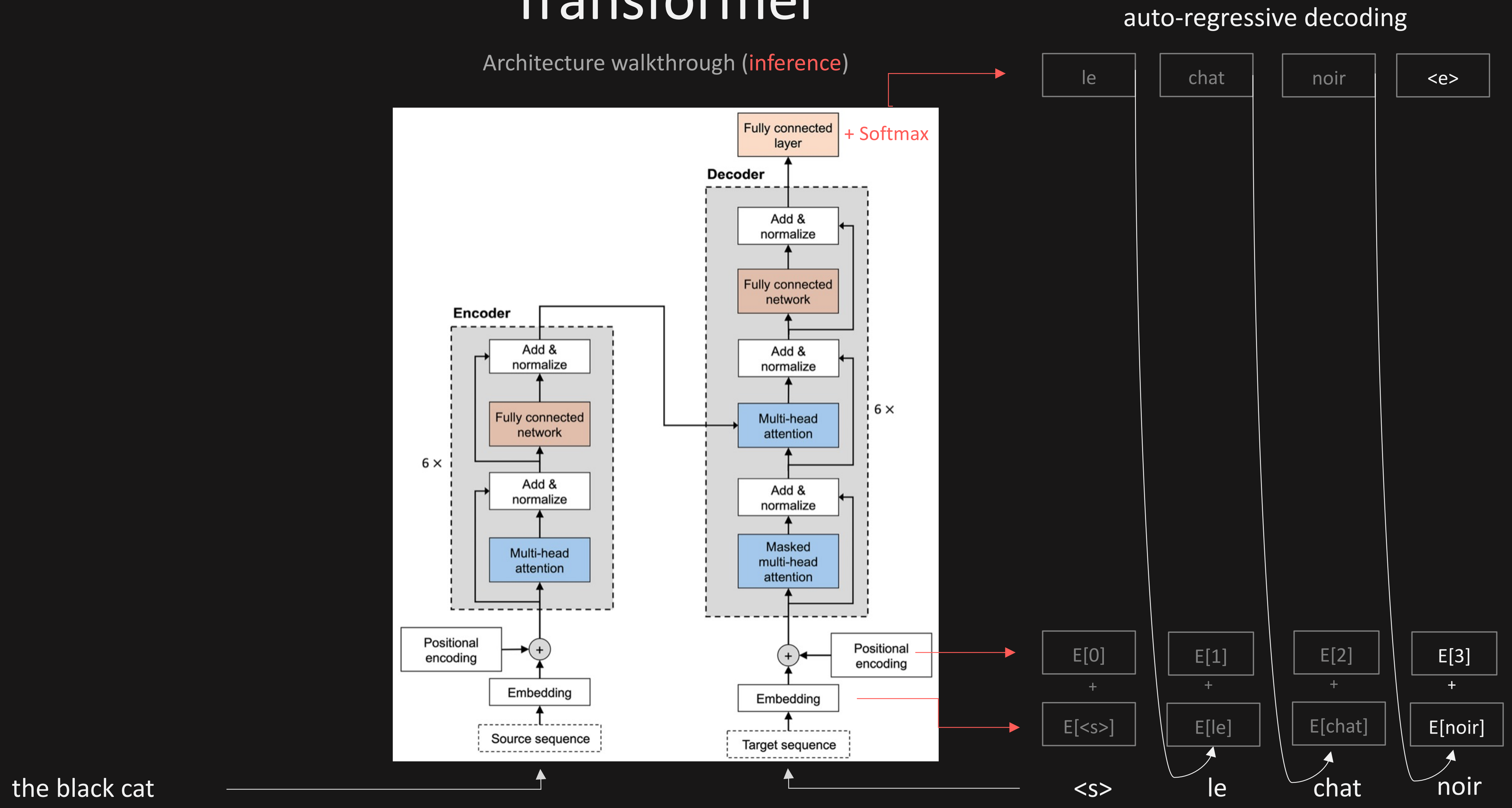
Transformer

Architecture walkthrough (inference)



Transformer

Architecture walkthrough (inference)



QUIZ

Q 1: What was the ”model of choice” before Transformers?

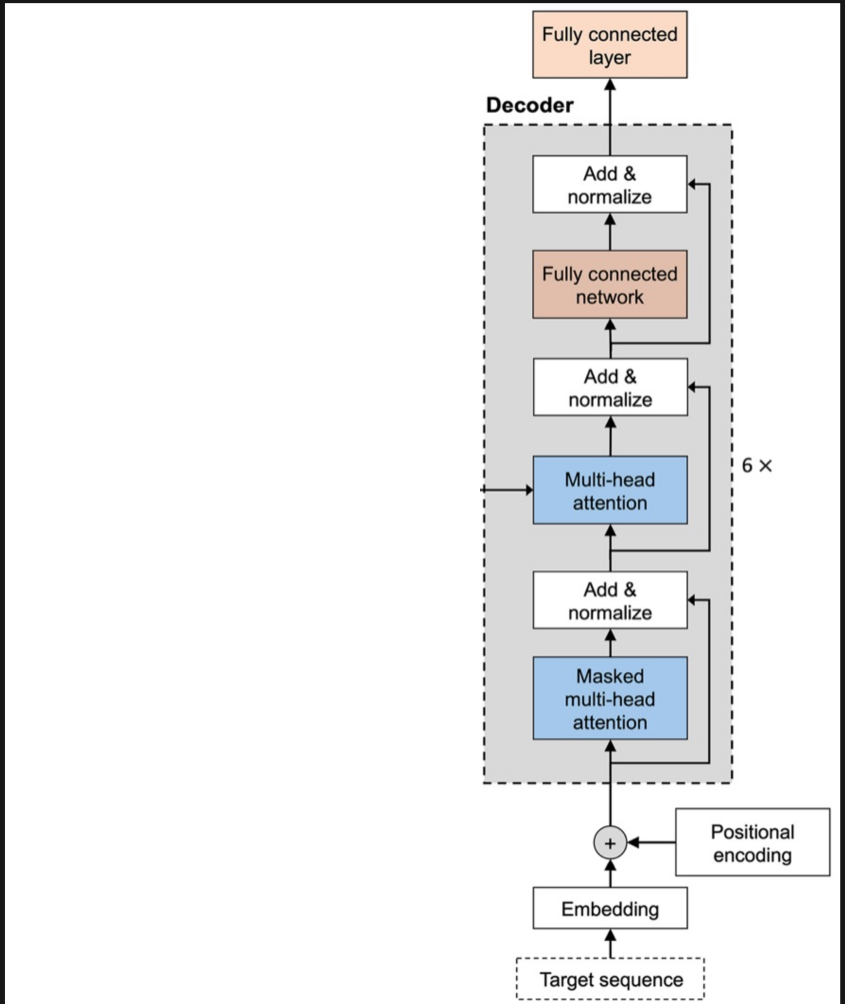
QUIZ

Q 2: What was the main issue Vaswani et al. 2017, wanted to address with the new Transformer architecture?

QUIZ

Q 2: Do you need both, an encoder and decoder?

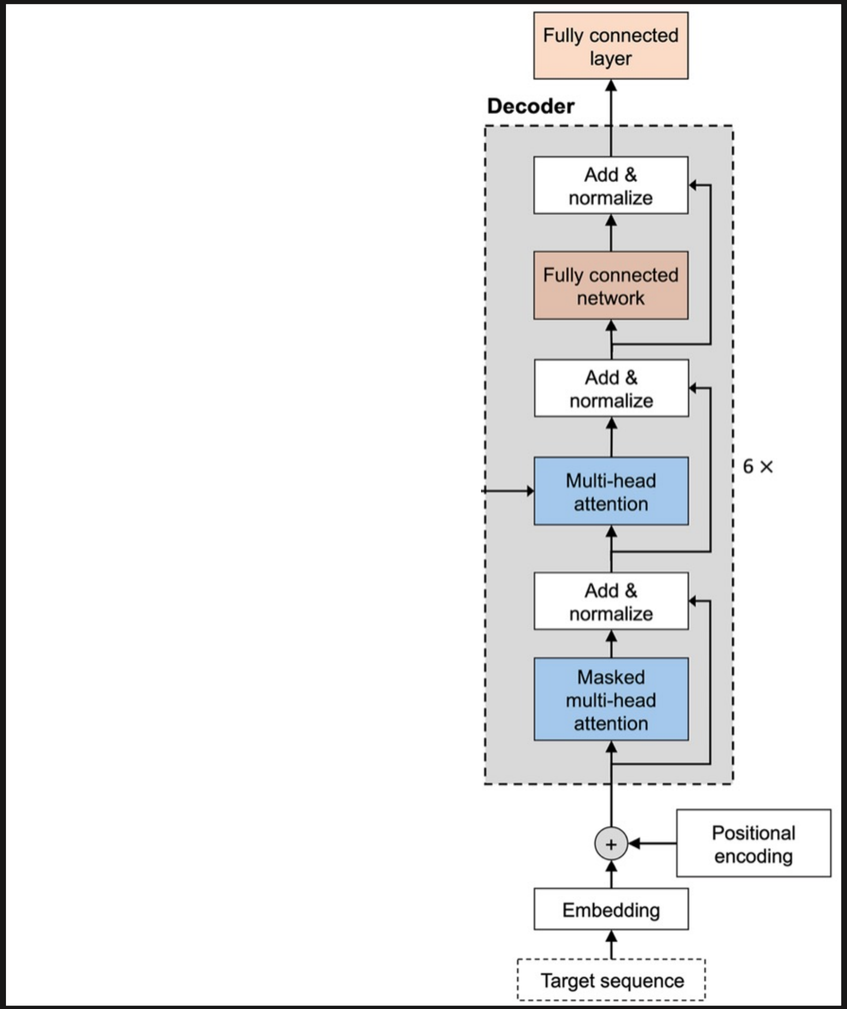
Decoder-only Transformer



e.g., OpenAI's GPT 1-4

ChatGPT

Decoder-only Transformer

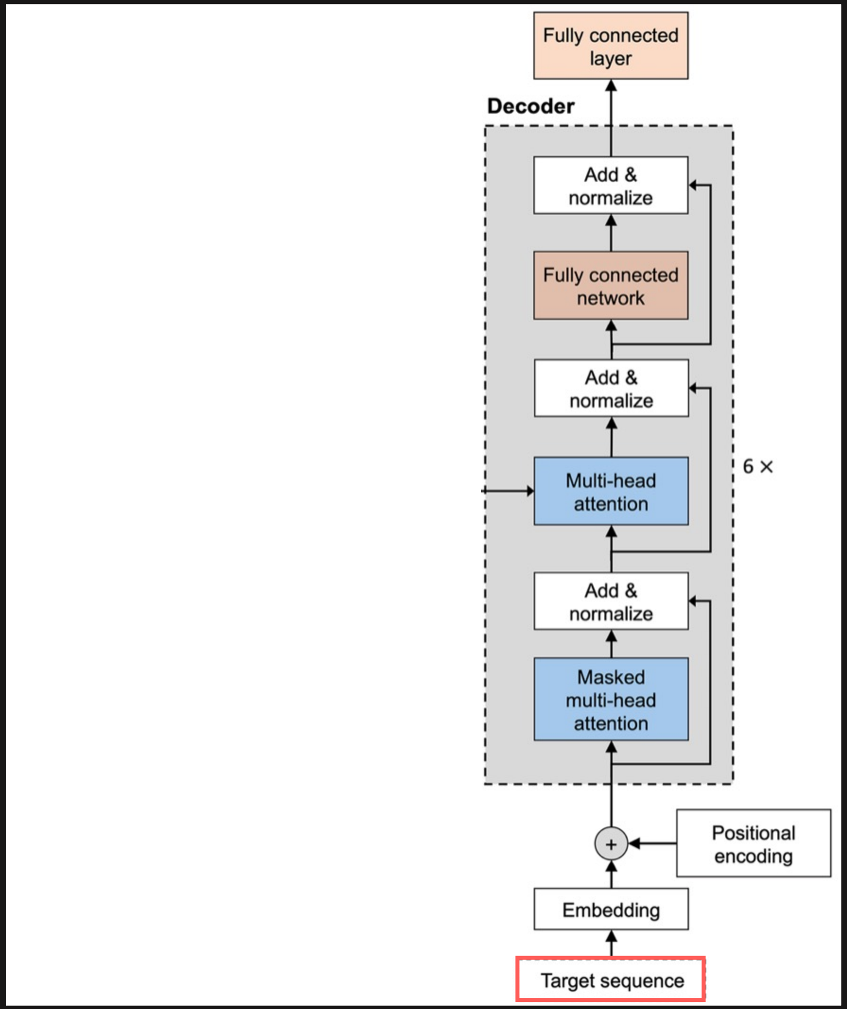


e.g., OpenAI's GPT 1-4

how does ChatGPT work?

Task: Predict next word(s)

Decoder-only Transformer



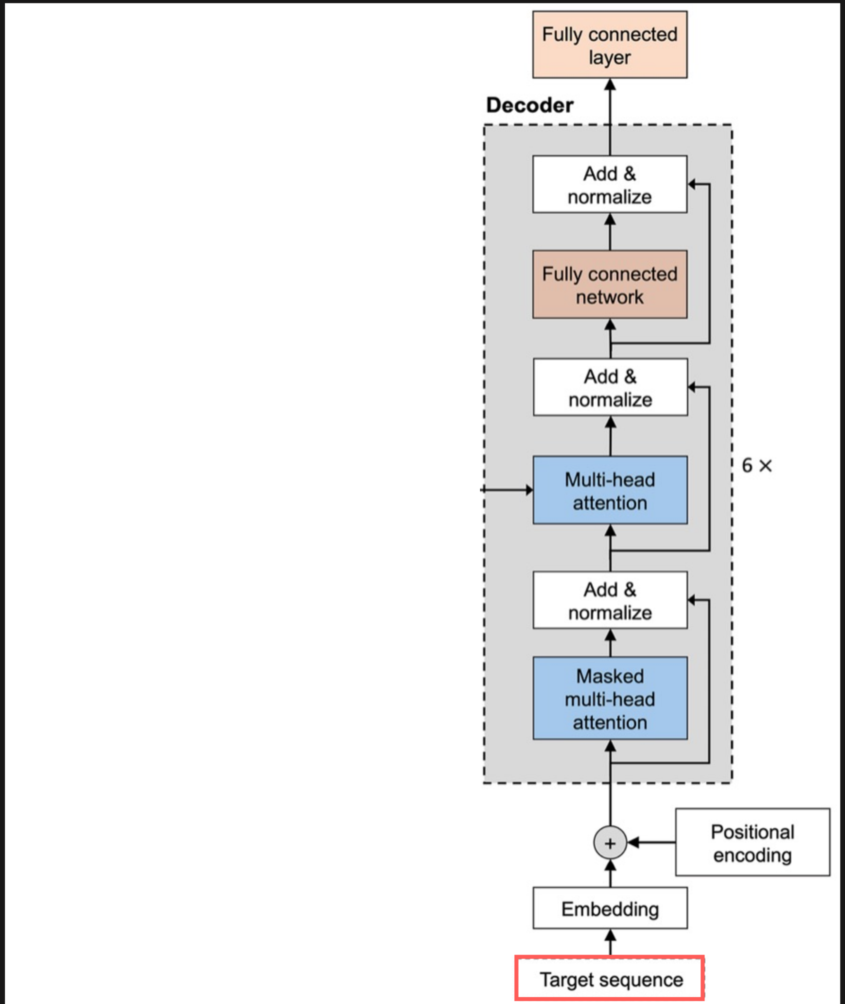
e.g., OpenAI's GPT 1-4

how does ChatGPT work?

Task: Predict next word(s)

~~<s>~~ What is a cat?

Decoder-only Transformer



e.g., OpenAI’s GPT 1-4

how does ChatGPT work?

Task: Predict next word(s)

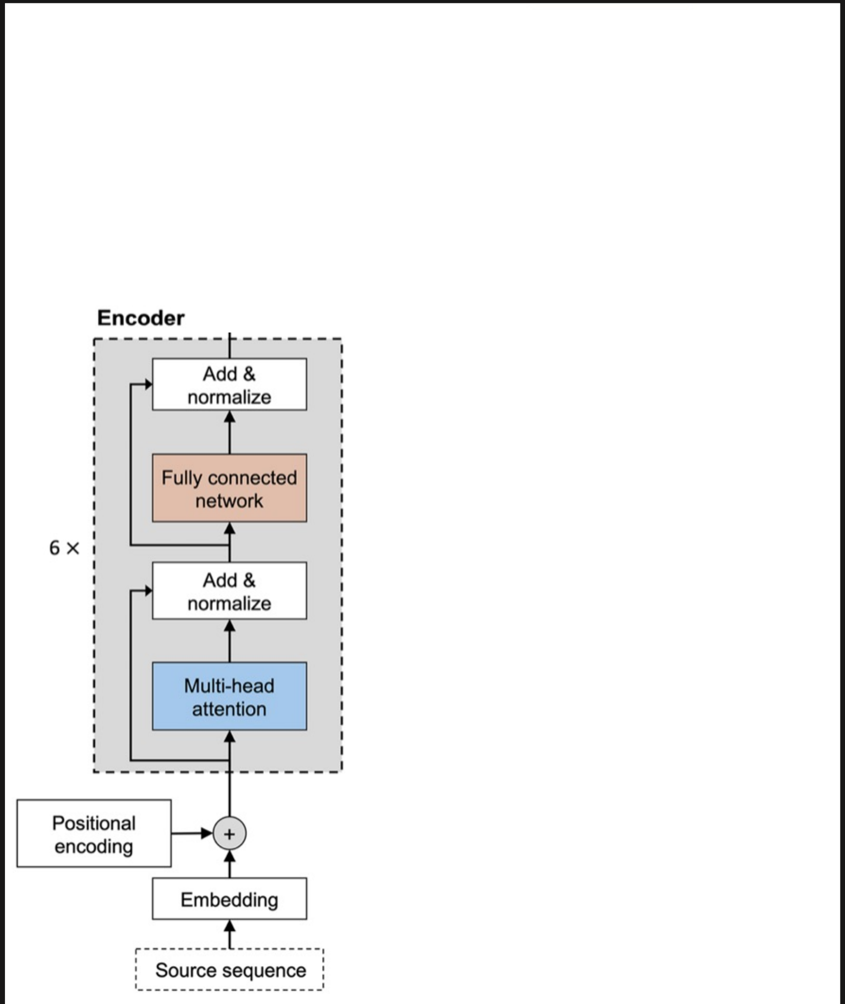
A cat is …

~~<s>~~ What is a cat?

Encoder-only Transformer

e.g., BERT

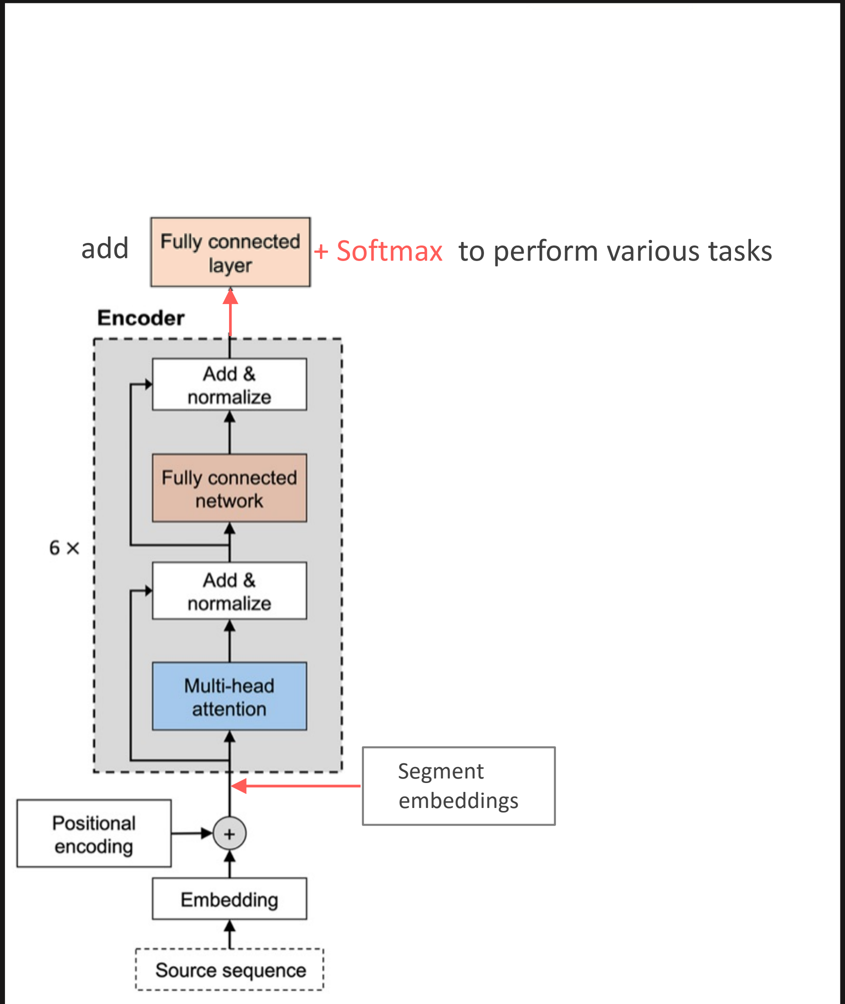
how does it work?



Encoder-only Transformer

e.g., BERT

how does it work?



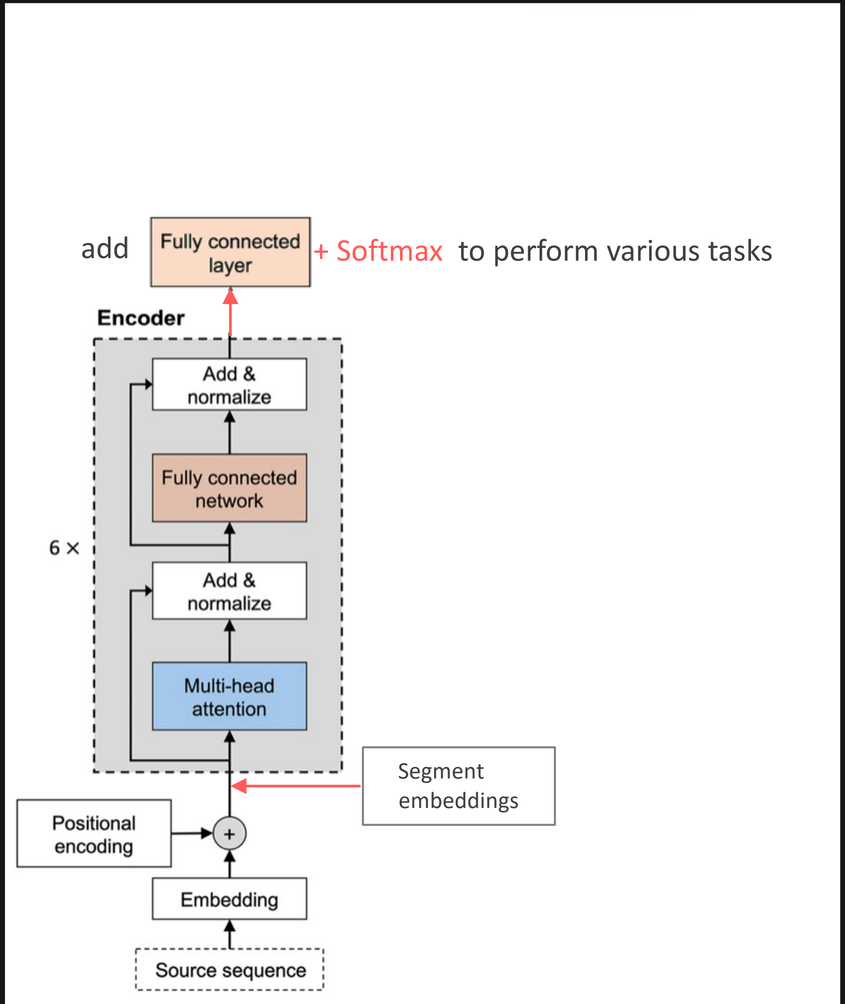
Encoder-only Transformer

e.g., BERT

how does it work?

Task 1: pre-train the model on large self-labeled data

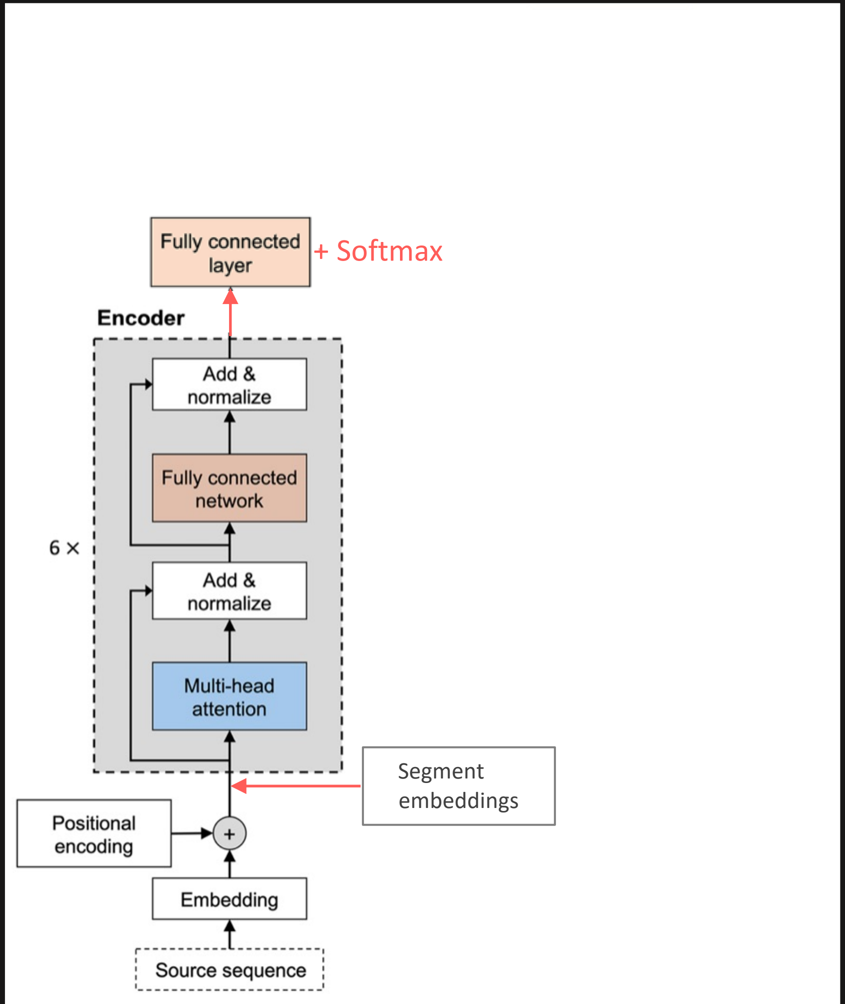
Task 2: fine-tune to specific task, e.g., sentiment analysis



BERT

Devlin et al., 2018

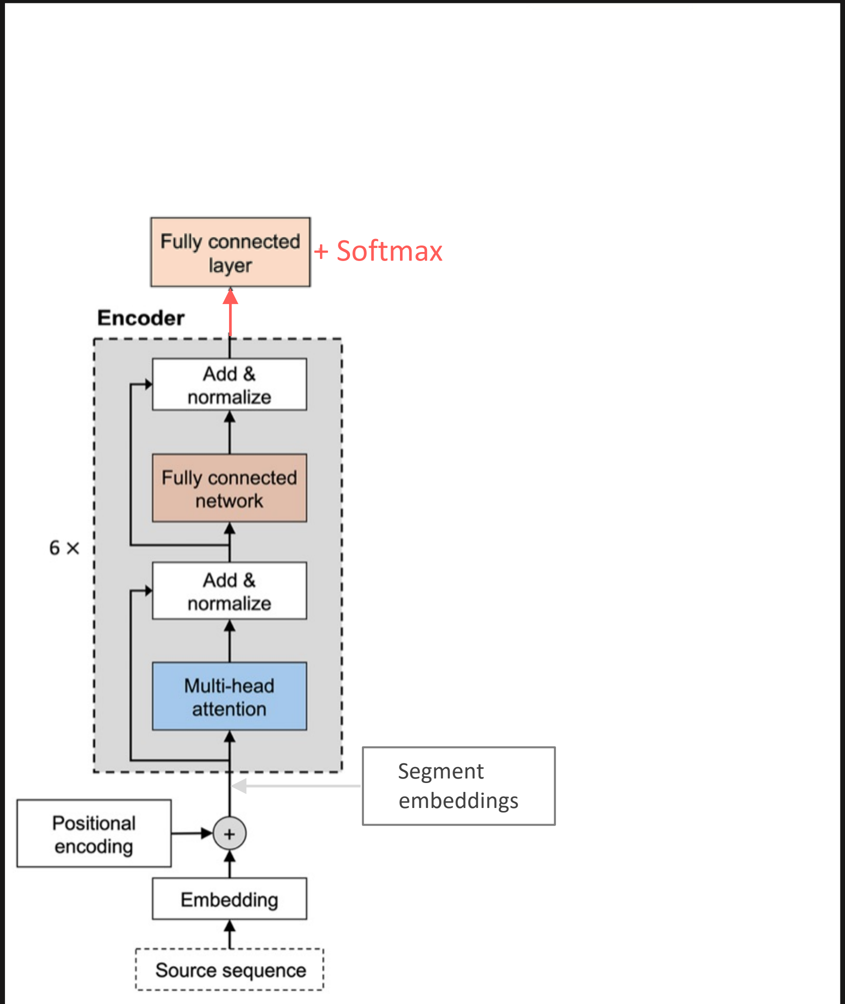
Task 1: pre-train the model on large self-labeled data



BERT

Devlin et al., 2018

Task 1: pre-train the model on large self-labeled data

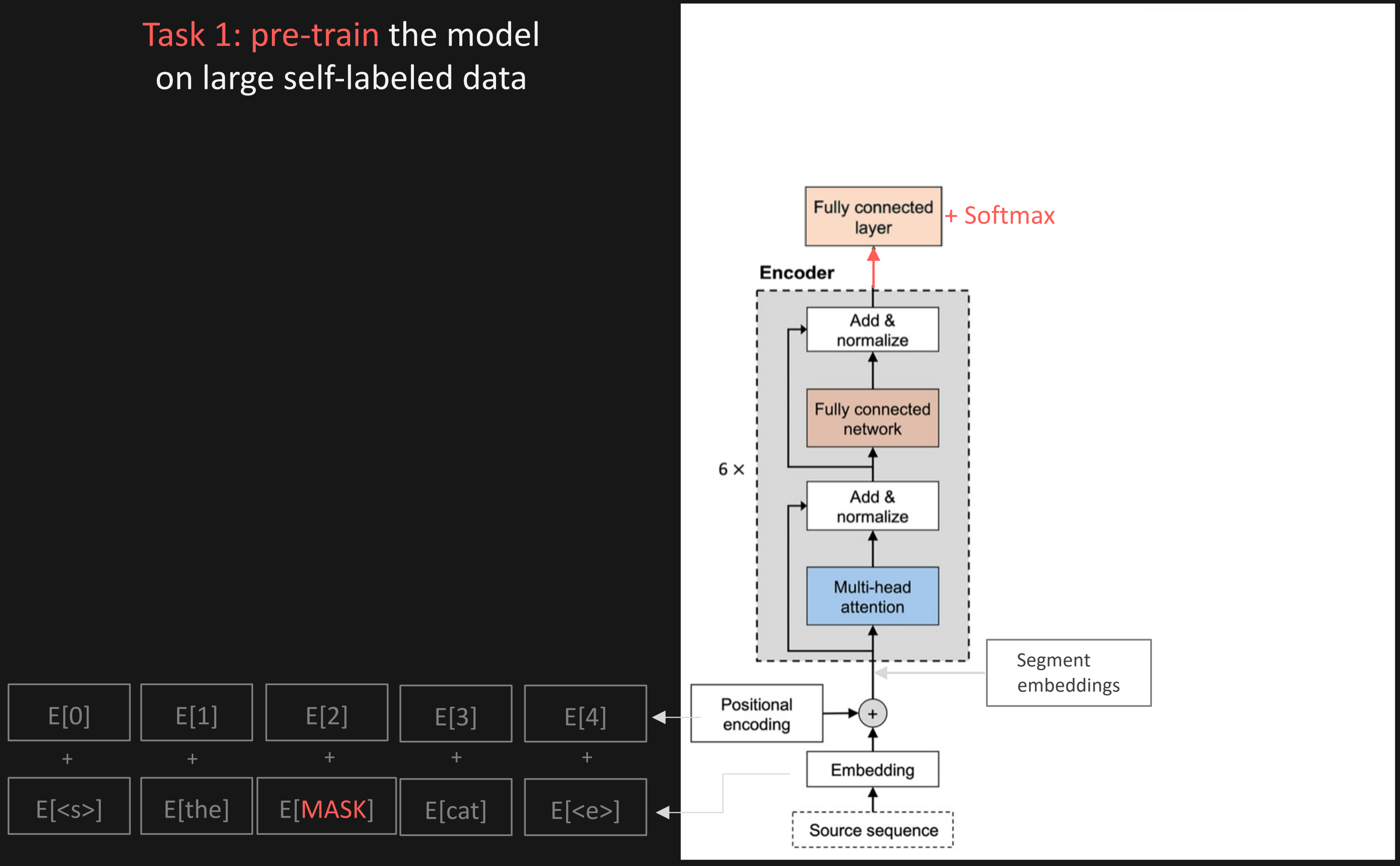


the MASK cat

BERT

Devlin et al., 2018

Task 1: pre-train the model on large self-labeled data

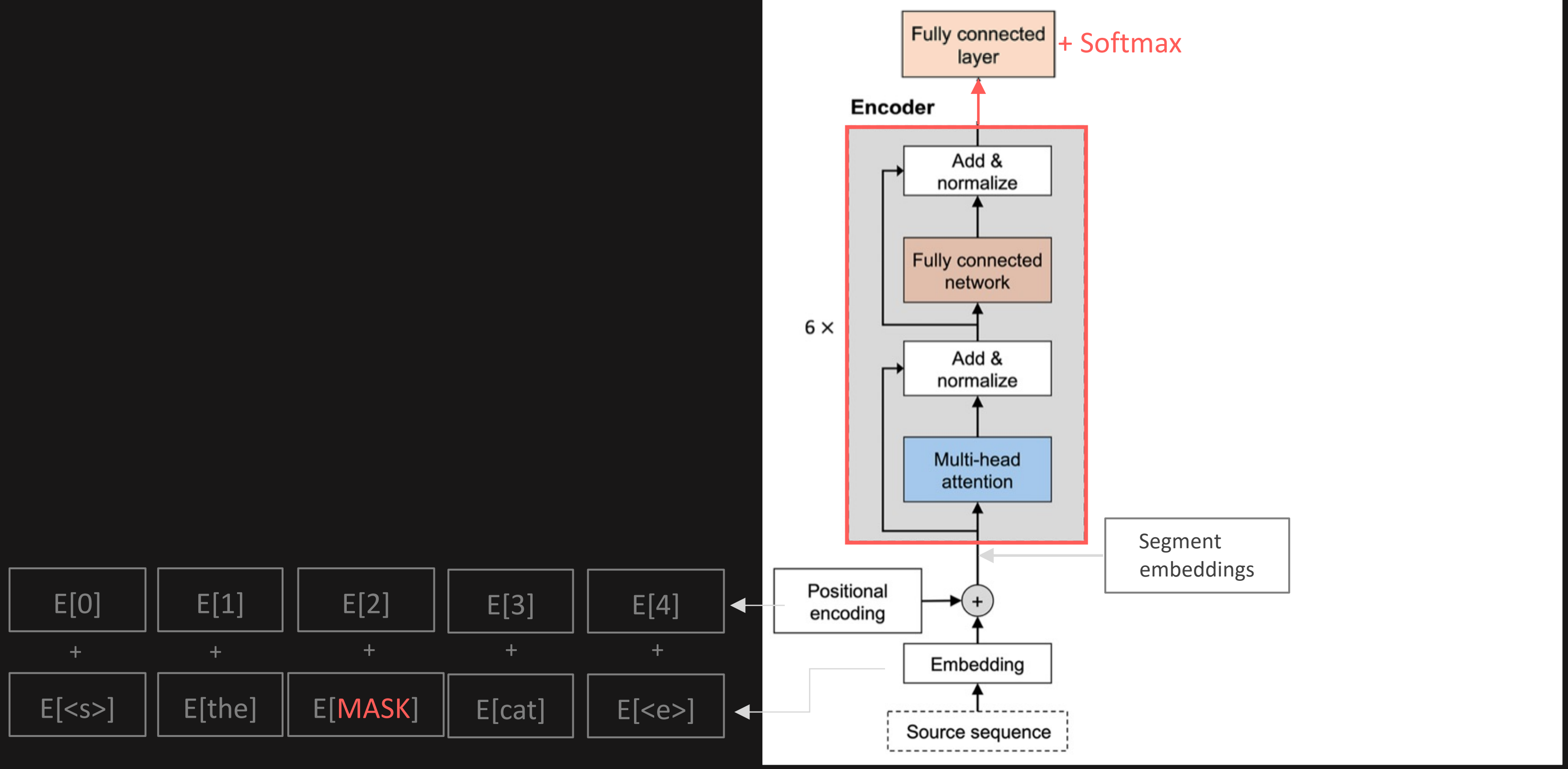


the MASK cat

BERT

Devlin et al., 2018

Task 1: pre-train the model
on large self-labeled data



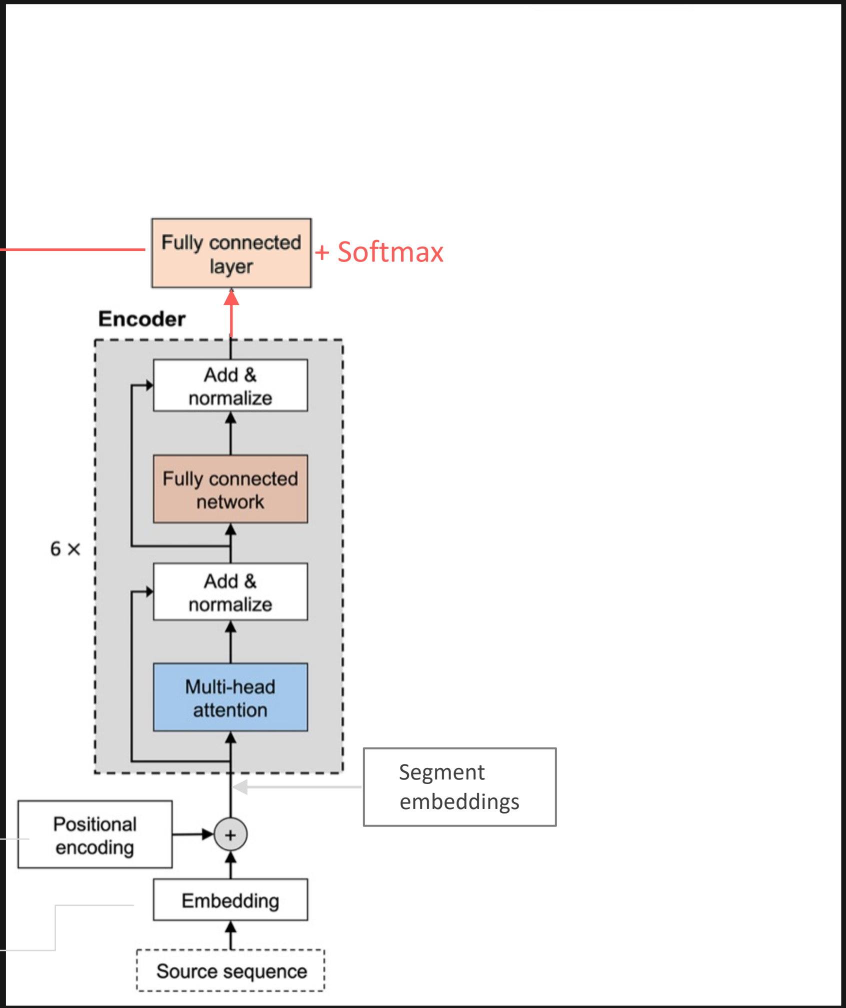
the MASK cat

BERT

Devlin et al., 2018

Task 1: pre-train the model
on large self-labeled data

P_{MASK}={...,black,...}



E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[MASK]	E[cat]	E[<e>]

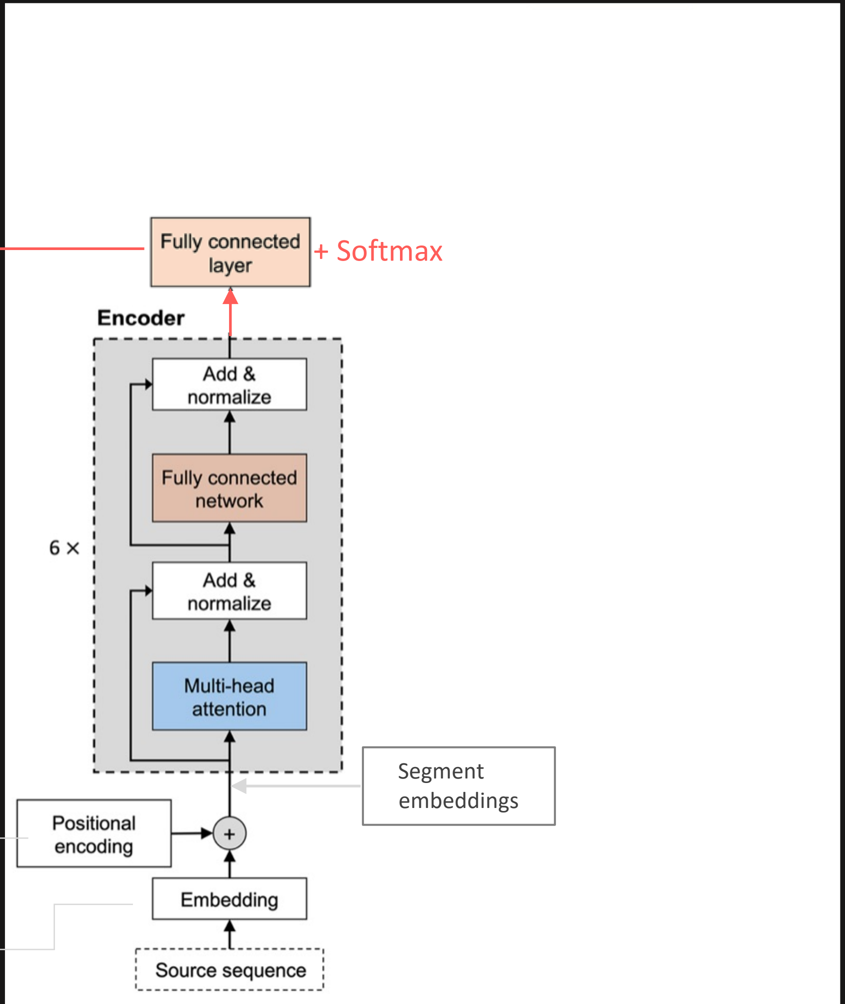
the MASK cat

BERT

Devlin et al., 2018

Task 1: pre-train the model on large self-labeled data

P_{MASK}={…,black,…}



Task 2: fine-tune to specific task, e.g., sentiment analysis

Sentiment prediction

E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[MASK]	E[cat]	E[<e>]

e.g., Movie reviews

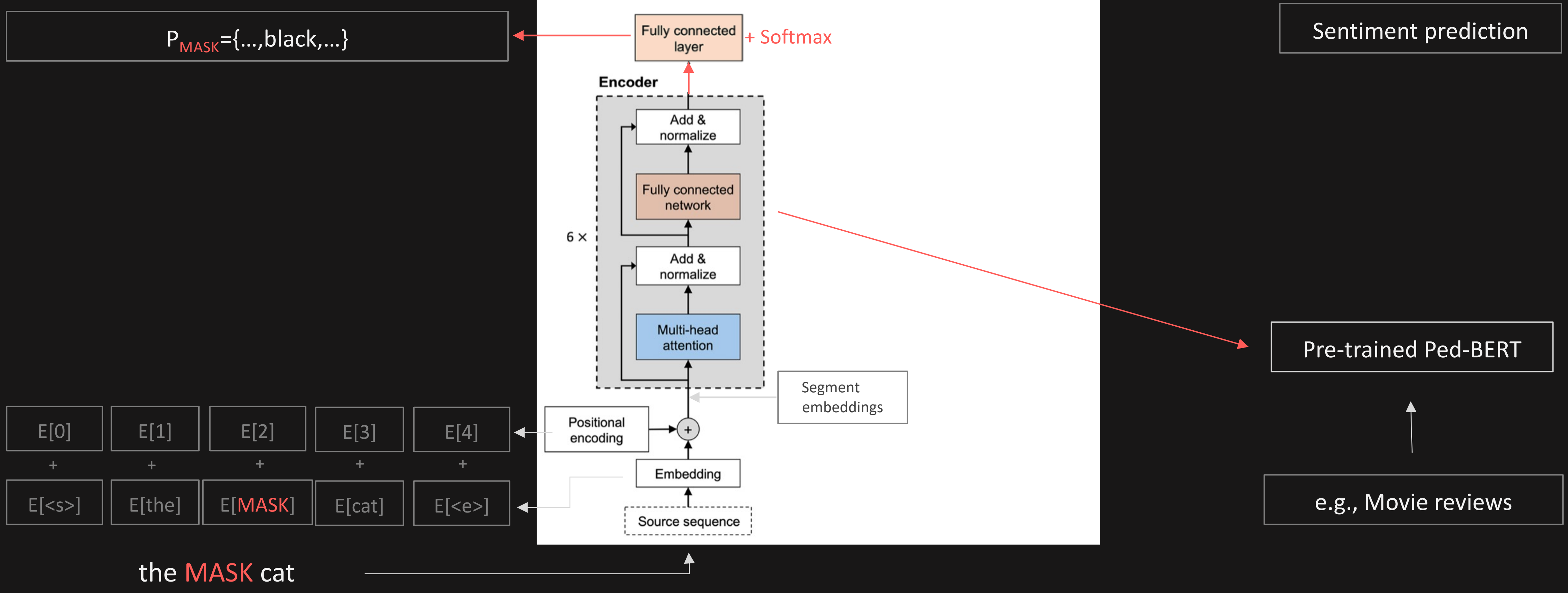
the MASK cat

BERT

Devlin et al., 2018

Task 1: pre-train the model on large self-labeled data

Task 2: fine-tune to specific task, e.g., sentiment analysis

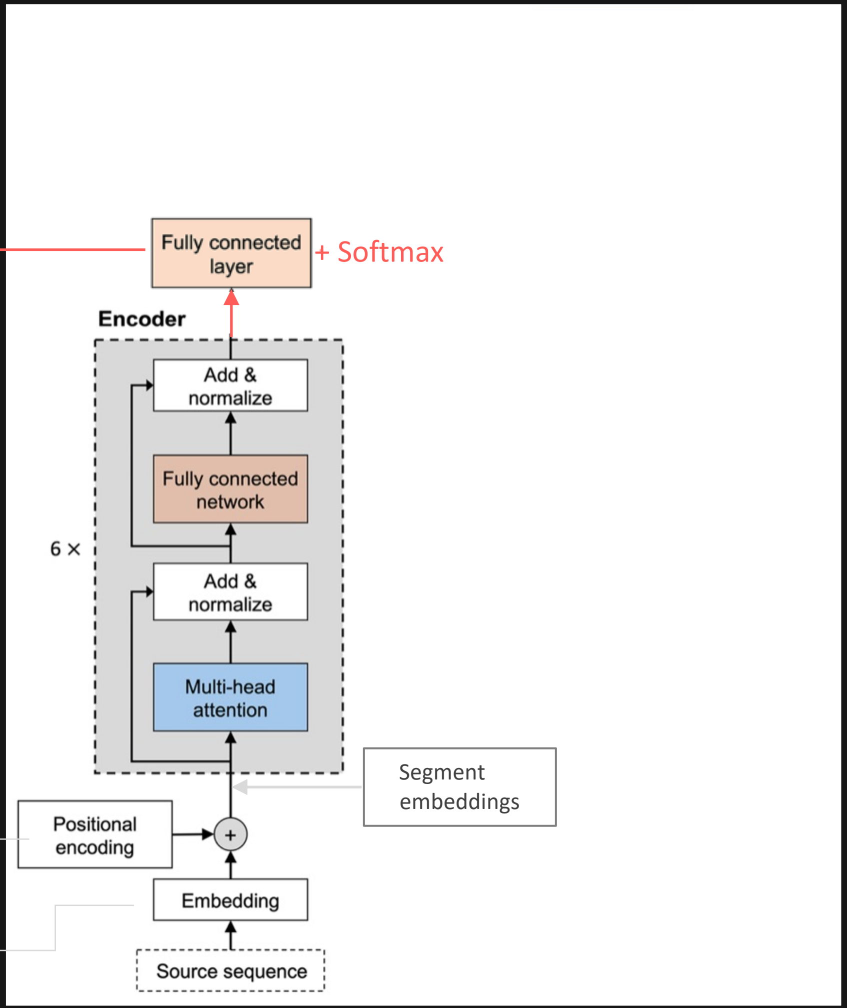


BERT

Devlin et al., 2018

Task 1: pre-train the model on large self-labeled data

P_{MASK}={…,black,…}



E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[the]	E[MASK]	E[cat]	E[<e>]

the MASK cat

Task 2: fine-tune to specific task, e.g., sentiment analysis

Sentiment prediction

output layer + softmax activation

+

FFNN

update all model parameters

+

Pre-trained Ped-BERT

e.g., Movie reviews

QUIZ

Q 4: What is the difference between pre-training and fine-tuning?

Ped-BERT: Early Detection of Disease for Pediatric Care

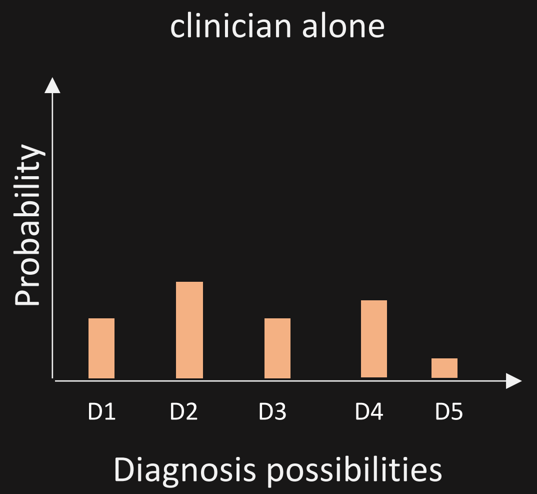
- Motivation
- Literature review
- Data
- Methods
- Results

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

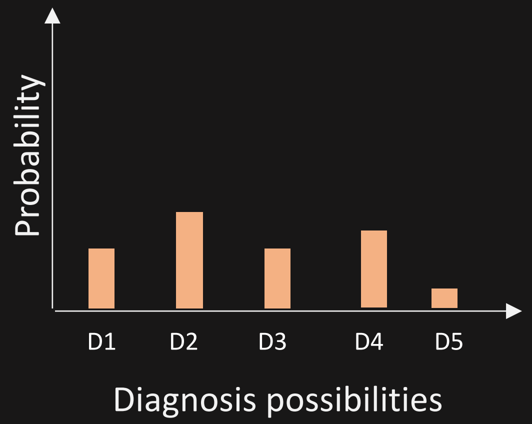
- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure



- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

Can we augment clinical decision making with ML?

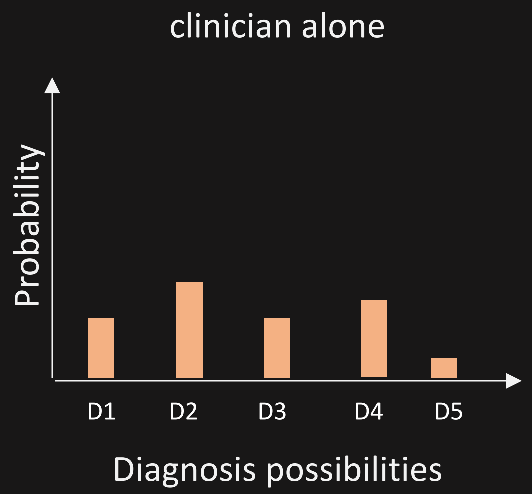
clinician alone



- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

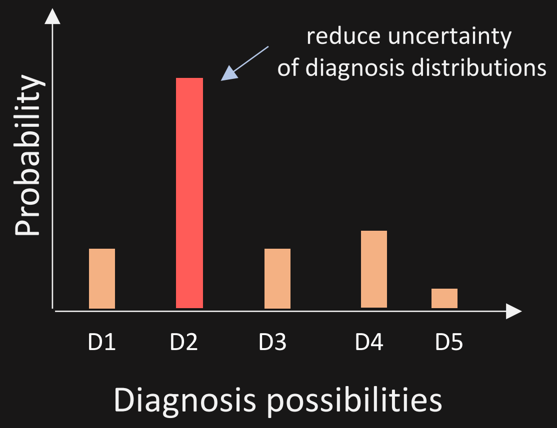
Not replace

Can we AUGMENT clinical decision making with ML?



- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

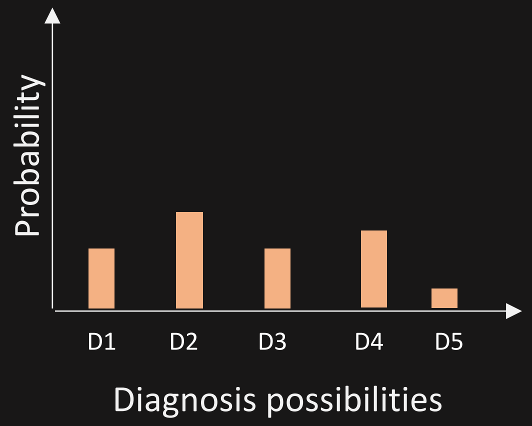
clinician + ML



Not replace

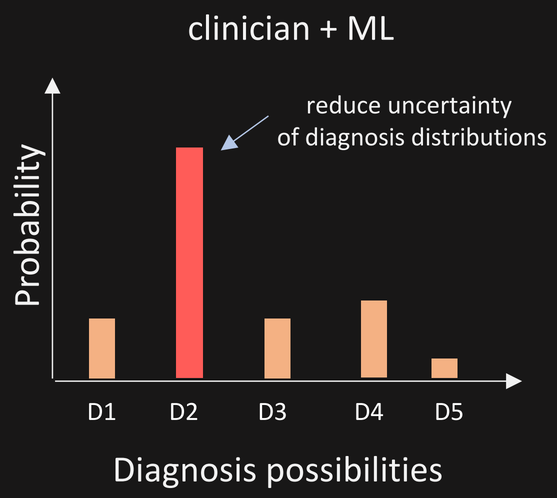
Can we AUGMENT clinical decision making with ML?

clinician alone



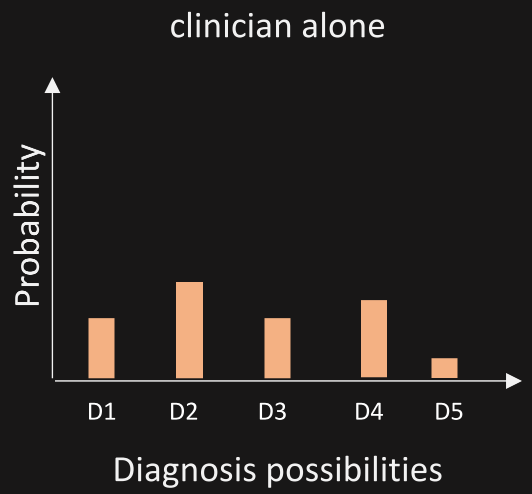
- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

Even more relevant for pediatric patients



Not replace

Can we AUGMENT clinical decision making with ML?



Literature

	Our study	Previous studies
Age	Pediatrics	Adults (mostly)

Literature

	Our study	Previous studies
Age	Pediatrics	Adults (mostly)
Data	US (+7000 hospitals in California)	UK (mostly)

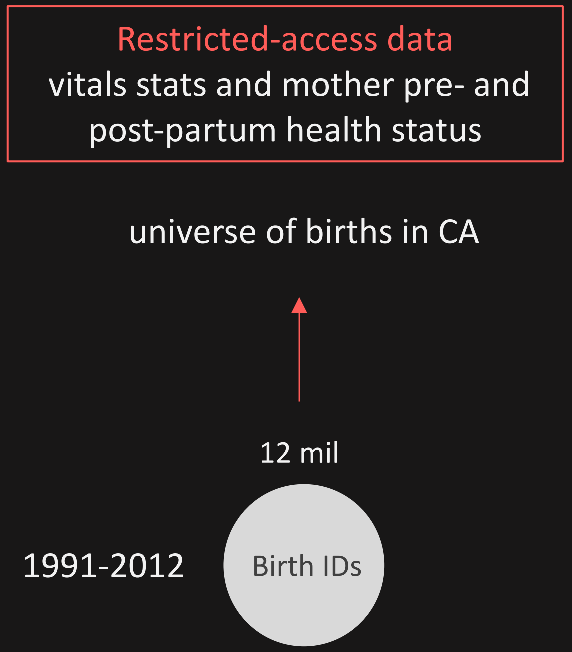
Literature

	Our study	Previous studies
Age	Pediatrics	Adults (mostly)
Data	US (+7000 hospitals in California)	UK (mostly)
Rare diseases	Yes	No

Literature

	Our study	Previous studies
Age	Pediatrics	Adults (mostly)
Data	US (+7000 hospitals in California)	UK (mostly)
Rare diseases	Yes	No
Modeling	Ped-BERT (incl., zip location)	Logistic, RNN/LSTM, BERT

Ped-BERT Data



Ped-BERT Data

Restricted-access data
vitals stats and mother pre- and post-partum health status

universe of births in CA

12 mil

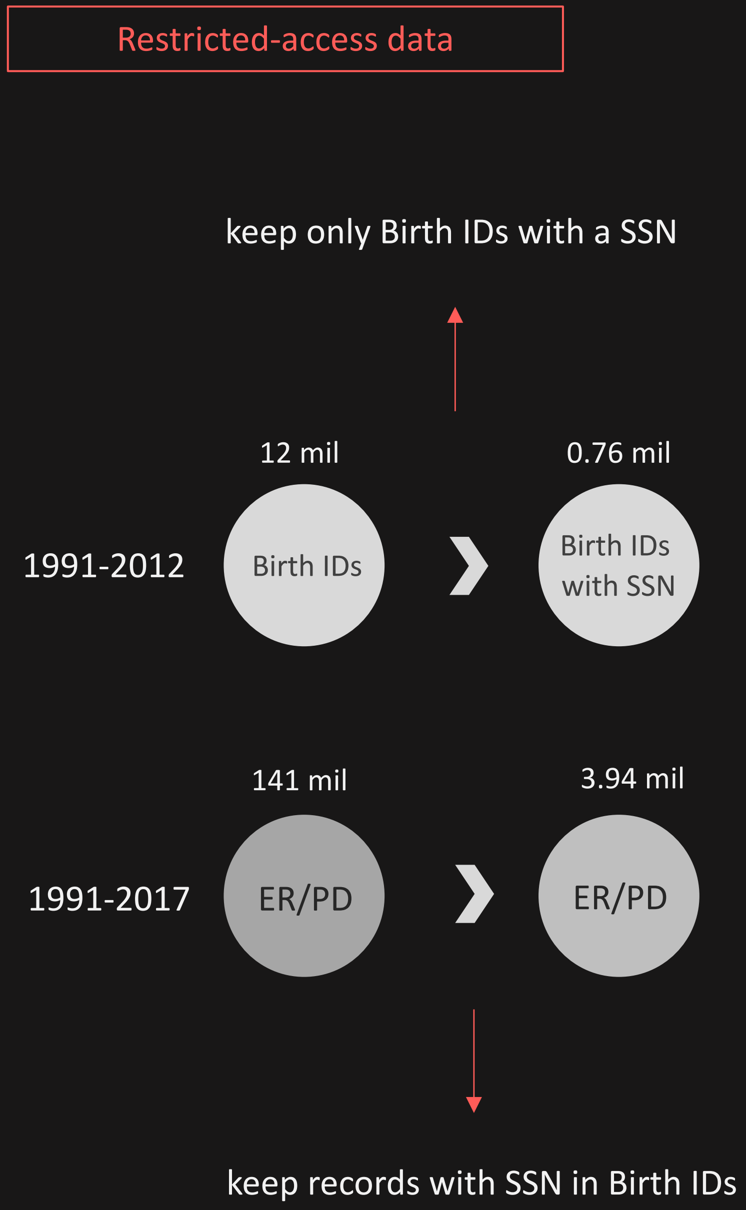
1991-2012 Birth IDs

141 mil

1991-2017 ER/PD

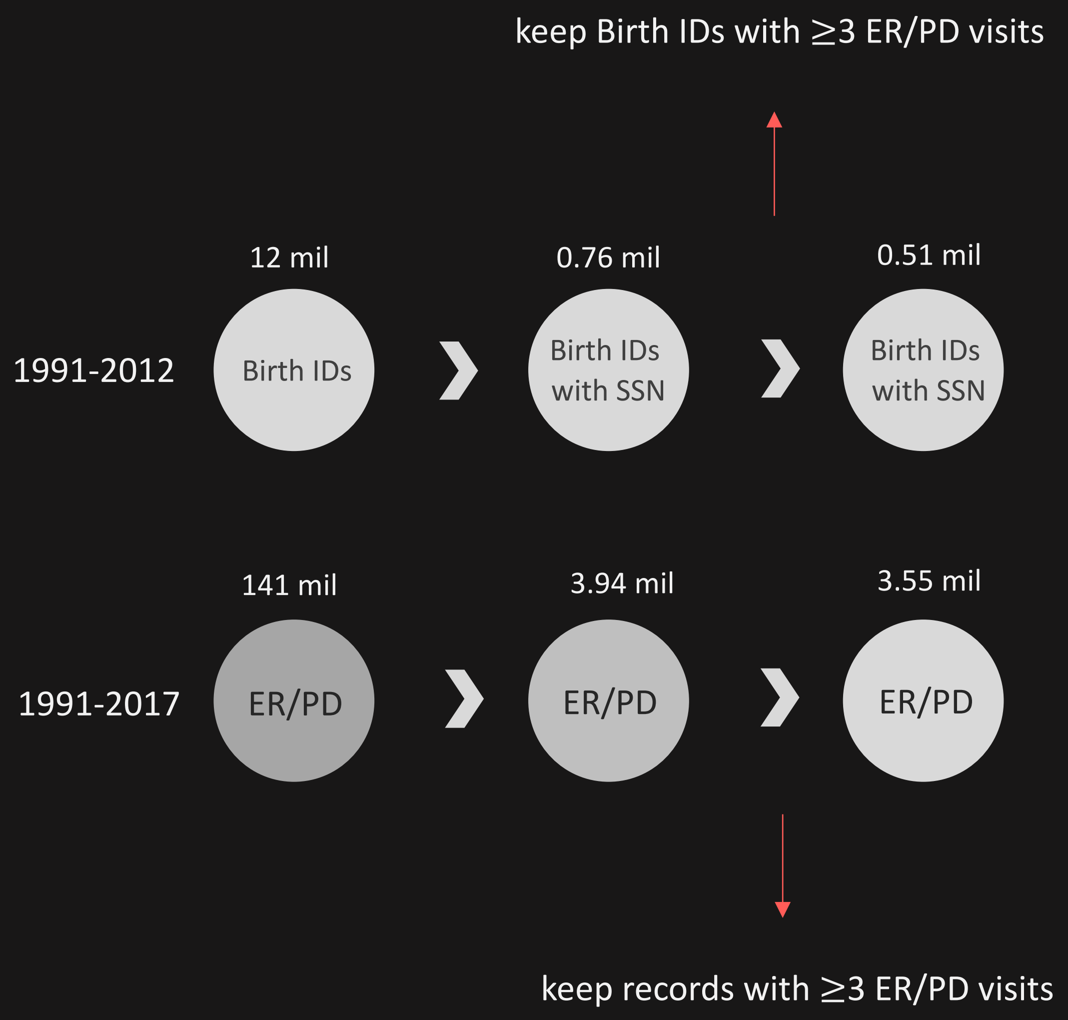
ER and Patient Discharges (PD) in CA
(diagnosis codes)

Ped-BERT Data



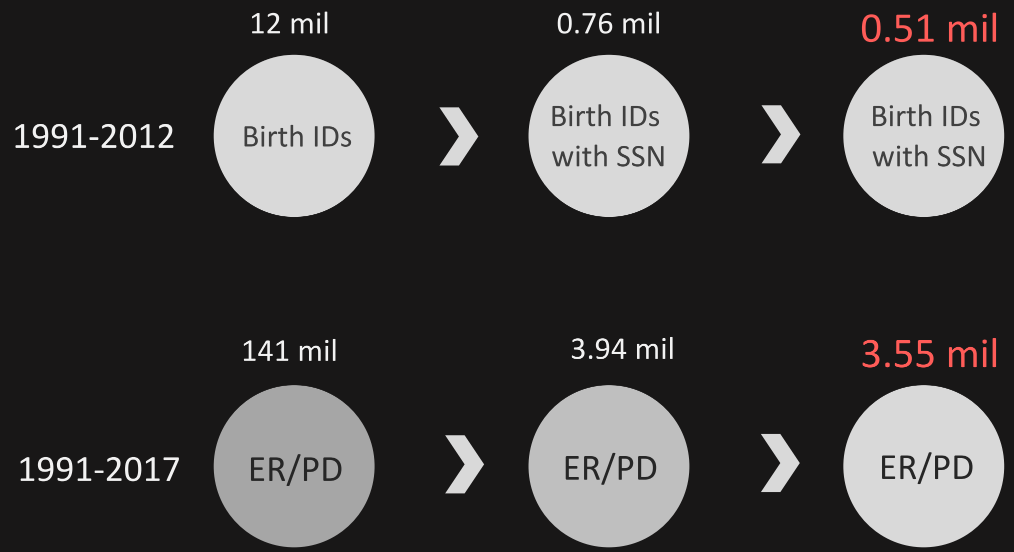
Ped-BERT Data

Restricted-access data



Ped-BERT Data

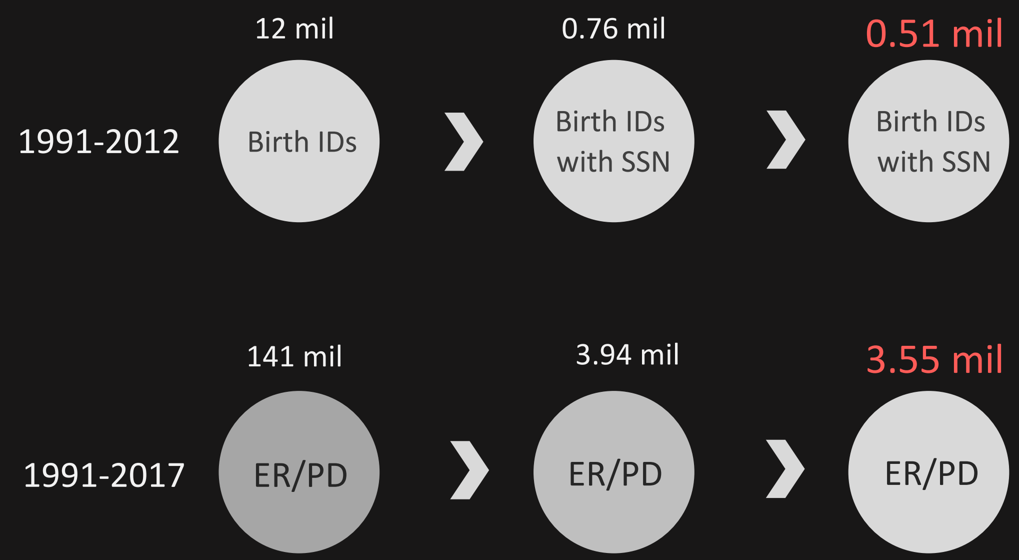
Restricted-access data



Ped-BERT Data

Restricted-access data

Task 2: fine-tune to specific task, diagnosis code prediction

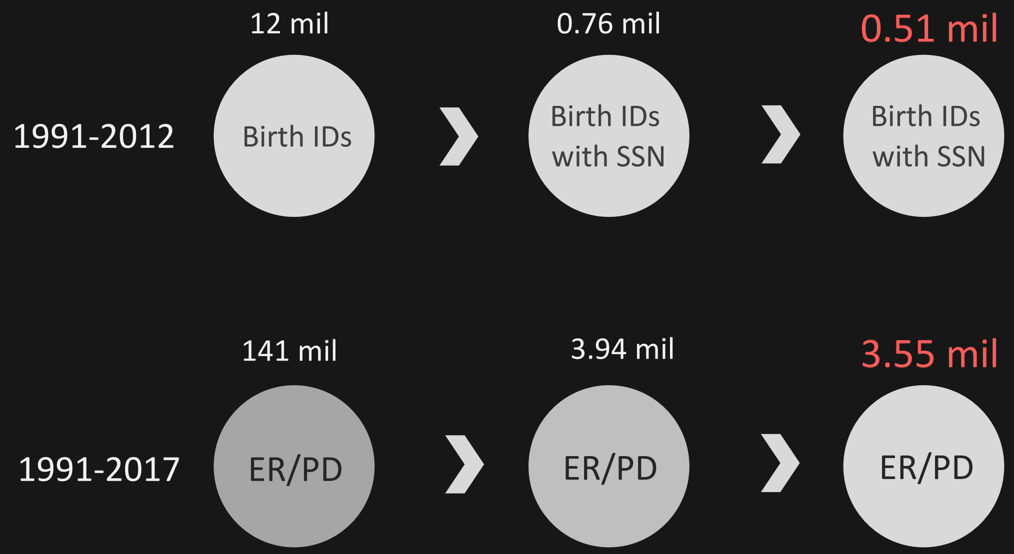


Ped-BERT Data

Restricted-access data

Task 1: pre-train the model on large self-labeled data

Task 2: fine-tune to specific task, diagnosis code prediction



Ped-BERT Data

Restricted-access data

Task 1: pre-train the model on large self-labeled data

Task 2: fine-tune to specific task, diagnosis code prediction

1991-2012: 12 mil Birth IDs ❯ 0.76 mil Birth IDs with SSN ❯ 0.51 mil Birth IDs with SSN

1991-2017: 141 mil ER/PD ❯ 3.94 mil ER/PD ❯ 3.55 mil ER/PD

drop all patients from the fine-tuning stage

drop all patients with < 3 ER/PD visits

Ped-BERT Data

Restricted-access data

Task 1: pre-train the model on large self-labeled data

20 mil

1991-2017 ER/PD

drop all patients from the fine-tuning stage

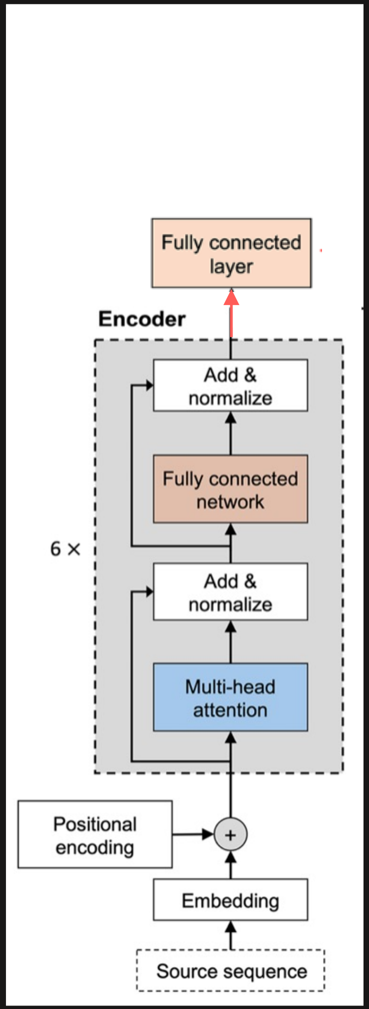
drop all patients with < 3 ER/PD visits

Task 2: fine-tune to specific task, diagnosis code prediction

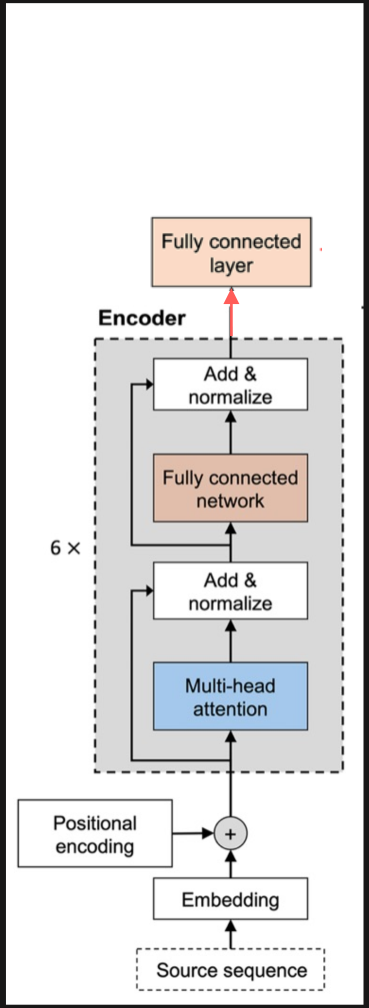
1991-2012: 12 mil Birth IDs ❯ 0.76 mil Birth IDs with SSN ❯ 0.51 mil Birth IDs with SSN

1991-2017: 141 mil ER/PD ❯ 3.94 mil ER/PD ❯ 3.55 mil ER/PD

Ped-BERT Architecture



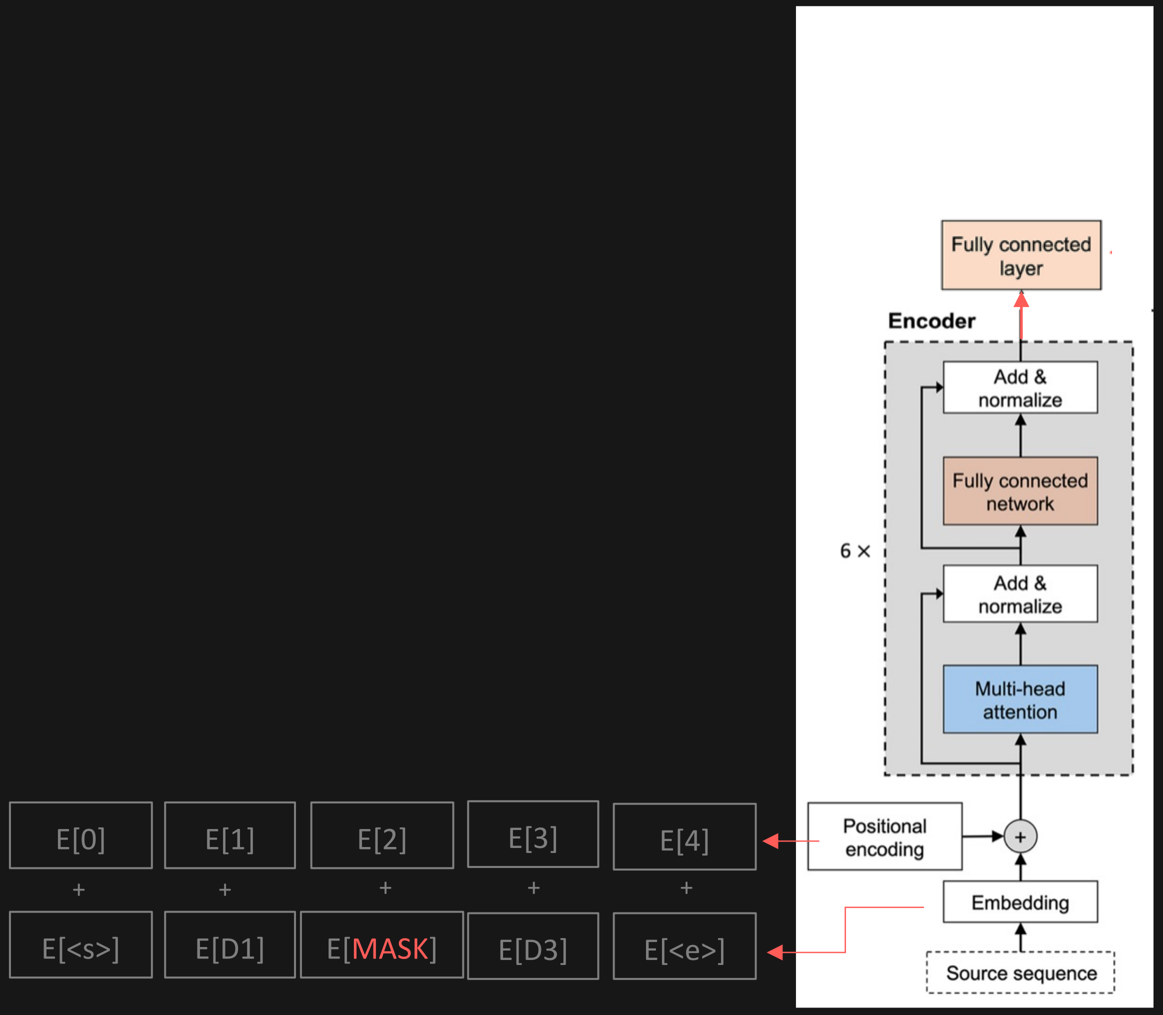
Ped-BERT Pre-training



[D1, MASK, D3]

visit 1 visit 2 visit 3

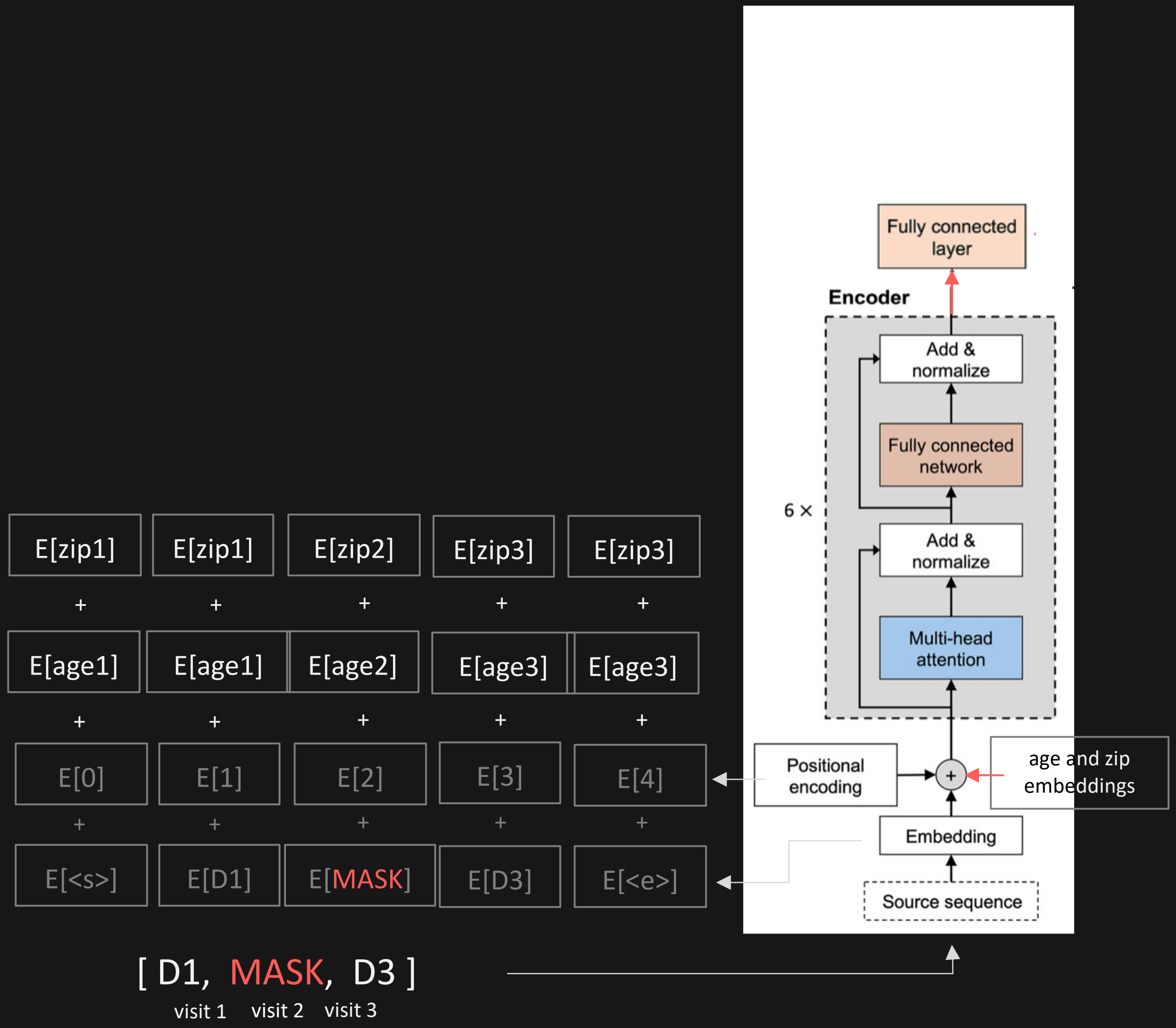
Ped-BERT Pre-training



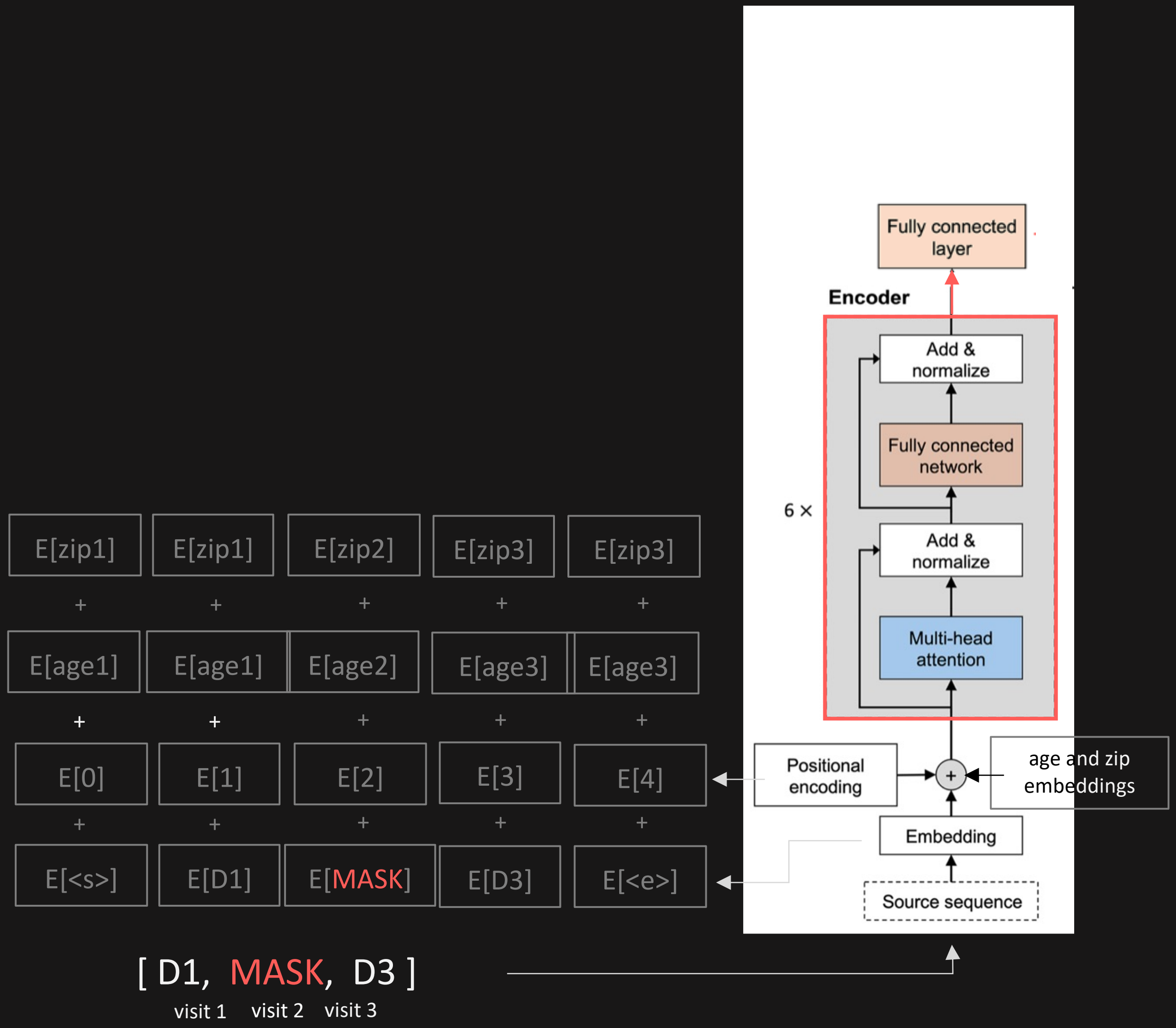
[D1, MASK, D3]

visit 1 visit 2 visit 3

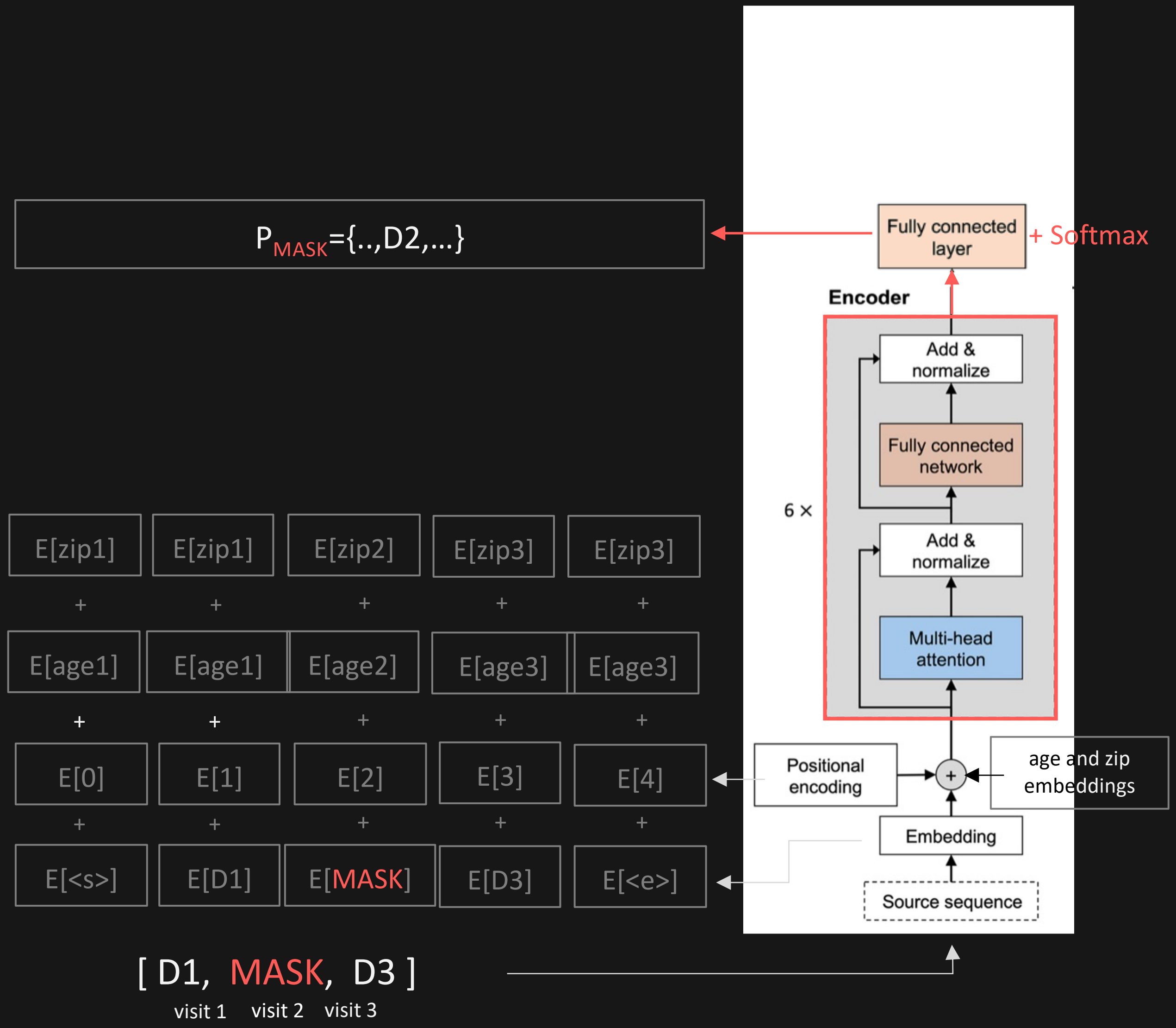
Ped-BERT Pre-training



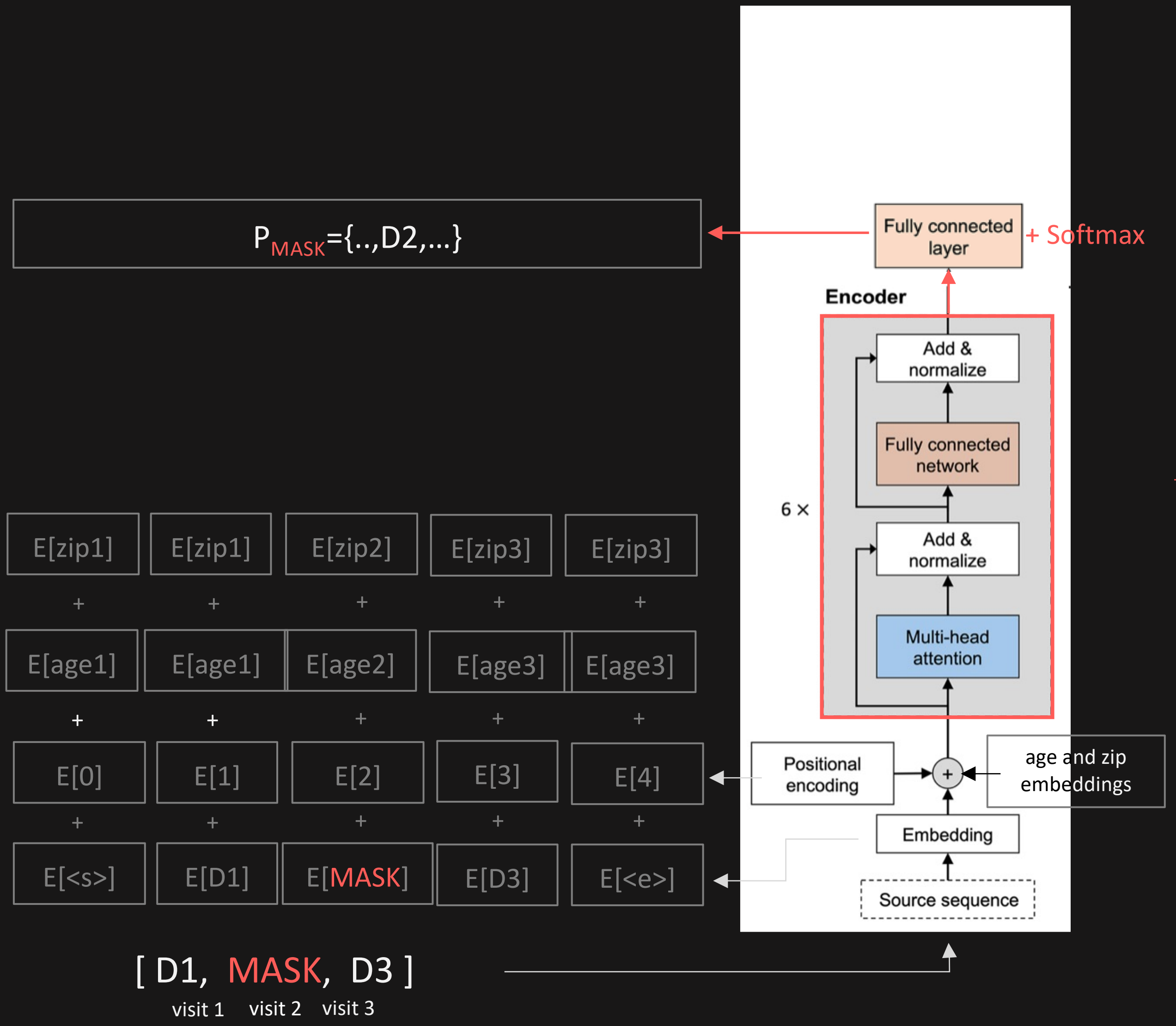
Ped-BERT Pre-training



Ped-BERT Pre-training

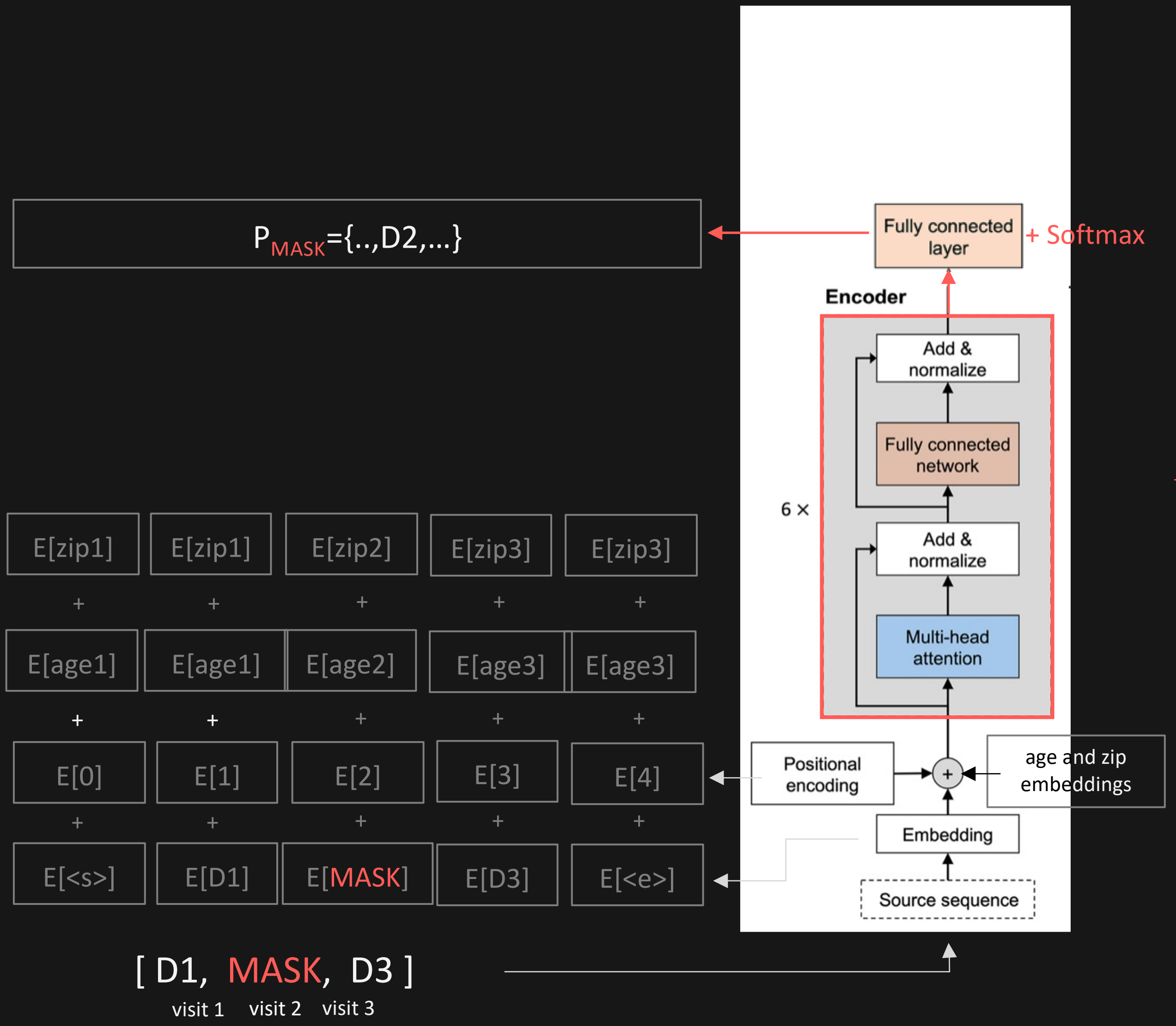


Ped-BERT Pre-training



Is it hard to code the Encoder block?

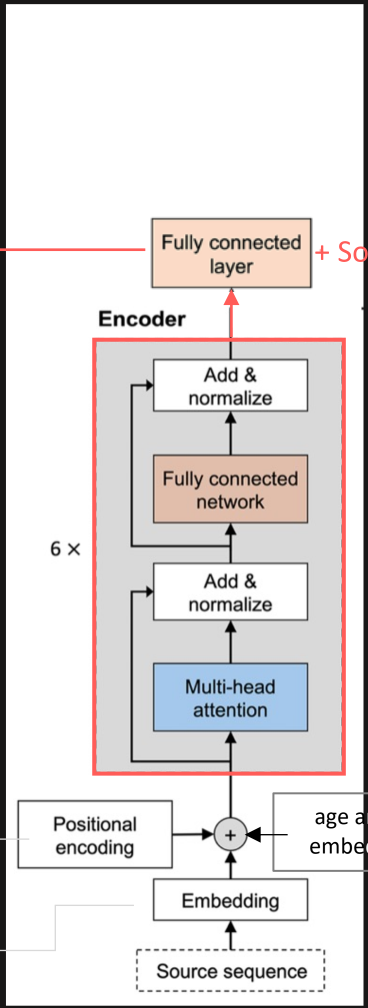
Ped-BERT Pre-training




```python
## ENCODER ##
#############
encoder_output = embeddings
for i in range(config.NUM_LAYERS): # only 1 layer

    ## Multi-Head self-attention ##
    ###############################
    query =  encoder_output
    key = encoder_output
    value = encoder_output

    # add multihead attention layer
    attention_output = layers.MultiHeadAttention(
        num_heads=config.NUM_HEAD,  # implement 8 attention layers
        key_dim=config.EMBED_DIM // config.NUM_HEAD,
        name="encoder_{}/att".format(i),
    )(query, key, value)

    # add dropout layer
    attention_output = layers.Dropout(
        0.1, name="encoder_{}/att_dropout".format(i))(
        attention_output
    )

    # add normalization layer
    attention_output = layers.LayerNormalization(
        epsilon=1e-6, name="encoder_{}/att_layernormalization".format(i)
    )(query + attention_output)

    ## Feed Forward layer ##
    ########################
    # fully connected layer
    ffn = keras.Sequential(
        [layers.Dense(config.FF_DIM, activation="relu"),
        layers.Dense(config.EMBED_DIM)],
        name="encoder_{}/ffn".format(i),
    )
    ffn_output = ffn(attention_output)
    # add dropout layer
    ffn_output = layers.Dropout(
        config.RATE, name="encoder_{}/ffn_dropout".format(i))(
        ffn_output
    )

    # add normalization layer
    sequence_output = layers.LayerNormalization(
        epsilon=1e-6, name="encoder_{}/ffn_layernormalization".format(i)
    )(attention_output + ffn_output)

    encoder_output = sequence_output
```

Ped-BERT Pre-training

P_{MASK}={..,D2,...}



+ Softmax

age and zip embeddings

```
# includes the multi-head attention, and fully connected layer
mlm_output = layers.Dense(
    len(vectorize_layer['diag'].get_vocabulary()),
    name="mlm_cls",
    activation="softmax")(encoder_output)
```

E[zip1]	E[zip1]	E[zip2]	E[zip3]	E[zip3]
+	+	+	+	+
E[age1]	E[age1]	E[age2]	E[age3]	E[age3]
+	+	+	+	+
E[0]	E[1]	E[2]	E[3]	E[4]
+	+	+	+	+
E[<s>]	E[D1]	E[MASK]	E[D3]	E[<e>]

[D1, MASK, D3]

visit 1 visit 2 visit 3

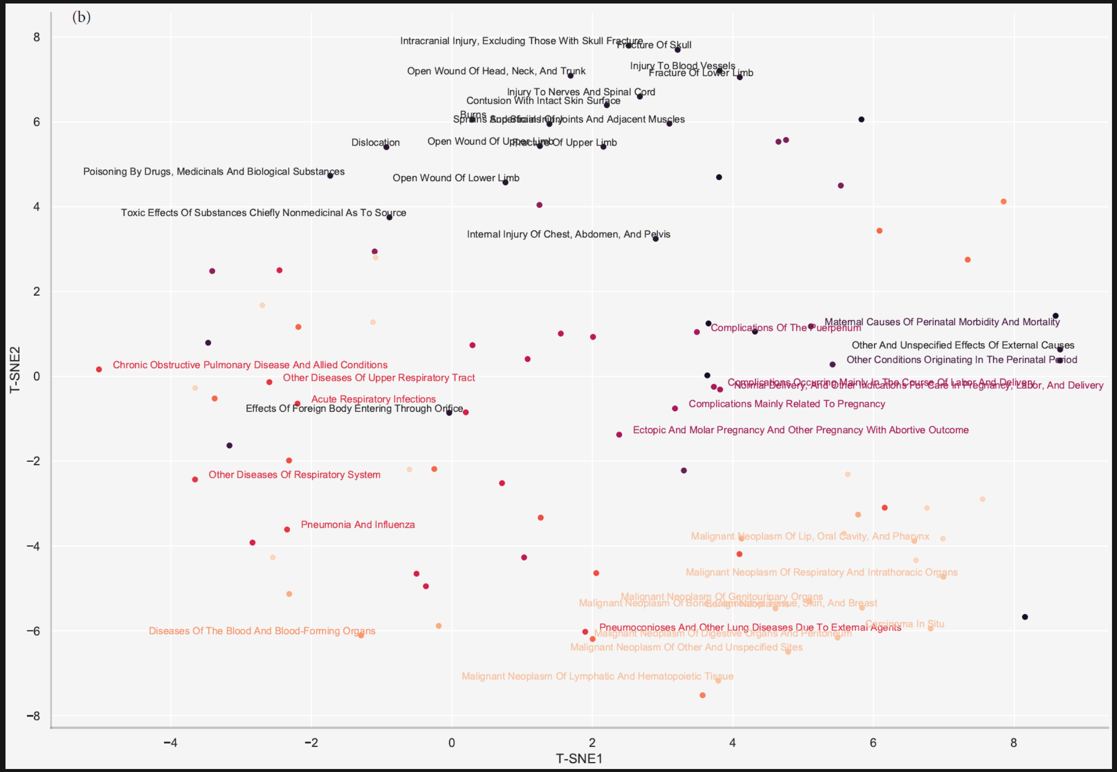
Ped-BERT Pre-training

Did we learn useful embeddings?

Let’s project embeddings in 2D space…

Ped-BERT Pre-training

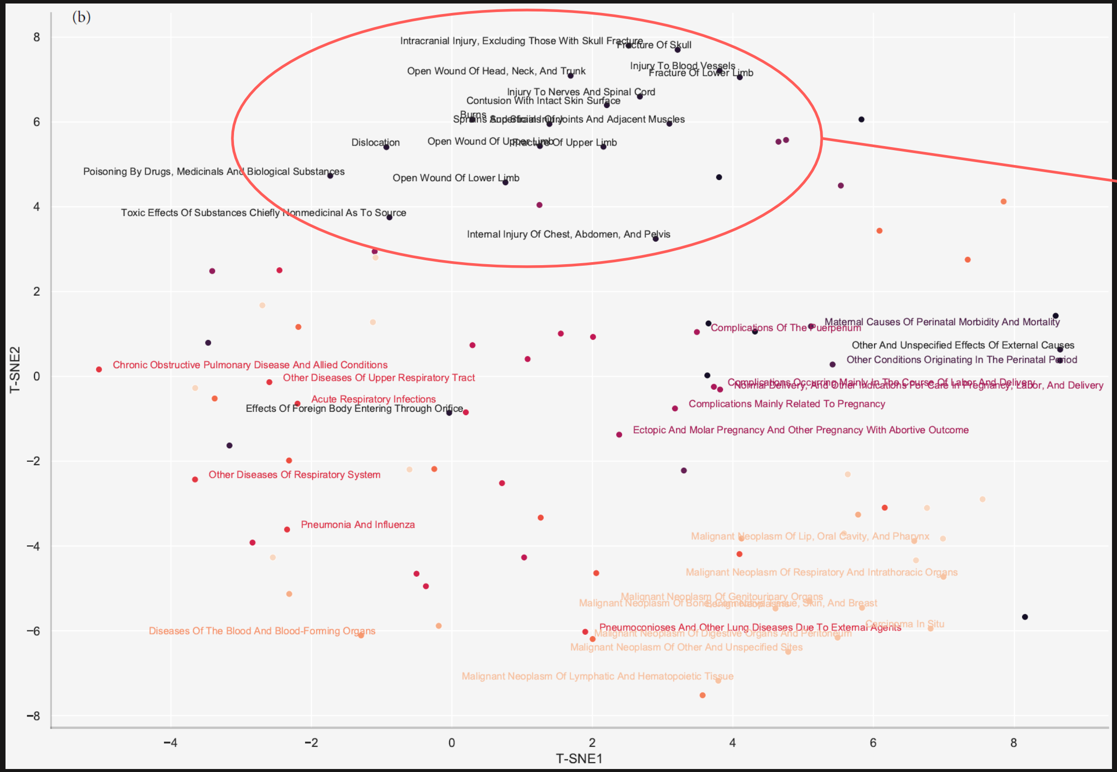
T-SNE2

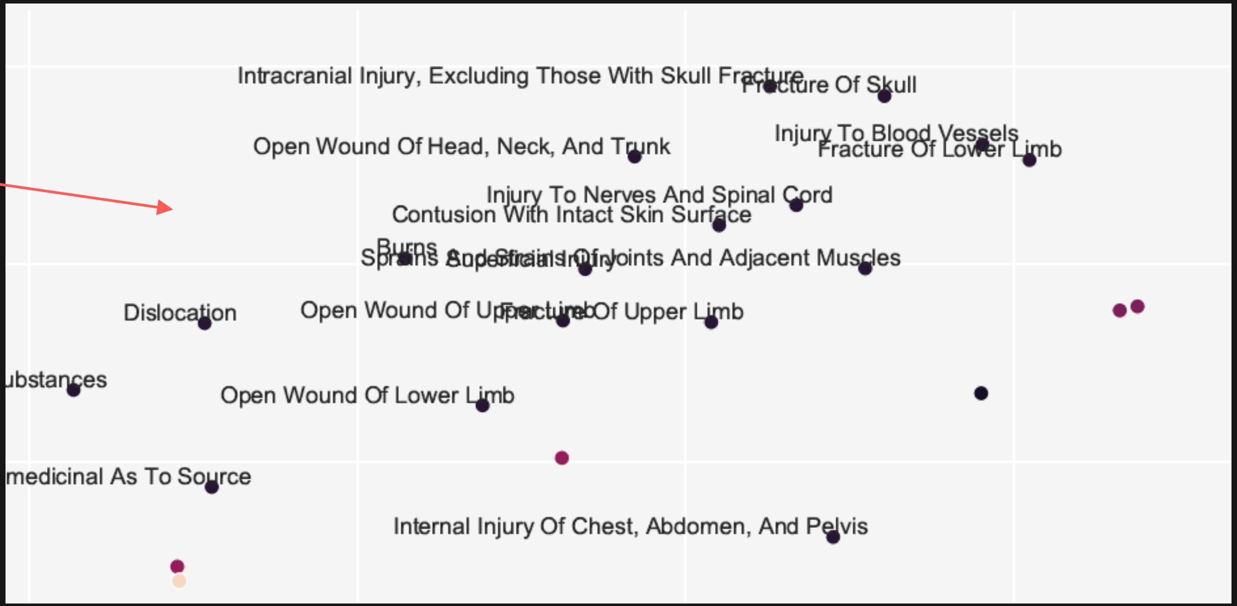


T-SNE1

Ped-BERT Pre-training

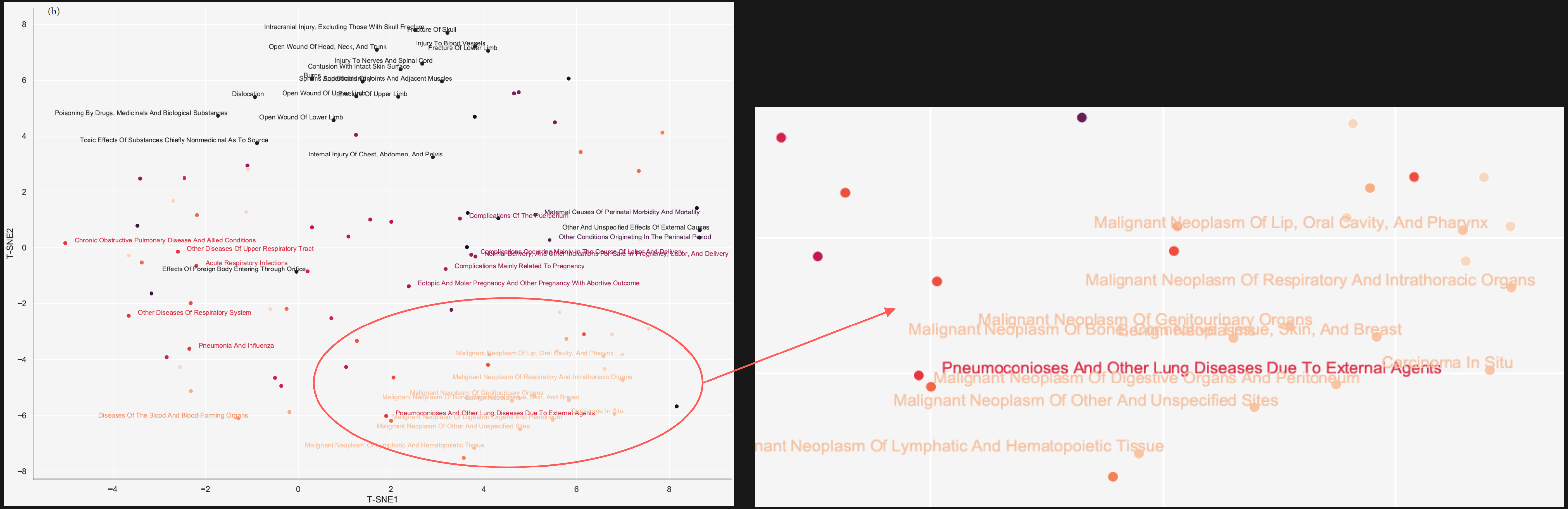
T-SNE2





T-SNE1

Ped-BERT Pre-training

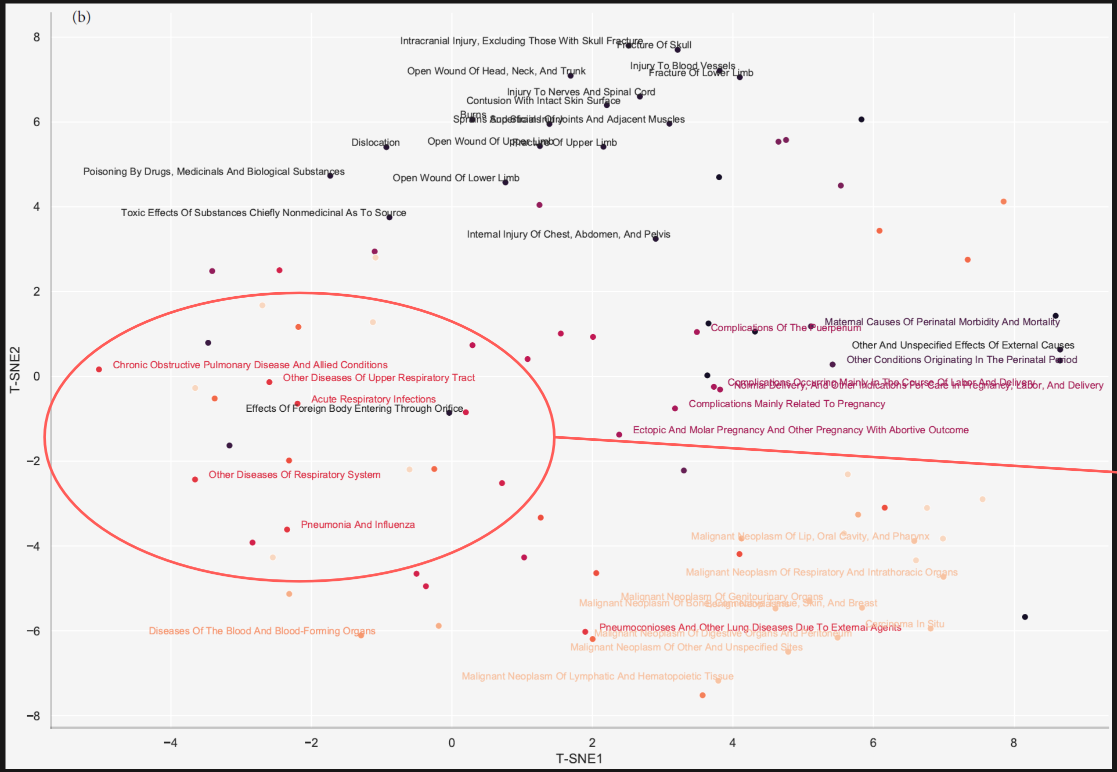


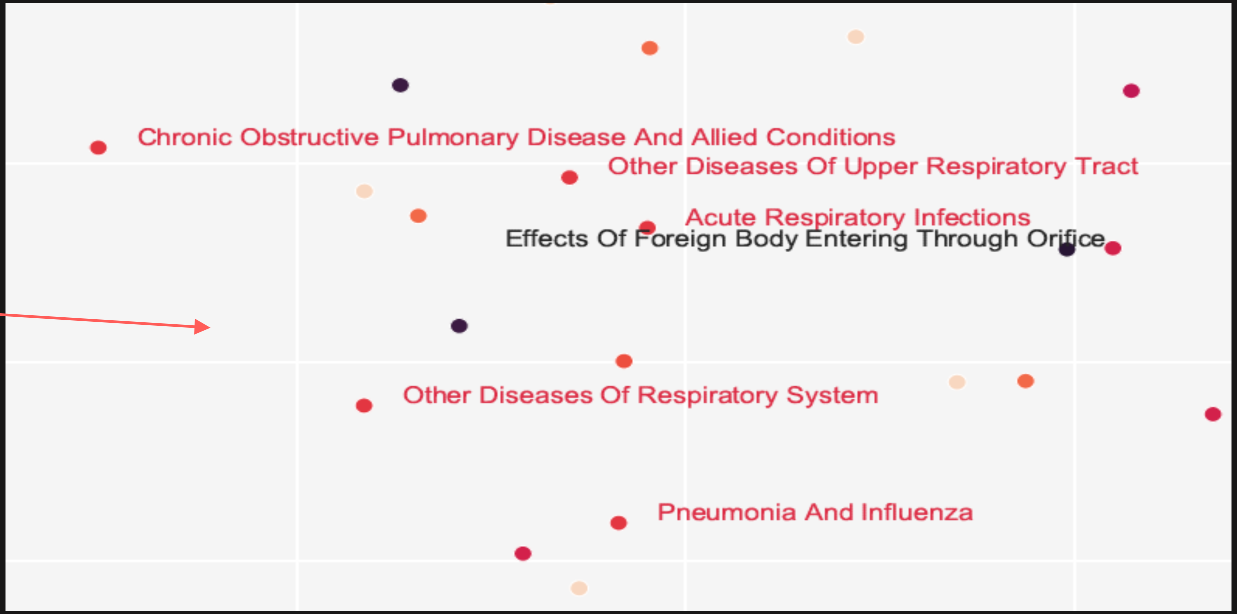
T-SNE2

T-SNE1

Ped-BERT Pre-training

T-SNE2

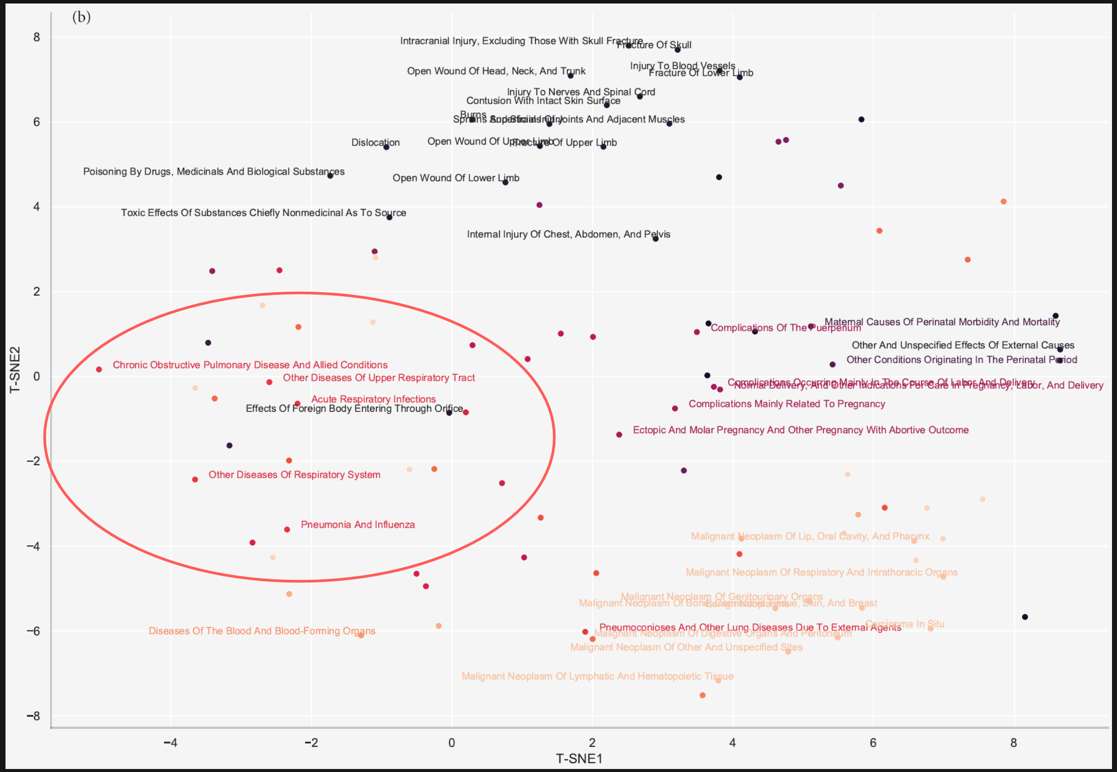




T-SNE1

Ped-BERT Pre-training

T-SNE2



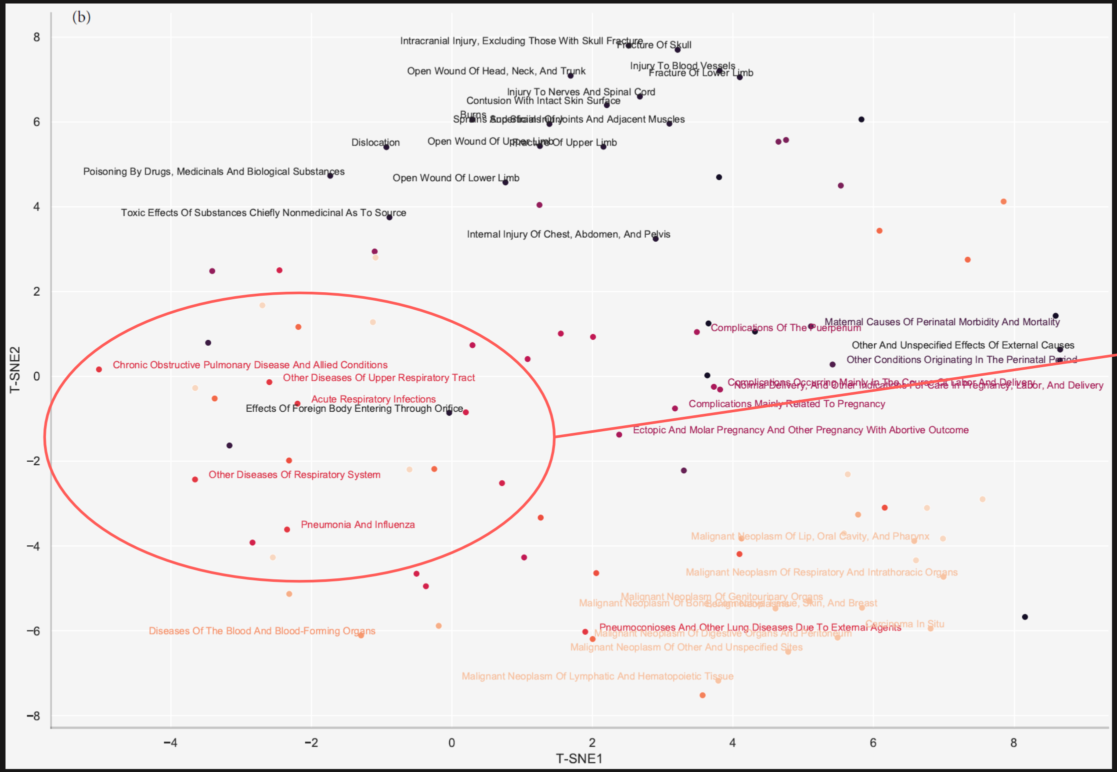
T-SNE1

International Classification of Diseases (ICD Codes)

001-139 Infectious and Parasitic Diseases
140-239 Neoplasms
240-279 Endocrine, Nutritional and Metabolic Diseases, And Immunity Disorders
280-289 Diseases Of The Blood And Blood-Forming Organs
290-319 Mental Disorders
320-389 Diseases Of the Nervous System and Sense Organs
390-459 Diseases Of The Circulatory System
460-519 Diseases of the Respiratory System
520-579 Diseases Of The Digestive System
580-629 Diseases Of The Genitourinary System
630-679 Complications Of Pregnancy, Childbirth, And The Puerperium
680-709 Diseases Of The Skin and Subcutaneous Tissue
710-739 Diseases Of The Musculoskeletal System And Connective Tissue
740-759 Congenital Anomalies
760-779 Certain Conditions Originating In The Perinatal Period
780-799 Symptoms, Signs, And Ill-Defined Conditions
800-999 Injury And Poisoning

Ped-BERT Pre-training

T-SNE2



T-SNE1

International Classification of Diseases (ICD Codes)

001-139	Infectious and Parasitic Diseases
140-239	Neoplasms
240-279	Endocrine, Nutritional and Metabolic Diseases, And Immunity Disorders
280-289	Diseases Of The Blood And Blood-Forming Organs
290-319	Mental Disorders
320-389	Diseases Of the Nervous System and Sense Organs
390-459	Diseases Of The Circulatory System
460-519	Diseases of the Respiratory System
520-579	Diseases Of The Digestive System
580-629	Diseases Of The Genitourinary System
630-679	Complications Of Pregnancy, Childbirth, And The Puerperium
680-709	Diseases Of The Skin and Subcutaneous Tissue
710-739	Diseases Of The Musculoskeletal System And Connective Tissue
740-759	Congenital Anomalies
760-779	Certain Conditions Originating In The Perinatal Period
780-799	Symptoms, Signs, And Ill-Defined Conditions
800-999	Injury And Poisoning

Ped-BERT Pre-training

T-SNE2

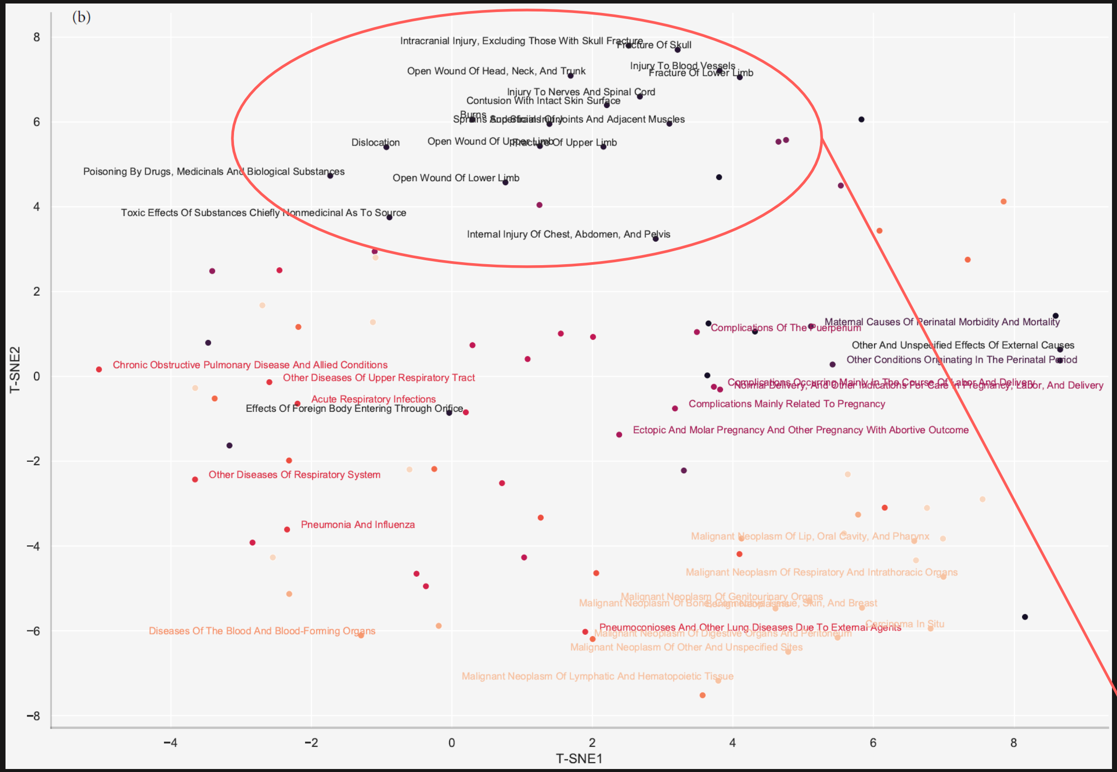
T-SNE1

International Classification of Diseases (ICD Codes)

001-139 Infectious and Parasitic Diseases
140-239 Neoplasms
240-279 Endocrine, Nutritional and Metabolic Diseases, And Immunity Disorders
280-289 Diseases Of The Blood And Blood-Forming Organs
290-319 Mental Disorders
320-389 Diseases Of the Nervous System and Sense Organs
390-459 Diseases Of The Circulatory System
460-519 Diseases of the Respiratory System
520-579 Diseases Of The Digestive System
580-629 Diseases Of The Genitourinary System
630-679 Complications Of Pregnancy, Childbirth, And The Puerperium
680-709 Diseases Of The Skin and Subcutaneous Tissue
710-739 Diseases Of The Musculoskeletal System And Connective Tissue
740-759 Congenital Anomalies
760-779 Certain Conditions Originating In The Perinatal Period
780-799 Symptoms, Signs, And Ill-Defined Conditions
800-999 Injury And Poisoning

Ped-BERT Pre-training

T-SNE2

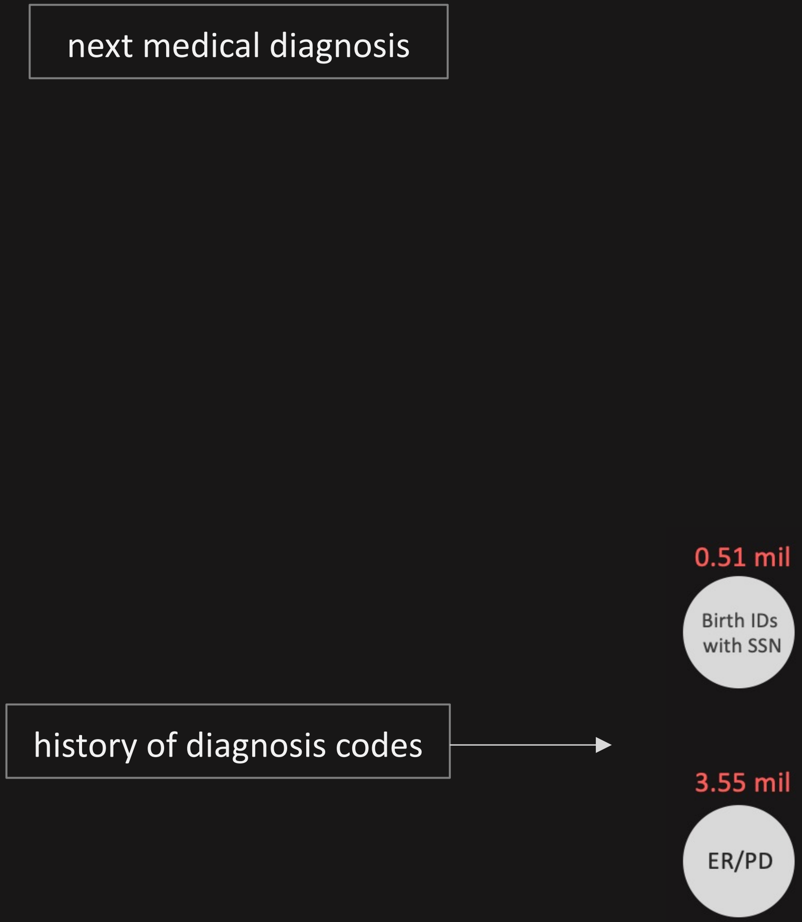


T-SNE1

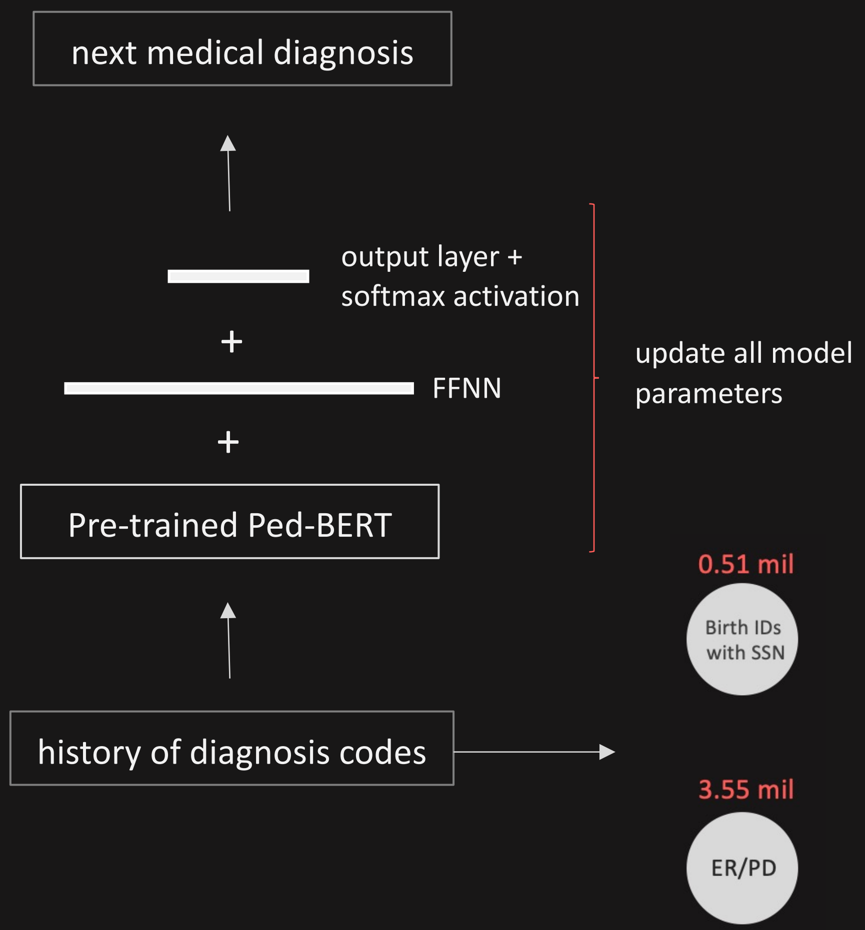
International Classification of Diseases (ICD Codes)

001-139	Infectious and Parasitic Diseases
140-239	Neoplasms
240-279	Endocrine, Nutritional and Metabolic Diseases, And Immunity Disorders
280-289	Diseases Of The Blood And Blood-Forming Organs
290-319	Mental Disorders
320-389	Diseases Of the Nervous System and Sense Organs
390-459	Diseases Of The Circulatory System
460-519	Diseases of the Respiratory System
520-579	Diseases Of The Digestive System
580-629	Diseases Of The Genitourinary System
630-679	Complications Of Pregnancy, Childbirth, And The Puerperium
680-709	Diseases Of The Skin and Subcutaneous Tissue
710-739	Diseases Of The Musculoskeletal System And Connective Tissue
740-759	Congenital Anomalies
760-779	Certain Conditions Originating In The Perinatal Period
780-799	Symptoms, Signs, And Ill-Defined Conditions
800-999	Injury And Poisoning

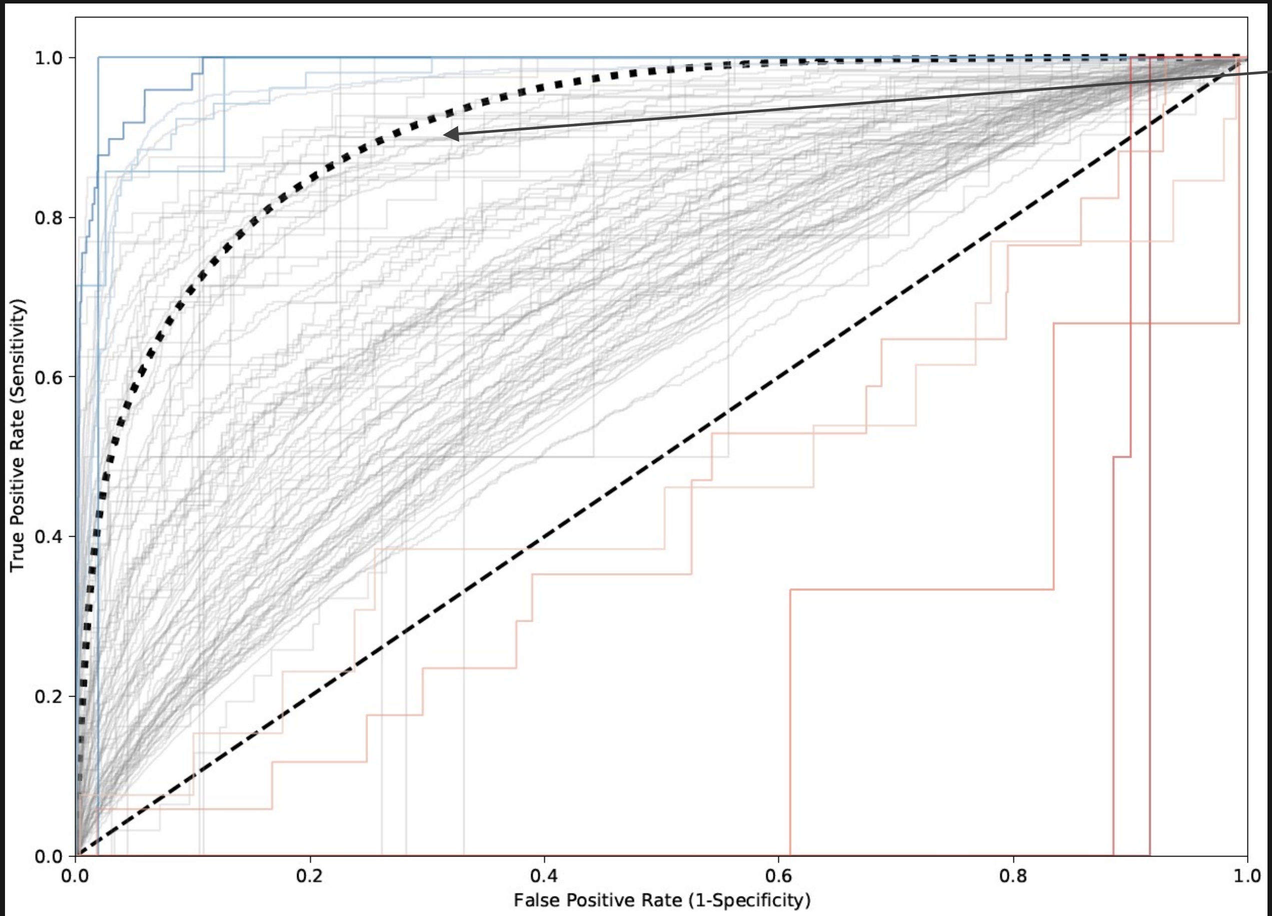
Ped-BERT Fine-tuning



Ped-BERT Fine-tuning

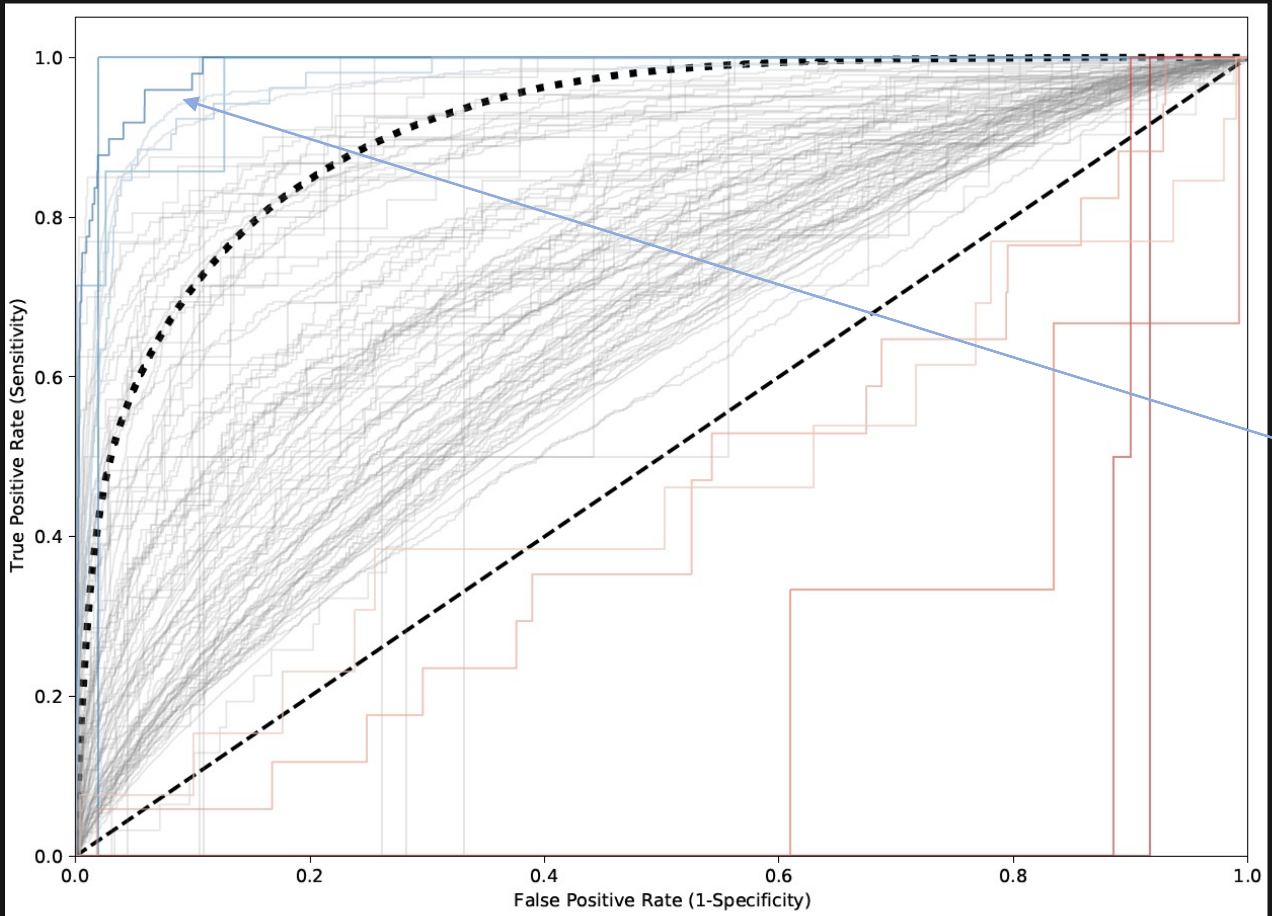


Ped-BERT Results



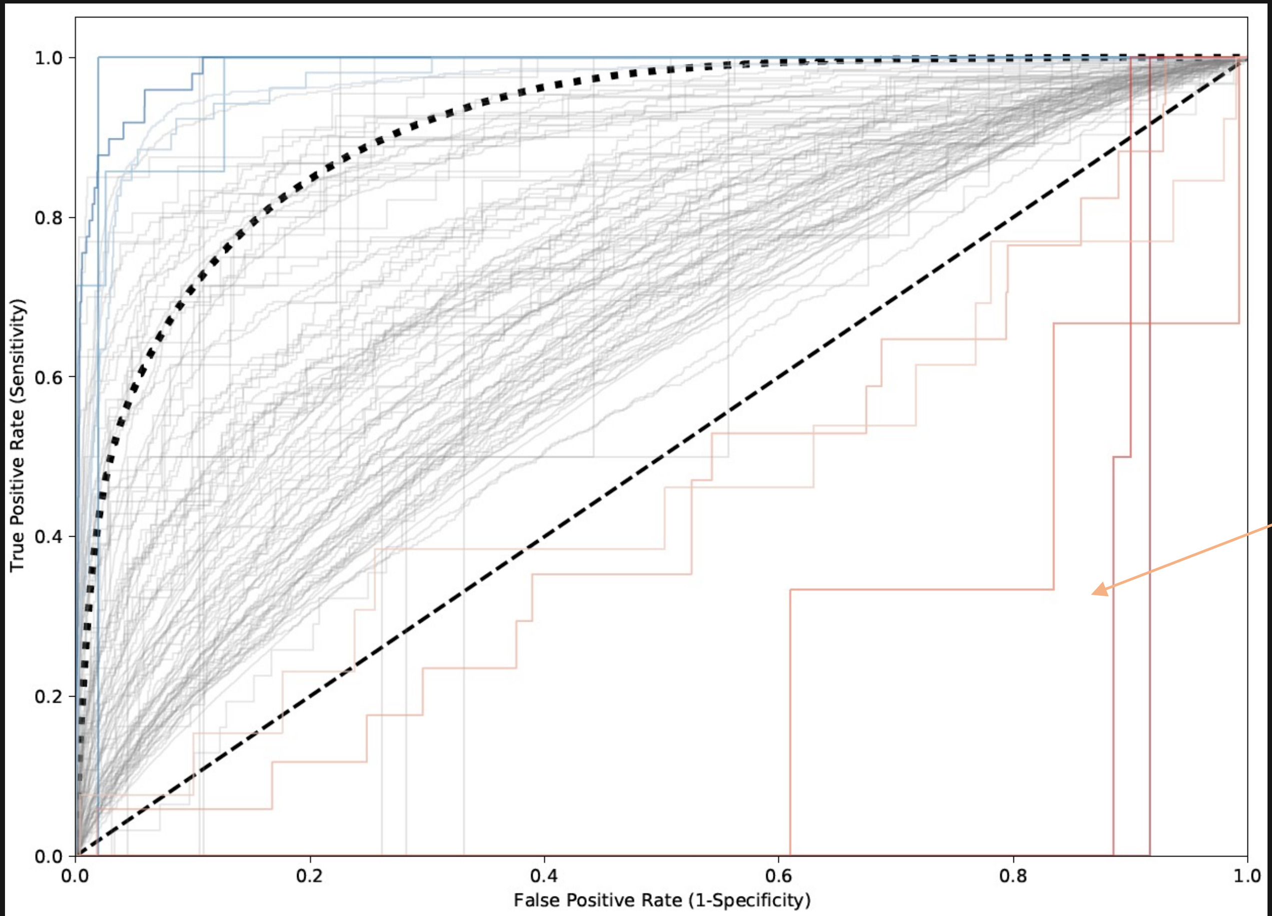
Sample-average ROC AUC curve: 0.911

Ped-BERT Results



Top 4 diagnosis predictions by ROC AUC
Maternal Causes of Perinatal Morbidity And Mortality (AUC = 0.988)
Disorders Of Thyroid Gland (AUC = 0.980)
Malignant Neoplasm of Genitourinary Organs (AUC = 0.978)
Complications of the Puerperium (AUC = 0.965)

Ped-BERT Results

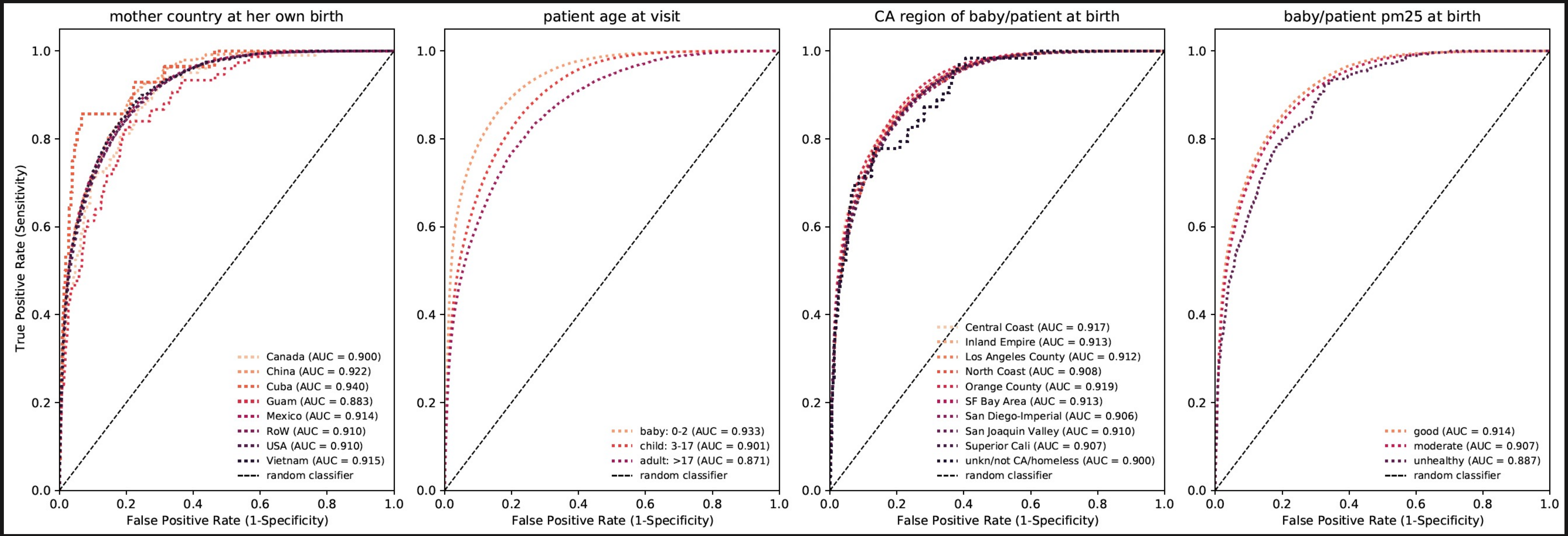


Least 4 diagnosis predictions by ROC AUC
Persons With Potential Health Hazards Related to Personal And Family History (AUC = 0.433)
Zoonotic Bacterial Diseases (AUC = 0.188)
Malignant Neoplasm of Respiratory And Intrathoracic Organs (AUC = 0.107)
Subclinical Iodine-Deficiency Hyperthyroidism (AUC = 0.084)

QUIZ

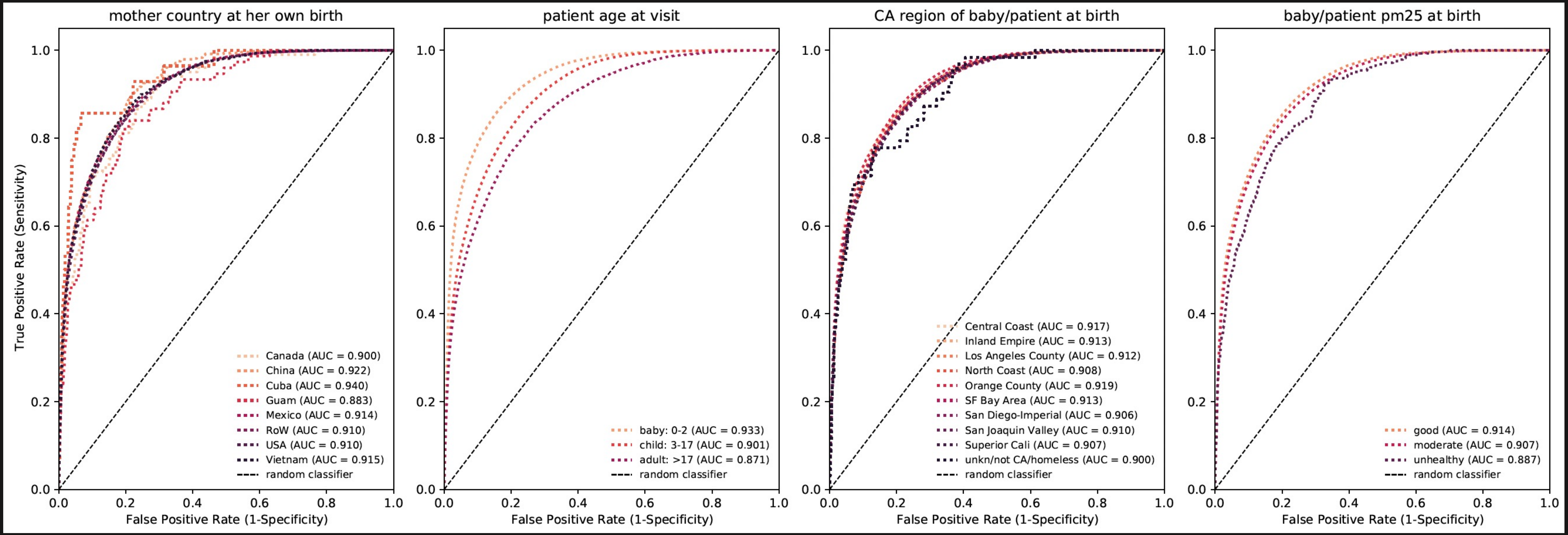
Q 5: What other model performance questions are important to look at?

Ped-BERT Fairness



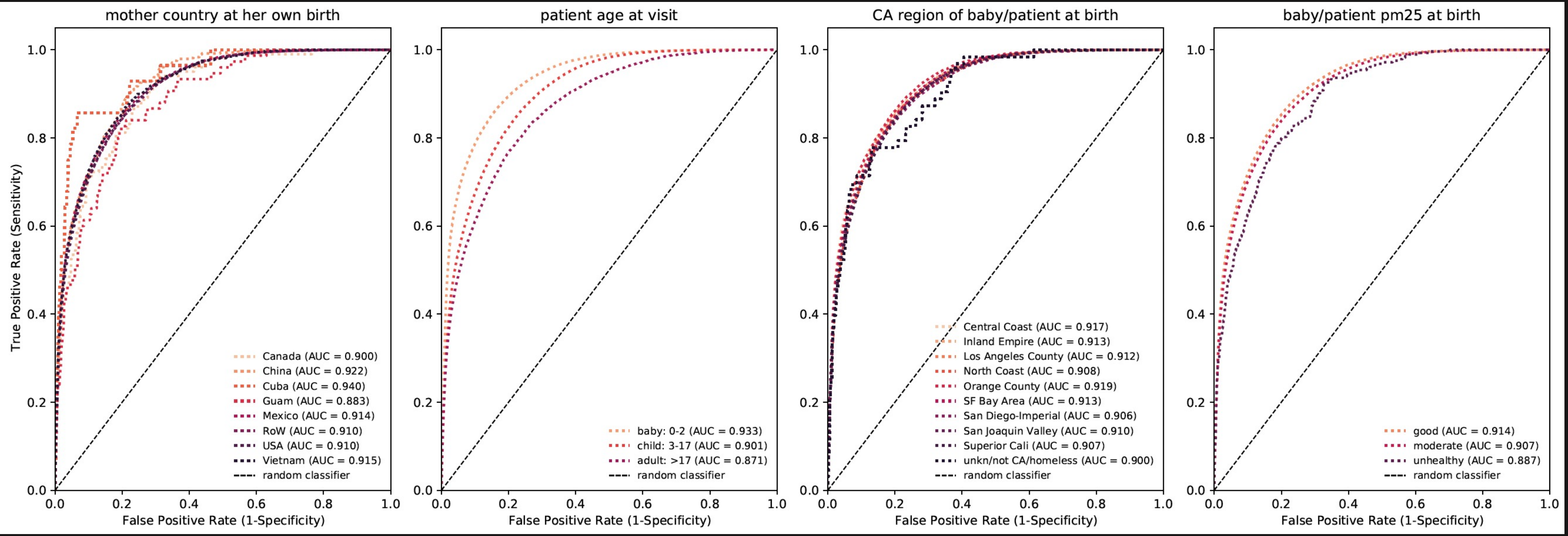
Less stable ROC AUC results across mother place of birth

Ped-BERT Fairness



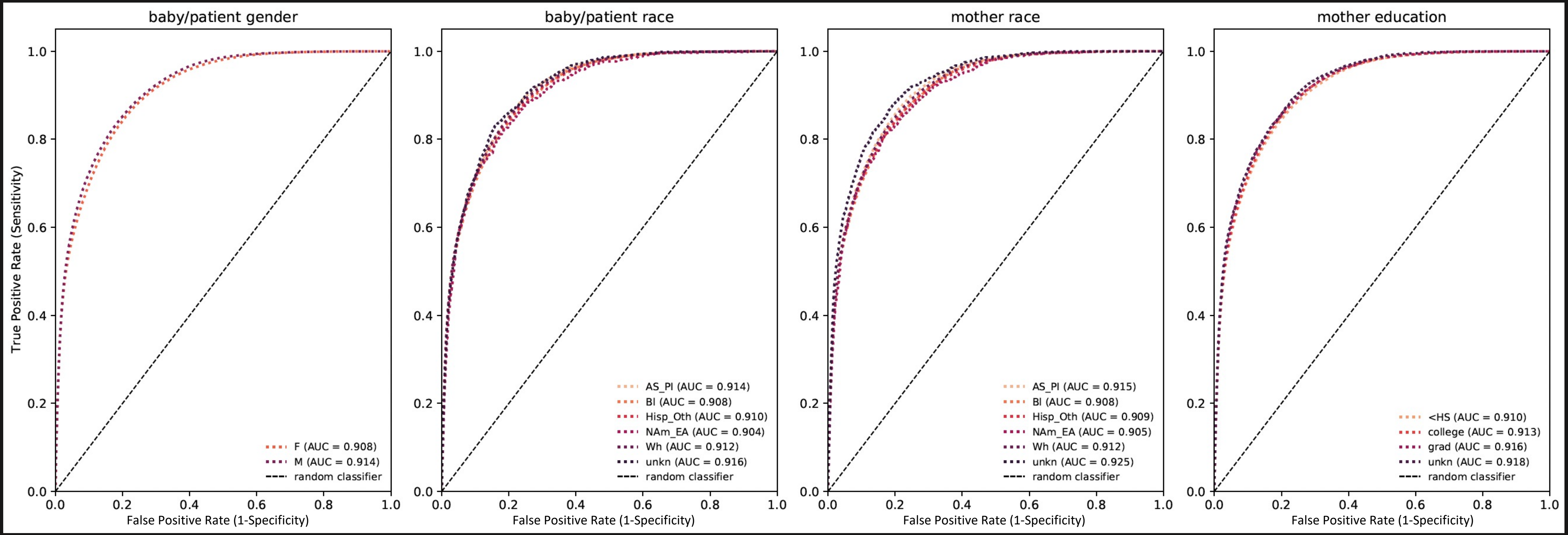
Less stable ROC AUC results for older patients

Ped-BERT Fairness



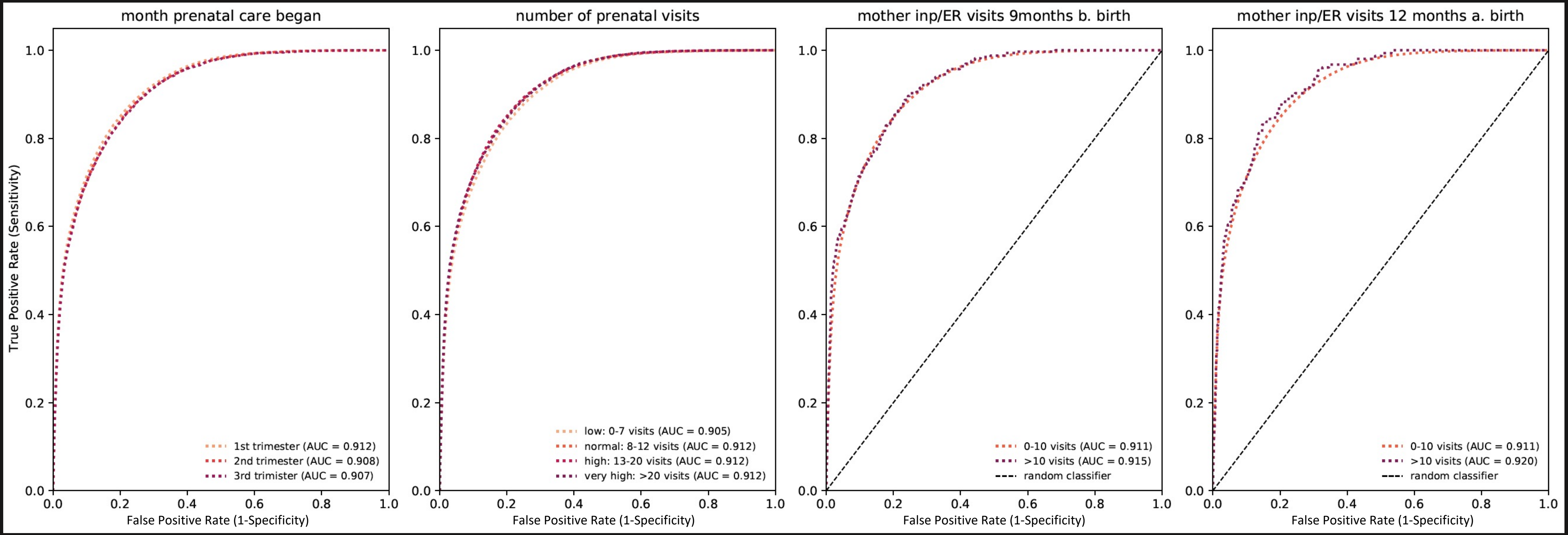
Less stable ROC AUC results for patients with higher level of zip-code pollution exposure

Ped-BERT Fairness



Stable ROC AUC results across these dimensions

Ped-BERT Fairness



Stable ROC AUC results across these dimensions

QUIZ

Q 6: Given the data and the results presented, can you think of other areas of exploration?

Ped-BERT Work in Progress

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

Ped-BERT Work in Progress

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

after 12 (!) hospital visits…

CANDLE syndrome, a rare, usually genetic, autoinflammatory condition

Ped-BERT Work in Progress

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

after 12 (!) hospital visits…

80% of rare diseases, usually genetic, onset during the first years of life (pediatric patients)

CANDLE syndrome, a rare, usually genetic, autoinflammatory condition

Ped-BERT Work in Progress

- recurrent fevers
- progressive fat loss
- swollen eyelids
- first seizure

Can we use Ped-BERT to detect rare diseases earlier?

after 12 (!) hospital visits…

80% of rare diseases, usually genetic, onset during the first years of life (pediatric patients)

CANDLE syndrome, a rare, usually genetic, autoinflammatory condition

Conclusions

- Transformers: history and architecture
- BERT (Transformer-encoder): architecture
- Application: predict diagnosis outcomes for pediatric patients using Ped-BERT

THANKS