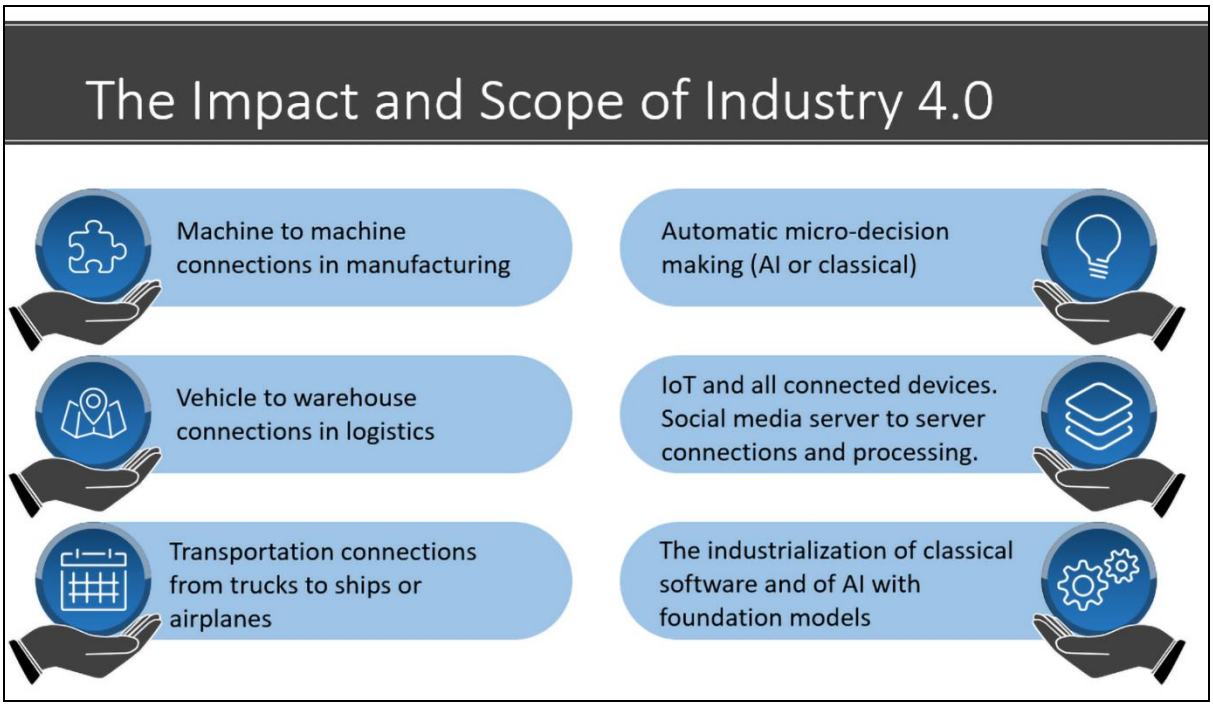
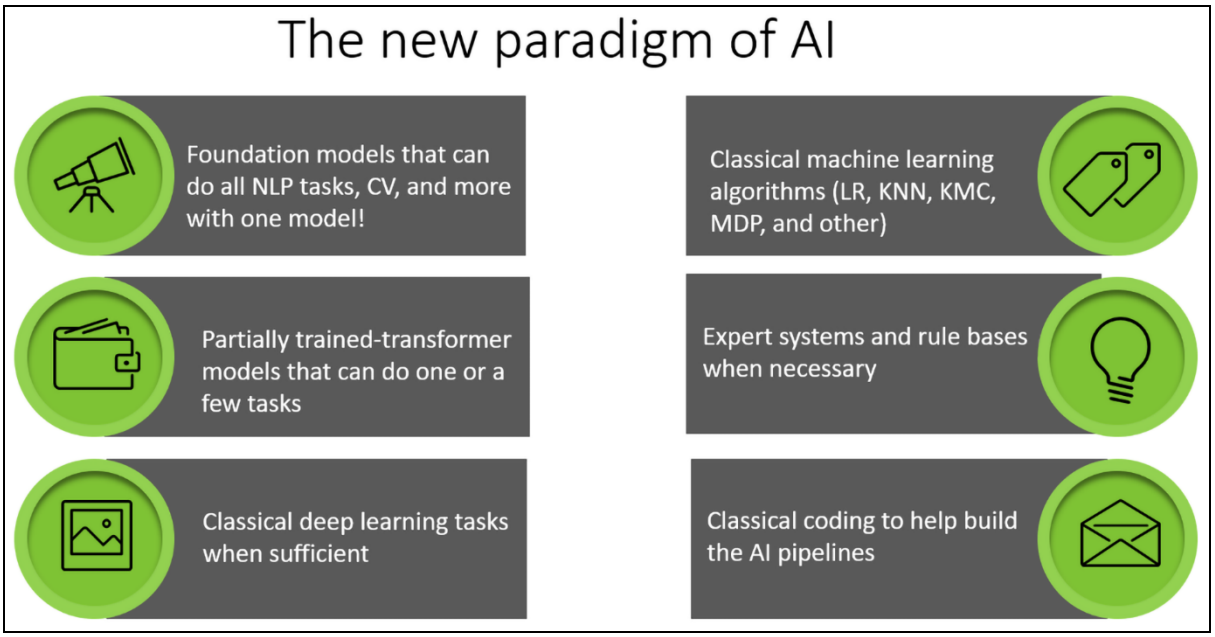
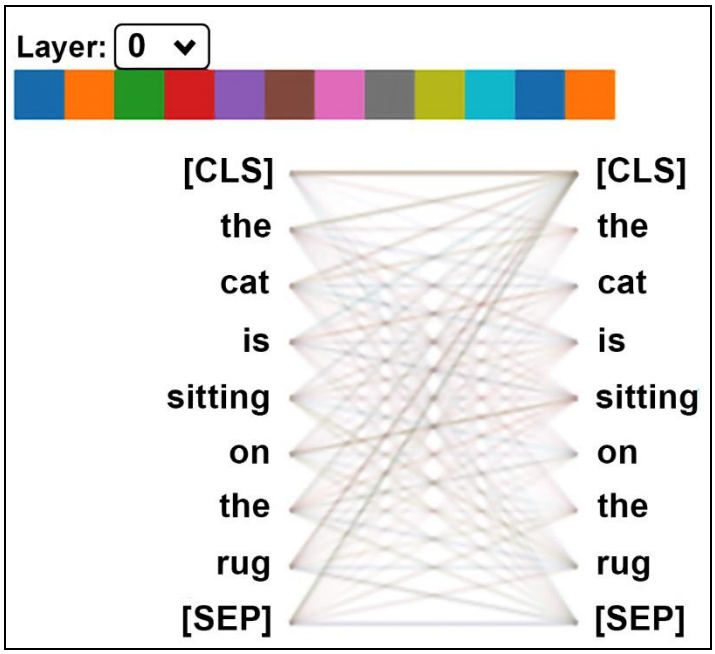
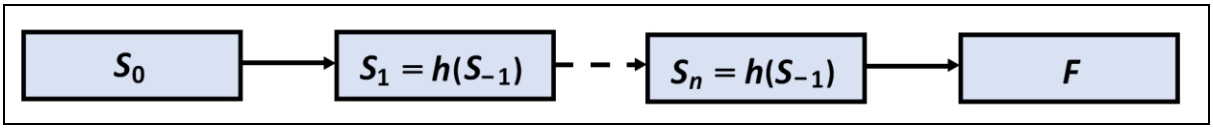
Chapter 1: What Are Transformers?




The new paradigm of AI
Foundation models that can do all NLP tasks, CV, and more with one model!
Classical machine learning algorithms (LR, KNN, KMC, MDP, and other)
Partially trained-transformer models that can do one or a few tasks
Expert systems and rule bases when necessary
Classical deep learning tasks when sufficient
Classical coding to help build the AI pipelines


Layer: 0
[CLS]
the
cat
is
sitting
on
the
rug
[SEP]
[CLS]
the
cat
is
sitting
on
the
rug
[SEP]



Document

A user visited the AllenNLP website, tried a transformer model, and found it interesting.

Run Model

Model Output

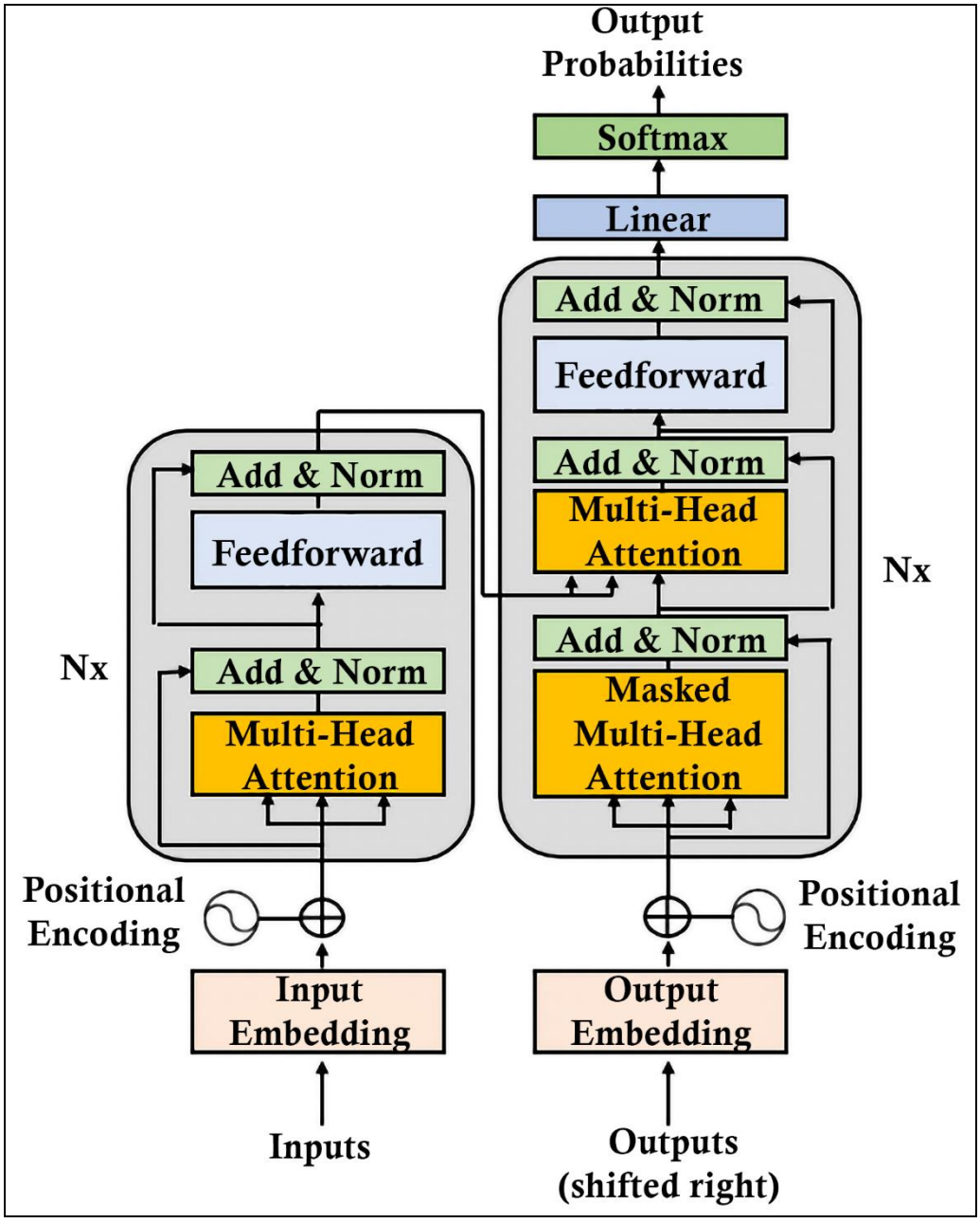
A user visited the AllenNLP website, tried 0 a transformer model , and found 0 it interesting.

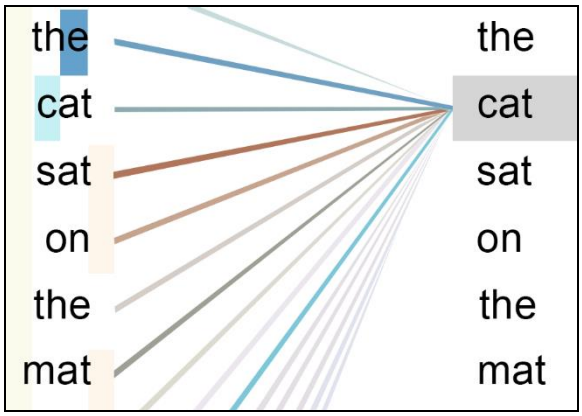
English ⇄ French

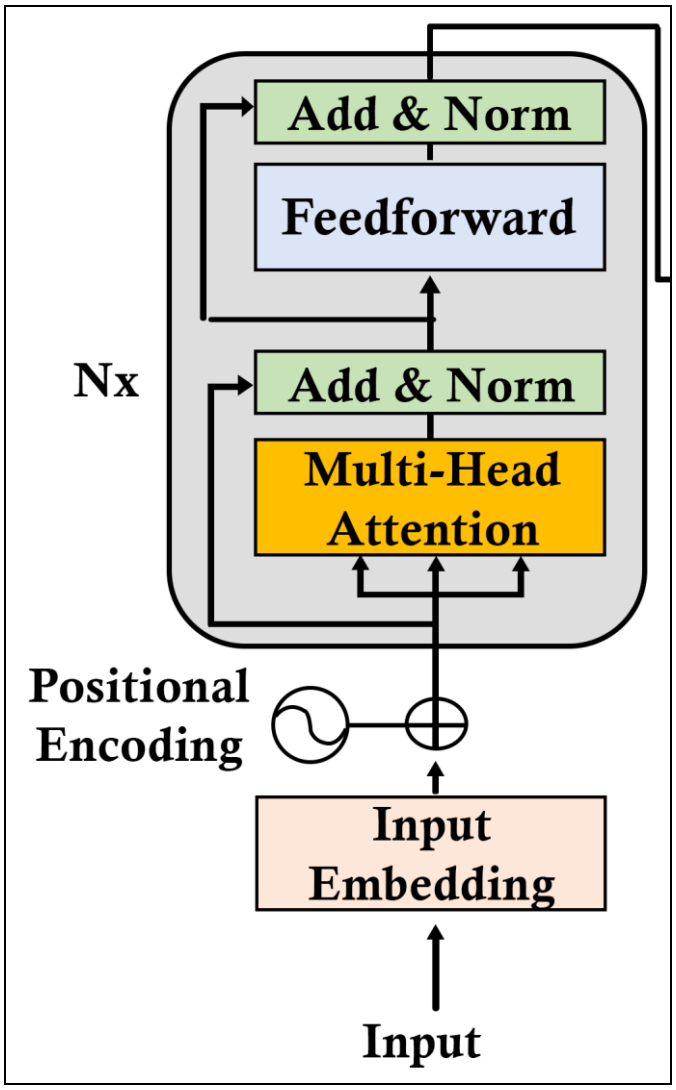
A user visited the AllenNLP website, tried a transformer model, and found it interesting.

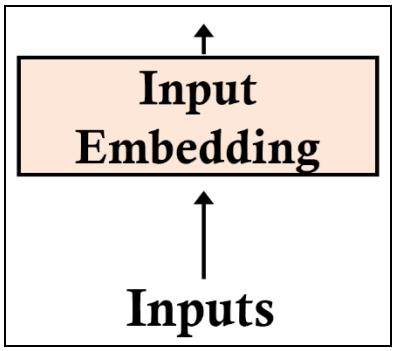
Un utilisateur a visité le site Web AllenNLP, a essayé un modele de transformateur et l'a trouvé intéressant.

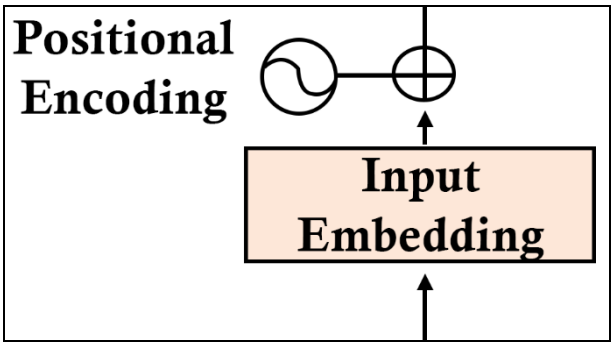
Chapter 2: Getting Started with the Architecture of the Transformer Model

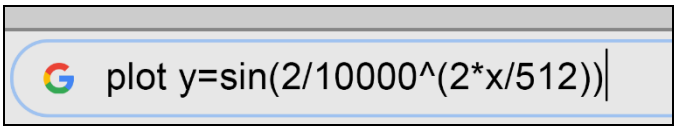


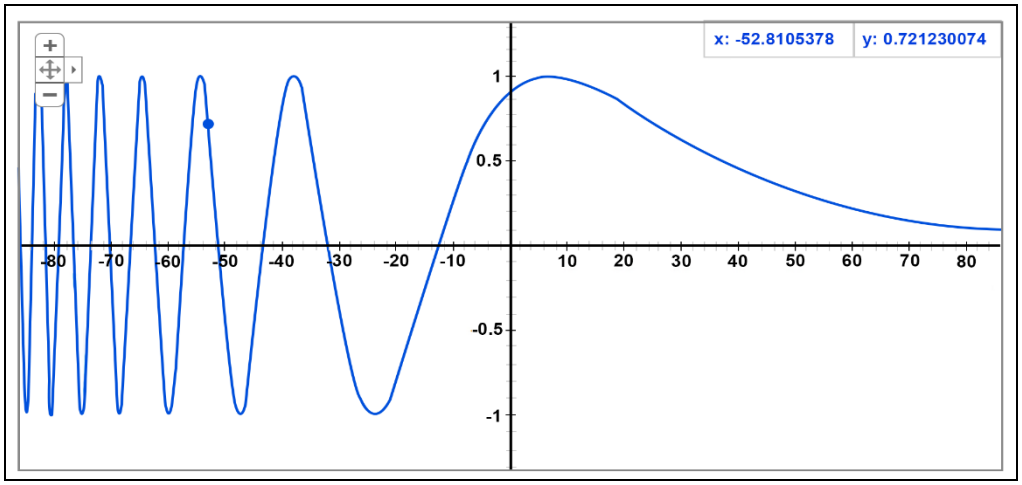

the
cat
sat
on
the
mat
the
cat
sat
on
the
mat

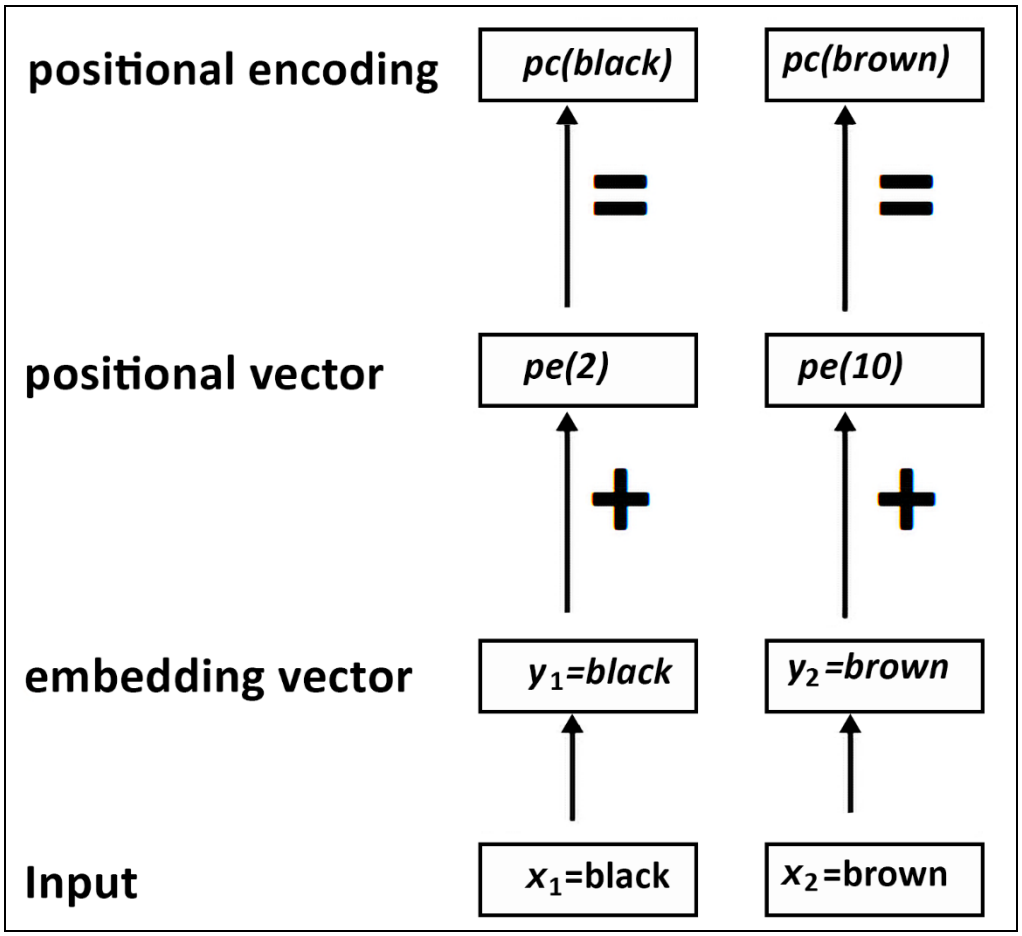

Add & Norm
Feedforward
Nx
Add & Norm
Multi-Head Attention
Positional Encoding
Input Embedding
Input

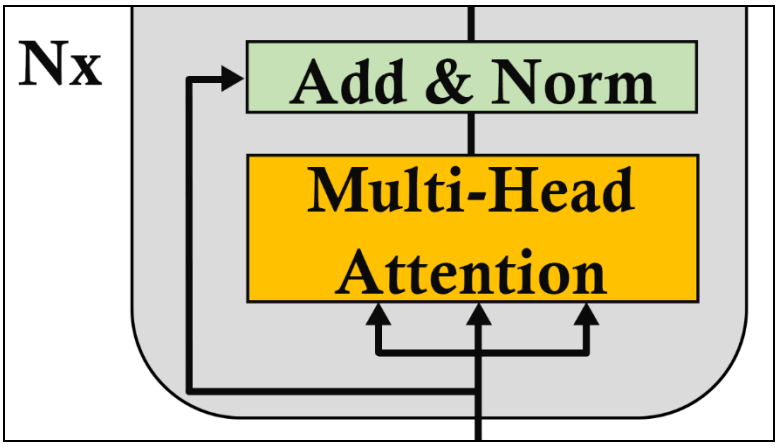

Input
Embedding
Inputs

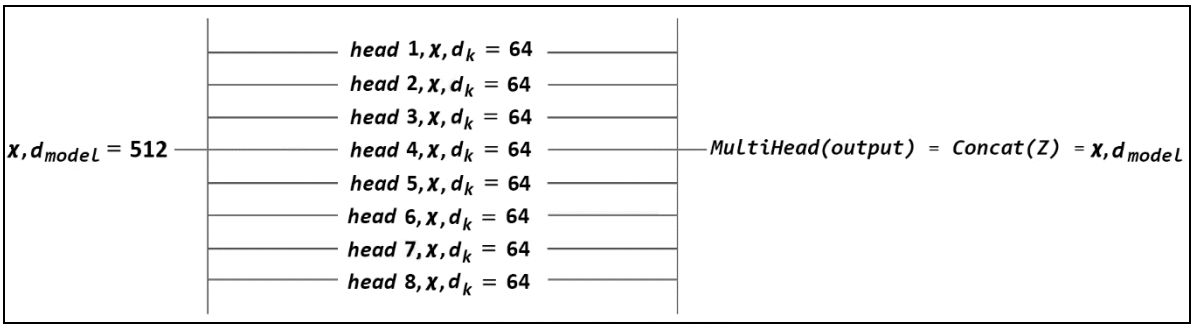

Positional
Encoding
Input
Embedding

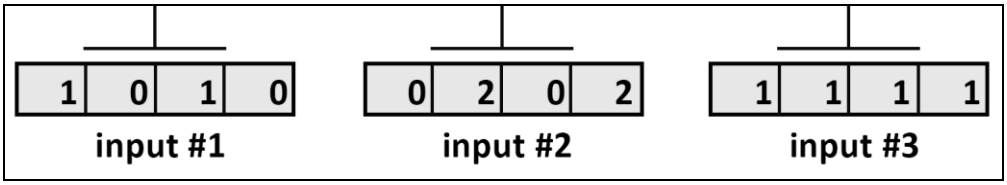

plot y=sin(2/10000^(2*x/512))

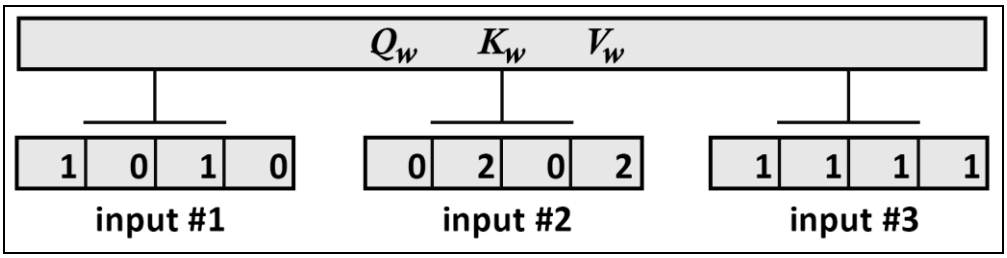

x: -52.8105378
y: 0.721230074
1
0.5
-0.5
-1
-80
-70
-60
-50
-40
-30
-20
-10
10
20
30
40
50
60
70
80

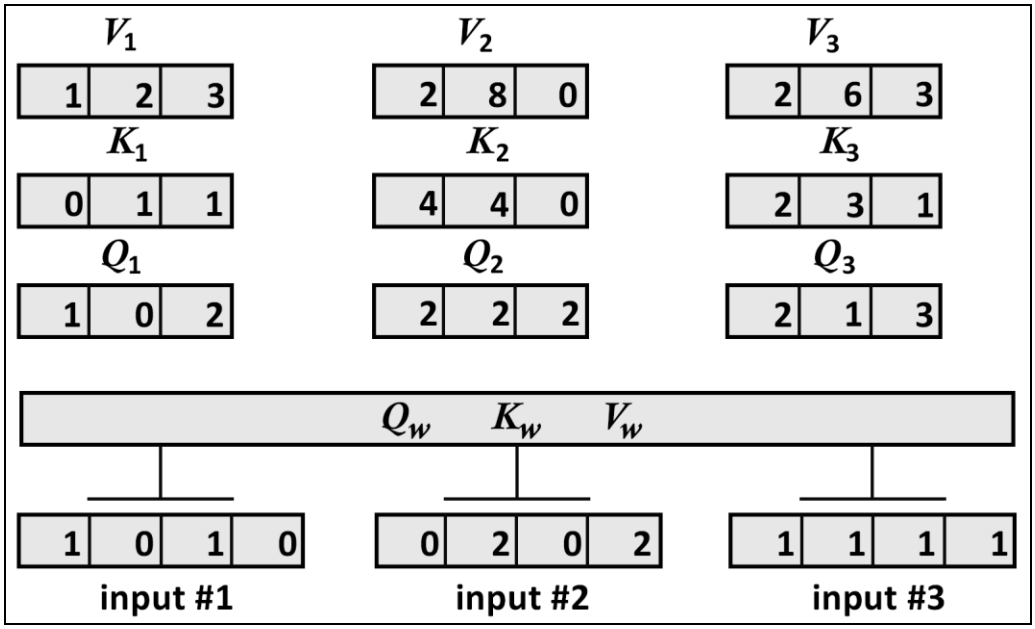

positional encoding
pc(black)
pc(brown)
positional vector
pe(2)
pe(10)
embedding vector
y_1=black
y_2=brown
Input
x_1=black
x_2=brown

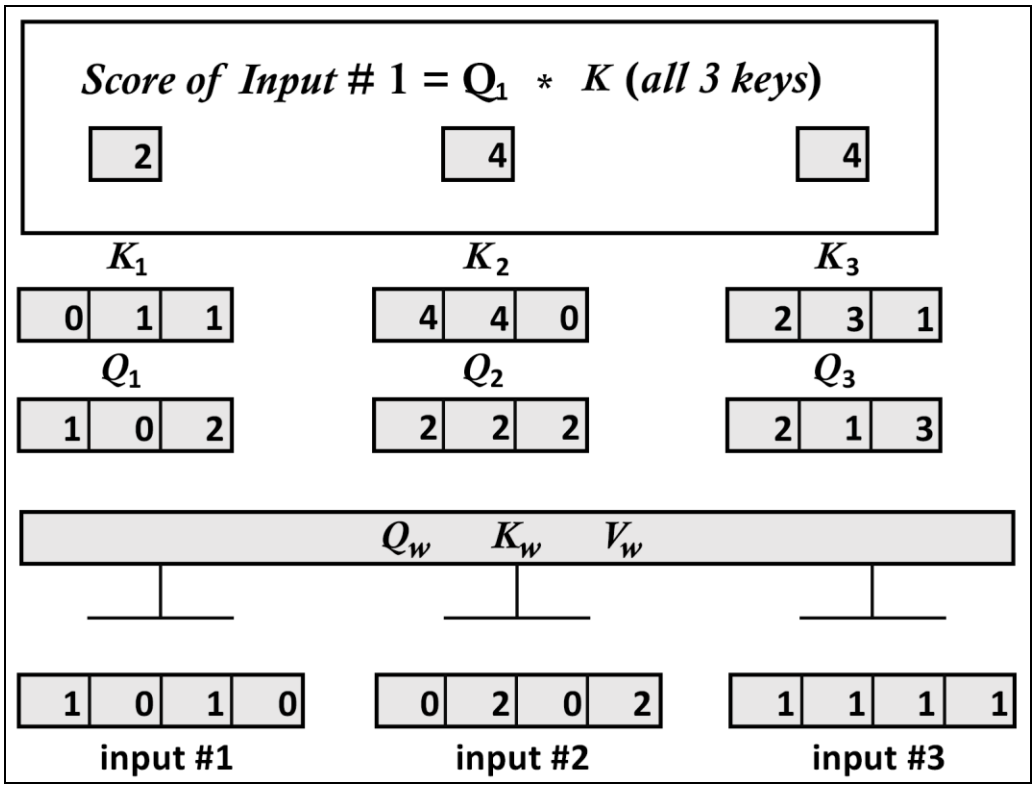

Nx
Add & Norm
Multi-Head Attention

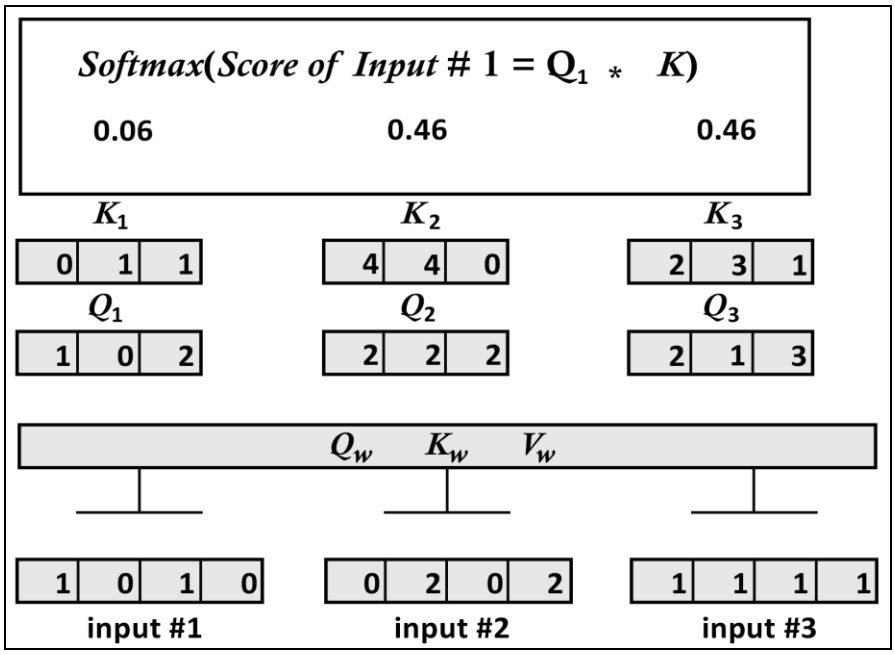

x, d_model = 512
head 1, x, d_k = 64
head 2, x, d_k = 64
head 3, x, d_k = 64
head 4, x, d_k = 64
head 5, x, d_k = 64
head 6, x, d_k = 64
head 7, x, d_k = 64
head 8, x, d_k = 64
MultiHead(output) = Concat(Z) = x, d_model

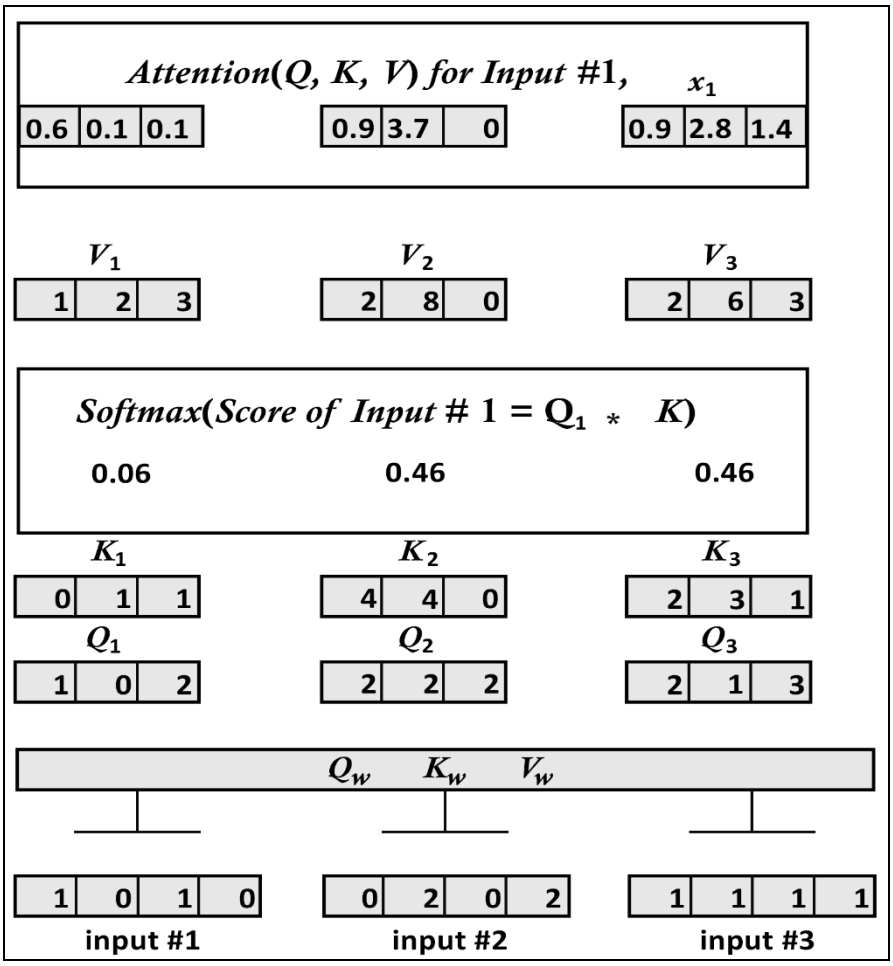

1 0 1 0
input #1
0 2 0 2
input #2
1 1 1 1
input #3

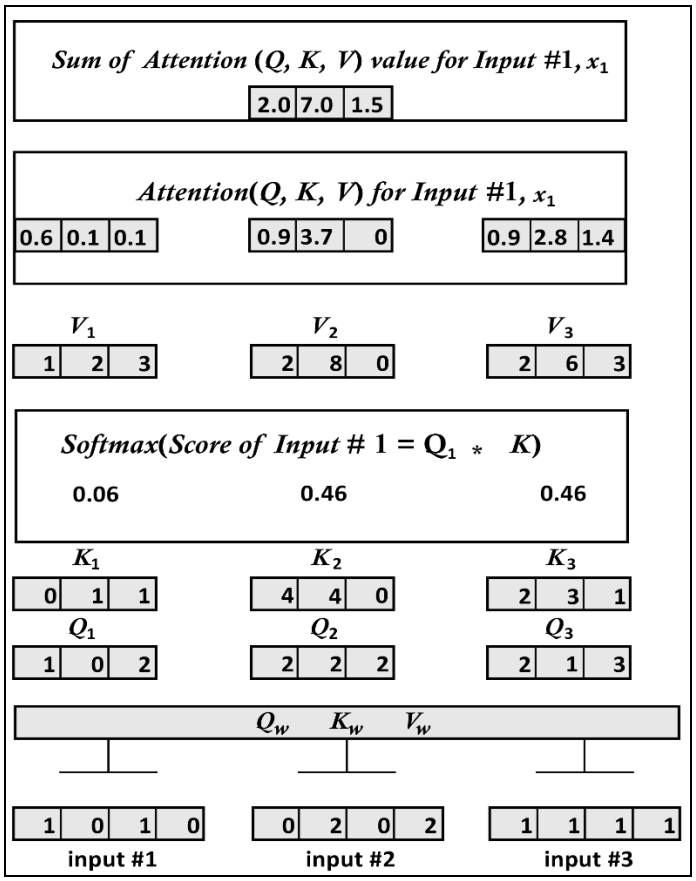

Q_w K_w V_w
1 0 1 0
input #1
0 2 0 2
input #2
1 1 1 1
input #3

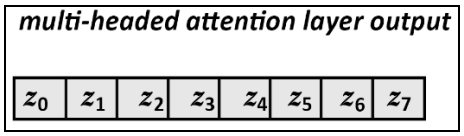

V_1
1 2 3
V_2
2 8 0
V_3
2 6 3
K_1
0 1 1
K_2
4 4 0
K_3
2 3 1
Q_1
1 0 2
Q_2
2 2 2
Q_3
2 1 3
Q_w K_w V_w
1 0 1 0
input #1
0 2 0 2
input #2
1 1 1 1
input #3

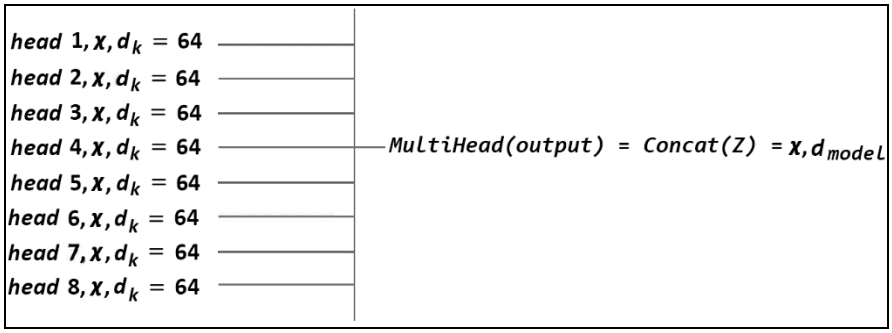

Score of Input # 1 = Q1 * K (all 3 keys)
2
4
4
K1
K2
K3
0 1 1
4 4 0
2 3 1
Q1
Q2
Q3
1 0 2
2 2 2
2 1 3
Qw Kw Vw
1 0 1 0
0 2 0 2
1 1 1 1
input #1
input #2
input #3

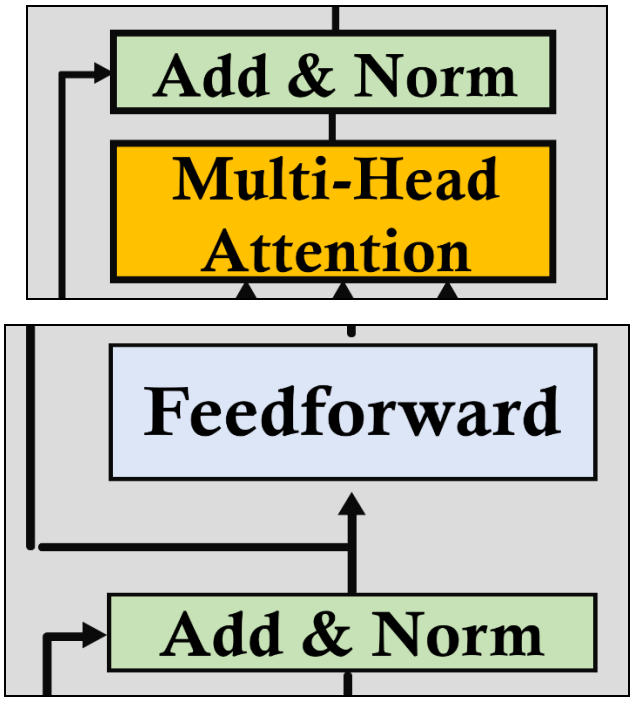

Softmax(Score of Input # 1 = Q1 * K)
0.06
0.46
0.46
K1
0 1 1
K2
4 4 0
K3
2 3 1
Q1
1 0 2
Q2
2 2 2
Q3
2 1 3
Qw Kw Vw
1 0 1 0
0 2 0 2
1 1 1 1
input #1
input #2
input #3

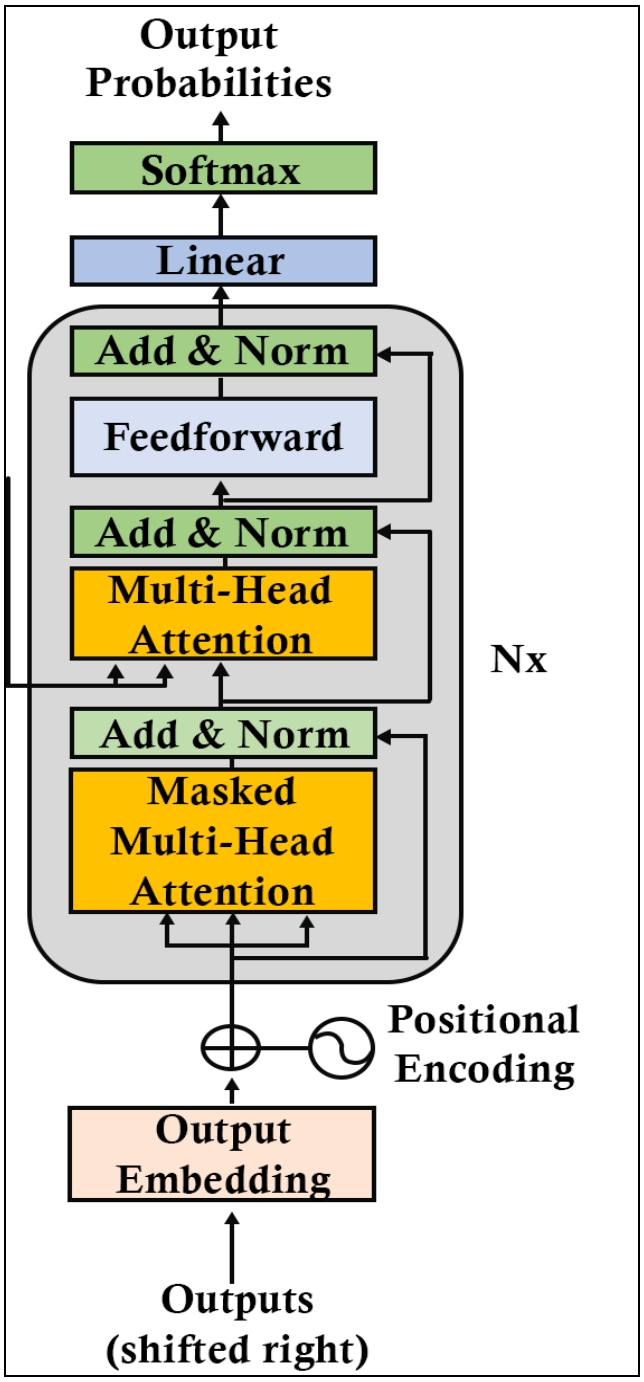

Attention(Q, K, V) for Input #1,
x_1
0.6 0.1 0.1
0.9 3.7 0
0.9 2.8 1.4
V_1
1 2 3
V_2
2 8 0
V_3
2 6 3
Softmax(Score of Input # 1 = Q_1 * K)
0.06
0.46
0.46
K_1
0 1 1
K_2
4 4 0
K_3
2 3 1
Q_1
1 0 2
Q_2
2 2 2
Q_3
2 1 3
Q_w K_w V_w
1 0 1 0
0 2 0 2
1 1 1 1
input #1
input #2
input #3


Sum of Attention (Q, K, V) value for Input #1, x_1
2.0 7.0 1.5
Attention(Q, K, V) for Input #1, x_1
0.6 0.1 0.1
0.9 3.7 0
0.9 2.8 1.4
V_1
1 2 3
V_2
2 8 0
V_3
2 6 3
Softmax(Score of Input # 1 = Q_1 * K)
0.06
0.46
0.46
K_1
0 1 1
K_2
4 4 0
K_3
2 3 1
Q_1
1 0 2
Q_2
2 2 2
Q_3
2 1 3
Q_w K_w V_w
1 0 1 0
0 2 0 2
1 1 1 1
input #1
input #2
input #3

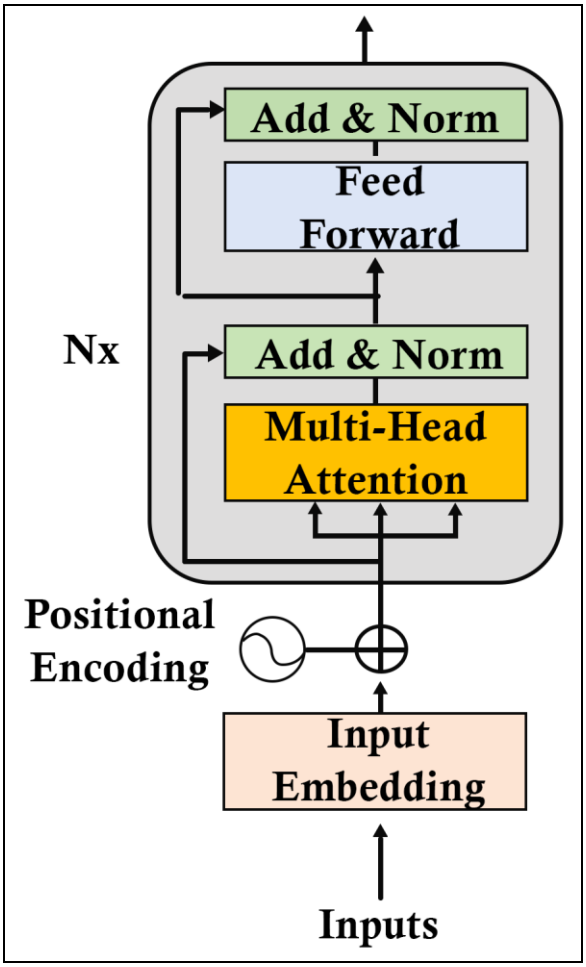

multi-headed attention layer output
z_0 z_1 z_2 z_3 z_4 z_5 z_6 z_7

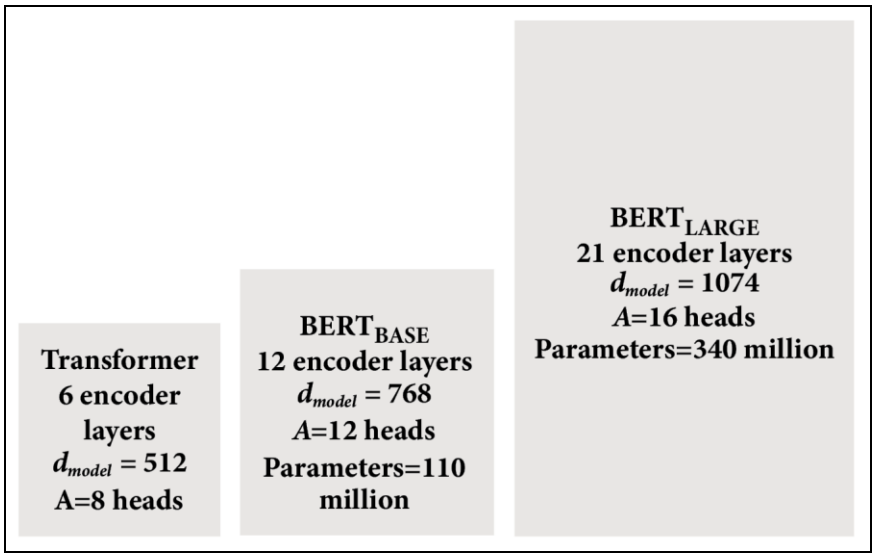

head 1, X, d_k = 64
head 2, X, d_k = 64
head 3, X, d_k = 64
head 4, X, d_k = 64
head 5, X, d_k = 64
head 6, X, d_k = 64
head 7, X, d_k = 64
head 8, X, d_k = 64
MultiHead(output) = Concat(Z) = X, d_{model}

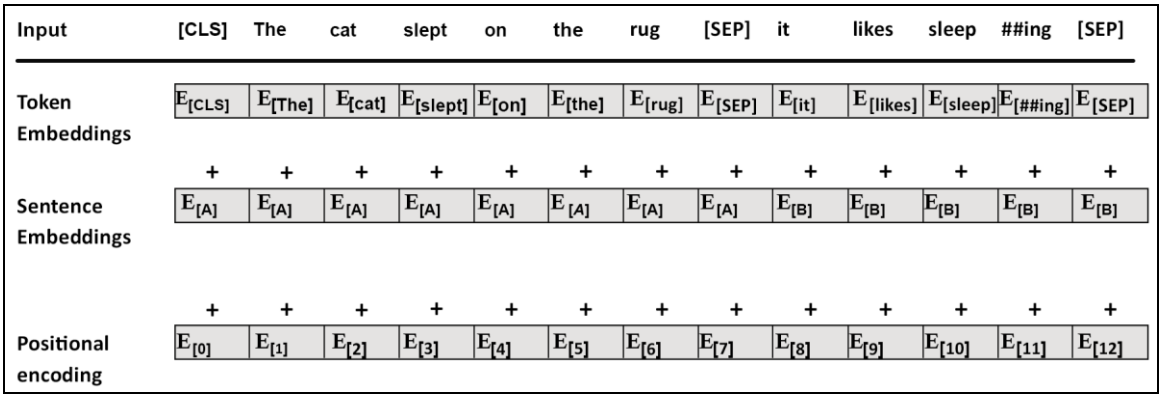

Add & Norm
Multi-Head Attention
Feedforward
Add & Norm

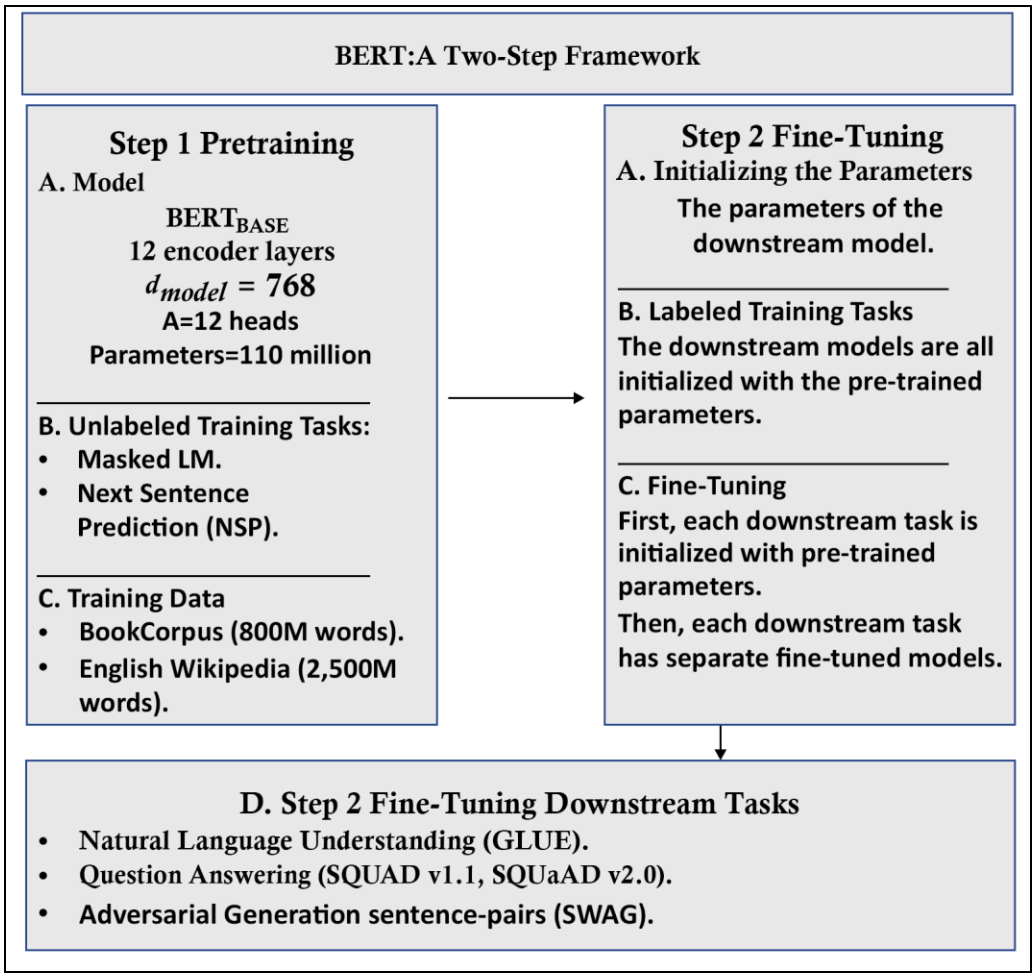

Output Probabilities
Softmax
Linear
Add & Norm
Feedforward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Positional Encoding
Output Embedding
Outputs (shifted right)

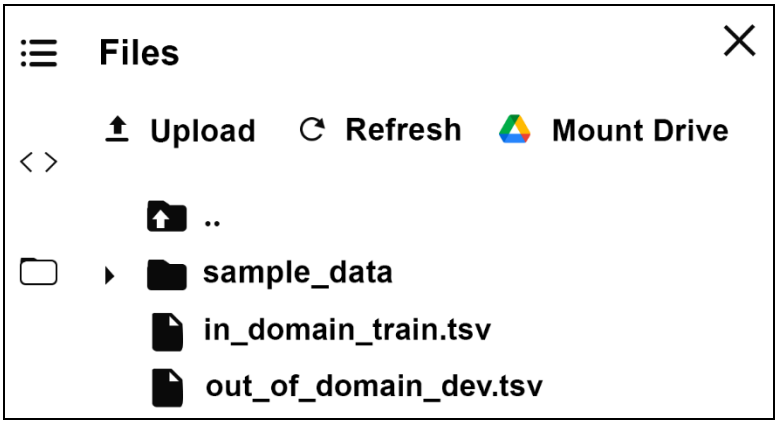
Chapter 3: Fine-Tuning BERT Models

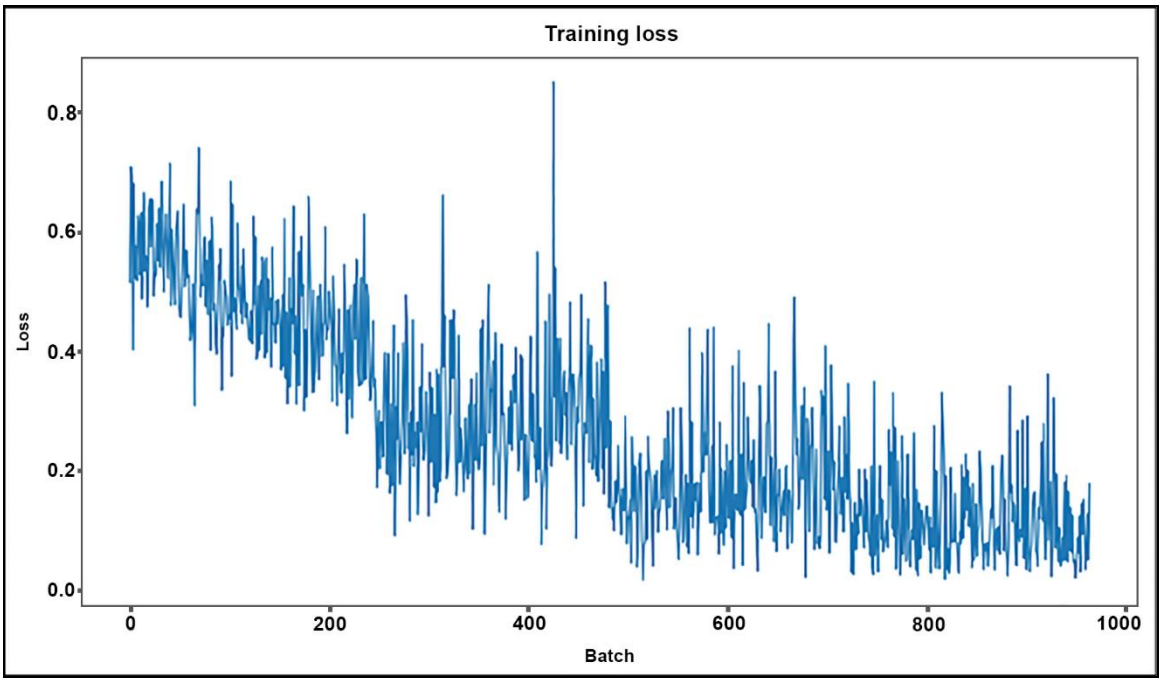



Transformer
6 encoder layers
d_{model} = 512
A=8 heads
BERT$_{BASE}$
12 encoder layers
d_{model} = 768
A=12 heads
Parameters=110 million
BERT$_{LARGE}$
21 encoder layers
d_{model} = 1074
A=16 heads
Parameters=340 million

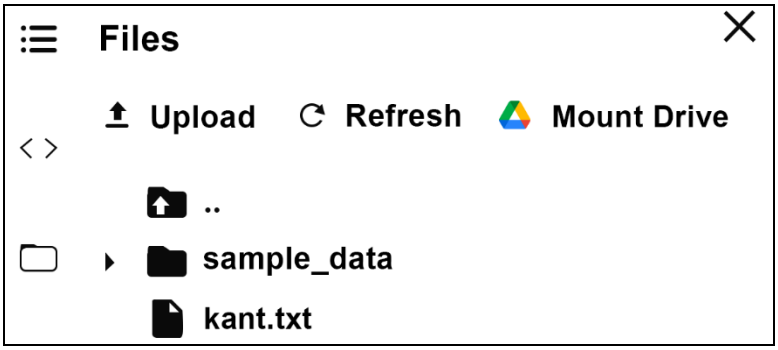

Input
[CLS] The cat slept on the rug [SEP] it likes sleep ##ing [SEP]
Token Embeddings
E[CLS] E[The] E[cat] E[slept] E[on] E[the] E[rug] E[SEP] E[it] E[likes] E[sleep] E[##ing] E[SEP]
Sentence Embeddings
E[A] E[A] E[A] E[A] E[A] E[A] E[A] E[A] E[B] E[B] E[B] E[B] E[B]
Positional encoding
E[0] E[1] E[2] E[3] E[4] E[5] E[6] E[7] E[8] E[9] E[10] E[11] E[12]

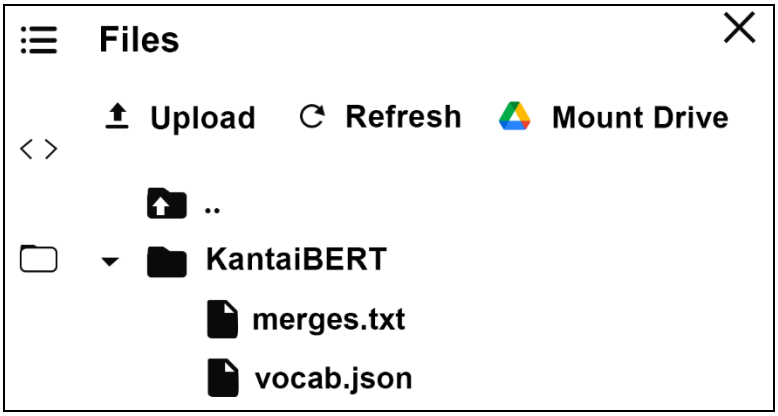

BERT:A Two-Step Framework
Step 1 Pretraining
A. Model
$BERT_{BASE}$
12 encoder layers
d_{model} = 768
A=12 heads
Parameters=110 million
B. Unlabeled Training Tasks:
• Masked LM.
• Next Sentence Prediction (NSP).
C. Training Data
• BookCorpus (800M words).
• English Wikipedia (2,500M words).
Step 2 Fine-Tuning
A. Initializing the Parameters
The parameters of the downstream model.
B. Labeled Training Tasks
The downstream models are all initialized with the pre-trained parameters.
C. Fine-Tuning
First, each downstream task is initialized with pre-trained parameters.
Then, each downstream task has separate fine-tuned models.
D. Step 2 Fine-Tuning Downstream Tasks
• Natural Language Understanding (GLUE).
• Question Answering (SQUAD v1.1, SQUaAD v2.0).
• Adversarial Generation sentence-pairs (SWAG).


Files
Upload
Refresh
Mount Drive
..
sample_data
in_domain_train.tsv
out_of_domain_dev.tsv


Training loss
Loss
0.8
0.6
0.4
0.2
0.0
0
200
400
600
800
1000
Batch

Chapter 4: Pretraining a RoBERTa Model from Scratch

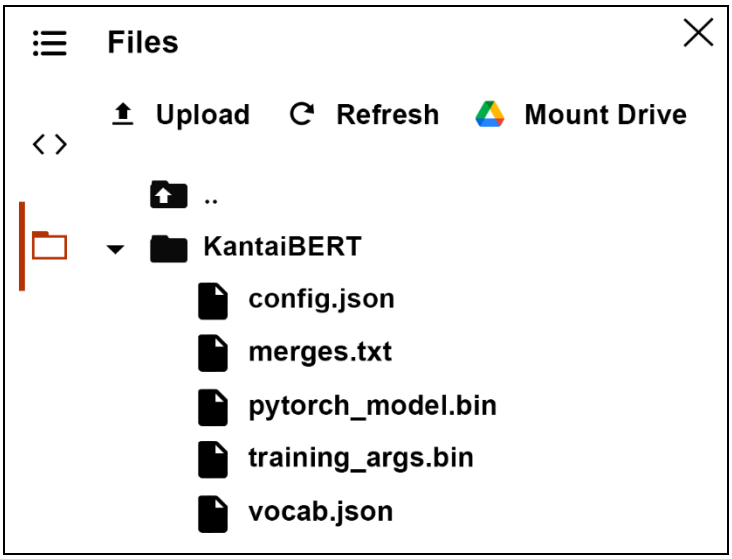




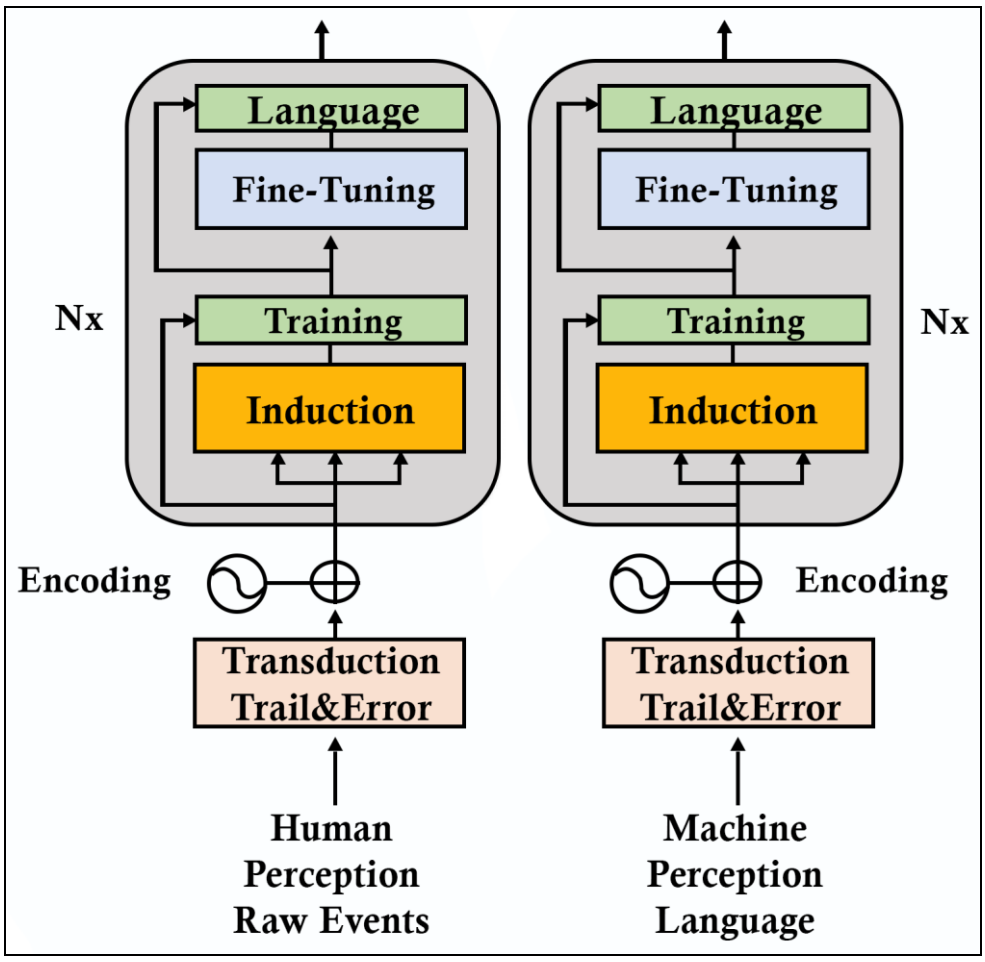

```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 440.82       Driver Version: 418.67       CUDA Version: 10.1     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|===============================+======================+======================|
|   0  Tesla K80           Off  | 00000000:00:04.0 Off |                    0 |
| N/A   49C    P0    63W / 149W |   9707MiB / 11441MiB |      0%      Default |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                       GPU Memory |
|  GPU       PID   Type   Process name                             Usage      |
|=============================================================================|
+-----------------------------------------------------------------------------+
```

list: LP

[Parameter with shape torch.Size([52000, 768]), Parameter with shape torch.Size([514, 768]), Parameter with shape torch.Size([1, 768]), Parameter with shape torch.Size([768])), Parameter with shape torch.Size([768]), ...] (108 items total)



Chapter 5: Downstream NLP Tasks with Transformers



Rank	Name	Model	URL	Score
1	Microsoft Alexander v-team	Turing NLR v5		91.2
2	ERNIE Team - Baidu	ERNIE		91.1
3	AliceMind & DIRL	StructBERT + CLEVER		91.0
4	liangzhu ge	DEBERTa + CLEVER		90.9
5	DeBERTa Team - Microsoft	DeBERTa / TuringNLR		90.8
6	HFL iFLYTEK	MacALBERT + DKM		90.7
17	GLUE Human Baselines	GLUE Human Baselines		87.1

	Rank	Name	Model
	1	SuperGLUE Human Baselines	SuperGLUE Human Baselines
+	2	T5 Team - Google	T5
+	3	Huawei Noah's Ark Lab	NEZHA-Plus

Rank	Name	Model	URL	Score
1	Microsoft Alexander v-team	Turing NLR v5		90.9
2	ERNIE Team - Baidu	ERNIE 3.0		90.6
3	Zirui Wang	T5 + UDG, Single Model (Google Brain		90.4
4	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3
5	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8

SuperGLUE Tasks

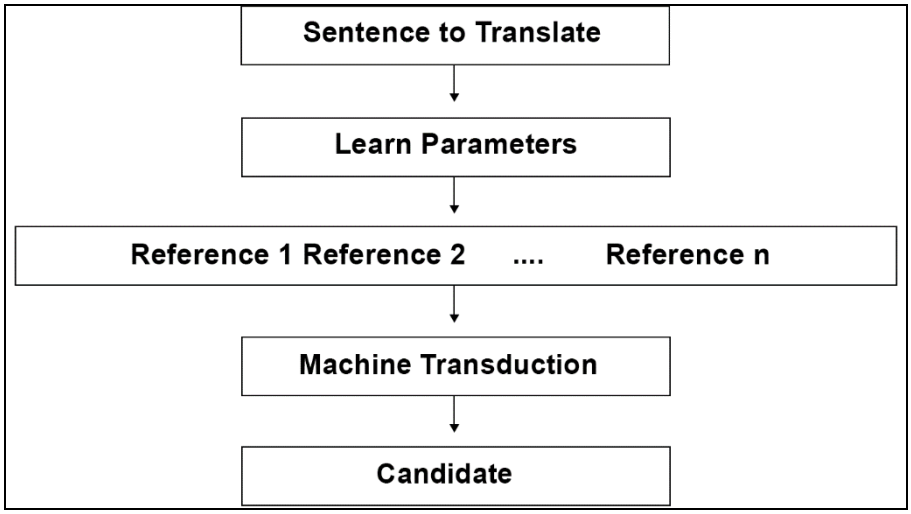
Name	Identifier	Download	More Info	Metric
Broadcoverage Diagnostics	AX-b			Matthew's Corr
CommitmentBank	CB			Avg. F1 / Accuracy
Choice of Plausible Alternatives	COPA			Accuracy
Multi-Sentence Reading Comprehension	MultiRC			F1a / EM
Recognaing Textual Entailment	RTE			Accuracy
Words in Context	WiC			Accuracy
The Winograd Schema Challenge	WSC			Accuracy
BoolQ	BoolQ			Accuracy
Reading Comprehension with Commonsense Reasoning	ReCoRD			F1 / Accuracy
Winogender Schema Diagnostics	AX-g			Gender Parity / Accuracy

DOWNLOAD ALL DATA

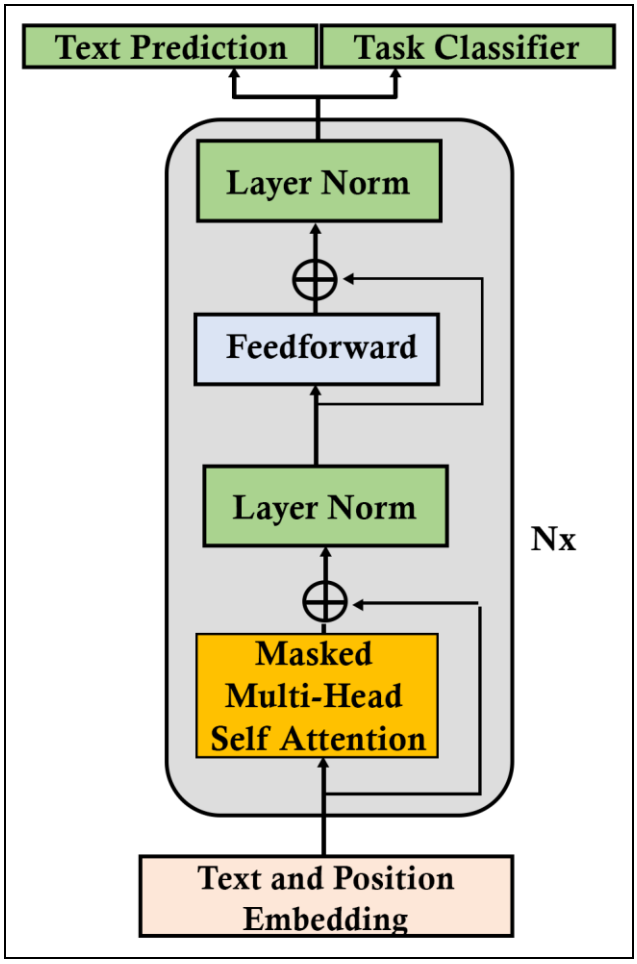
Score	BoolQ	CB	COPA	MultiRC	ReCoRD	RTE	WiC	WSC	AX-b	AX-g
89.8	89.0	95.8/98.9	100.0	81.8/51.9	91.7/91.3	93.6	80.0	100.0	76.6	99.3/99.7

Rank	Name	Model	URL	Score	COPA
1	Microsoft Alexander v-team	Turing NLR v5		90.9	98.2
2	ERNIE Team - Baidu	ERNIE 3.0		90.6	97.4
3	Zirui Wang	T5 + UDG, Single Model (Google Brain)		90.4	98.0
4	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4		90.3	98.4
5	SuperGLUE Human Baselines	SuperGLUE Human Baselines		89.8	100.0

Chapter 6: Machine Translation with the Transformer

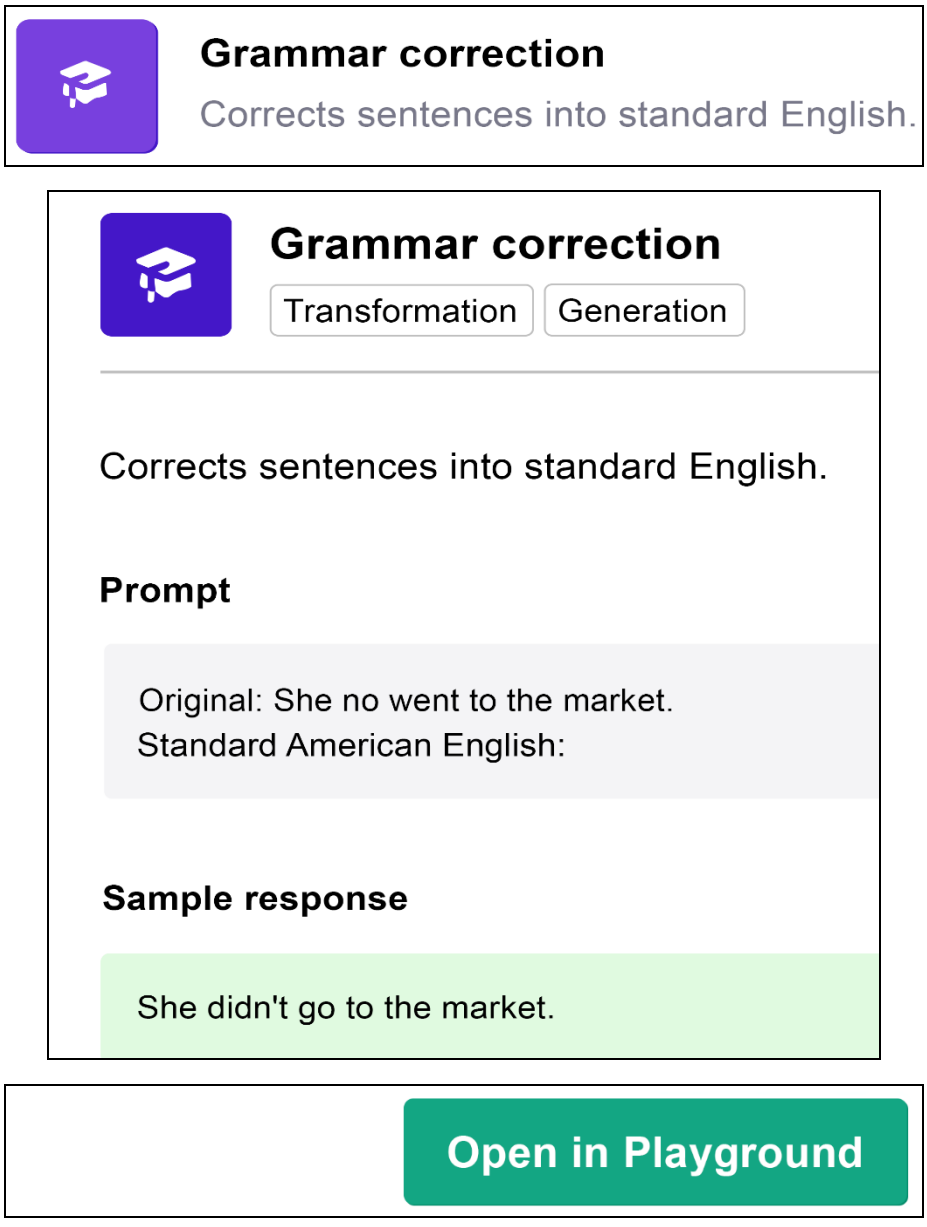


Chapter 7: The Rise of Suprahuman Transformers with GPT-3 Engines



prompt >>>

prompt >>>


Grammar correction
Corrects sentences into standard English.
Grammar correction
Transformation
Generation
Corrects sentences into standard English.
Prompt
Original: She no went to the market.
Standard American English:
Sample response
She didn't go to the market.
Open in Playground

API request

davinci ▾ python ▾ Copy

```python
import os
import openai

openai.api_key = os.getenv("OPENAI_API_KEY")

response = openai.Completion.create(
  engine="davinci",
  prompt="Original: She no went to the market.\nStandard
  temperature=0,
  max_tokens=60,
  top_p=1.0,
  frequency_penalty=0.0,
  presence_penalty=0.0,
  stop=["\n"]
)
```

Back to Future: 👦👴🚗🕒

Batman: 🤵🦇

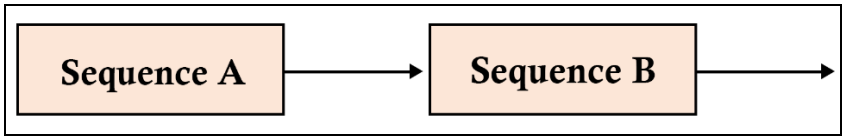
Transformers: 🚗🤖

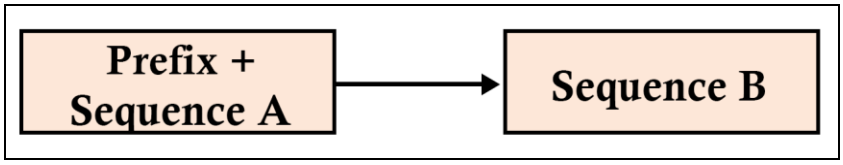
Mission **Impossible:** 🚕👨🏭🚕

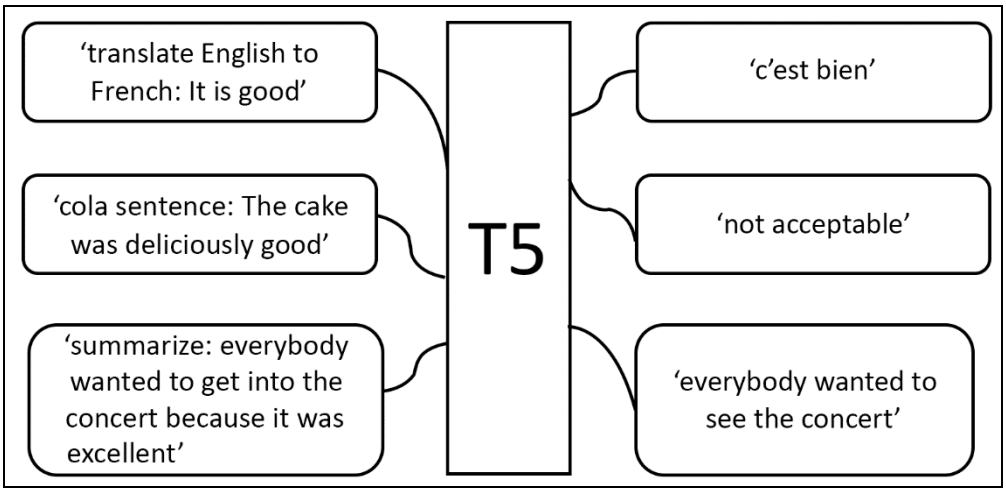
```
WARNING: The following packages were previously imported in this runtime:
  [pandas]
You must restart the runtime in order to use newly installed versions.
```

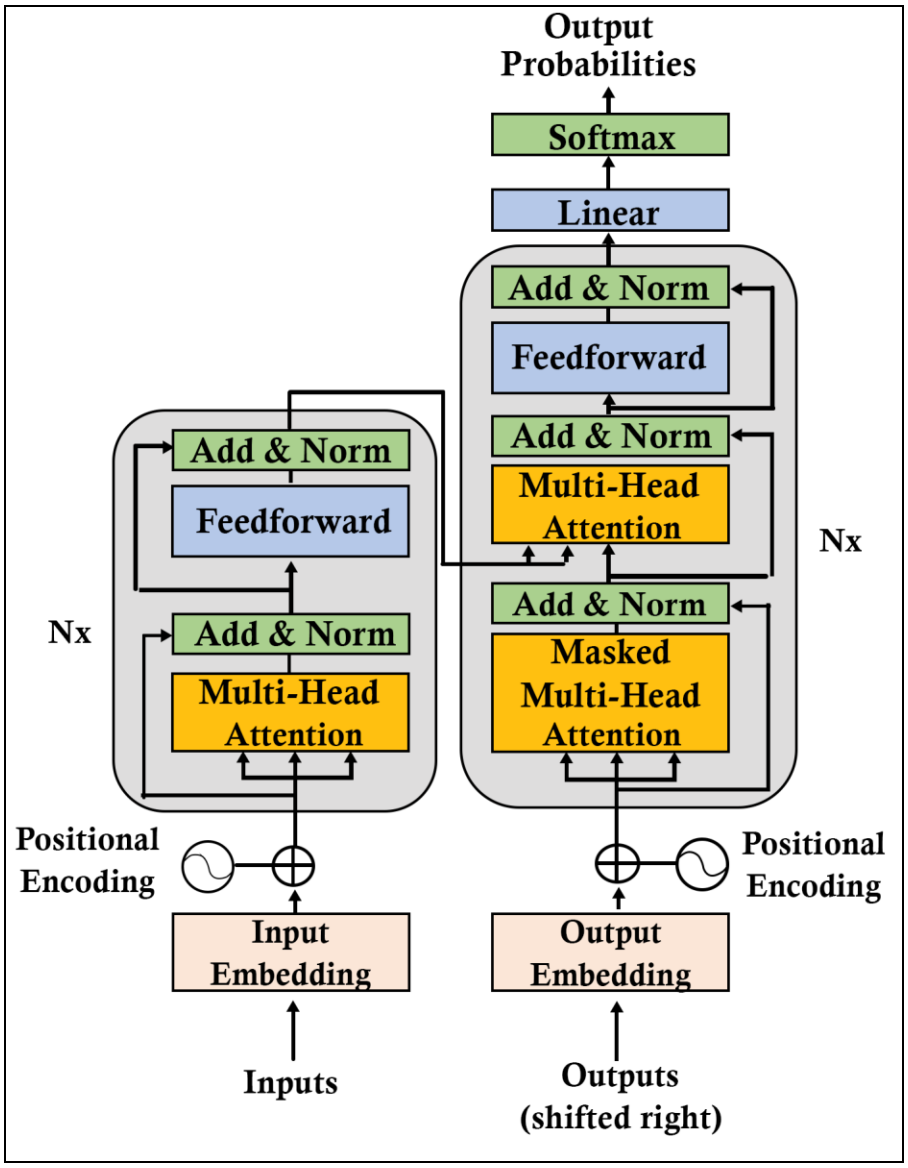
RESTART RUNTIME

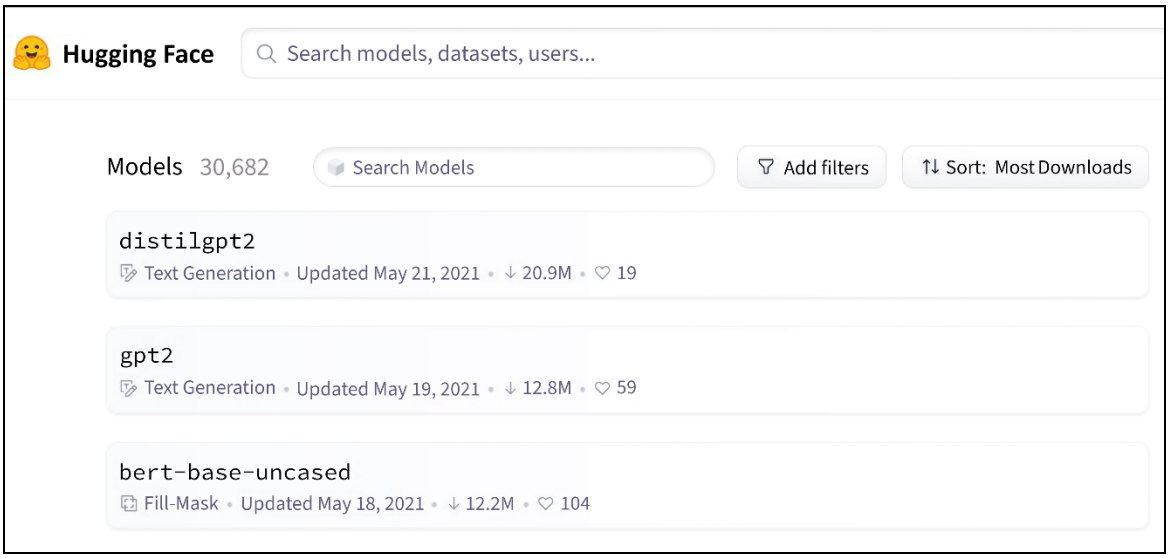
Chapter 8: Applying Transformers to Legal and Financial Documents for AI Text Summarization

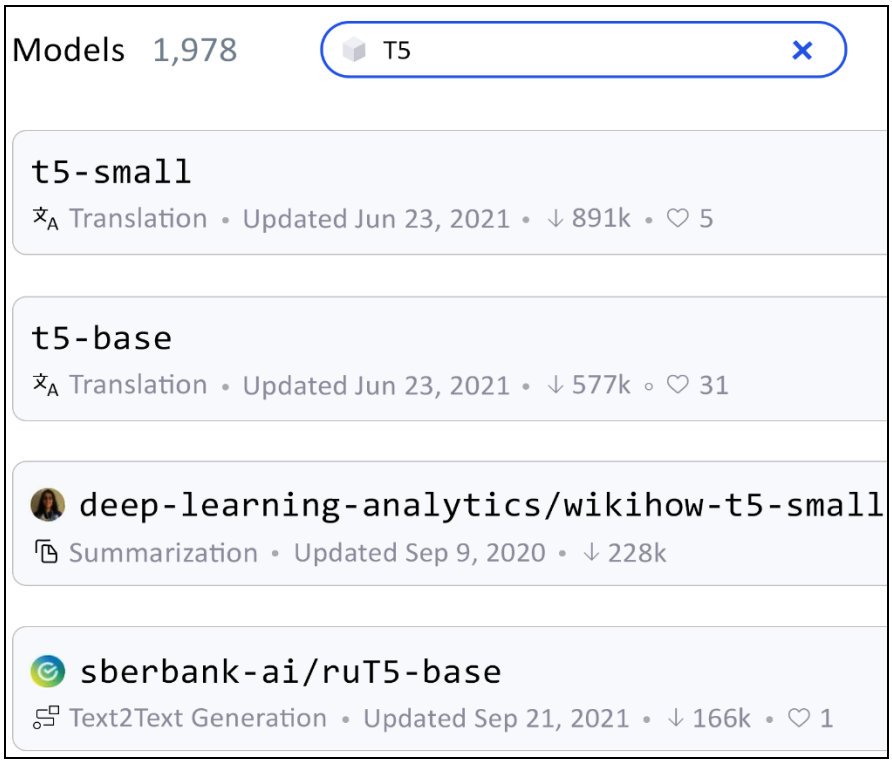







Output Probabilities
Softmax
Linear
Add & Norm
Feedforward
Add & Norm
Multi-Head Attention
Add & Norm
Masked Multi-Head Attention
Nx
Add & Norm
Feedforward
Add & Norm
Multi-Head Attention
Nx
Positional Encoding
Positional Encoding
Input Embedding
Output Embedding
Inputs
Outputs (shifted right)


Hugging Face
Search models, datasets, users...
Models 30,682
Search Models
Add filters
Sort: Most Downloads
distilgpt2
Text Generation • Updated May 21, 2021 • ↓ 20.9M • ♡ 19
gpt2
Text Generation • Updated May 19, 2021 • ↓ 12.8M • ♡ 59
bert-base-uncased
Fill-Mask • Updated May 18, 2021 • ↓ 12.2M • ♡ 104


Models 1,978
T5
t5-small
Translation • Updated Jun 23, 2021 • ↓ 891k • ♡ 5
t5-base
Translation • Updated Jun 23, 2021 • ↓ 577k • ♡ 31
deep-learning-analytics/wikihow-t5-small
Summarization • Updated Sep 9, 2020 • ↓ 228k
sberbank-ai/ruT5-base
Text2Text Generation • Updated Sep 21, 2021 • ↓ 166k • ♡ 1

How to use from the 🤗/transformers library

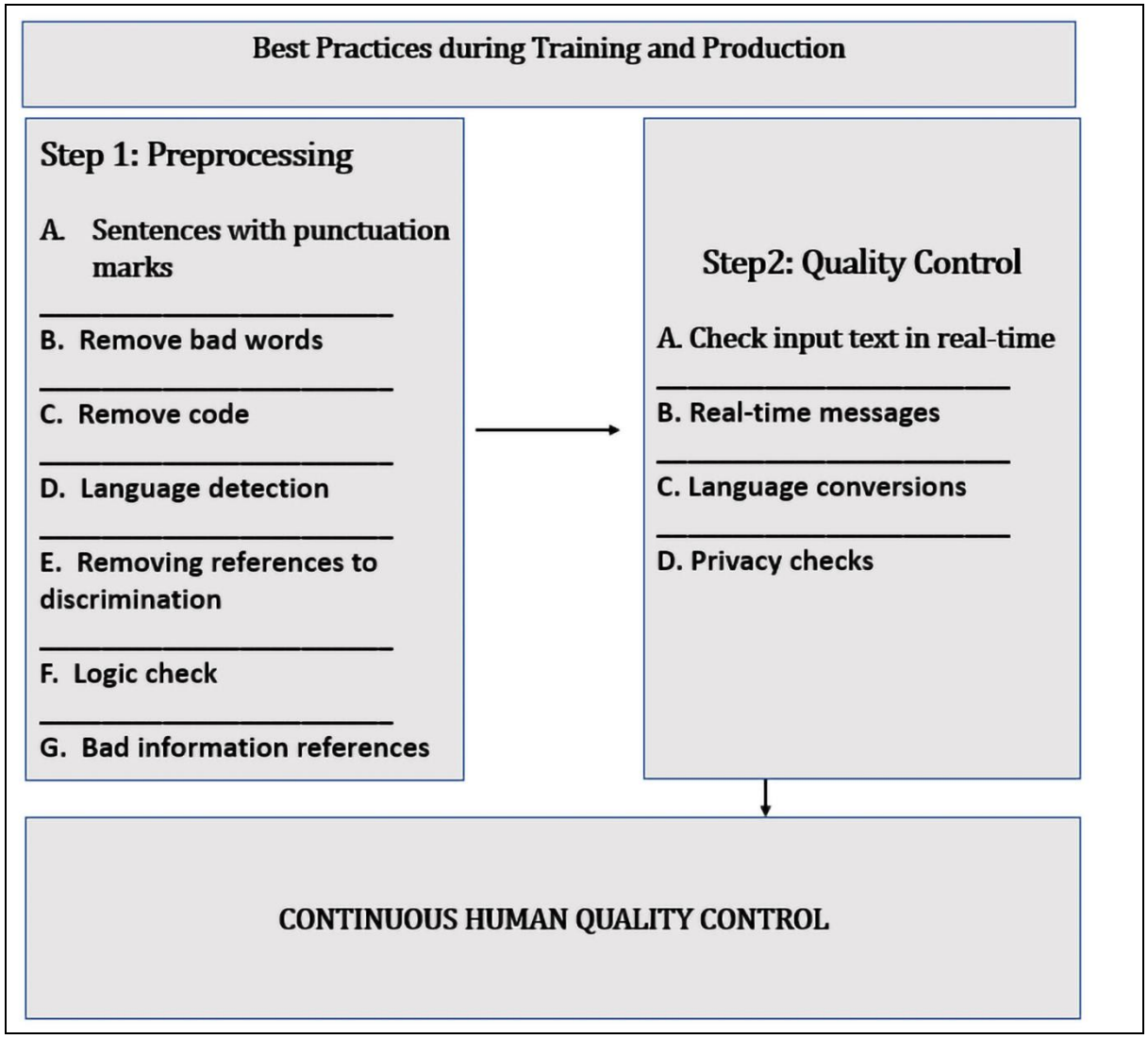
```
from transformers import AutoTokenizer, AutoModelWithLMHead

tokenizer = AutoTokenizer.from_pretrained("t5-large")

model = AutoModelWithLMHead.from_pretrained("t5-large")
```

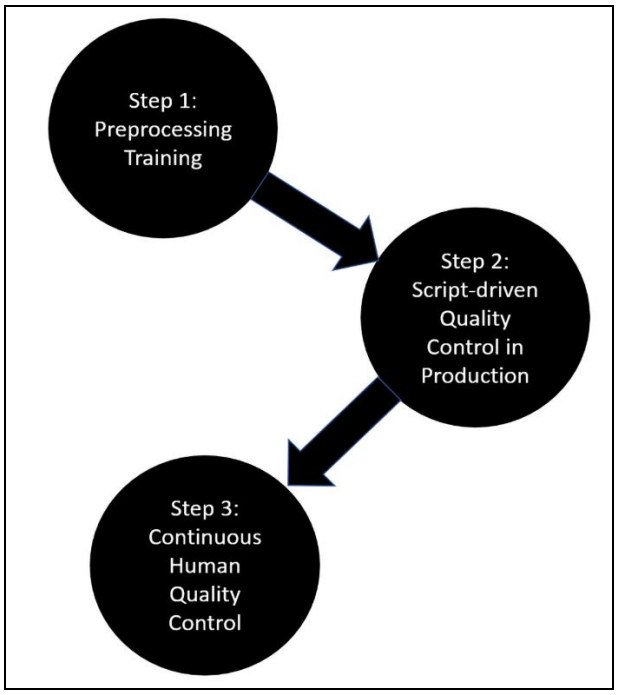
Copy

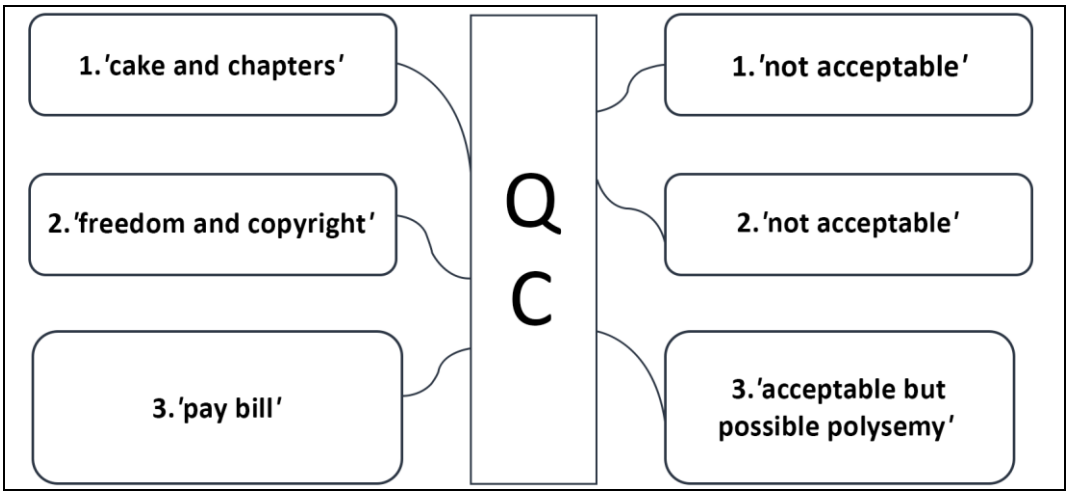
Summarize for a 2nd grader

Translates difficult text into simpler concepts.

Chapter 9: Machine Tokenizers and Datasets



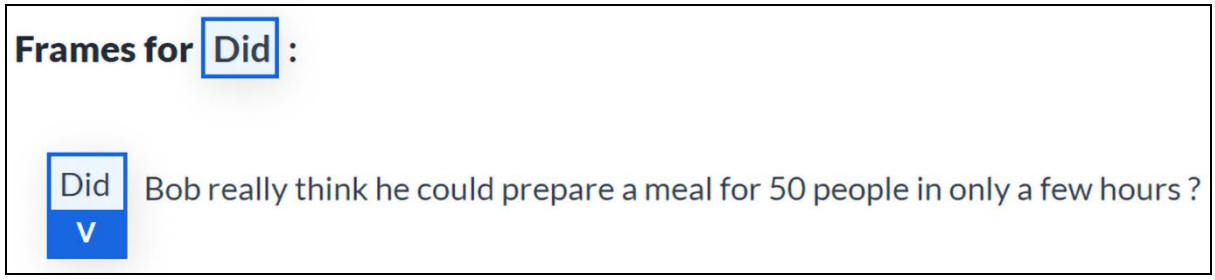

Step 1:
Preprocessing
Training
Step 2:
Script-driven
Quality
Control in
Production
Step 3:
Continuous
Human
Quality
Control

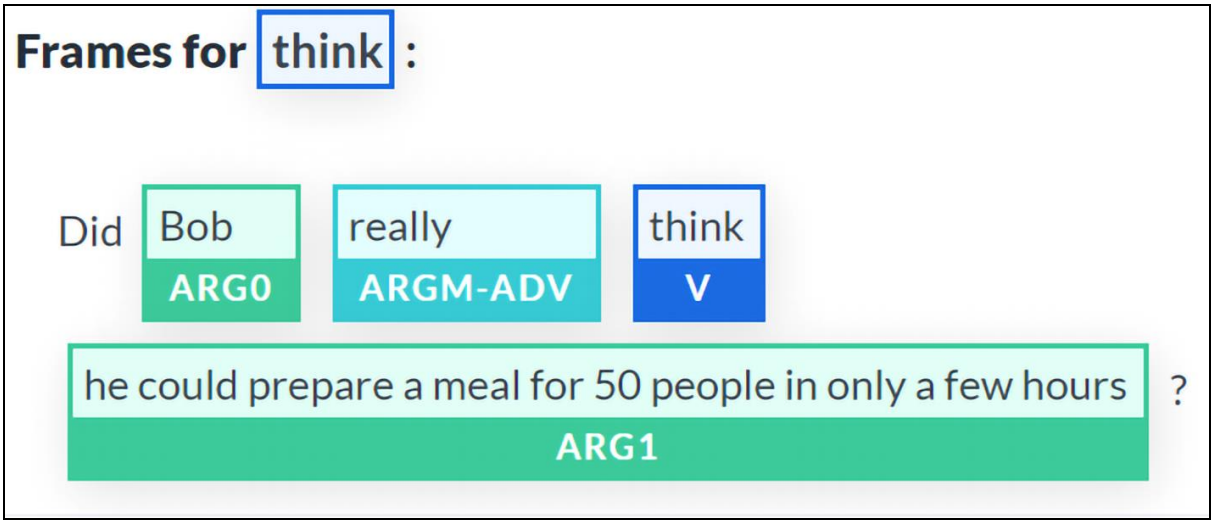

1.'cake and chapters'
2.'freedom and copyright'
3.'pay bill'
QC
1.'not acceptable'
2.'not acceptable'
3.'acceptable but possible polysemy'

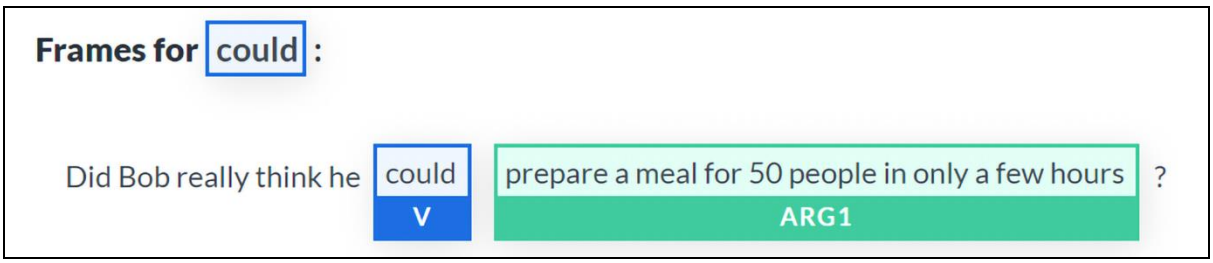
Playground

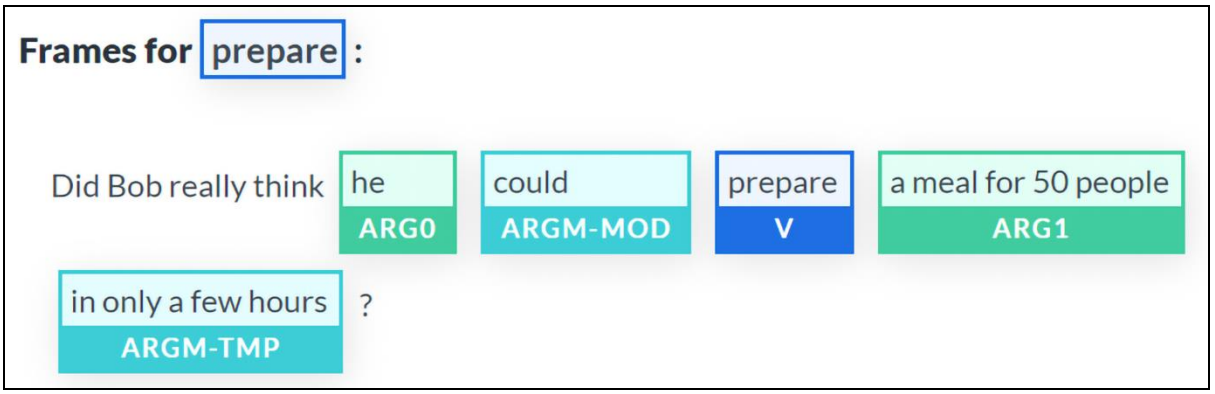
Q: What is the definition of amoeboid?

Chapter 10: Semantic Role Labeling with BERT-Based Transformers



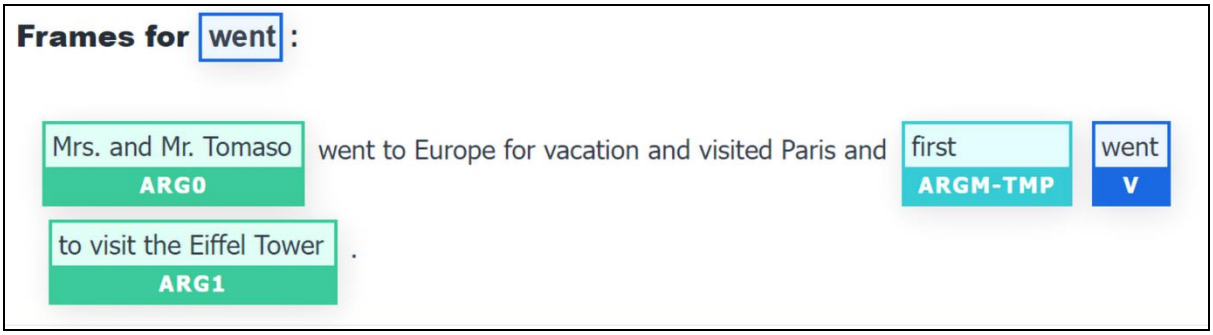

Frames for think :
Did
Bob
ARG0
really
ARGM-ADV
think
V
he could prepare a meal for 50 people in only a few hours
ARG1
?

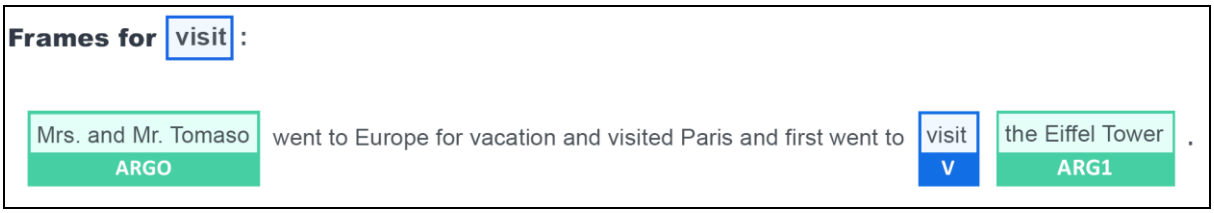

Frames for could :
Did Bob really think he
could
V
prepare a meal for 50 people in only a few hours
ARG1
?


Frames for prepare :
Did Bob really think
he
ARG0
could
ARGM-MOD
prepare
V
a meal for 50 people
ARG1
in only a few hours
ARGM-TMP
?

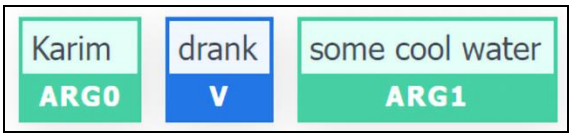
Frames for went :
Mrs. and Mr. Tomaso
ARG0
went
V
to Europe
ARG4
for vacation
ARGM-PRP

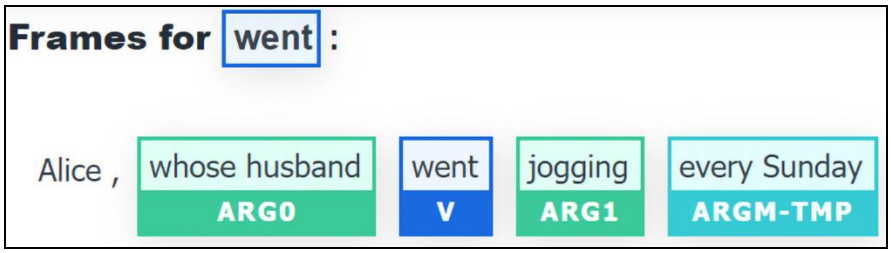
Frames for visited :
Mrs. and Mr. Tomaso
ARG0
went to Europe for vacation and
visited
V
Paris
ARG1

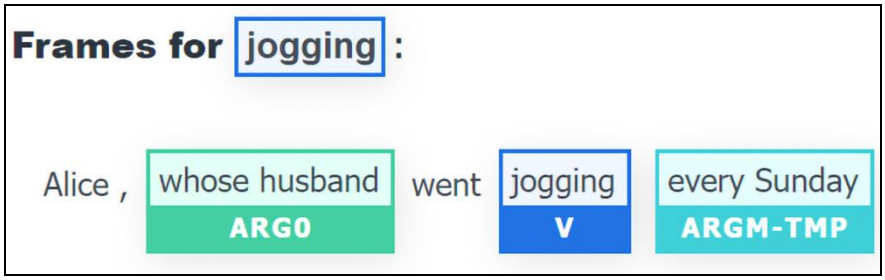

Frames for went :
Mrs. and Mr. Tomaso
ARG0
went to Europe for vacation and visited Paris and
first
ARGM-TMP
went
V
to visit the Eiffel Tower
ARG1
.

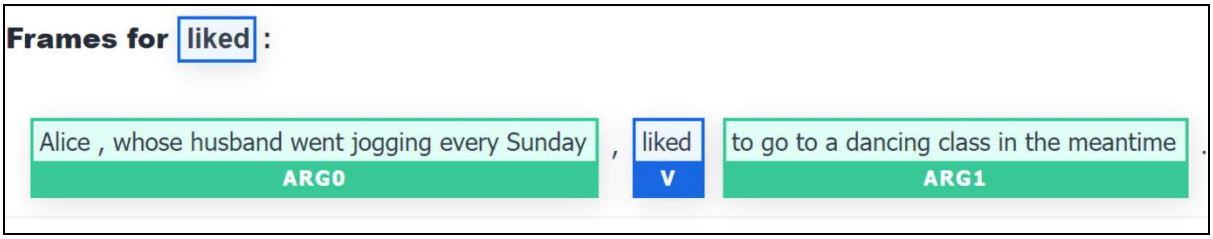

Frames for visit :
Mrs. and Mr. Tomaso
ARGO
went to Europe for vacation and visited Paris and first went to
visit
V
the Eiffel Tower
ARG1
.

Frames for wanted :
John
ARG0
wanted
V
to drink tea
ARG1


Karim
ARG0
drank
V
some cool water
ARG1


Frames for went :
Alice ,
whose husband
ARG0
went
V
jogging
ARG1
every Sunday
ARGM-TMP

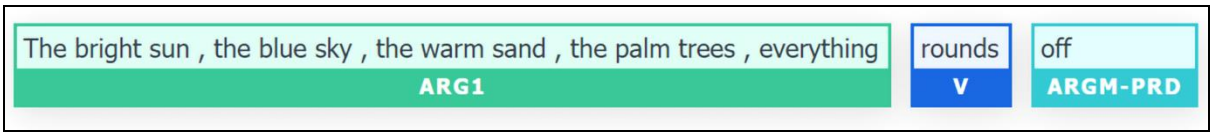

Frames for jogging :
Alice ,
whose husband
ARG0
went
jogging
V
every Sunday
ARGM-TMP

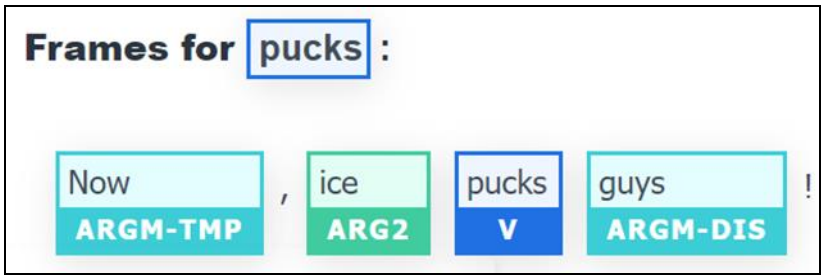

Frames for liked :
Alice , whose husband went jogging every Sunday
ARG0
,
liked
V
to go to a dancing class in the meantime
ARG1
.

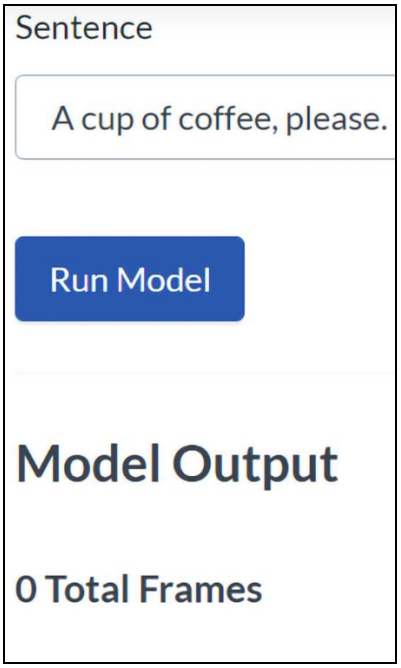
Frames for go :
Alice , whose husband went jogging every Sunday
ARG0
, liked to
go
V
to a dancing class
ARG4
in the meantime
ARGM-TMP
.

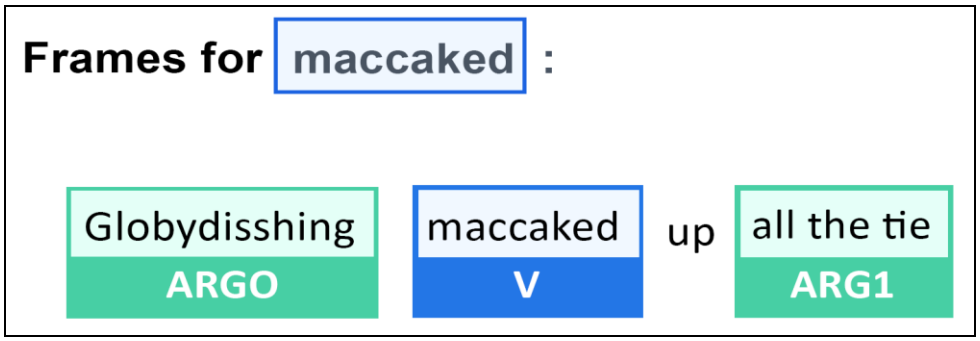
to a dancing class
ARG4

everything
ARG1
round
V
off
ARGM-PRD

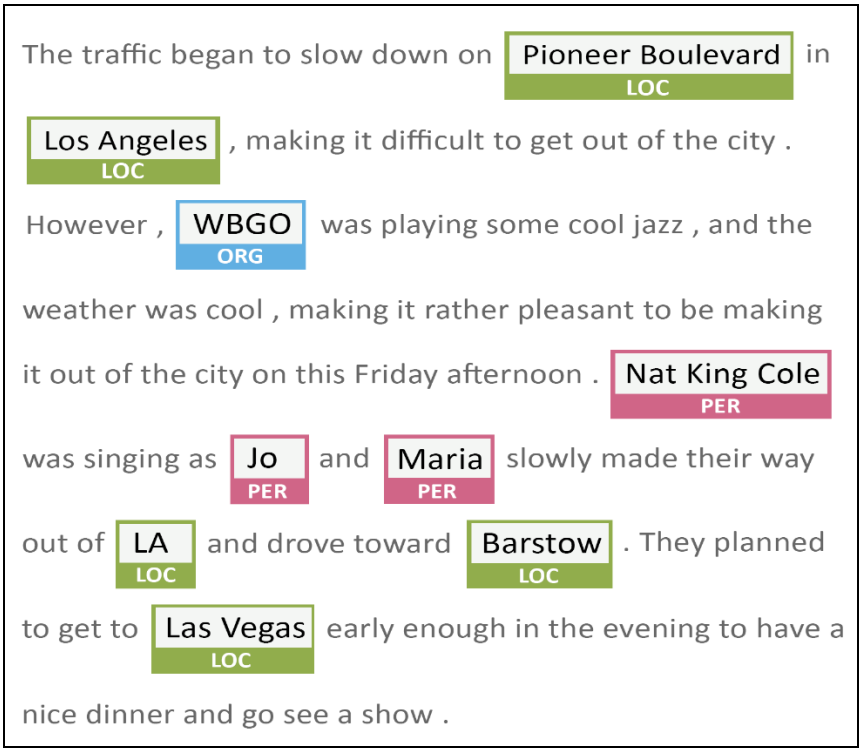

The bright sun , the blue sky , the warm sand , the palm trees , everything
ARG1
rounds
V
off
ARGM-PRD


Frames for pucks :
Now
ARGM-TMP
,
ice
ARG2
pucks
V
guys
ARGM-DIS
!


Sentence
A cup of coffee, please.
Run Model
Model Output
0 Total Frames


Frames for maccaked :
Globydisshing
ARGO
maccaked
V
up
all the tie
ARG1

Chapter 11: Let Your Data Do the Talking: Story, Questions and Answers



Frames for drove :

The traffic began to slow down on Pioneer Boulevard in Los Angeles , making it difficult to get out of the city . However , WBGO was playing some cool jazz , and the weather was cool , making it rather pleasant to be making it out of the city on this Friday afternoon . Nat King Cole was singing as Jo and Maria [ARG0] slowly [ARGM-MNR] made their way out of LA and drove [V] toward Barstow [ARGM-DIR] . They planned to get to Las Vegas early

Original Sequence	Masked tokens		Sample		Prediction
Nat	[MASK]	Generator	Nat	Discriminator ELECTRA	original
King	King		King		original
Cole	Cole		Cole		original
was	was		was		original
singing	[MASK]		driving		**replaced**

The traffic began [V]

to slow down on Pioneer Boulevard in Los Angeles , making it difficult to get out of the city

[ARG1]

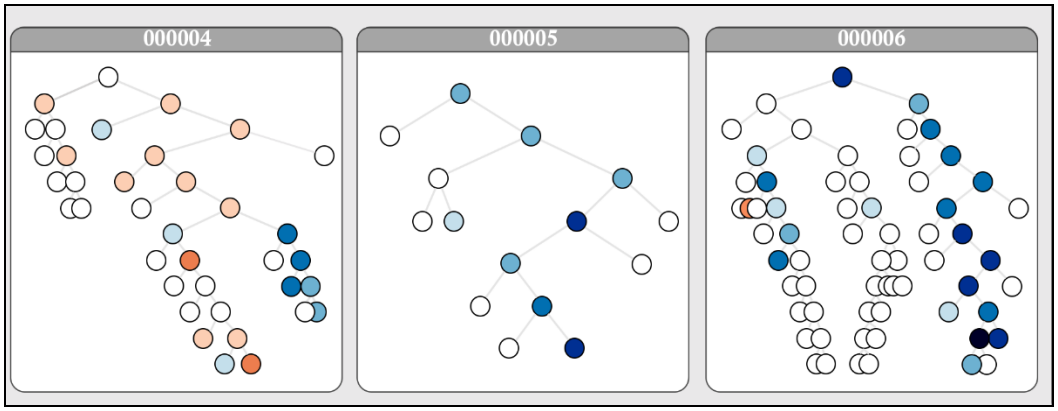
Frames for playing :

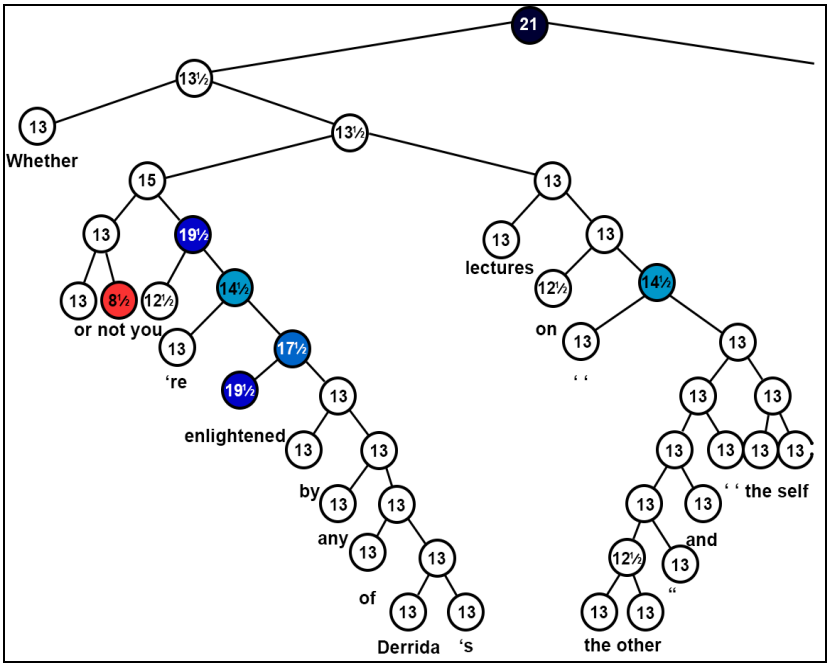
The traffic began to slow down on Pioneer Boulevard in Los Angeles , making it difficult to get out of the city . However ,

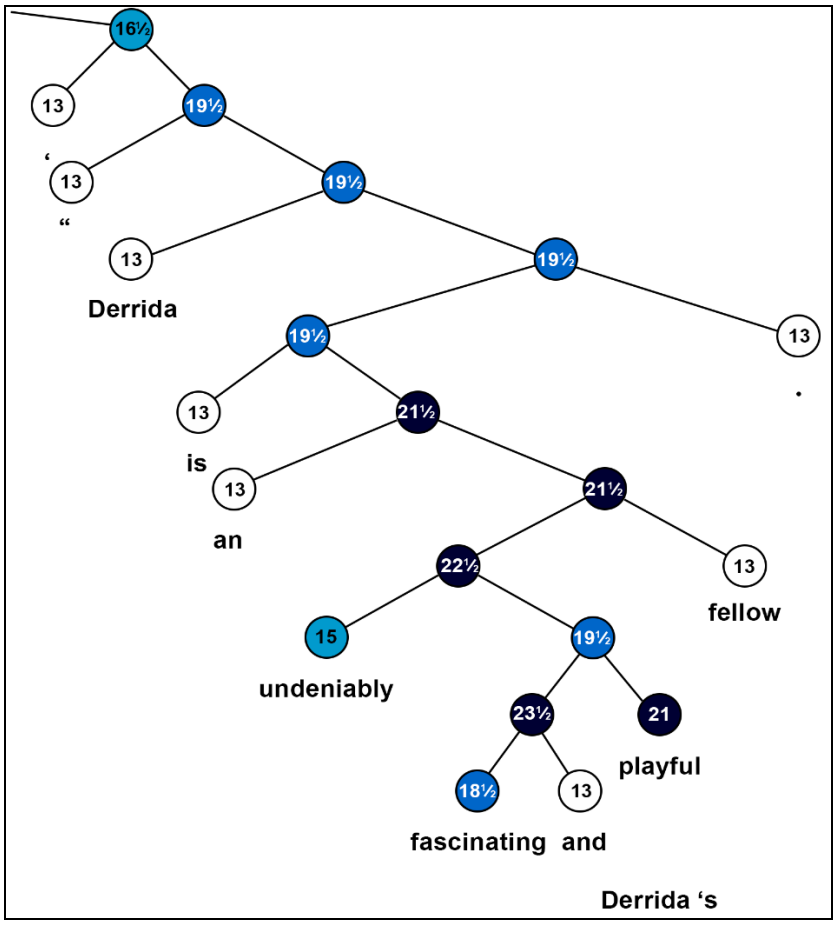
WBGO (ARG0) was playing (V) some cool jazz (ARG1) , and the weather LA and drove toward Barstow . They (ARG0) planned to get to Las Vegas early enough in the evening to have a nice dinner and go (V) see a show (ARG1) .

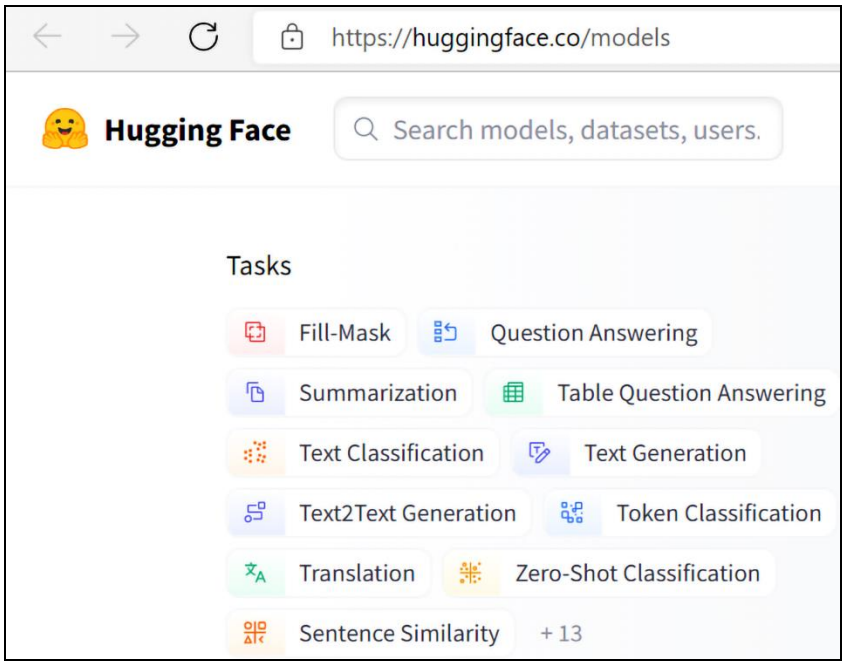
The traffic began to slow down on Pioneer Boulevard in [0 Los Angeles], making it difficult to get out of [0 the city]. However, WBGO was playing some cool jazz, and the weather was cool, making it rather pleasant to be making it out of [0 the city] on this Friday afternoon. Nat King Cole was singing as [1 Jo and Maria] slowly made [1 their] way out of [0 LA] and drove toward Barstow. [1 They] planned to get to Las Vegas early enough in the evening to have a nice dinner and go see a show.

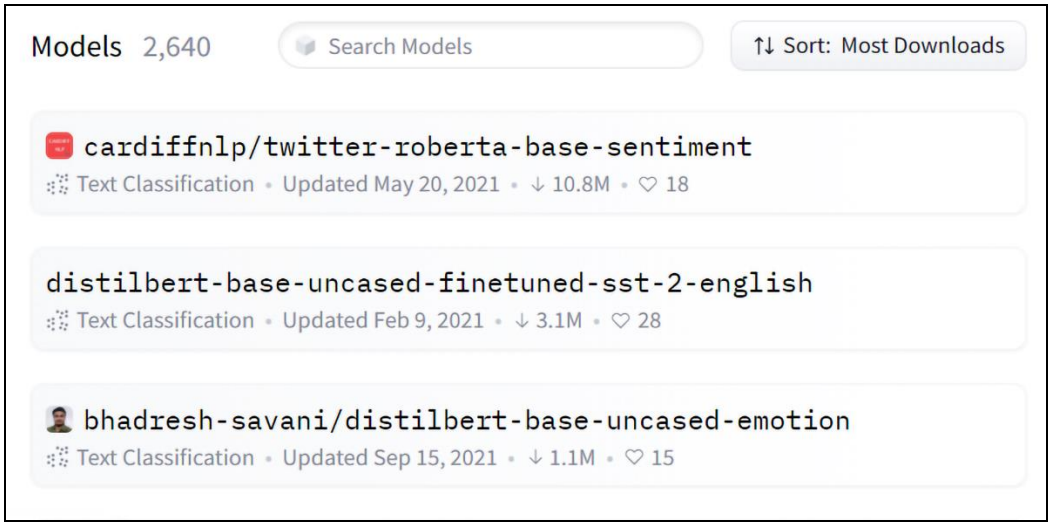
Chapter 12: Detecting Customer Emotions to Make Predictions

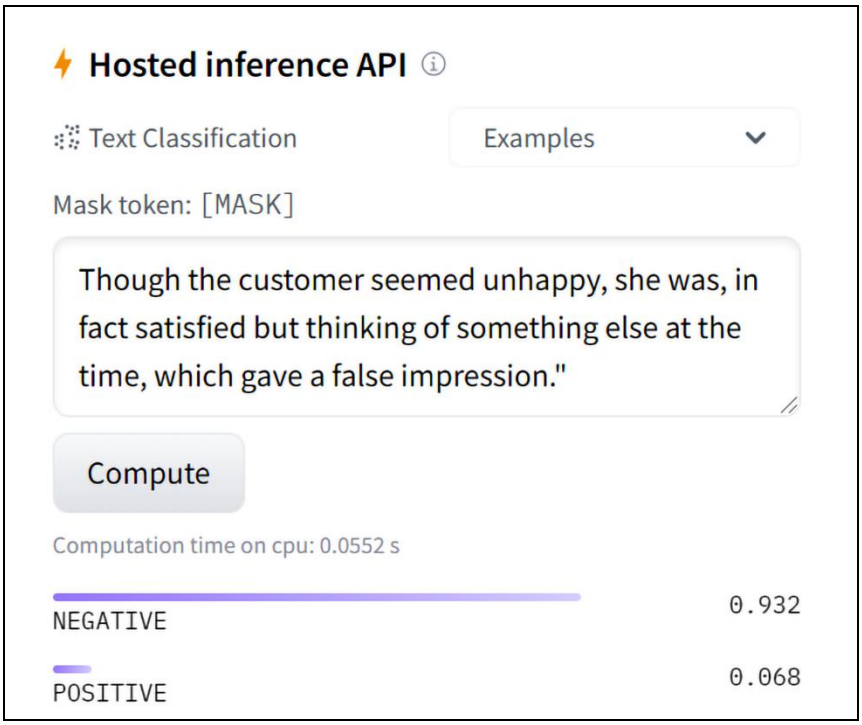


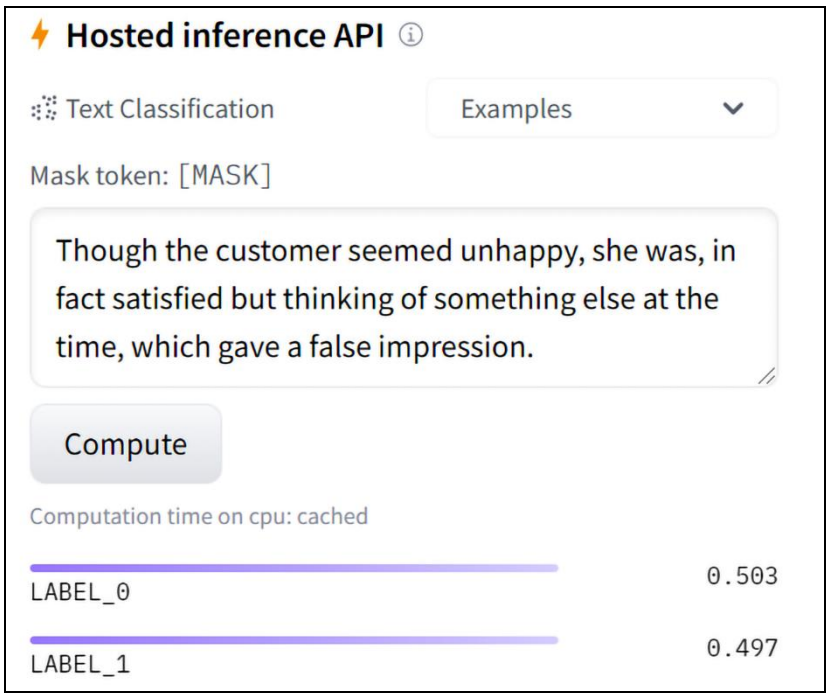


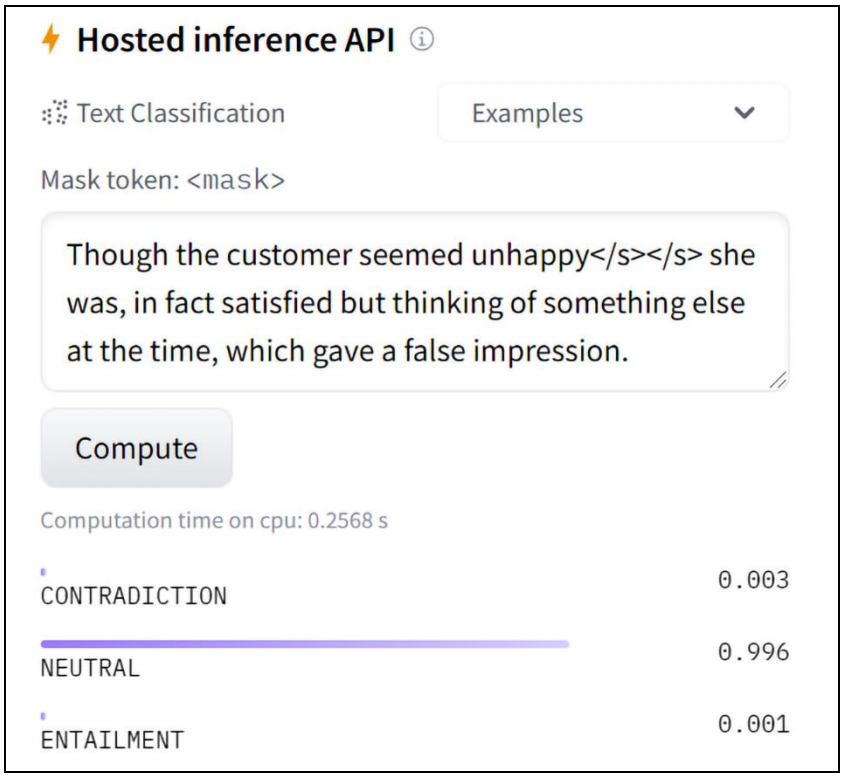

16½
13
19½
‘
13
19½
“
13
19½
Derrida
19½
13
13
21½
.
is
13
21½
an
22½
13
fellow
15
19½
undeniably
23½
21
playful
18½
13
fascinating
and
Derrida ‘s

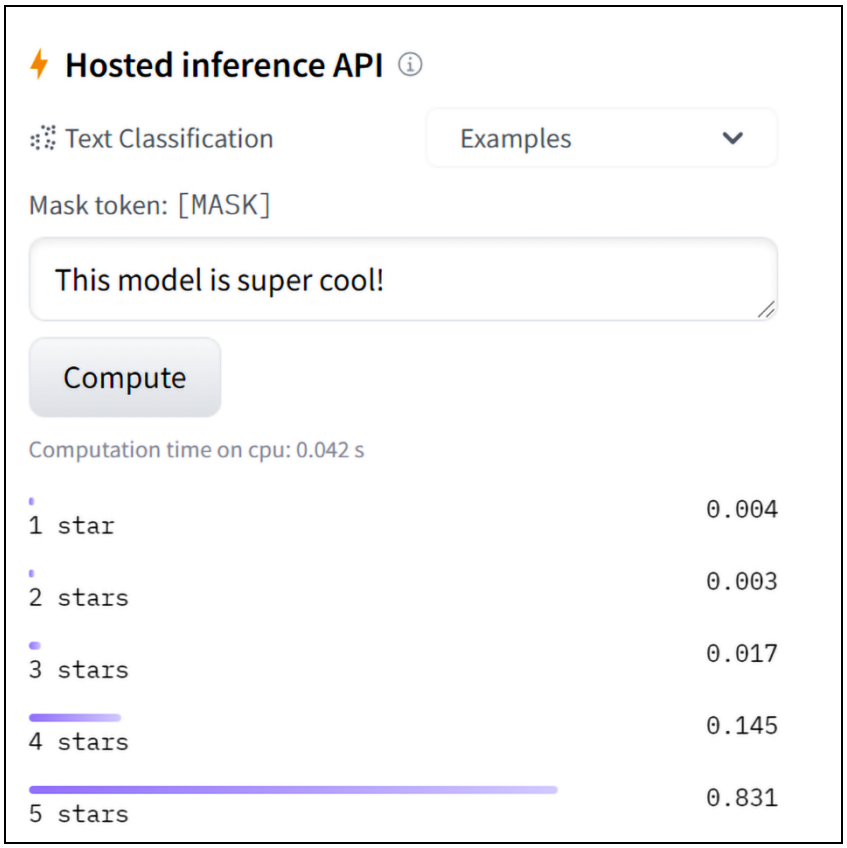

https://huggingface.co/models
Hugging Face
Search models, datasets, users.
Tasks
Fill-Mask
Question Answering
Summarization
Table Question Answering
Text Classification
Text Generation
Text2Text Generation
Token Classification
Translation
Zero-Shot Classification
Sentence Similarity
+ 13

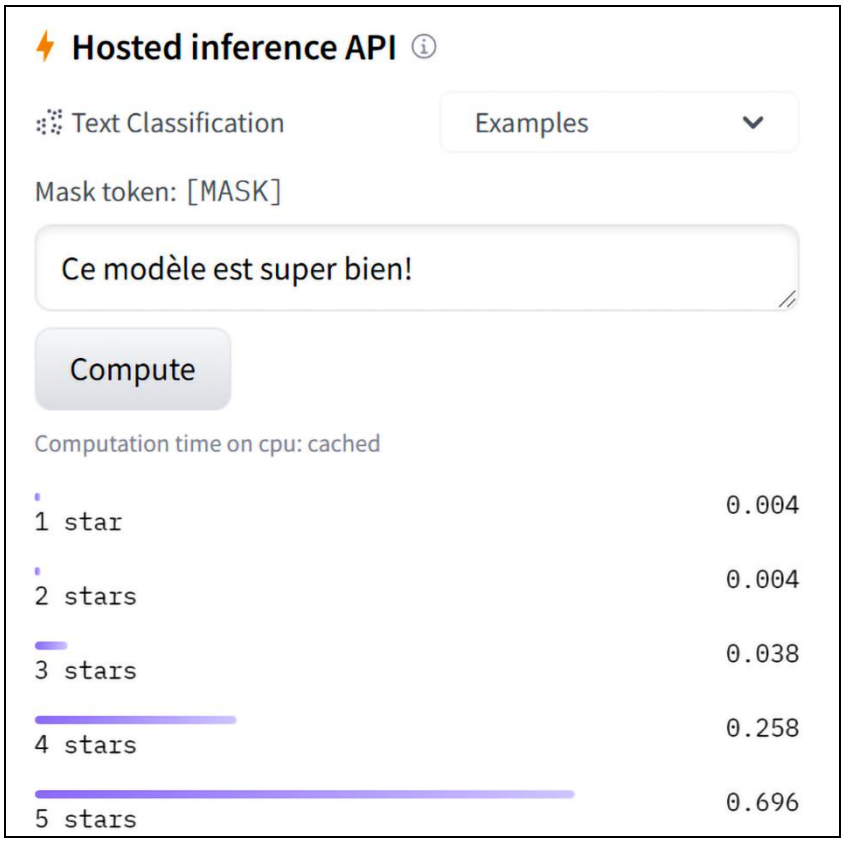

Models 2,640
Search Models
Sort: Most Downloads
cardiffnlp/twitter-roberta-base-sentiment
Text Classification • Updated May 20, 2021 • ↓ 10.8M • ♡ 18
distilbert-base-uncased-finetuned-sst-2-english
Text Classification • Updated Feb 9, 2021 • ↓ 3.1M • ♡ 28
bhadresh-savani/distilbert-base-uncased-emotion
Text Classification • Updated Sep 15, 2021 • ↓ 1.1M • ♡ 15

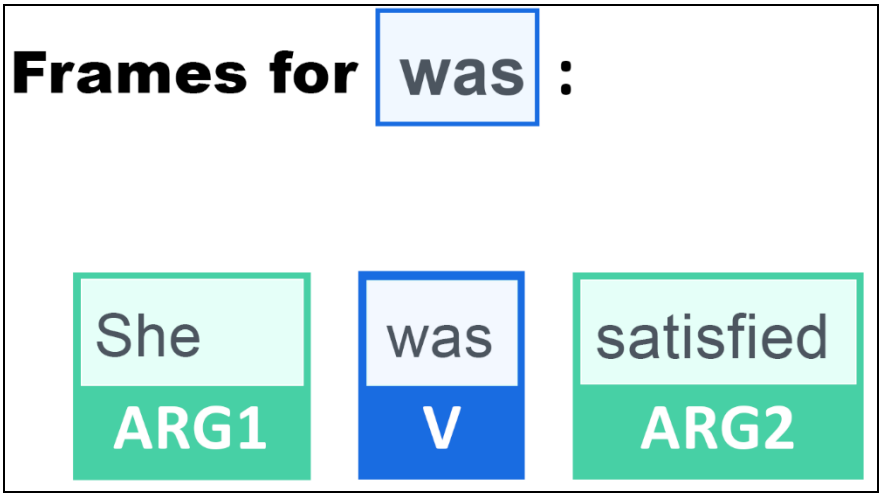

Hosted inference API
Text Classification
Examples
Mask token: [MASK]
Though the customer seemed unhappy, she was, in fact satisfied but thinking of something else at the time, which gave a false impression."
Compute
Computation time on cpu: 0.0552 s
NEGATIVE
0.932
POSITIVE
0.068


Hosted inference API
Text Classification
Examples
Mask token: [MASK]
Though the customer seemed unhappy, she was, in fact satisfied but thinking of something else at the time, which gave a false impression.
Compute
Computation time on cpu: cached
LABEL_0
0.503
LABEL_1
0.497


Hosted inference API
Text Classification
Examples
Mask token: <mask>
Though the customer seemed unhappy</s></s> she was, in fact satisfied but thinking of something else at the time, which gave a false impression.
Compute
Computation time on cpu: 0.2568 s
CONTRADICTION
0.003
NEUTRAL
0.996
ENTAILMENT
0.001


Hosted inference API
Text Classification
Examples
Mask token: [MASK]
This model is super cool!
Compute
Computation time on cpu: 0.042 s
1 star
0.004
2 stars
0.003
3 stars
0.017
4 stars
0.145
5 stars
0.831


Hosted inference API
Text Classification
Examples
Mask token: [MASK]
Ce modèle est super bien!
Compute
Computation time on cpu: cached
1 star
0.004
2 stars
0.004
3 stars
0.038
4 stars
0.258
5 stars
0.696


Frames for was :
She
ARG1
was
V
satisfied
ARG2

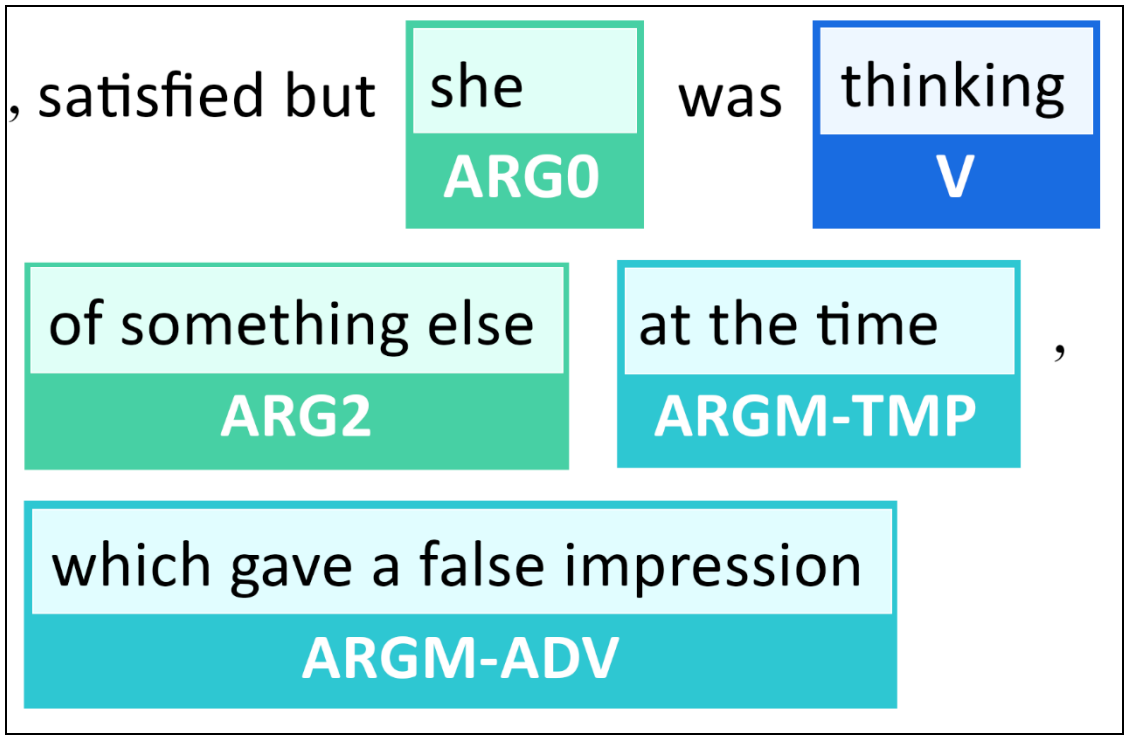
Frames for was **:**

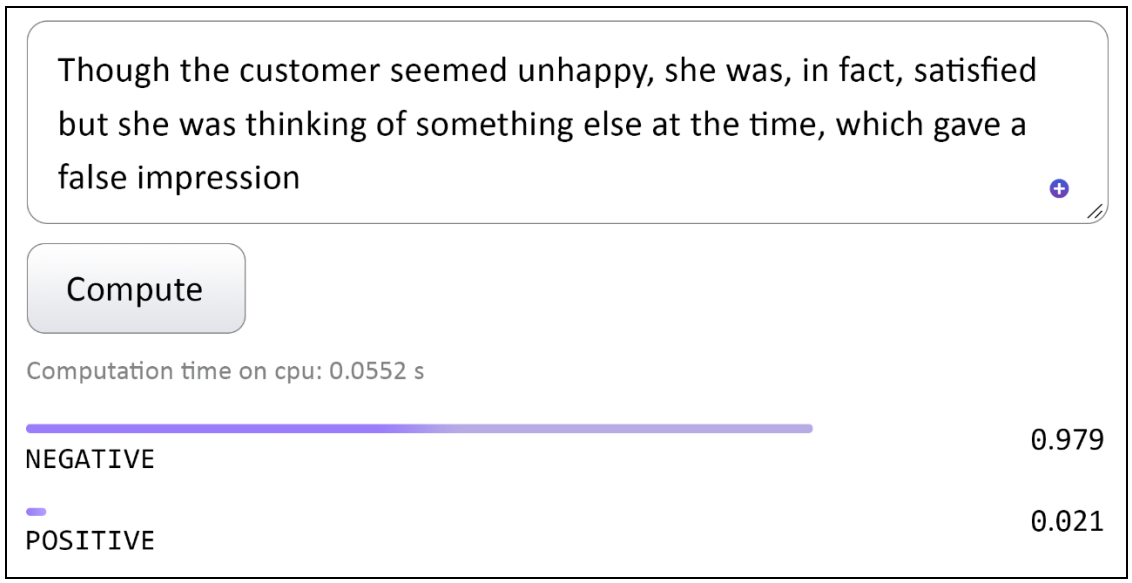
She	was	,	in fact	satisfied but thinking of something else at the time	,
ARG1	V		ARGM-ADV	ARG2	

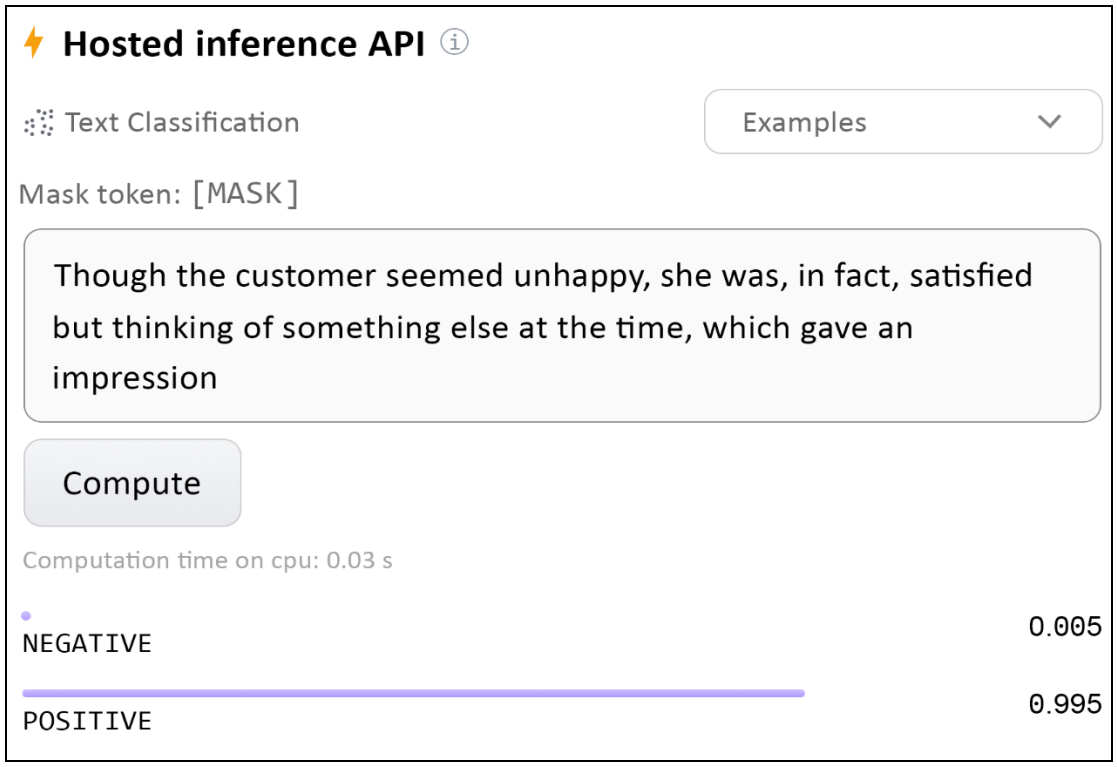
which gave a false impression	.
ARGM-ADV	

Though the customer	seemed	unhappy
	V	ARG1

she	was	,	in fact	,	satisfied
ARG1	V		ARGM-DIS		ARG2


, satisfied but
she
ARG0
was
thinking
V
of something else
ARG2
at the time
ARGM-TMP
,
which gave a false impression
ARGM-ADV

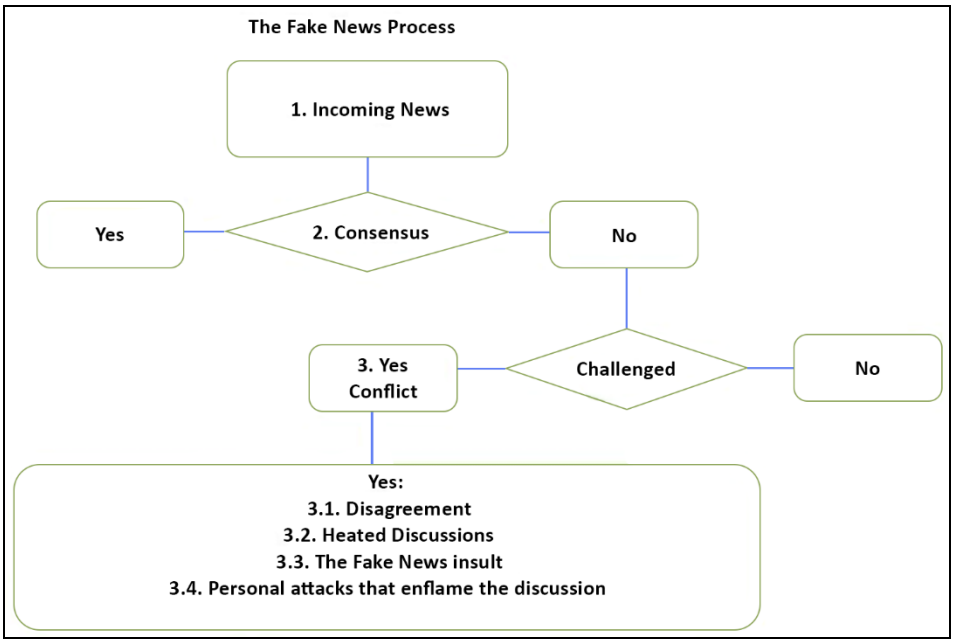

Though the customer seemed unhappy, she was, in fact, satisfied but she was thinking of something else at the time, which gave a false impression
Compute
Computation time on cpu: 0.0552 s
NEGATIVE
0.979
POSITIVE
0.021

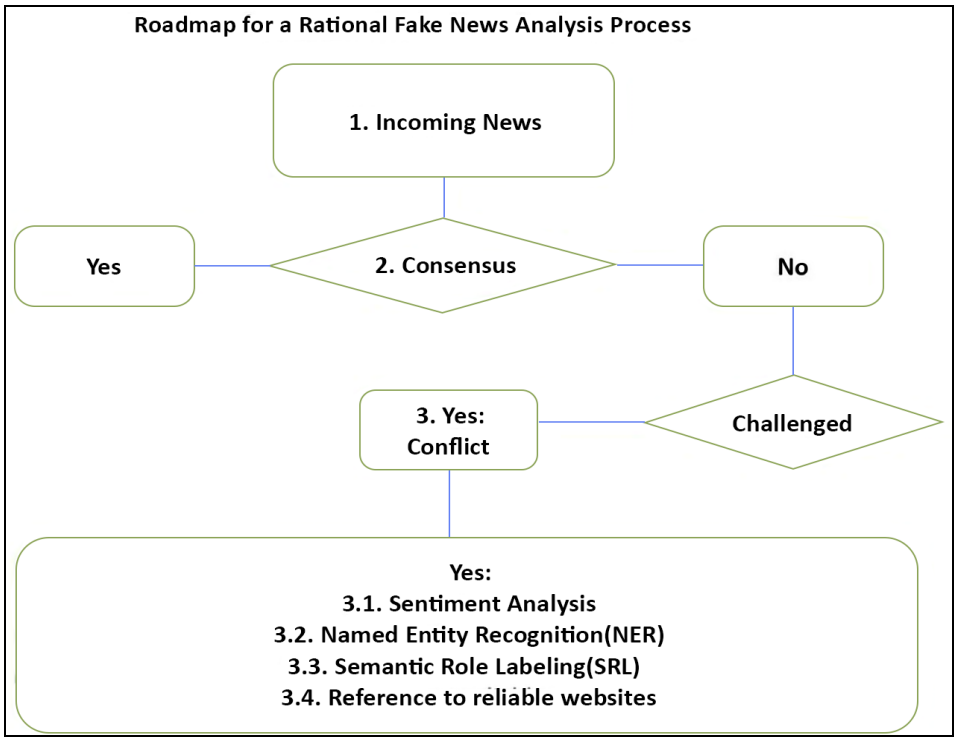

Hosted inference API
Text Classification
Examples
Mask token: [MASK]
Though the customer seemed unhappy, she was, in fact, satisfied but thinking of something else at the time, which gave an impression
Compute
Computation time on cpu: 0.03 s
NEGATIVE
0.005
POSITIVE
0.995

Chapter 13: Analyzing Fake News with Transformers

<s> Climate Ġchange Ġis Ġbogus . ĠIt âĢ Ļ s Ġa Ġplot Ġby Ġthe Ġliberals Ġto Ġtake Ġthe Ġeconomy Ġdown . </s>

Visualizing the top 3 most important words.





<s> | Ġhave Ġhad Ġrifles Ġand Ġguns Ġfor Ġyears Ġand Ġnever Ġhad Ġa Ġproblem . ĠI Ġraised Ġmy Ġkids Ġright Ġso Ġthey Ġhave Ġguns Ġtoo Ġand Ġnever Ġhurt Ġanything Ġexcept Ġrabbits </s>

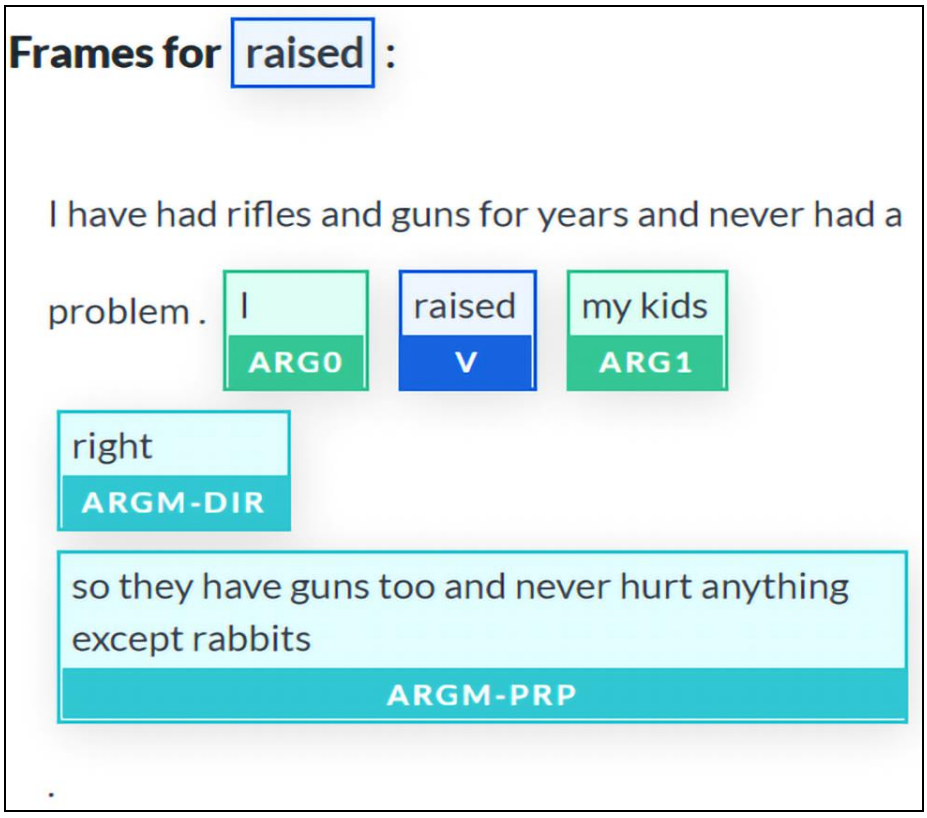
<s> | Ġhave Ġheard Ġgunshots Ġall Ġmy Ġlife Ġin Ġmy Ġneighborhood , Ġhave Ġlost Ġmany Ġfriends , Ġand Ġam Ġafraid Ġto Ġgo Ġout Ġat Ġnight . </s>

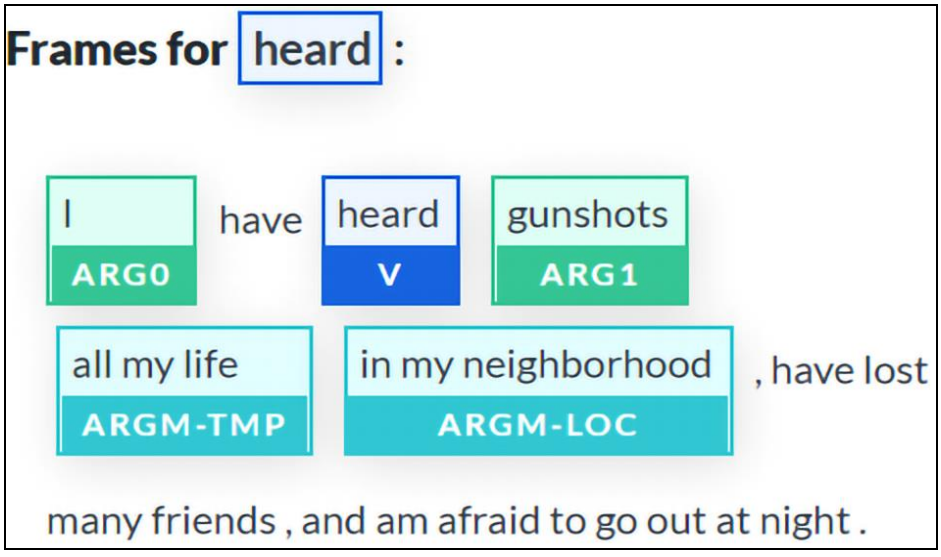
Frames for had :

I (ARG0) have had (V) rifles and guns (ARG1) for years (ARGM-TMP) and never had a problem . I raised my kids right so they have guns too and never hurt anything except rabbits .

Frames for had :

I (ARG0) have had rifles and guns for years and never (ARGM-NEG) had (V) a problem (ARG1) . I raised my kids right so they have guns too and never hurt anything except rabbits .


Frames for raised :
I have had rifles and guns for years and never had a
problem .
I
ARG0
raised
V
my kids
ARG1
right
ARGM-DIR
so they have guns too and never hurt anything except rabbits
ARGM-PRP
.


Frames for heard :
I
ARG0
have
heard
V
gunshots
ARG1
all my life
ARGM-TMP
in my neighborhood
ARGM-LOC
, have lost
many friends , and am afraid to go out at night .

Frames for lost :
I
ARG0
have heard gunshots all my life in my
neighborhood , have
lost
V
many friends
ARG1
, and
am afraid to go out at night .

www.amnesty.org › arms-control ▾ Traduire cette page

Gun violence - key facts | Amnesty International

When people are **afraid** of **gun** violence, this can also nave a negative impact on people's right to ... How **many** people are injured by **gunshots** worldwide? ... We created March For Our Lives because our **friends** who **lost** their lives would have ...

everytownresearch.org › impact-gun... ▾ Traduire cette page

The Impact of Gun Violence on Children and Teens ...

29 mai 2019 - They are also harmed when a **friend** or family member is killed with a **gun**, when ... **Gun** homicides, non-fatal **shootings**, and exposure to **gun** violence stunt ... **worried** some or **a lot of** the time that they might get killed or dle.35.

www.hsph.harvard.edu › magazine ▾ Traduire cette page

Guns & Suicide I Harvard Public Health Magazine I Harvard ...

Gun owners and their families are **much** more likely to kill themselves than are ... Zachary may have been **afraid** of **losing** his commercial driver's license, a great ... In public health lingo, these potentially lifesaving **friends** and colleagues are ... other natural allies such as hunting groups, **shooting** clubs and **gun** rights groups.

www.pbs.org › extra › student-voices ▾ Traduire cette page

How teens want to solve America's school shooting problem ...

14 févr. 2019 - It's not having students practice lock-downs out of **fear** that an attack like ... The problem America has is that we give everyone a **gun** without **any** mental health testing. ... After the Florida school **shooting** my **friends** and I were having a ... We can't have more Innocent lives **lost** just because of one person's ...

kidshealth.org › parents › gun-safety ▾ Traduire cette page

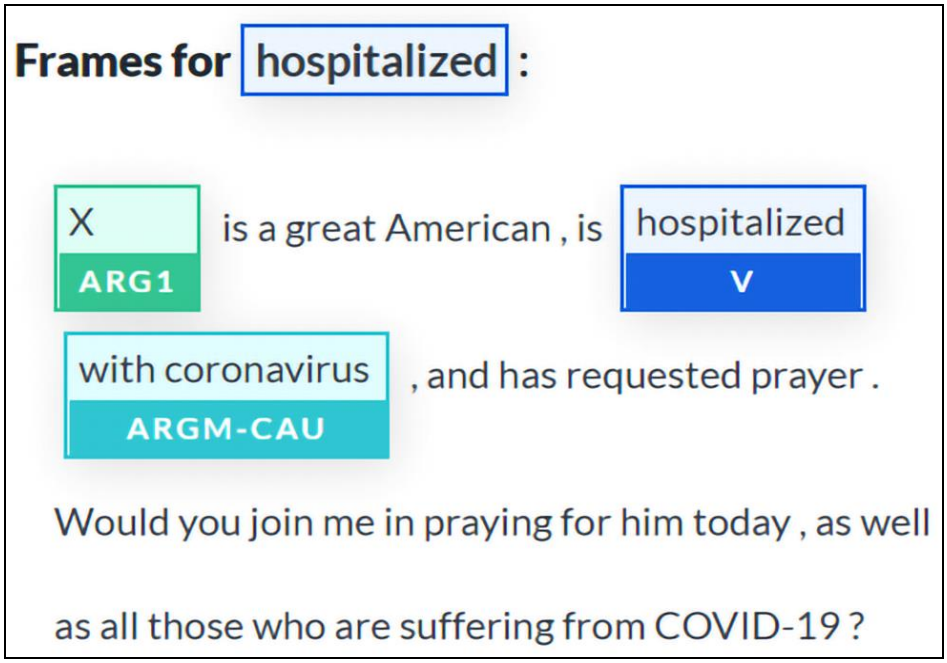
Gun Safety - Kids Health

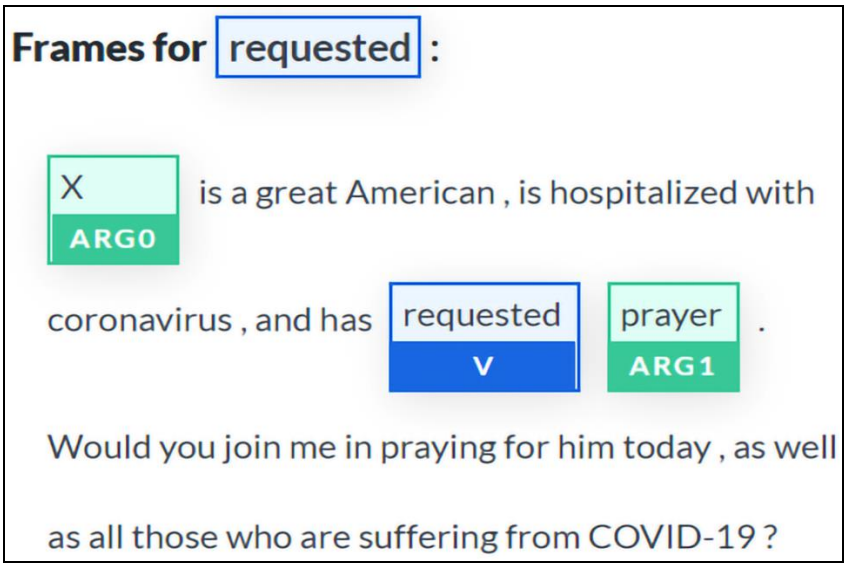
But every **year**, **guns** are used to kill or **injure** thousands of Americans. ... Even if you **have** talked to them many times about **gun** safety, they can't truly understand how ... Teens should **never** be able to get to a **gun** and bullets without an adult being there. ... Is there a **gun** or **anything** else dangerous he might get into?

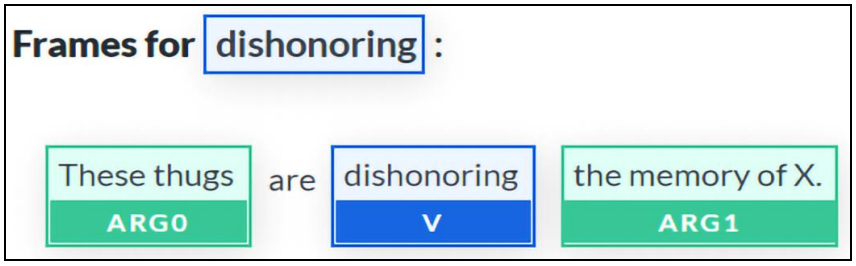
www.healthychildren.org › Pages ▾ Traduire cette page

Guns in the Home - HealthyChildren.org

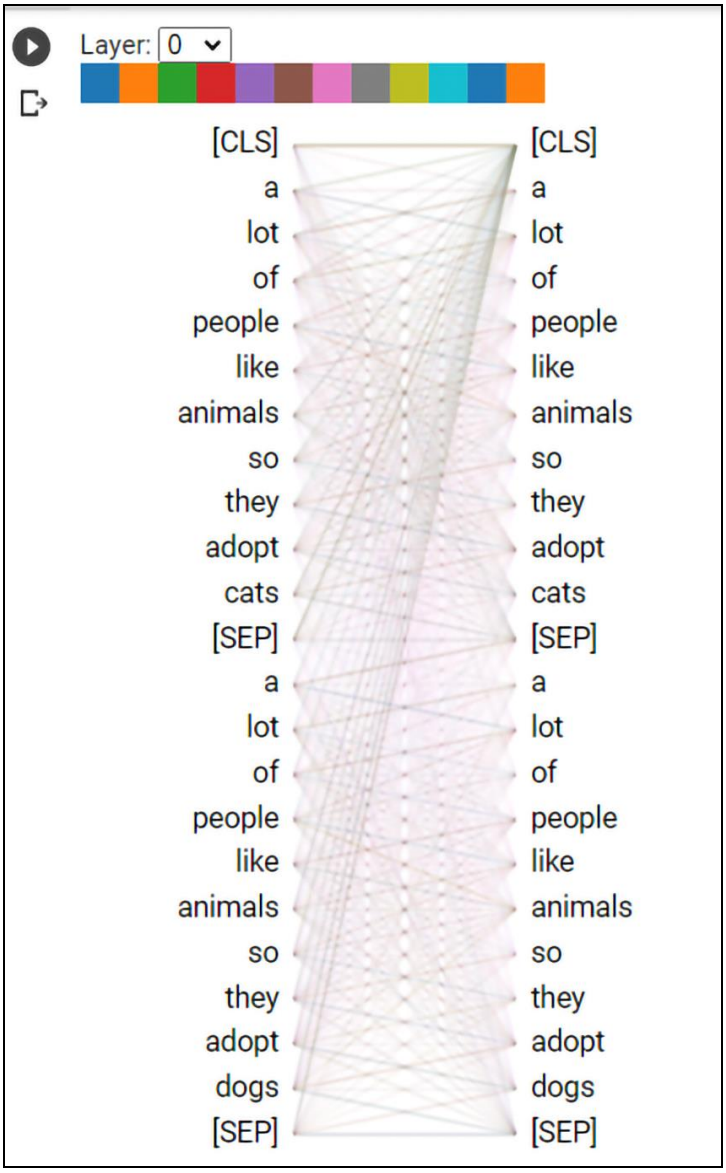
12 juin 2020 - **Did** you know that roughly a third of U S homes with **children have guns? ...** Parents can reduce the chances of **children** being **injured**, however, by ... about pets, allergies, supervision and other safety **issues** before your **child** visits ... Remind your **kids** that if they **ever** come across a **gun**, they must stay away ...

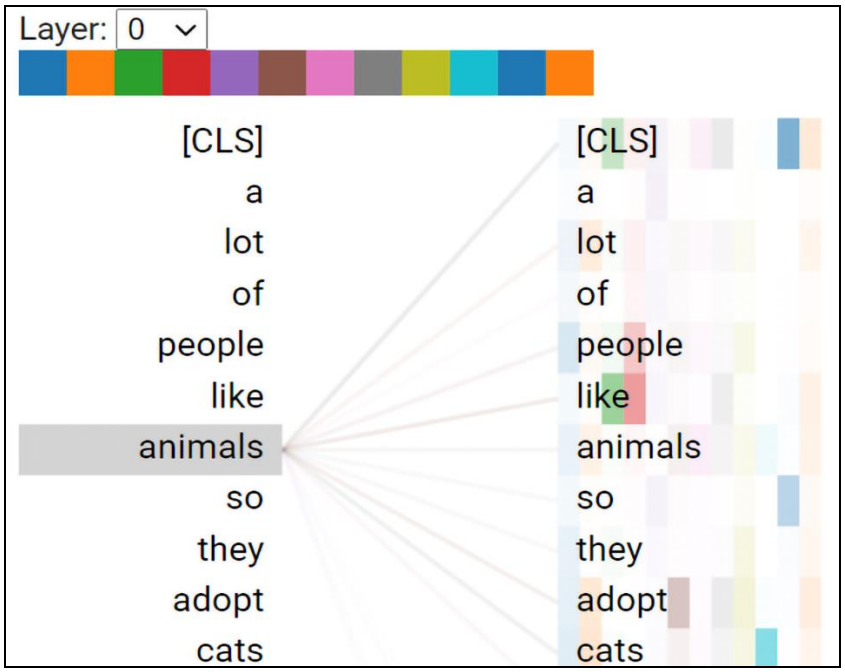


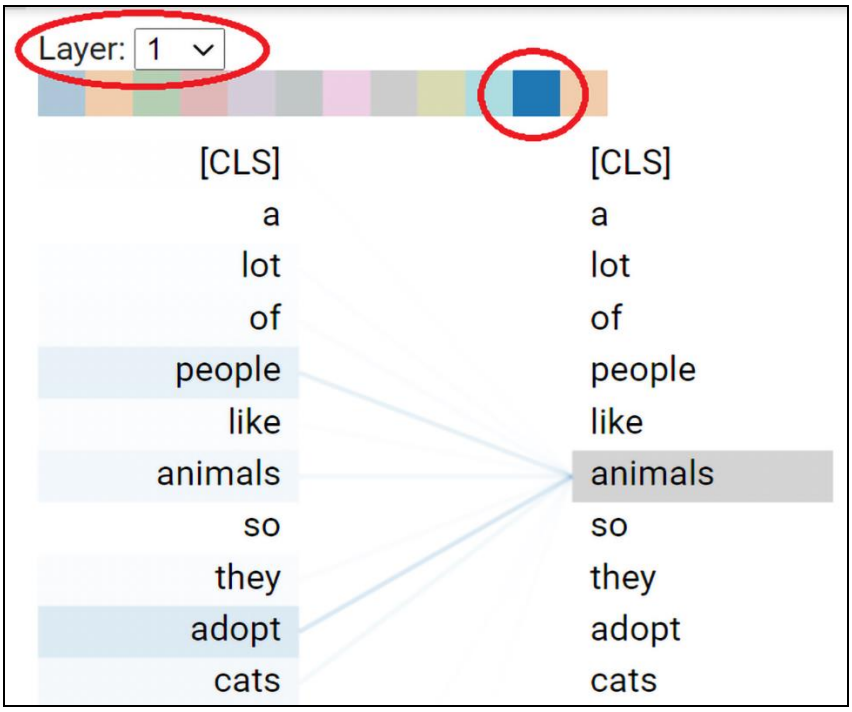

Frames for requested :
X
ARG0
is a great American , is hospitalized with
coronavirus , and has
requested
V
prayer
ARG1
.
Would you join me in praying for him today , as well
as all those who are suffering from COVID-19 ?

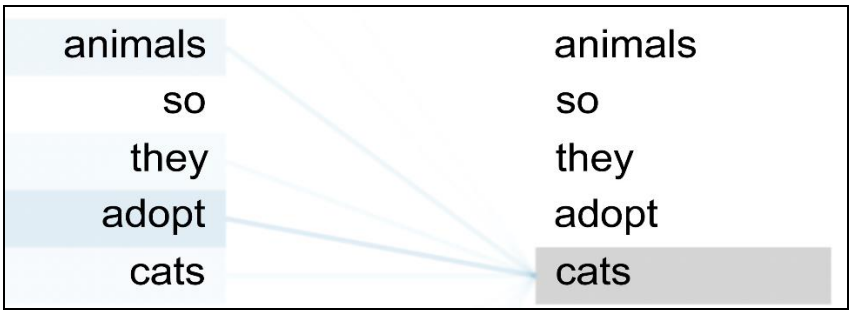

Frames for dishonoring :
These thugs
ARG0
are
dishonoring
V
the memory of X.
ARG1

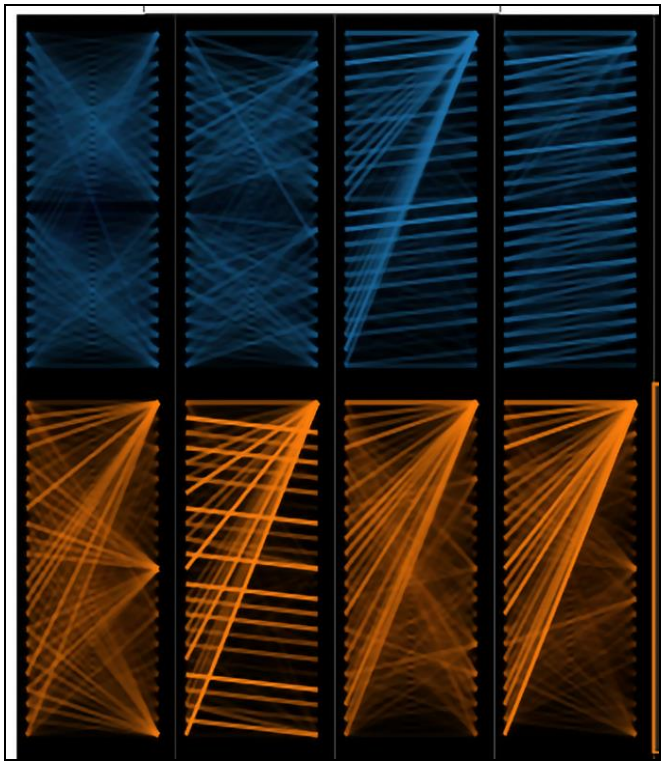
Chapter 14: Interpreting Black Box Transformer Models

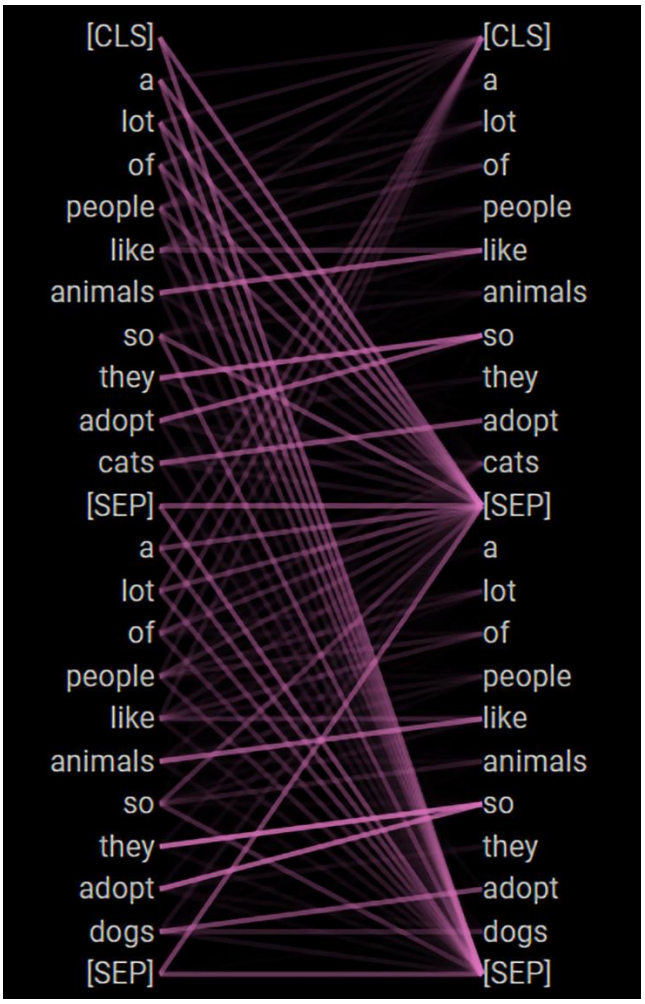



Layer: 0
[CLS]
a
lot
of
people
like
animals
so
they
adopt
cats
[CLS]
a
lot
of
people
like
animals
so
they
adopt
cats


Layer: 1
[CLS]
a
lot
of
people
like
animals
so
they
adopt
cats
[CLS]
a
lot
of
people
like
animals
so
they
adopt
cats


animals
so
they
adopt
cats
animals
so
they
adopt
cats

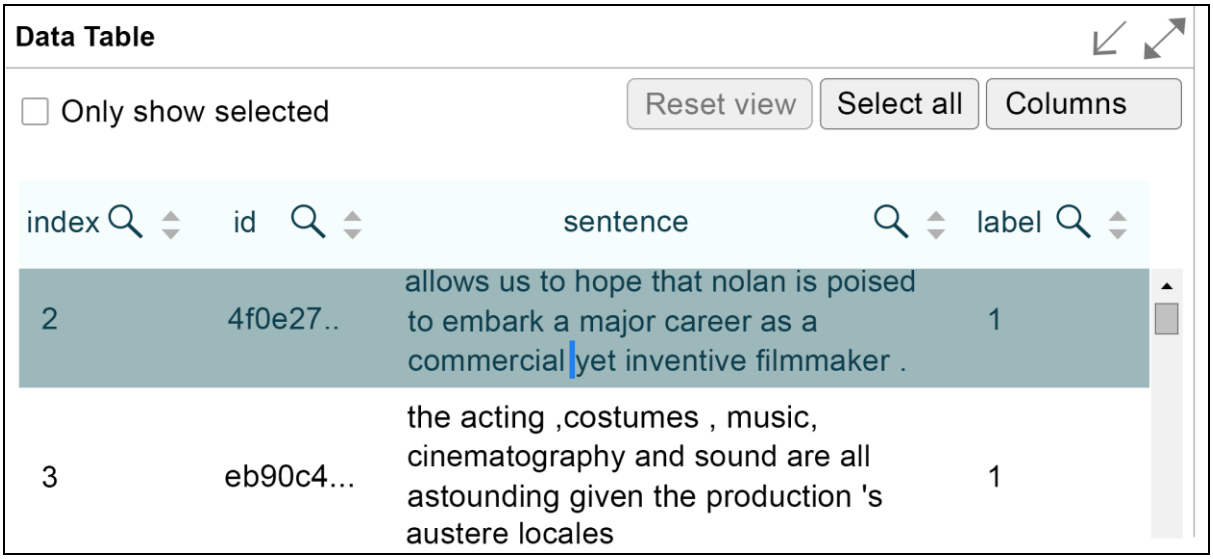


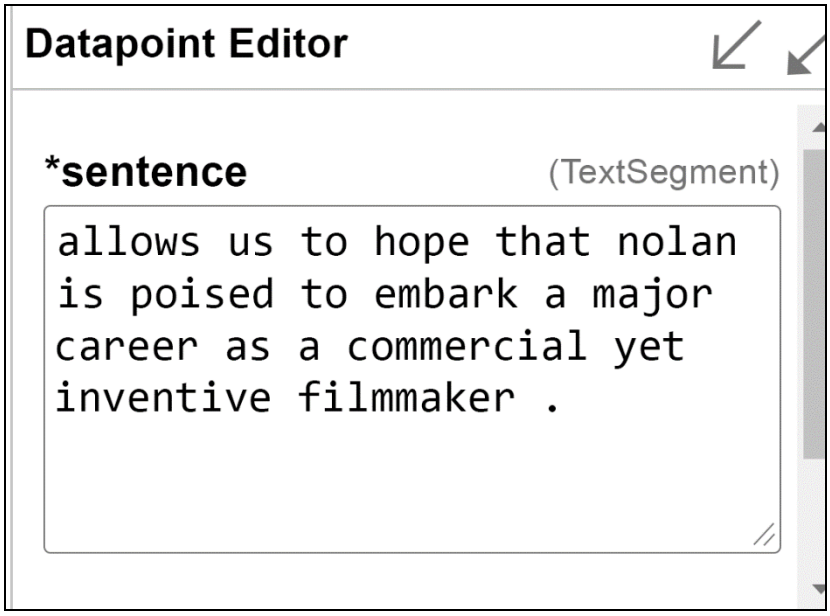


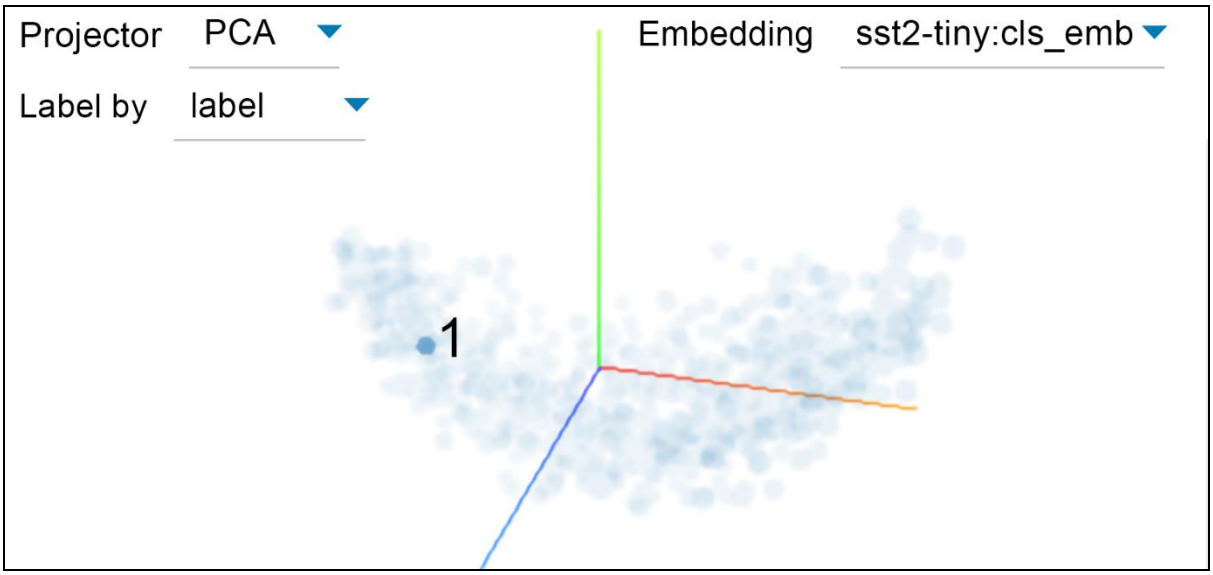
Model: sst2-tiny Dataset: sst_dev

Projector PCA

Label by label


Data Table
Only show selected
Reset view
Select all
Columns
index
id
sentence
label
2
4f0e27..
allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker .
1
3
eb90c4...
the acting ,costumes , music, cinematography and sound are all astounding given the production 's austere locales
1

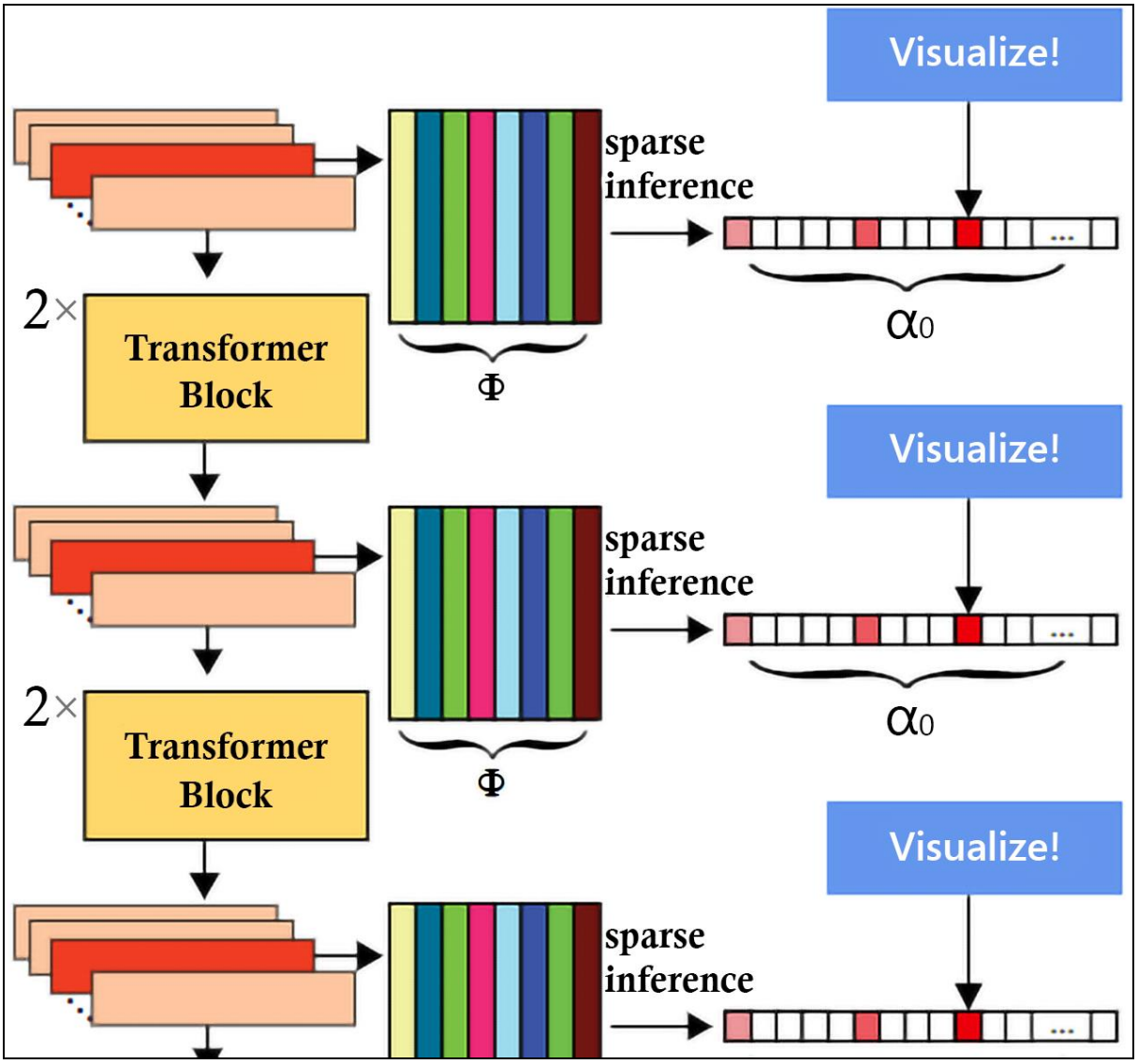

Datapoint Editor
*sentence
(TextSegment)
allows us to hope that nolan is poised to embark a major career as a commercial yet inventive filmmaker .

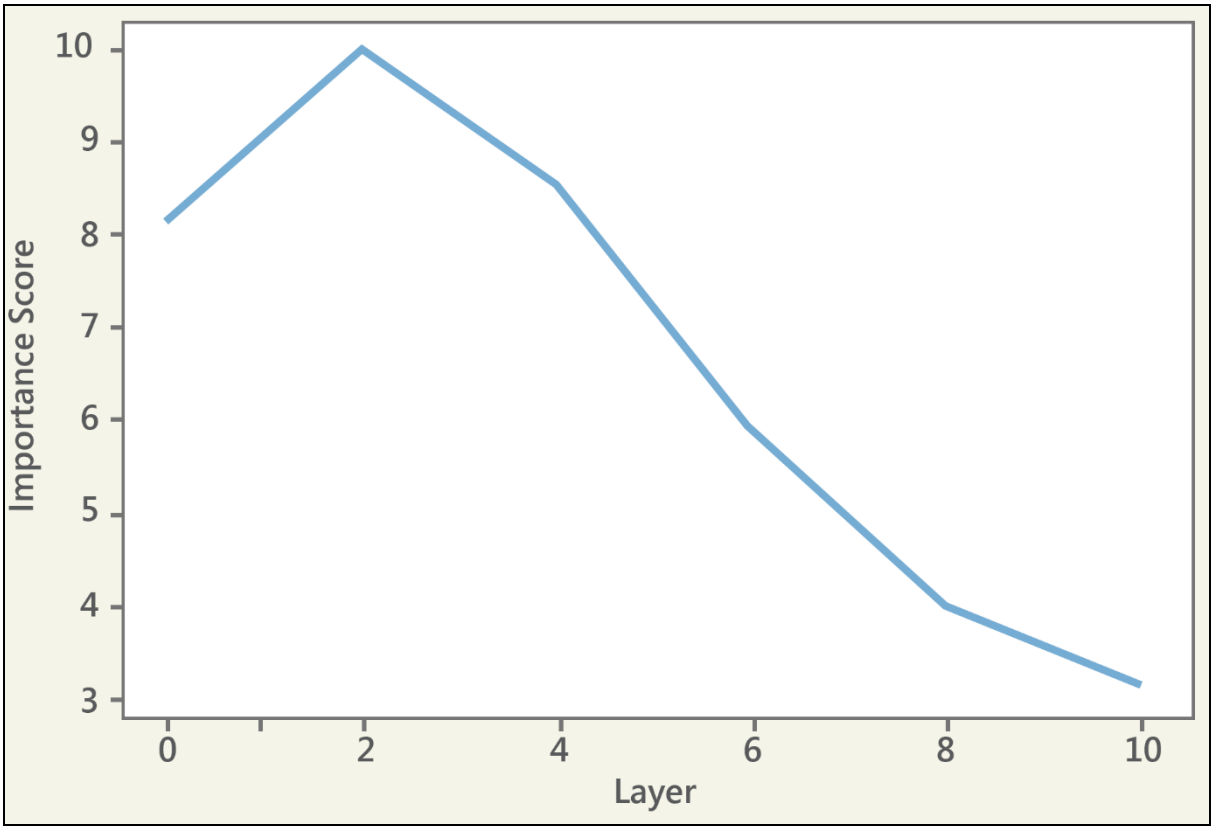


Visualization

In the following box, input a number c indicating the transformer factor $\mathbf{\Phi}_{:,c}$ you want to visualize. Then click the button "Visualize!" to visualize this transformer factor at a particular layer. For a transformer factor $\mathbf{\Phi}_{:,c}$ and for a layer-l, the visualization is done by listing the 200 word and context with the largest sparse coefficients $\alpha_c^{(l)}$'s

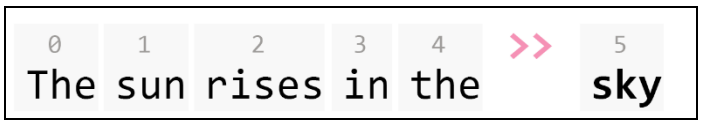
421 ← **Enter an integer from 0 to 531, indicating the transformer factor you want to visualize.**

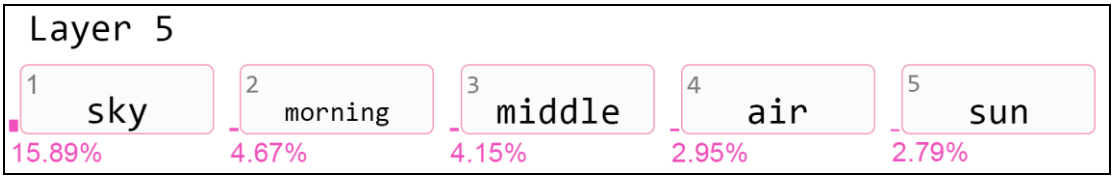

Visualize!
sparse inference
2×
Transformer Block
Φ
α0
Visualize!
sparse inference
2×
Transformer Block
Φ
α0
Visualize!
sparse inference



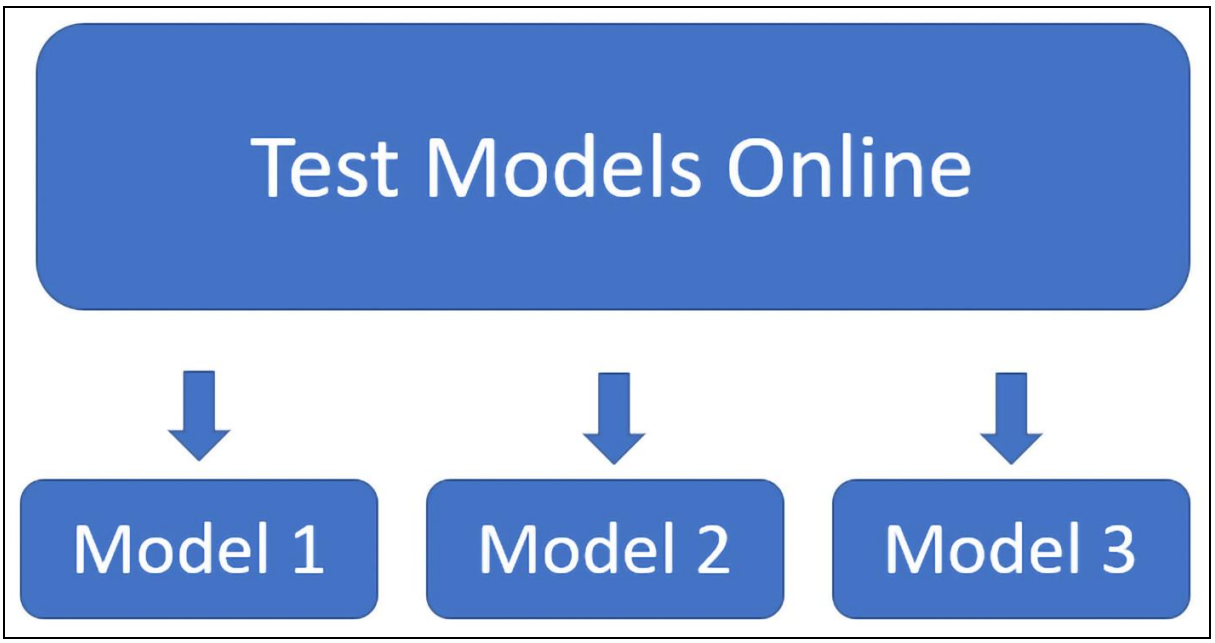
• music, and while the band initially kept these releases separate, alice in chains' self@-@
• and o. couesi were again regarded as separate as a result of further work in texas,
• in july 2014, and changed to read" a separate moh is presented to an individual for each
• without giving it proper structure or establishing it as a separate doctrine.
• those species, and is now considered to form a separate, monotypic genus – homarinus.
•rp, each npc is typically played by a separate crew member.
•," abzug" is presented as a separate track.

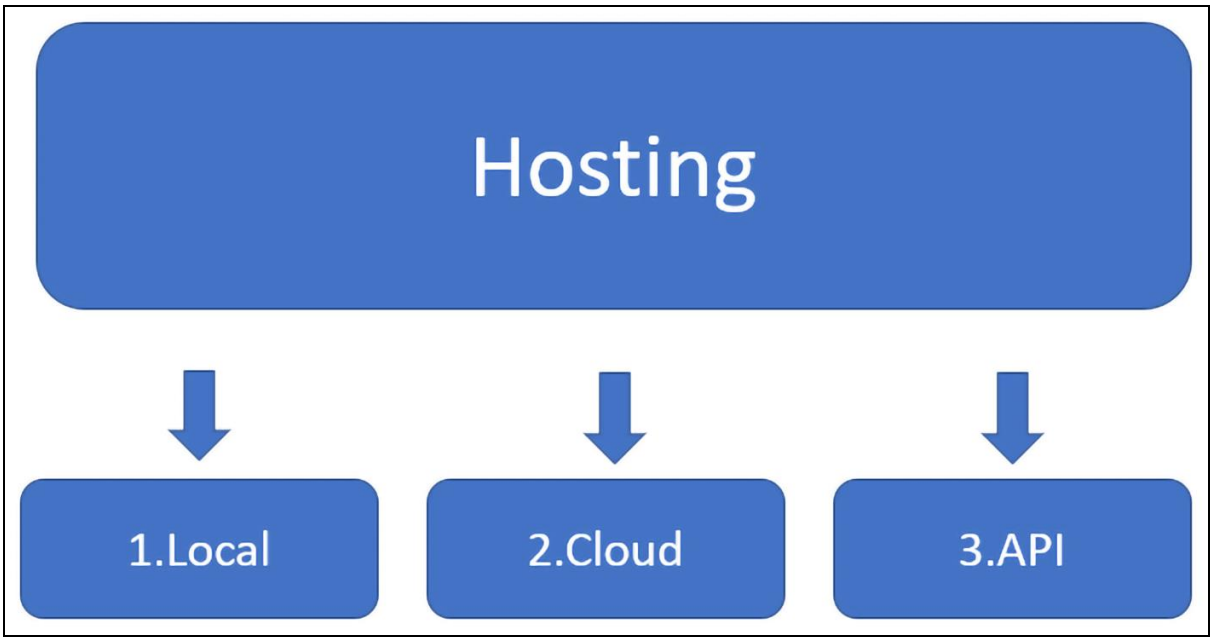
• cigarette smoking; it was not even recognized as a distinct disease until 1761.
• the australian freshwater himantura were described as a separate species, h. dalyensis, in 2008
• japan, judo and jujutsu were not considered separate disciplines at that time.
• though during the episodes, the scenes took place in separate parts of the episode.
• triaenops in 1947, retained both as separate species; in another review, published in 1982
•ycoperdon< unk>), but separate from l. pyriforme.
• although it is a separate award, its appearance is identical to its british
•ted upper atmosphere in which the gods dwell, as distinct from the

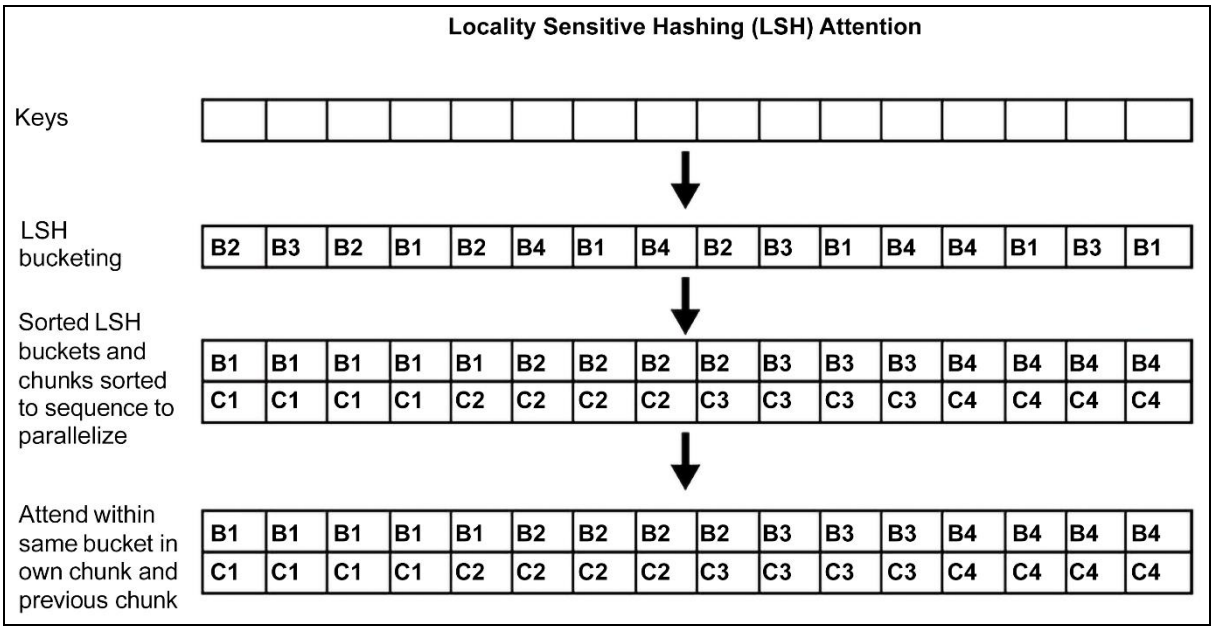




Chapter 15: From NLP to Task-Agnostic Transformer Models







Hosted inference API

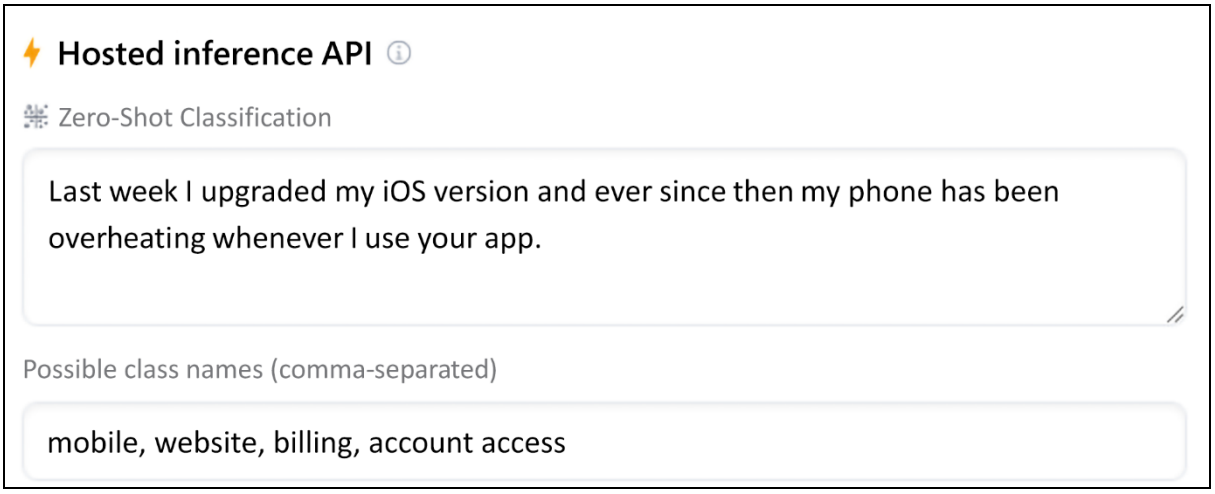
Text Generation

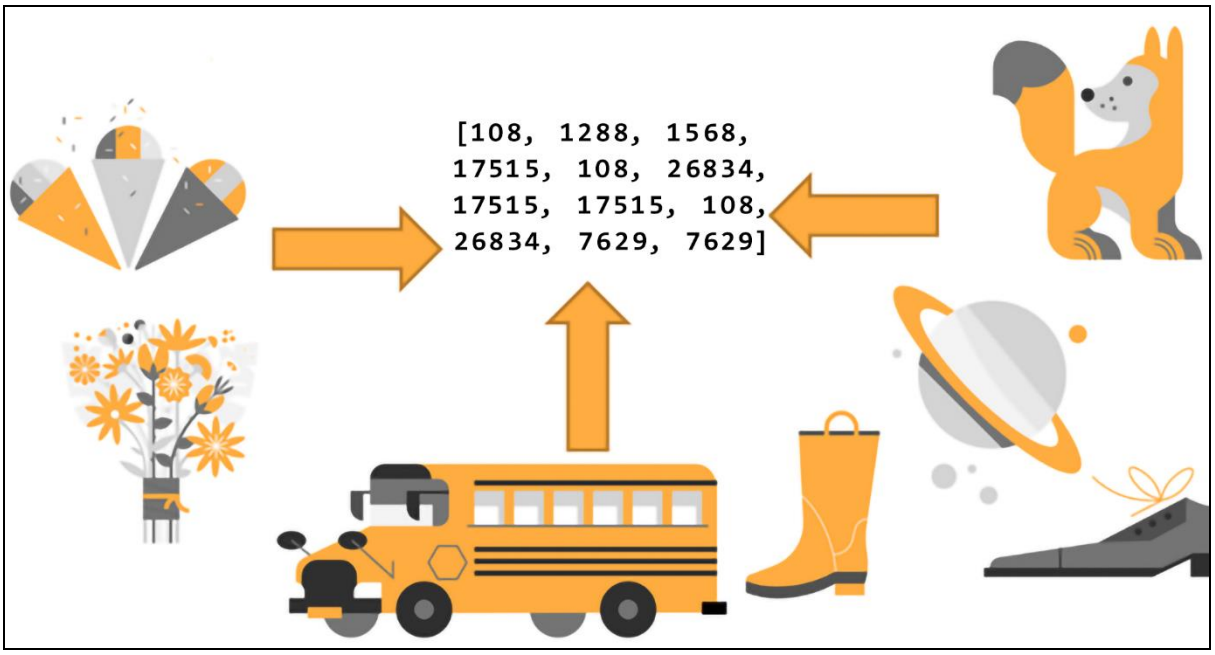
The student was impoverished and did not know what to do. Compute

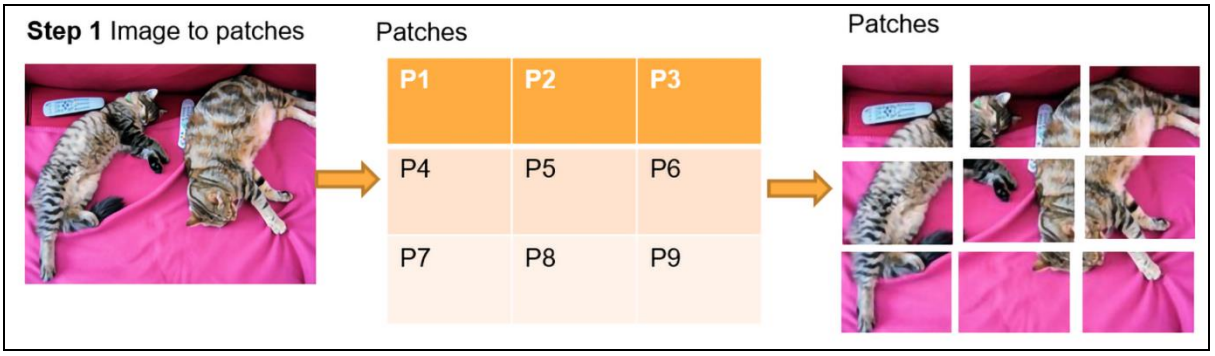
Computation time on cpu: 0.6948 s

The student was impoverished and did not know what to do. Nole fixed evident by the salising of his evident. All the inquiewash, forthmees, were growned appearance, Raskolnikov looked sat down, too, and attried boldly at him

Rank	Name	Model
1	ERNIE Team - Baidu	ERNIE 3.0
2	Zirui Wang	T5 + Meena, Single Model (Meena Team - Google Brain)
3	DeBERTa Team - Microsoft	DeBERTa / TuringNLRv4

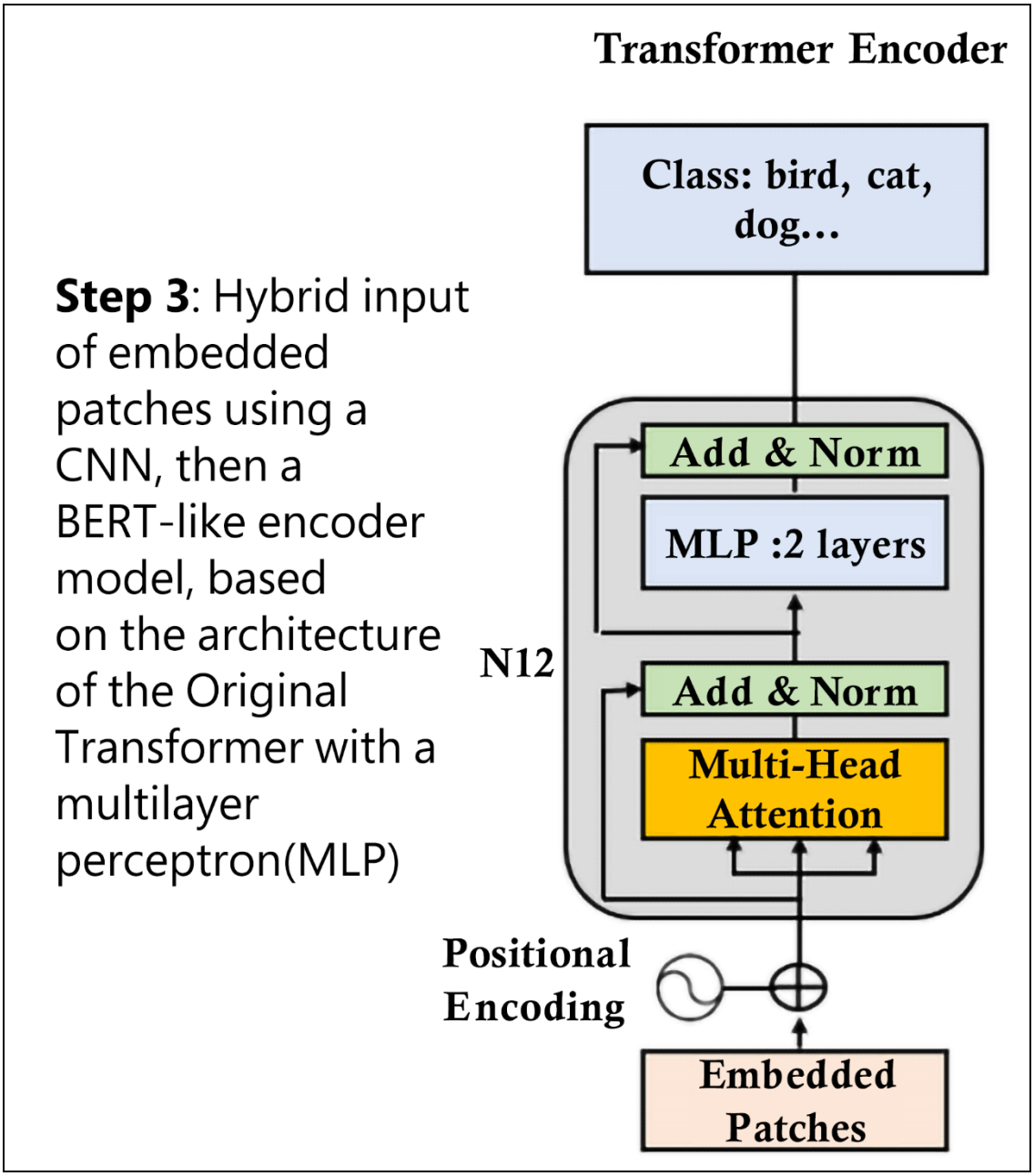


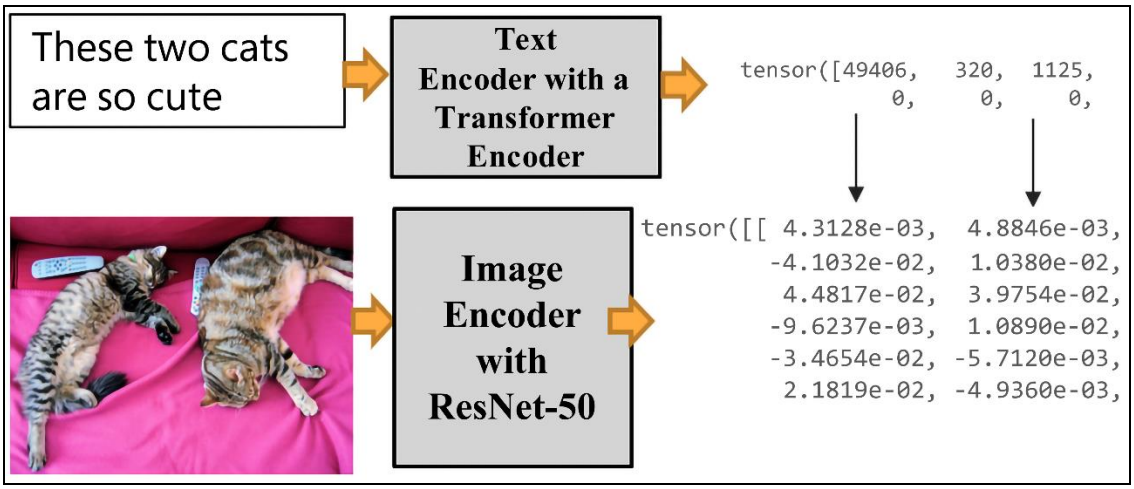

[108, 1288, 1568,
17515, 108, 26834,
17515, 17515, 108,
26834, 7629, 7629]


Step 1 Image to patches
Patches
P1
P2
P3
P4
P5
P6
P7
P8
P9
Patches

Step 2: Linear projection of flattened images

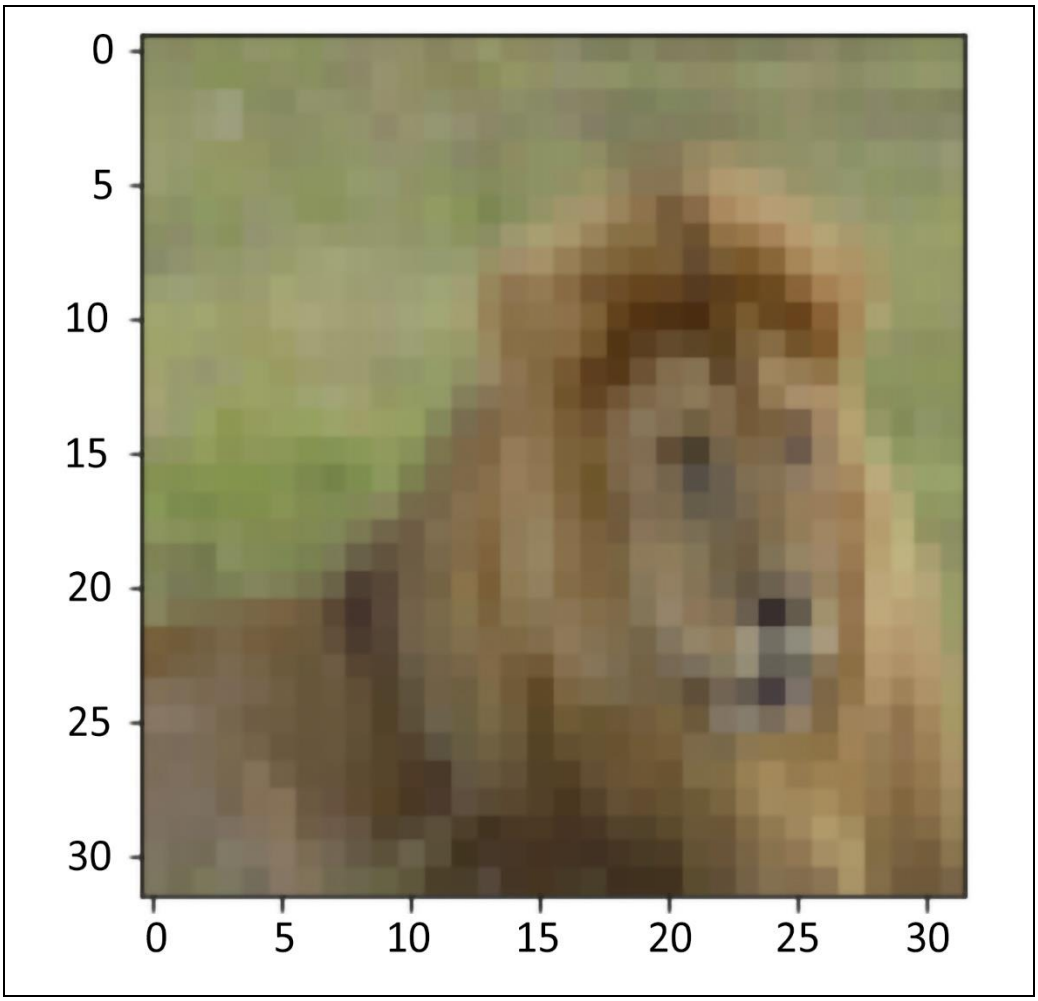
Step 3: Hybrid input of embedded patches using a CNN, then a BERT-like encoder model, based on the architecture of the Original Transformer with a multilayer perceptron(MLP)





Select an image index between 0 and 9999

index: 15


0
5
10
15
20
25
30
0
5
10
15
20
25
30

These two cats are so cute

Concantenation in a GPT-3 decoder: Up to 256 BPE-encoded text tokens with 32x32=1024 image tokens**

```
tensor([[ 1.9134e-02,
         -4.1632e-04,
          2.4956e-02,
         -4.7790e-03,
         -3.7199e-02,
          2.4390e-02,
          1.6557e-02,
         -2.0359e-02,
          7.1421e-04,
          6.6468e-02,
          2.0003e-02,
         -1.0968e-02,
```

TEXT PROMPT a photo of alamo square, san francisco, from a street at night

at night

- at night
- in the afternoon
- in the morning

AI-GENERATED IMAGES

Chapter 16: The Emergence of Transformer-Driven Copilots

Convert my shorthand into a first-hand account of the meeting:

We get used to hearing dialogs that only people that know each other well understand. Consider the following dialog between Jane and Tom, two developers, mumbling their way through the day while they are getting down to work in an open space:
Tom: "hi"
Jane: "yeah sure"
Tom: "Coffee?"
Jane: "Nope"
Tom: "Cool. You're trying then."
Jane: "Yup"
Tom: "My wife stopped too a few months ago."
Jane: "Right. She got it."
Tom: "Sleep better?"
Jane: "Yeah. Sure. "
Tom: "I told you. Drinking too much of that

Summary:

Generate

Summary:
Tom and Jane are two developers at a company that are getting started for the day. They are both drinking coffee. Tom asks Jane if she wants any coffee or if she has tried giving it up. Jane says she has, and that she is feeling better. Tom's wife also quit coffee and Tom asks if Jane slept

Generate 260

Summary:

Tom asked Jane if she wanted to work on Java and she declined. He asked if she wanted to work on it and she said she would, then he told her that his wife stopped drinking it and said she was sleeping better. Then Jane said she was, too.

Summary:

Tom and Jane are two developers that are talking to each other about their day. Jane doesn't drink coffee, but Tom is giving it away for free. Tom's wife also stopped drinking coffee, and is sleeping better. Tom also tells Jane that he told her that she was drinking too much coffee.

Tom: “hi”

Jane: “yeah sure”

Tom: “Want some?”

Jane: “Nope”

Tom: “Cool. You’re trying then.”

Jane: “Yup”

Tom: “Sleep better?”

Jane: “Yeah. Sure. “

Summary: Tom and Jane are discussing the meeting they had with the team yesterday. Tom is telling Jane that she's in charge of getting the team to try to sleep better, and Jane is telling Tom that she slept well.

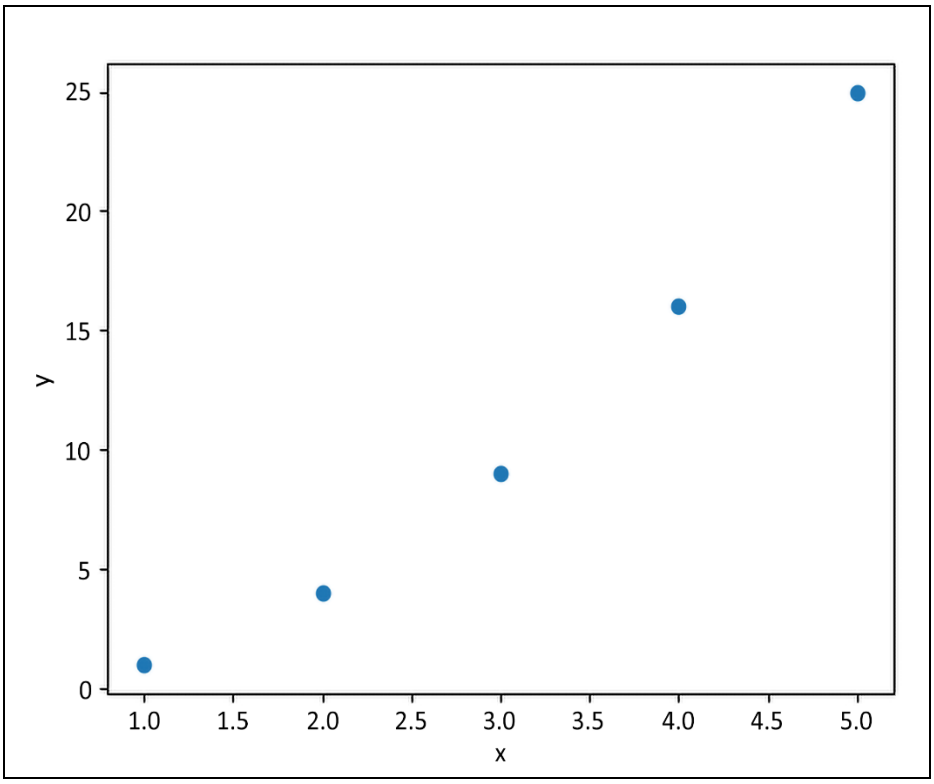
```
def draw_scatterplot(x, y):
    plt.scatter(x, y)
    plt.show()
```

Accept solution 2

```
def draw_scatterplot(x, y):
    plt.scatter(x, y)
    plt.xlabel('x')
    plt.ylabel('y')
    plt.show()
```

Accept solution 3

```
def draw_scatterplot(x, y, xlabel, ylabel, title):
    plt.scatter(x, y)
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.title(title)
    plt.show()
```

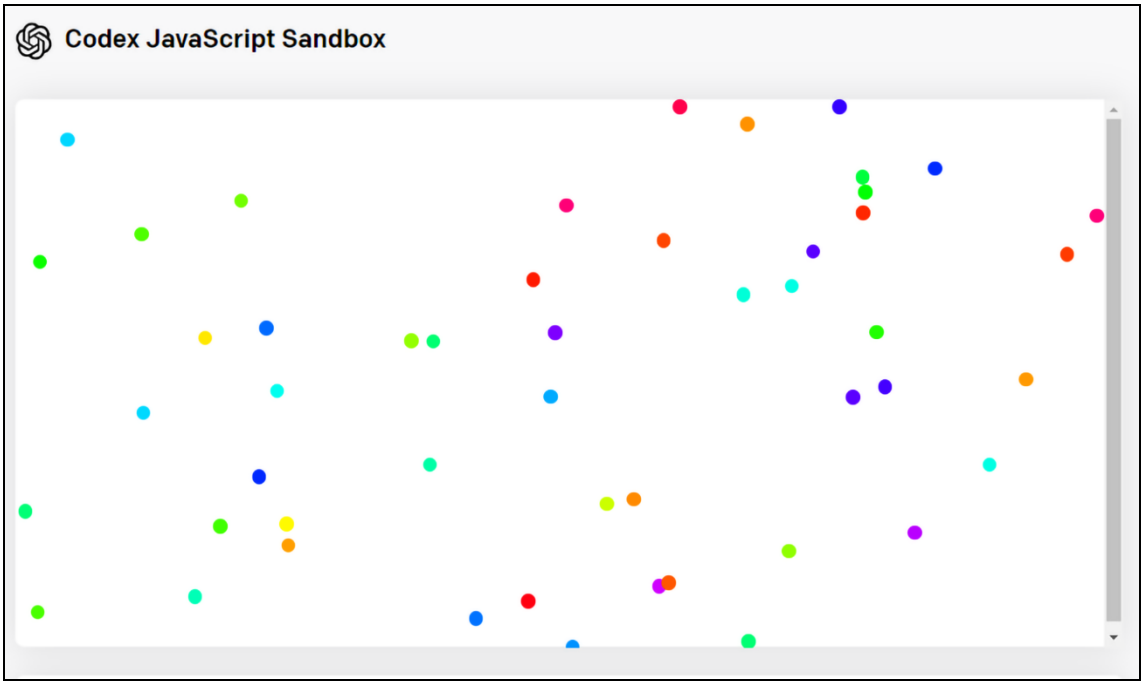


Codex JavaScript Sandbox

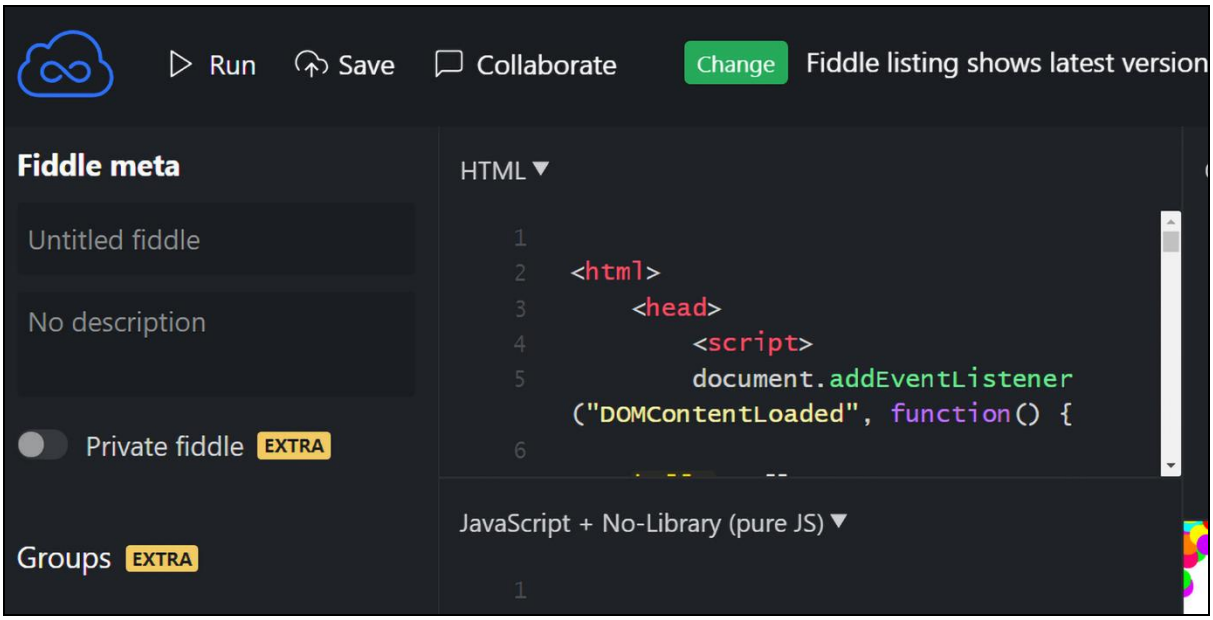
Tell Codex what to do below, and it will generate JavaScript that runs here.

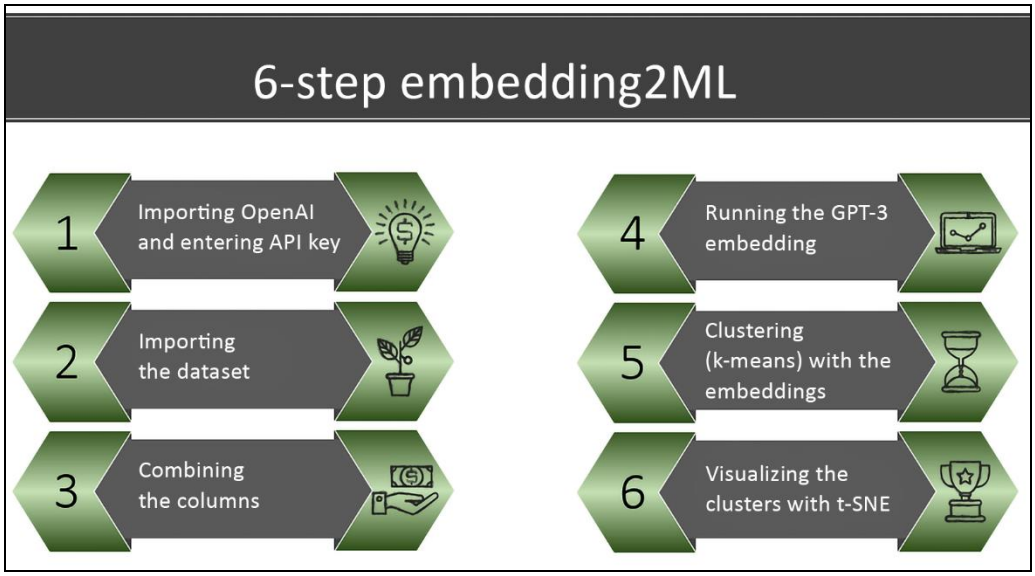
Provide instructions...

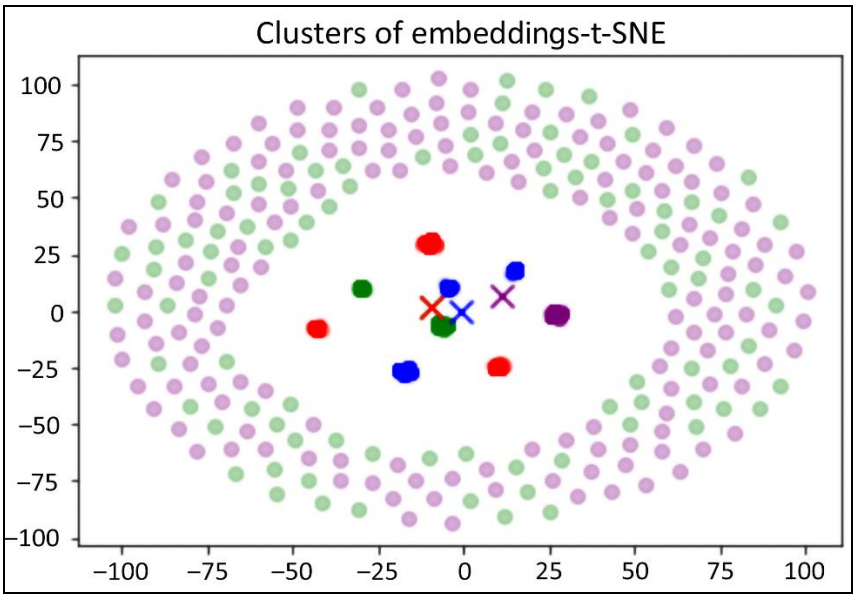
Codex JavaScript Sandbox

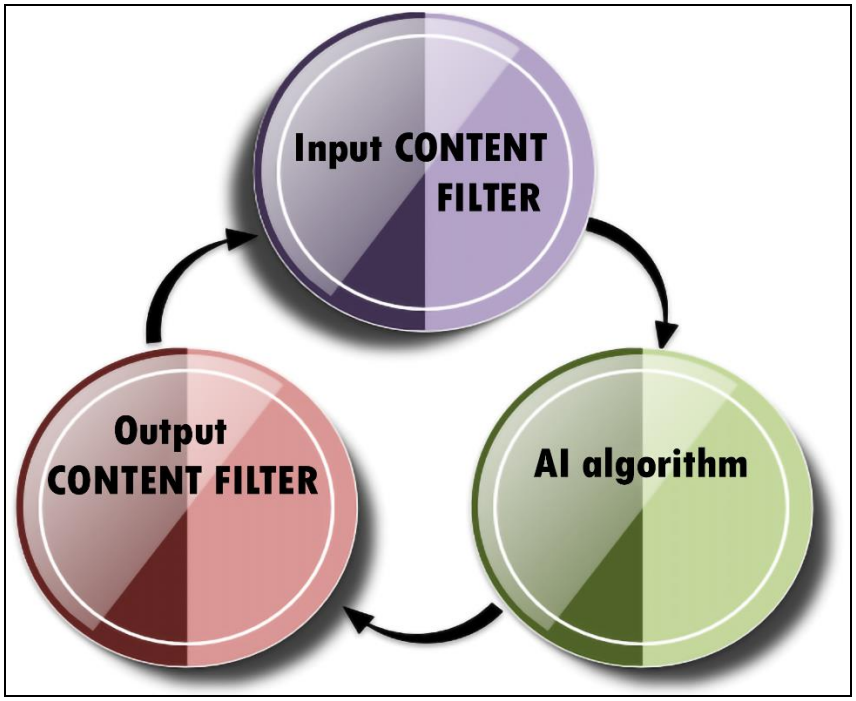


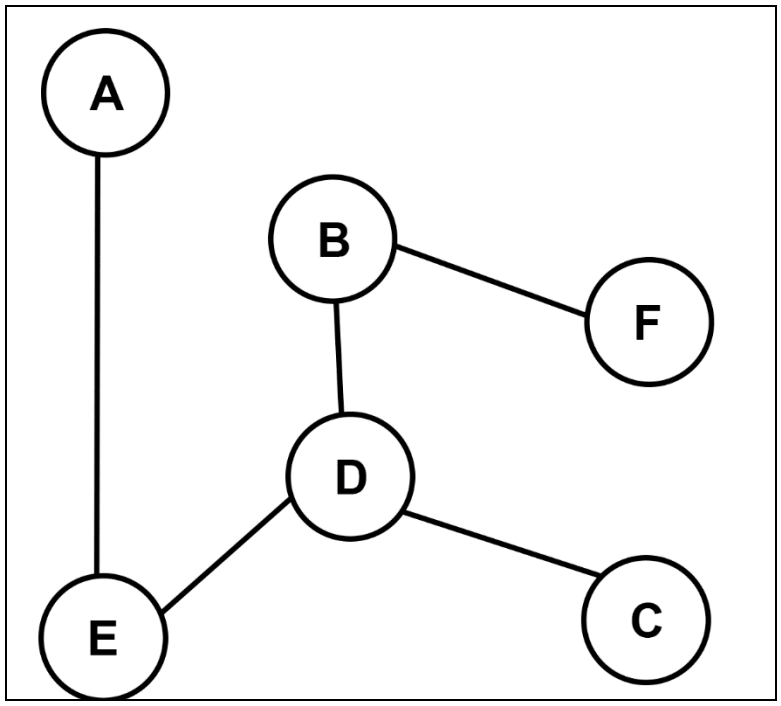
Export to JSFiddle

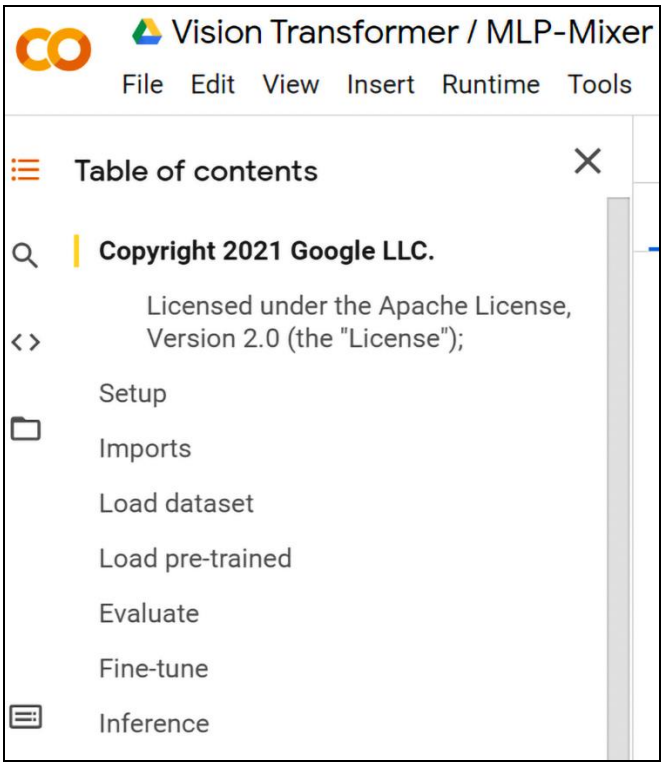

Run
Save
Collaborate
Change
Fiddle listing shows latest version
Fiddle meta
Untitled fiddle
No description
Private fiddle
EXTRA
Groups
EXTRA
HTML
<html>
<head>
<script>
document.addEventListener
("DOMContentLoaded", function() {
JavaScript + No-Library (pure JS)

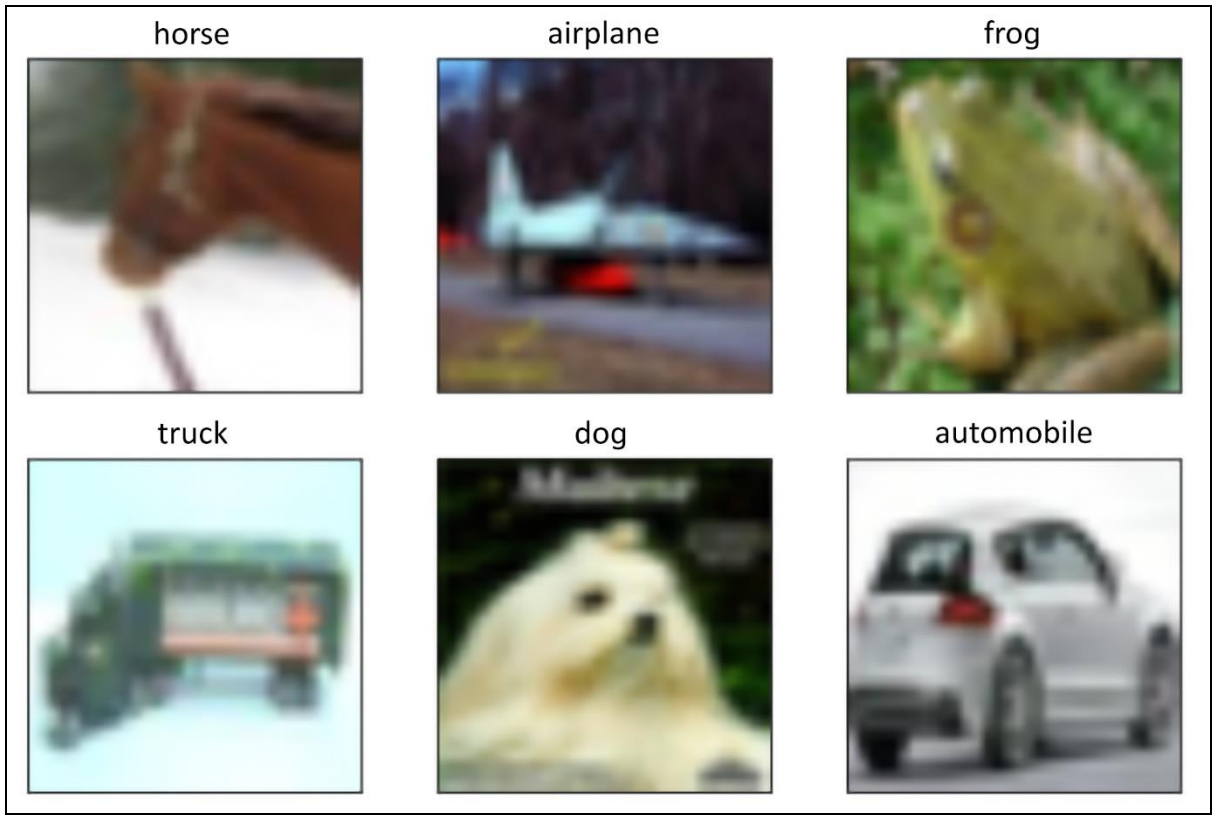

6-step embedding2ML
1
Importing OpenAI and entering API key
2
Importing the dataset
3
Combining the columns
4
Running the GPT-3 embedding
5
Clustering (k-means) with the embeddings
6
Visualizing the clusters with t-SNE


Clusters of embeddings-t-SNE
100
75
50
25
0
−25
−50
−75
−100
−100
−75
−50
−25
0
25
50
75
100


Input CONTENT FILTER
AI algorithm
Output CONTENT FILTER

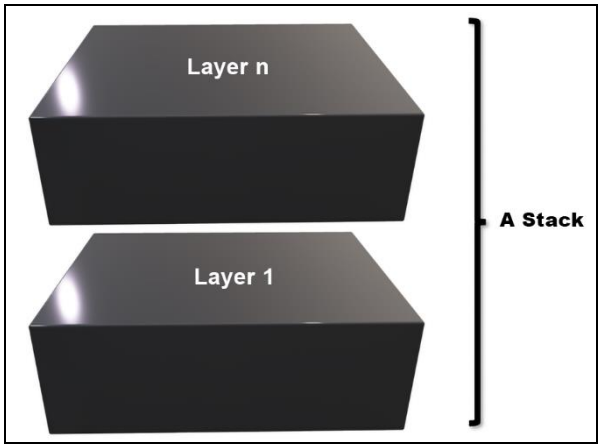

A
B
F
D
E
C

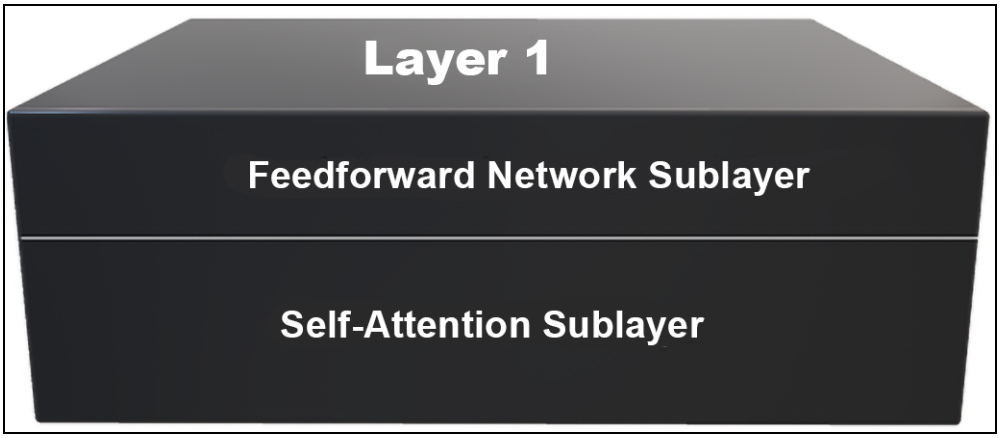

Vision Transformer / MLP-Mixer
File Edit View Insert Runtime Tools
Table of contents
Copyright 2021 Google LLC.
Licensed under the Apache License, Version 2.0 (the "License");
Setup
Imports
Load dataset
Load pre-trained
Evaluate
Fine-tune
Inference

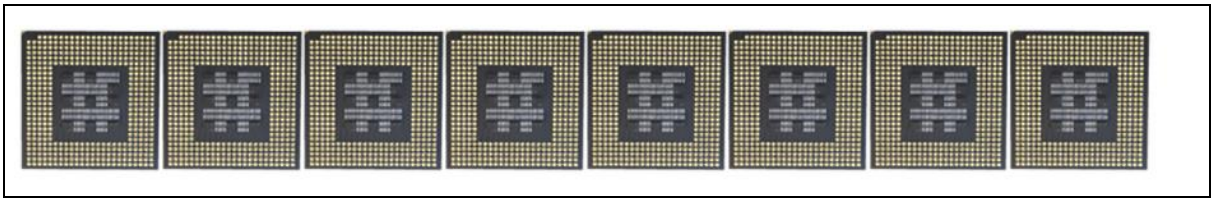

horse
airplane
frog
truck
dog
automobile

frog
dog
automobile
dog
frog
ship

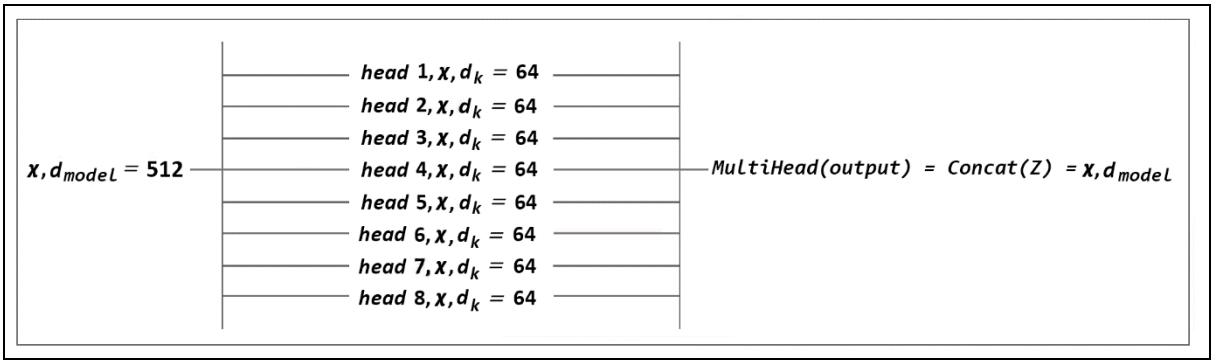
Appendix I: Terminology of Transformer Models

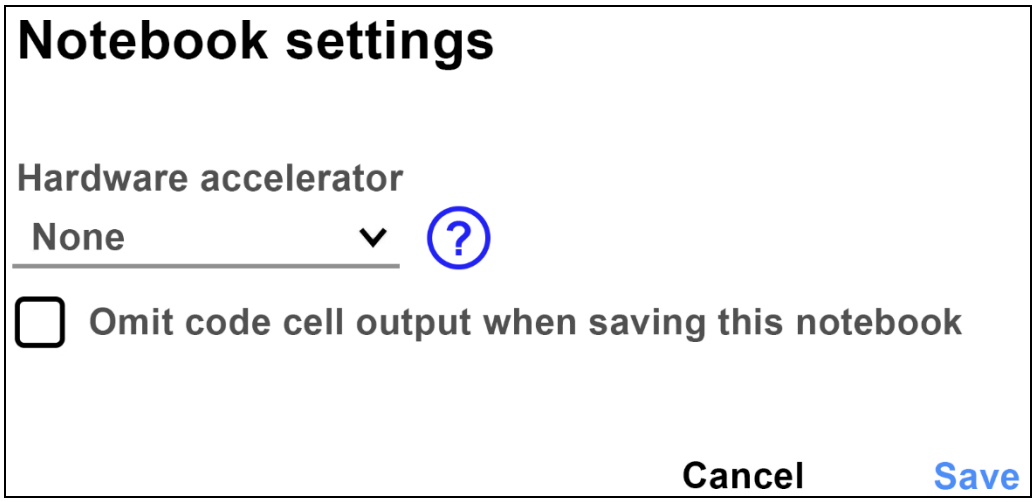






Appendix II: Hardware Constraints for Transformer Models





Notebook settings

Hardware accelerator

GPU

To get the most out of Colab, avoid using a GPU unless you need one. Learn more

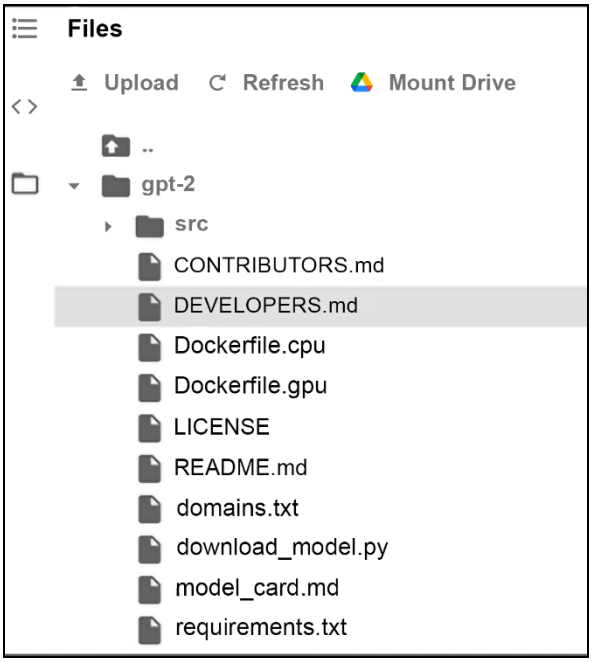
```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 495.44       Driver Version: 460.32.03    CUDA Version: 11.2     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla K80           Off  | 00000000:00:04.0 Off |                    0 |
| N/A   39C    P8    27W / 149W |      0MiB / 11441MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+

+-----------------------------------------------------------------------------+
| Processes:                                                                  |
|  GPU   GI   CI        PID   Type   Process name                  GPU Memory |
|        ID   ID                                                   Usage      |
|=============================================================================|
|  No running processes found                                                 |
+-----------------------------------------------------------------------------+
```

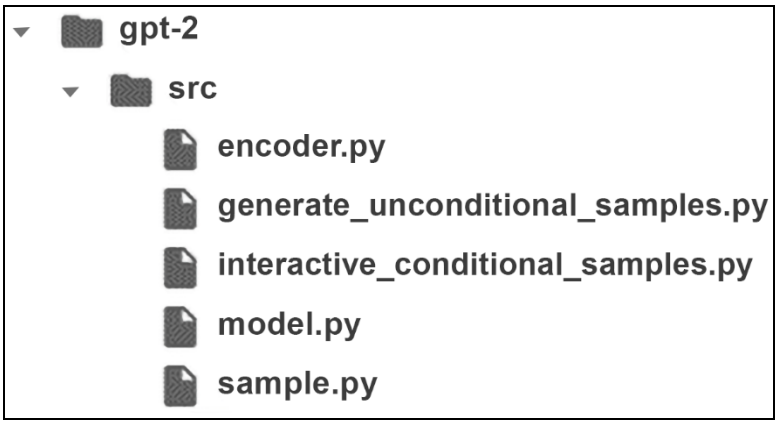
```
Epoch:   0%|          | 0/4 [00:00<?, ?it/s]Train los
Epoch:  25%|██        | 1/4 [04:58<14:56, 299.00s/it]
Train loss: 0.30048875815208026
Epoch:  50%|████      | 2/4 [09:58<09:58, 299.42s/it]
Train loss: 0.1783793037950498
Epoch:  75%|██████    | 3/4 [14:58<04:59, 299.55s/it]
Train loss: 0.11217724044973425
Epoch: 100%|██████████| 4/4 [19:58<00:00, 299.57s/it]
```

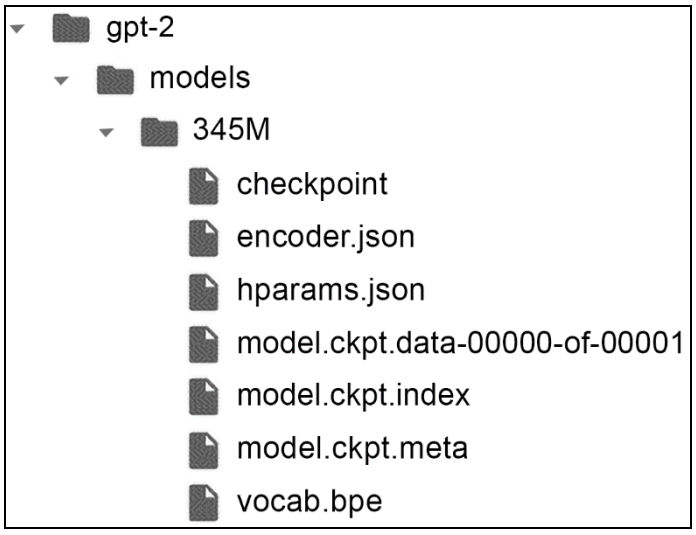
```
+-----------------------------------------------------------------------------+
| NVIDIA-SMI 495.44       Driver Version: 460.32.03    CUDA Version: 11.2     |
|-------------------------------+----------------------+----------------------+
| GPU  Name        Persistence-M| Bus-Id        Disp.A | Volatile Uncorr. ECC |
| Fan  Temp  Perf  Pwr:Usage/Cap|         Memory-Usage | GPU-Util  Compute M. |
|                               |                      |               MIG M. |
|===============================+======================+======================|
|   0  Tesla P100-PCIE...  Off  | 00000000:00:04.0 Off |                    0 |
| N/A   41C    P0    28W / 250W |      2MiB / 16280MiB |      0%      Default |
|                               |                      |                  N/A |
+-------------------------------+----------------------+----------------------+
```

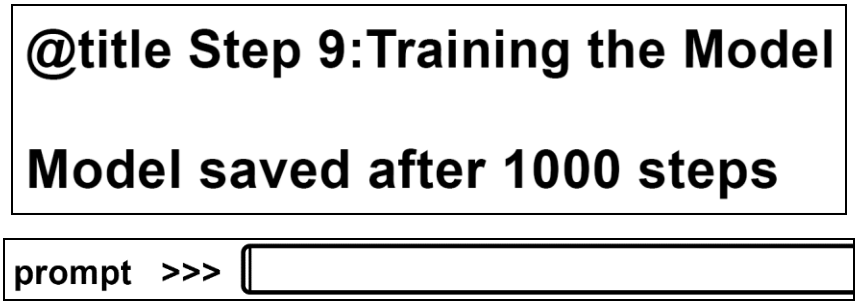
```
Epoch:   0%|          | 0/4 [00:00<?, ?it/s]Train lo
Epoch:  25%|██        | 1/4 [01:35<04:47, 95.71s/it]
Train loss: 0.3125095507168671
Epoch:  50%|████      | 2/4 [03:11<03:11, 95.57s/it]
Train loss: 0.18029312002646478
Epoch:  75%|██████    | 3/4 [04:46<01:35, 95.51s/it]
Train loss: 0.11255507657296678
Epoch: 100%|██████████| 4/4 [06:22<00:00, 95.53s/it]
```

Appendix III: Generic Text Completion with GPT-2




gpt-2
src
encoder.py
generate_unconditional_samples.py
interactive_conditional_samples.py
model.py
sample.py


gpt-2
models
345M
checkpoint
encoder.json
hparams.json
model.ckpt.data-00000-of-00001
model.ckpt.index
model.ckpt.meta
vocab.bpe

prompt >>>

Appendix IV: Custom Text Completion with GPT-2