Transforming Enterprise Machine Learning

@rajistics

raj@huggingface.co

Colab Notebook:
https://bit.ly/raj_ai

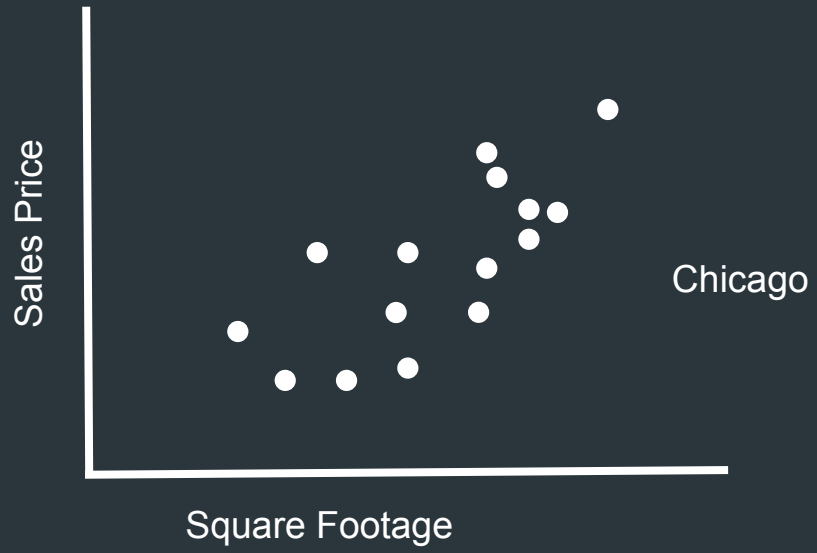
Plan

Methods

Use Cases

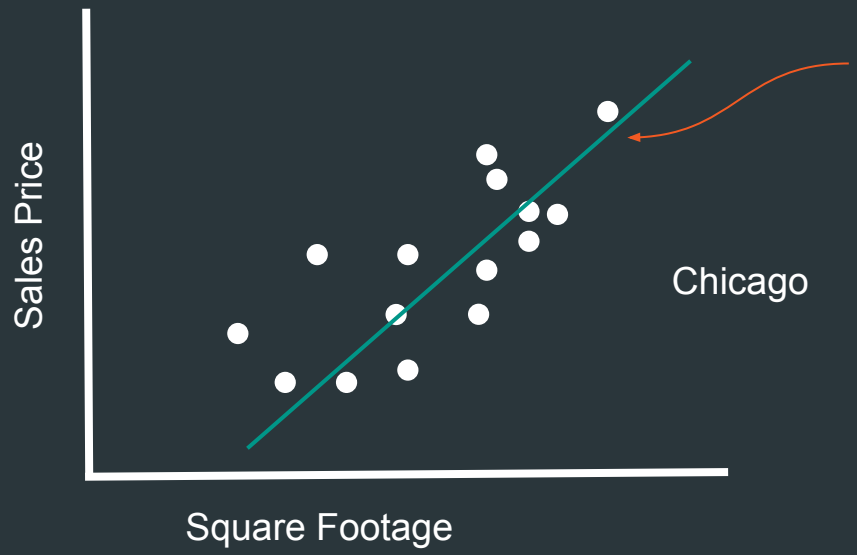
Coding

Analyzing Data



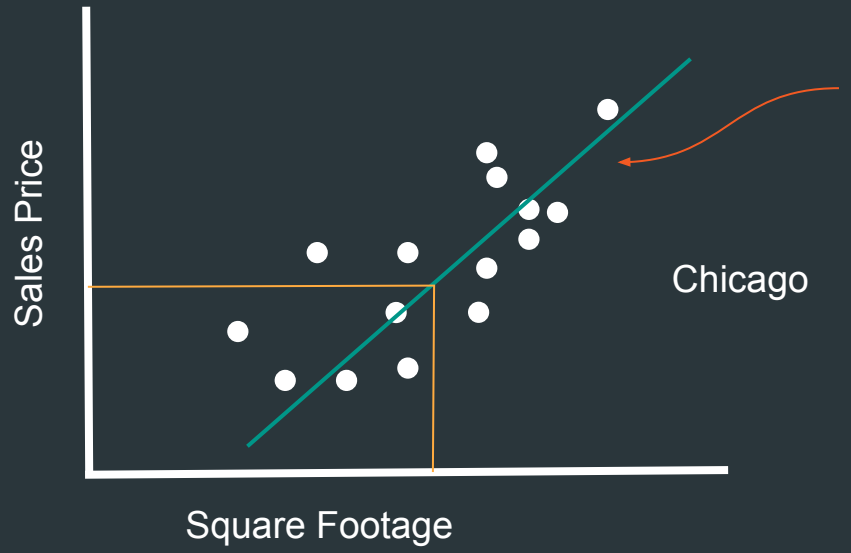
Collect as much data as possible

Machine Learning



The line is a model.

Machine Learning

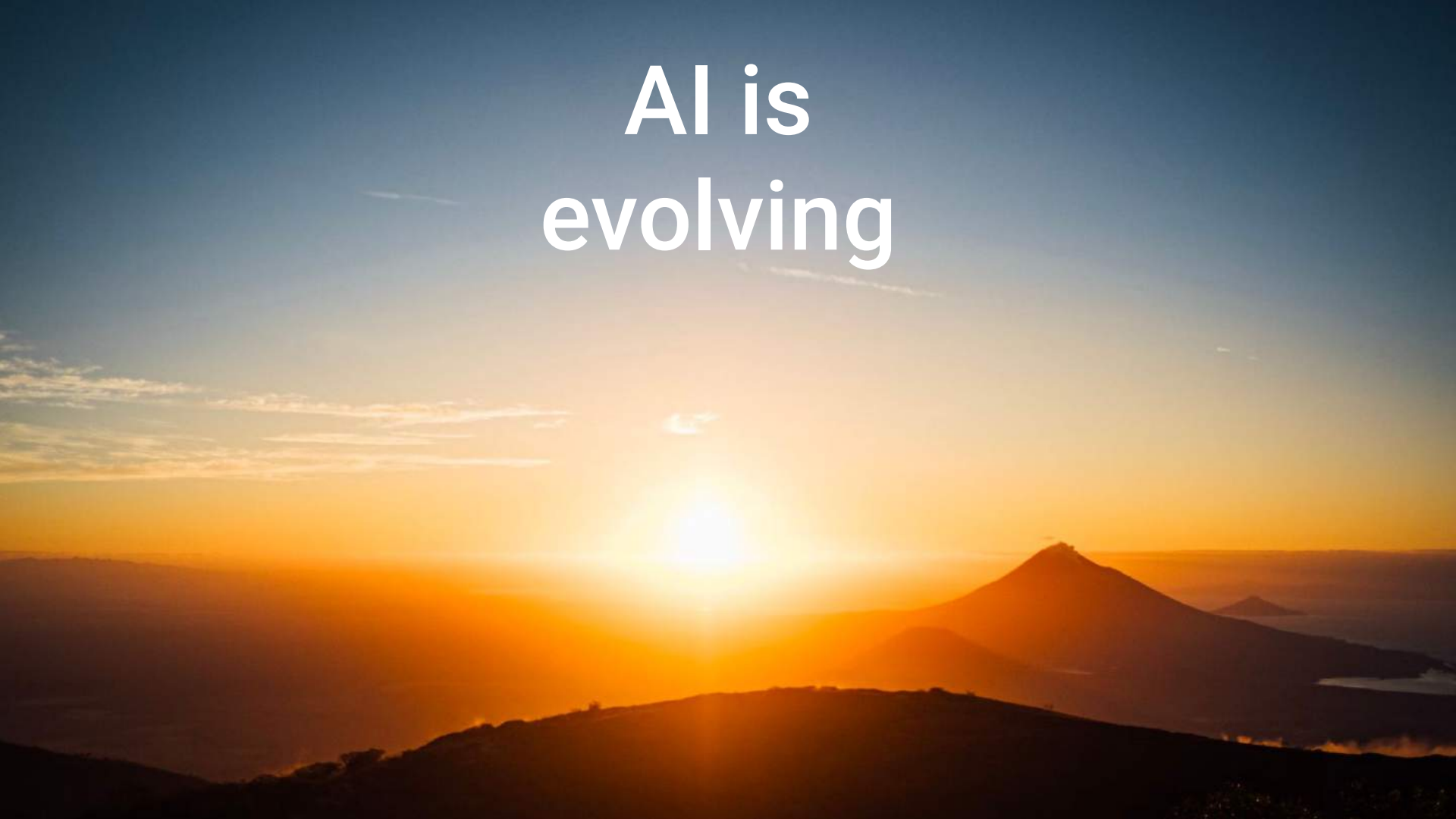


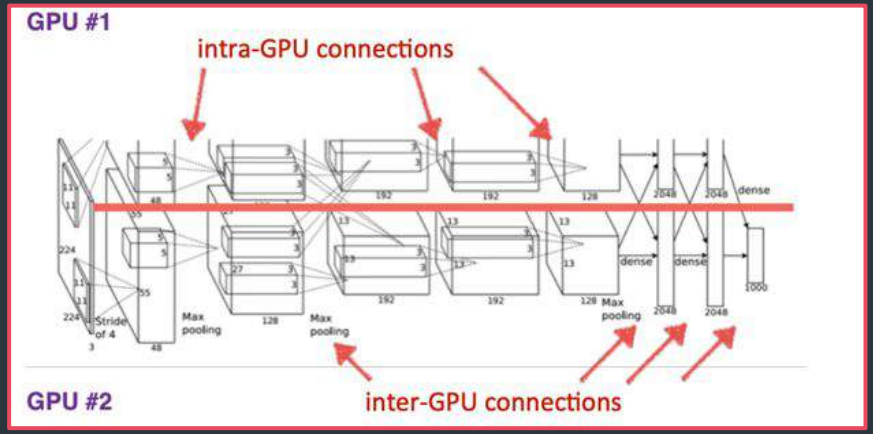
The line is a model. If you tell the model a square footage, it will make a prediction.

Traditional Supervised Learning

Chicago	Data Prep	Labeling	Feature Engineering	Training	Test Prediction
Phoenix	Data Prep	Labeling	Feature Engineering	Training	Test Prediction
. . .	Data Prep	Labeling	Feature Engineering	Training	Test Prediction

- Tabular features for the world
- Predict a number or category
- Long life cycle to a prediction
- Every task requires a new model


AI is
evolving


GPU #1
intra-GPU connections
GPU #2
inter-GPU connections
Stride of 4
Max pooling
Max pooling
Max pooling
dense
dense
dense

GTX 690
NVIDIA

Transformers Package

Model

Pre-Trained

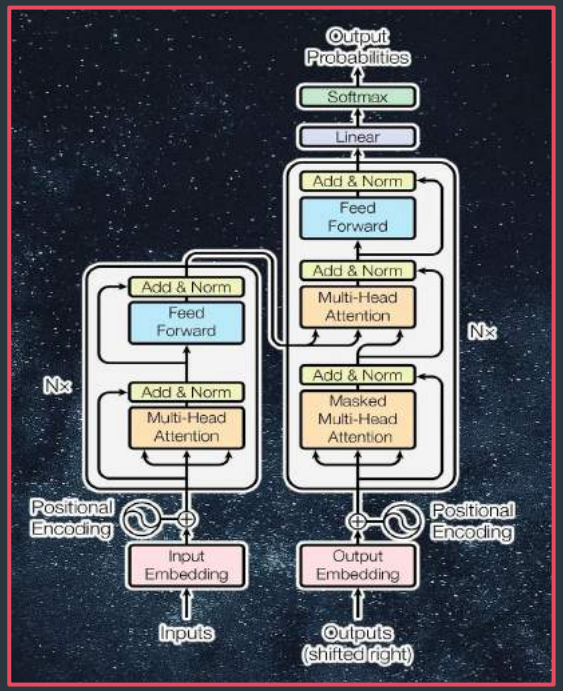
Fine Tune

Data

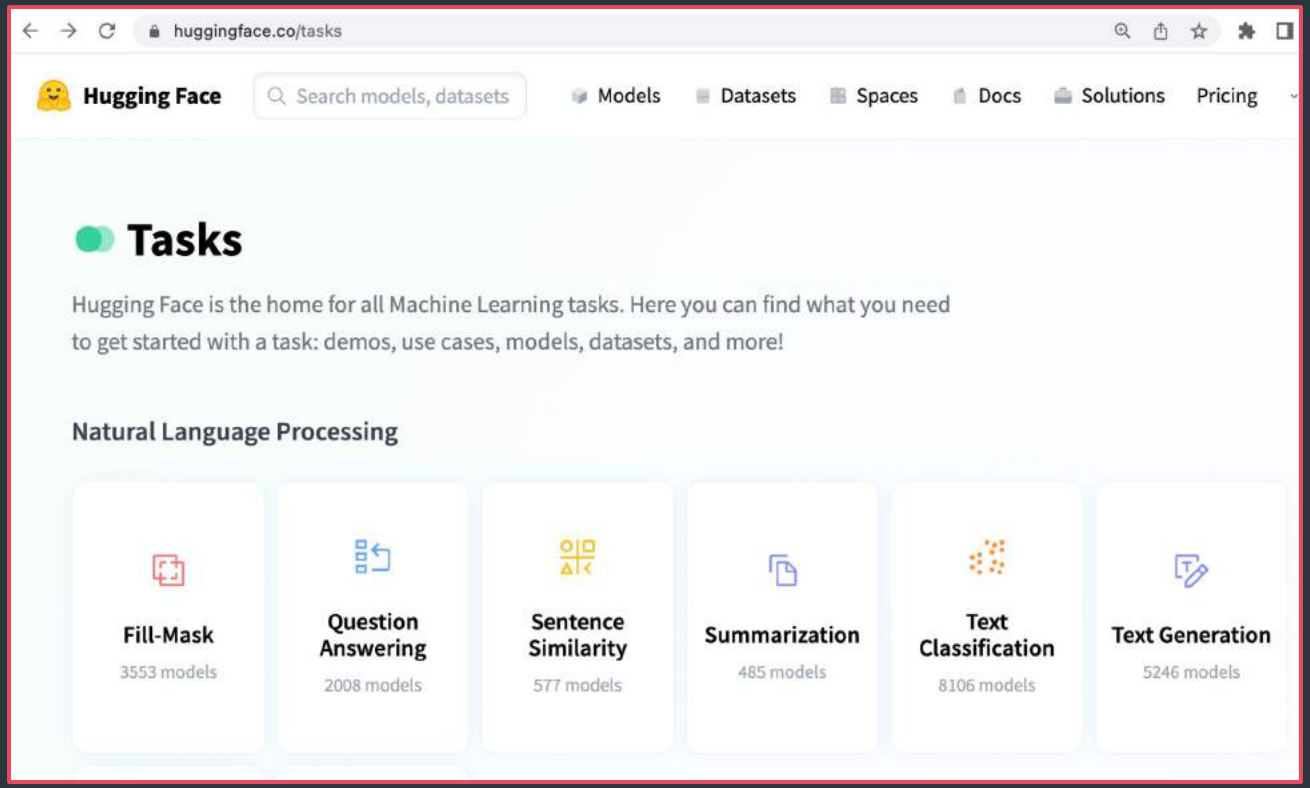
Multi Modal Models - Text, Image, Audio, Video, Code

Objective

Discriminative / Generative



Explore pretrained models at hf.co/tasks



Pre-Trained Model

Old Way

Data Prep

Labeling

Feature Engineering

Training

Test Prediction

Pre-Trained

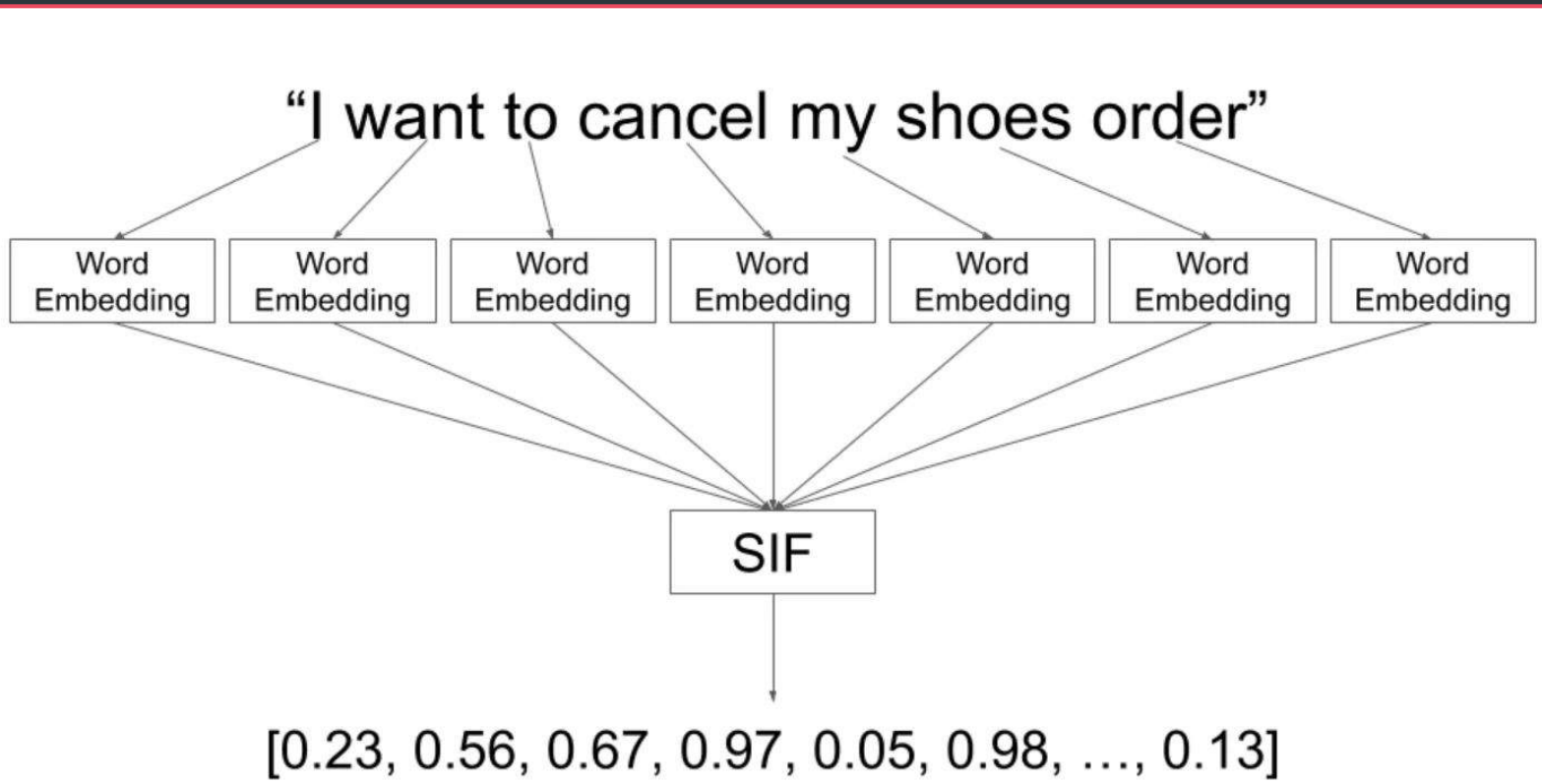
Test Prediction

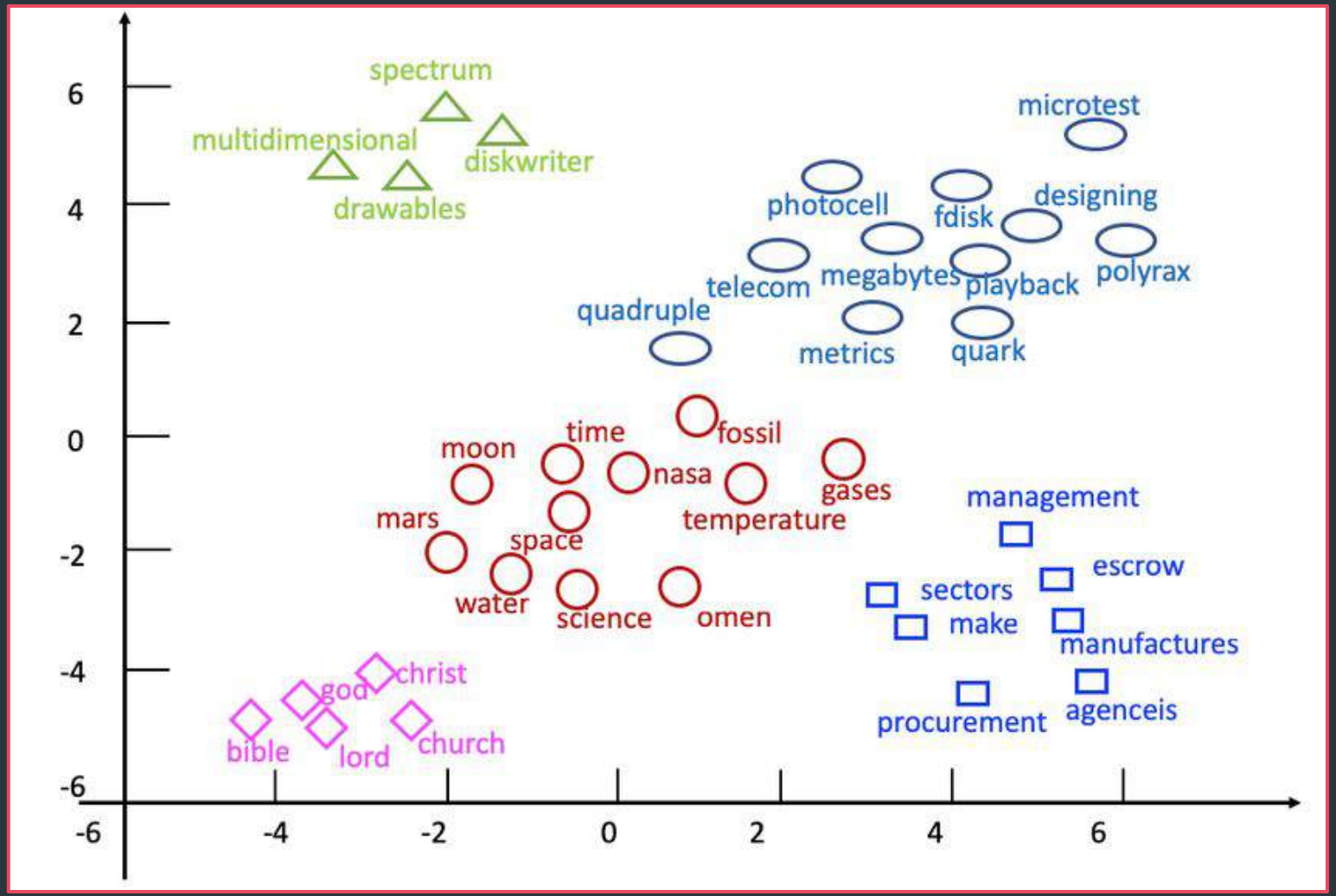
So Much Faster!

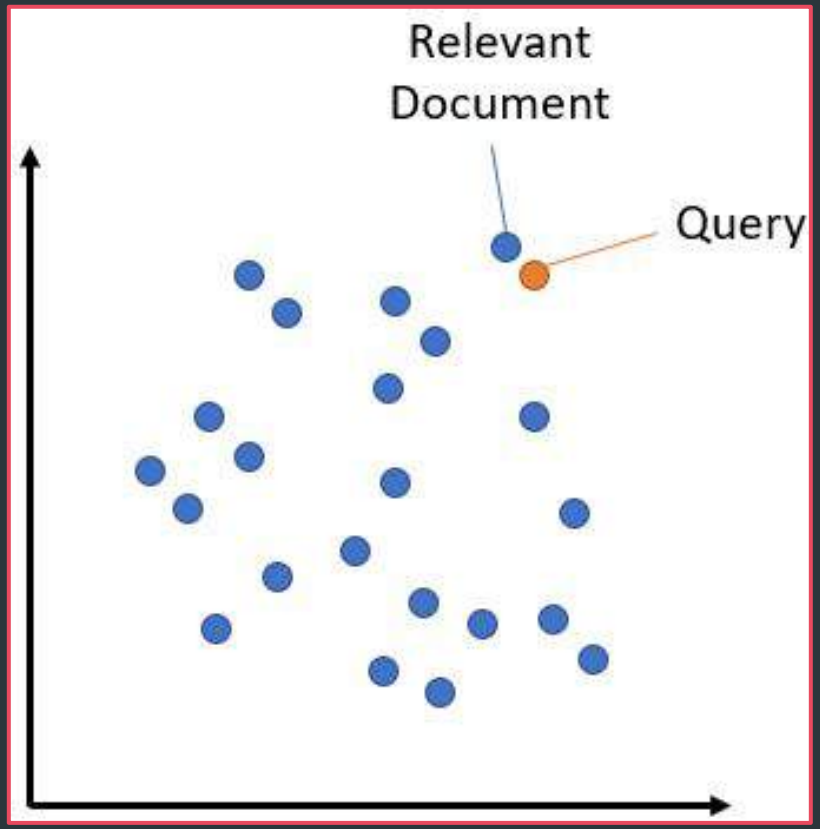
Use Cases

"eat better during xmas holidays"

"Alex Straney chats to Dr. Preeya Alexander about how to stay healthy over Christmas and about her letter to patients."


“I want to cancel my shoes order”
Word Embedding
Word Embedding
Word Embedding
Word Embedding
Word Embedding
Word Embedding
Word Embedding
SIF
[0.23, 0.56, 0.67, 0.97, 0.05, 0.98, …, 0.13]


spectrum
multidimensional
diskwriter
drawables
microtest
photocell
fdisk
designing
telecom
megabytes
playback
polyrax
quadruple
metrics
quark
moon
time
fossil
nasa
gases
mars
space
temperature
management
escrow
sectors
make
manufactures
water
science
omen
christ
god
bible
lord
church
procurement
agenceis
6
4
2
0
-2
-4
-6
-6
-4
-2
0
2
4
6



Semantic Search

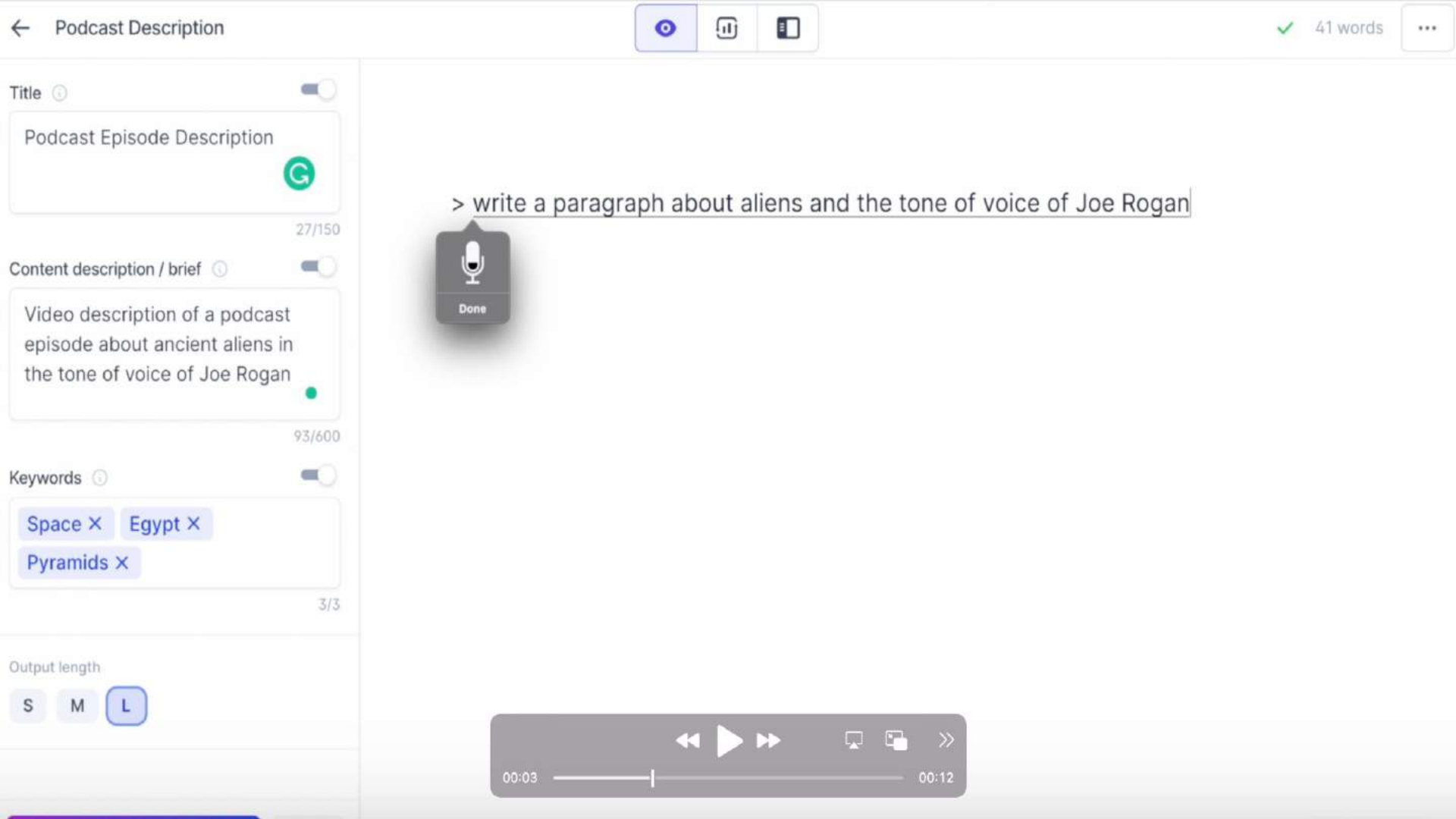
Query: “eat better”

Relevant Document:
“heathy”

parse_expenses.py | write_sql.go | sentiment.ts | addresses.rb

```
1 cl
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
```

Visual Studio Code
runtime.go
days_between_dates.js
find_files.sh
server.go
1 #!/
2
3
4
5
6
main
Ln 1 Col, 3


Podcast Description
41 words
Title
Podcast Episode Description
27/150
Content description / brief
Video description of a podcast episode about ancient aliens in the tone of voice of Joe Rogan
93/600
Keywords
Space ×
Egypt ×
Pyramids ×
3/3
Output length
S
M
L
> write a paragraph about aliens and the tone of voice of Joe Rogan
Done
00:03
00:12

Title

Podcast Episode Description

27/150

Content description / brief

Video description of a podcast episode about ancient aliens in the tone of voice of Joe Rogan

93/600

Keywords

Space × Egypt × Pyramids ×

3/3

Output length

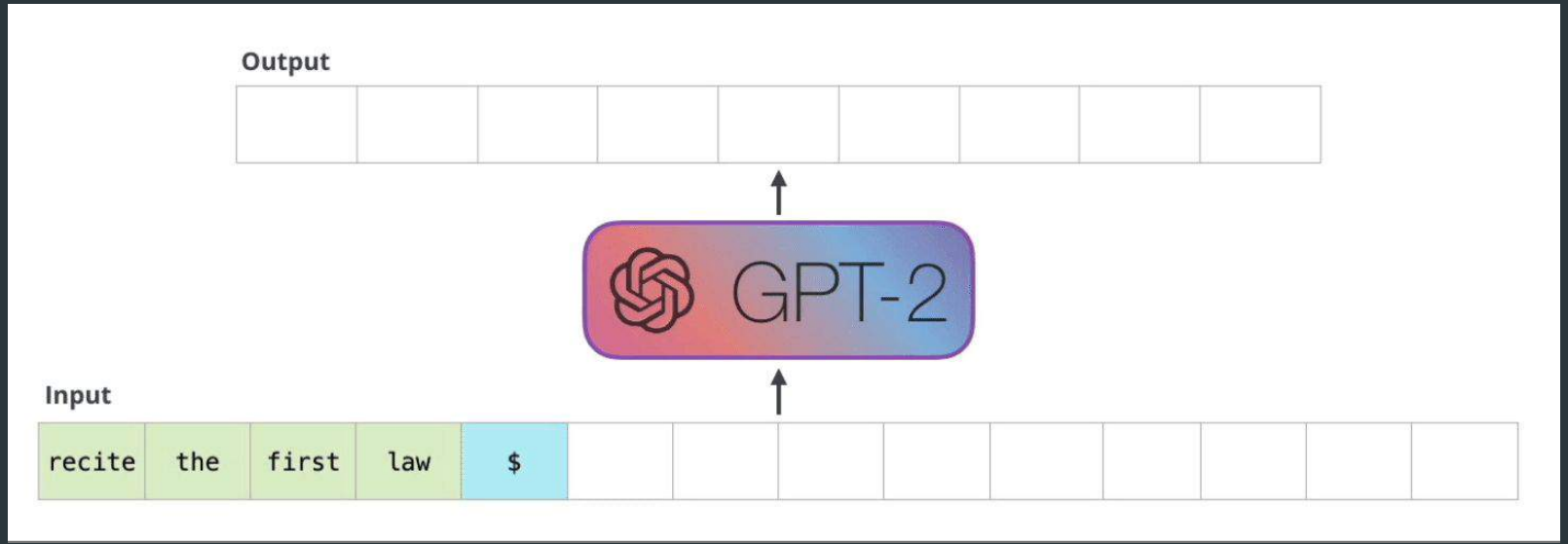
S M L

Aliens are cool man. I think they're cool. Not everyone believes in aliens, but those who do believe that there is a lot of evidence to suggest that they exist. To me, it seems like an interesting potential reality.

The idea that we're not alone in the universe is awesome.

But what about some of these? Like, some people think there was an ancient alien civilization helping to build the pyramids. And they have all kinds of "evidence" for it. These are called The Ancient Astron

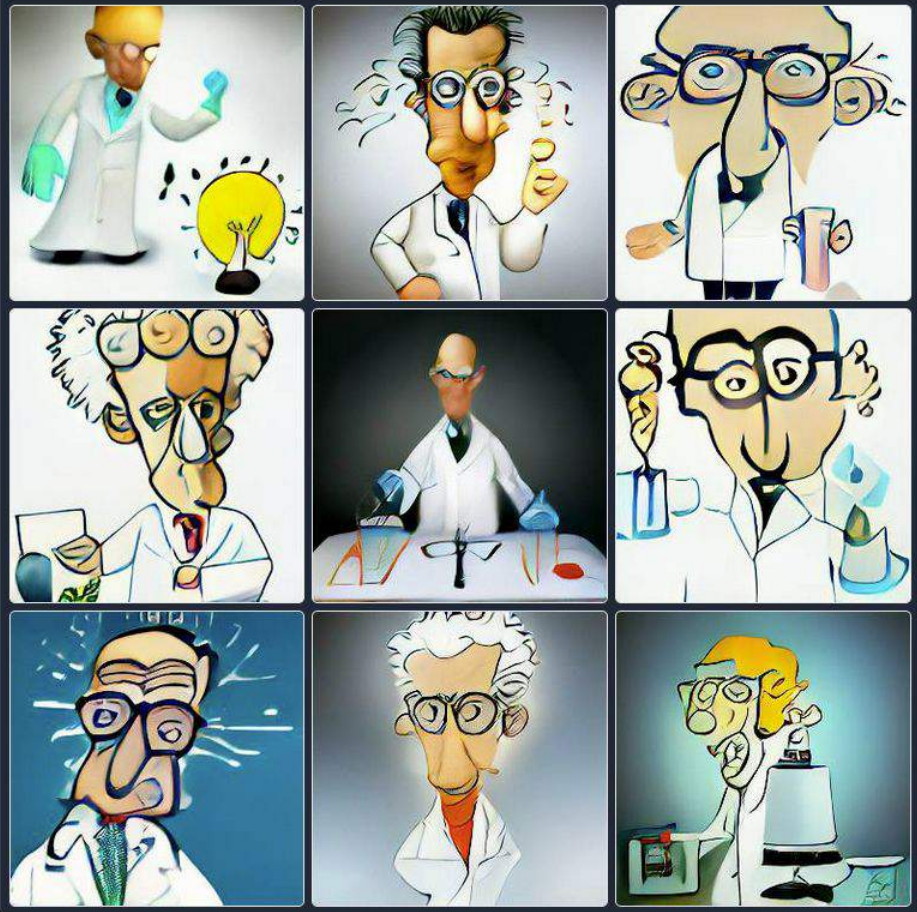
Generative AI



https://jalammar.github.io/illustrated-gpt2/

genius scientist

Run



INVOICE

LOGO

East Repair Inc.
1912 Harvest Lane
New York, NY 12210

BILL TO
John Smith
2 Court Square
New York, NY 12210

SHIP TO
John Smith
3787 Pineview Drive
Cambridge, MA 12210

INVOICE #	US-001
INVOICE DATE	11/02/2019
P.O.#	2312/2019
DUE DATE	26/02/2019

QTY	DESCRIPTION	UNIT PRICE	AMOUNT
1	Front and rear brake cables	100.00	100.00
2	New set of pedal arms	15.00	30.00
3	Labor 3hrs	5.00	15.00
		Subtotal	145.00
		Sales Tax 6.25%	9.06
		TOTAL	$154.06

John Smith

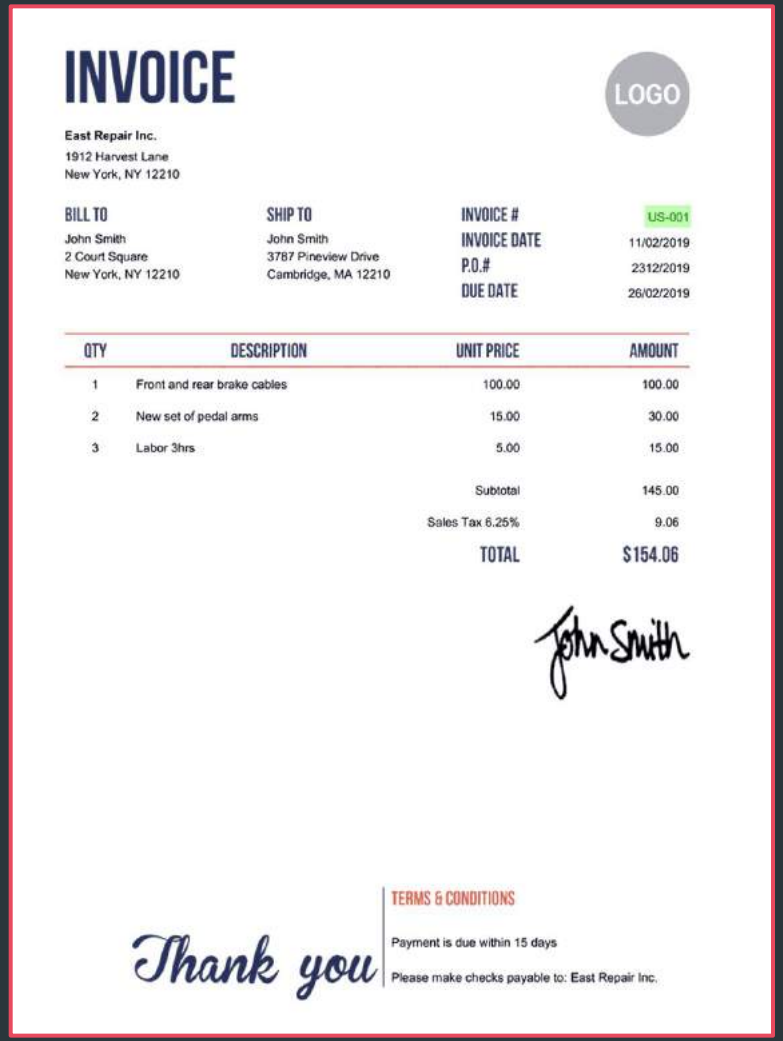
Thank you

TERMS & CONDITIONS

Payment is due within 15 days

Please make checks payable to: East Repair Inc.

What is the Invoice number?



INVOICE

LOGO

East Repair Inc.
1912 Harvest Lane
New York, NY 12210

BILL TO
John Smith
2 Court Square
New York, NY 12210

SHIP TO
John Smith
3787 Pineview Drive
Cambridge, MA 12210

INVOICE #	US-001
INVOICE DATE	11/02/2019
P.O.#	2312/2019
DUE DATE	26/02/2019

QTY	DESCRIPTION	UNIT PRICE	AMOUNT
1	Front and rear brake cables	100.00	100.00
2	New set of pedal arms	15.00	30.00
3	Labor 3hrs	5.00	15.00
		Subtotal	145.00
		Sales Tax 6.25%	9.06
		TOTAL	$154.06

John Smith

Thank you

TERMS & CONDITIONS

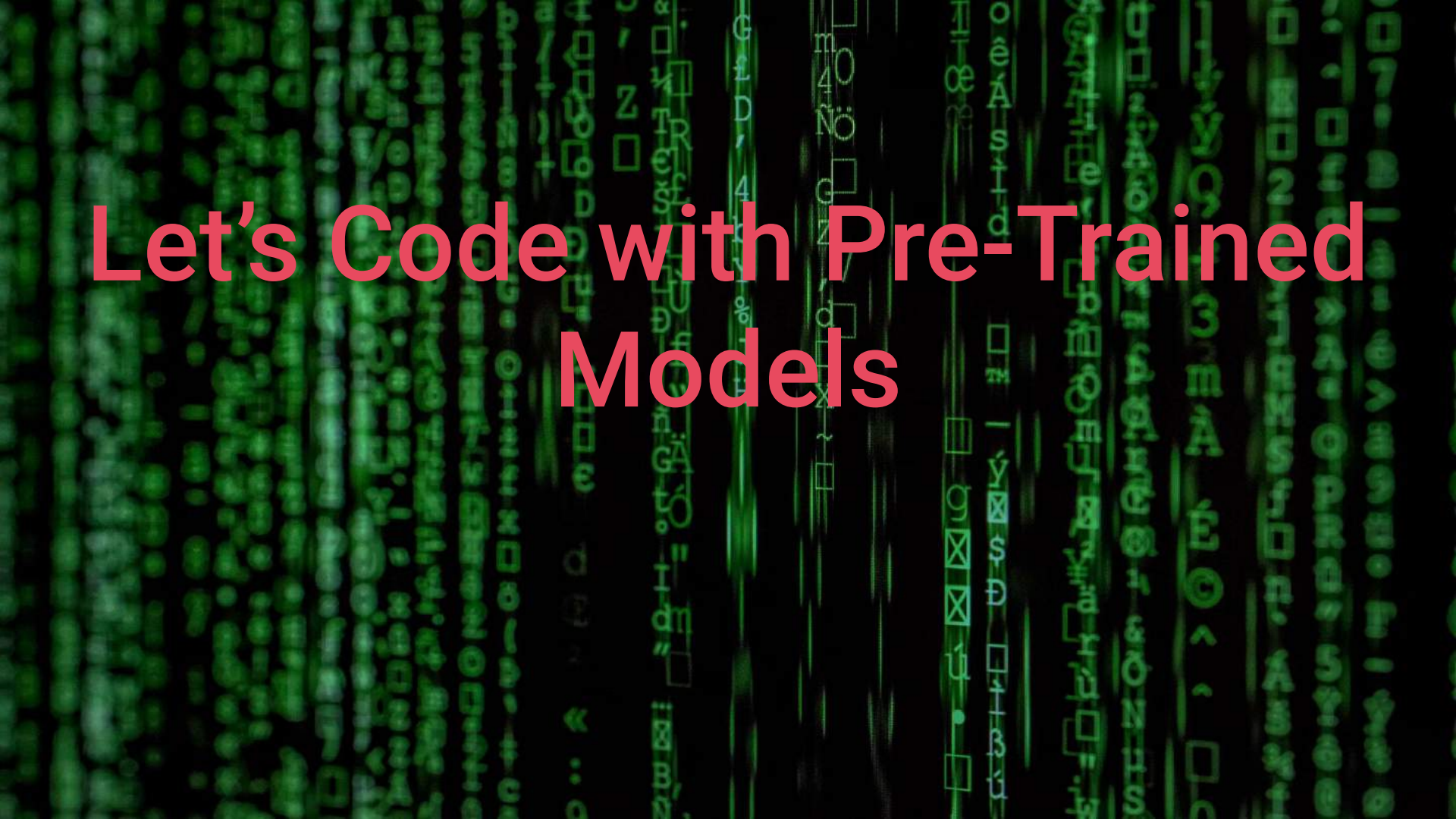
Payment is due within 15 days

Please make checks payable to: East Repair Inc.

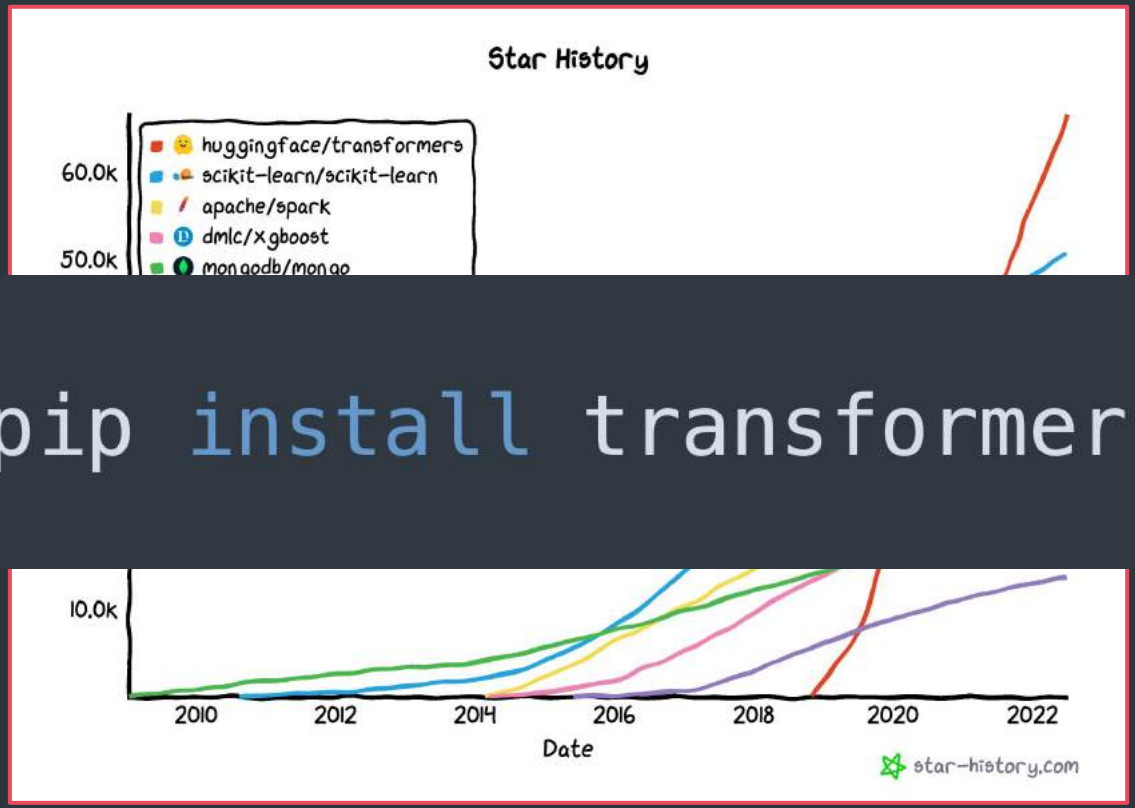
What is the Invoice number?

us-001

https://huggingface.co/spaces/impira/docquery


Let’s Code with Pre-Trained Models

Rise of Transformers




```
> pip install transformers
```

https://star-history.com/

Let’s use the transformer’s library in the Colab Notebook

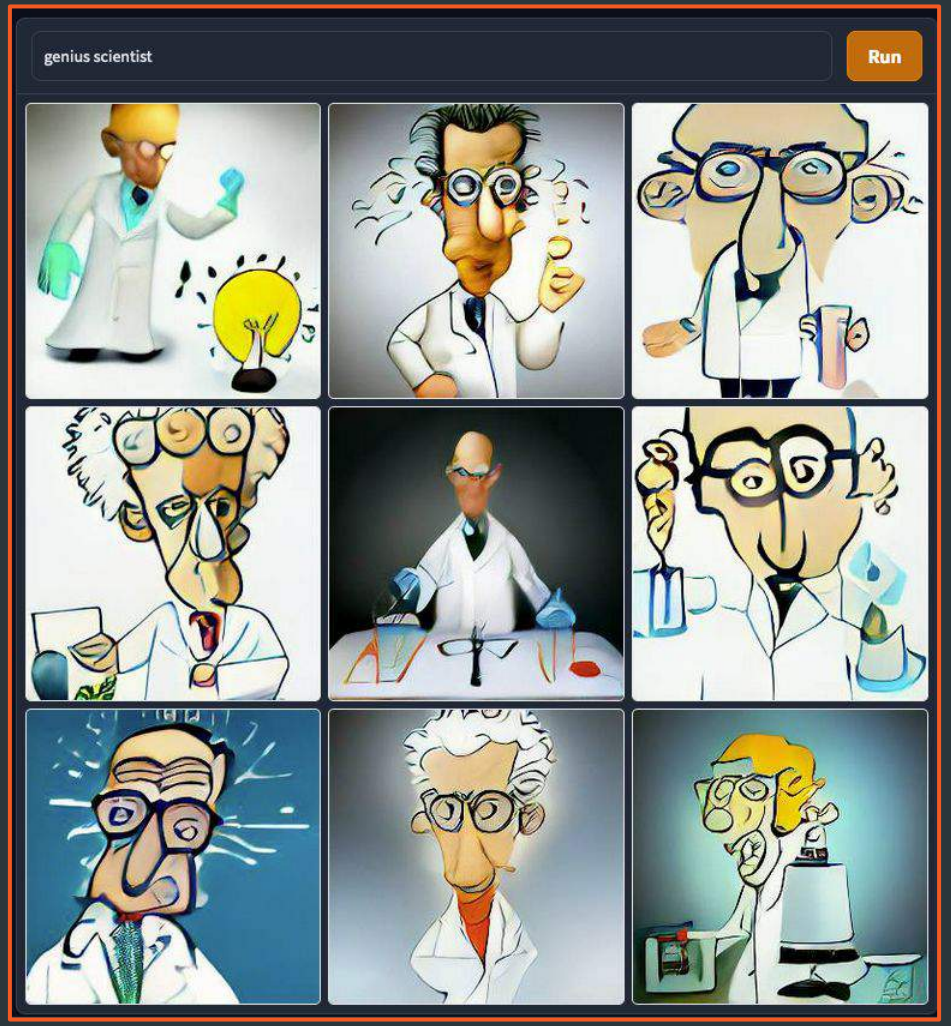
Colab Notebook: https://bit.ly/raj_ai

Concerns
Risks
Caveats
Dangers

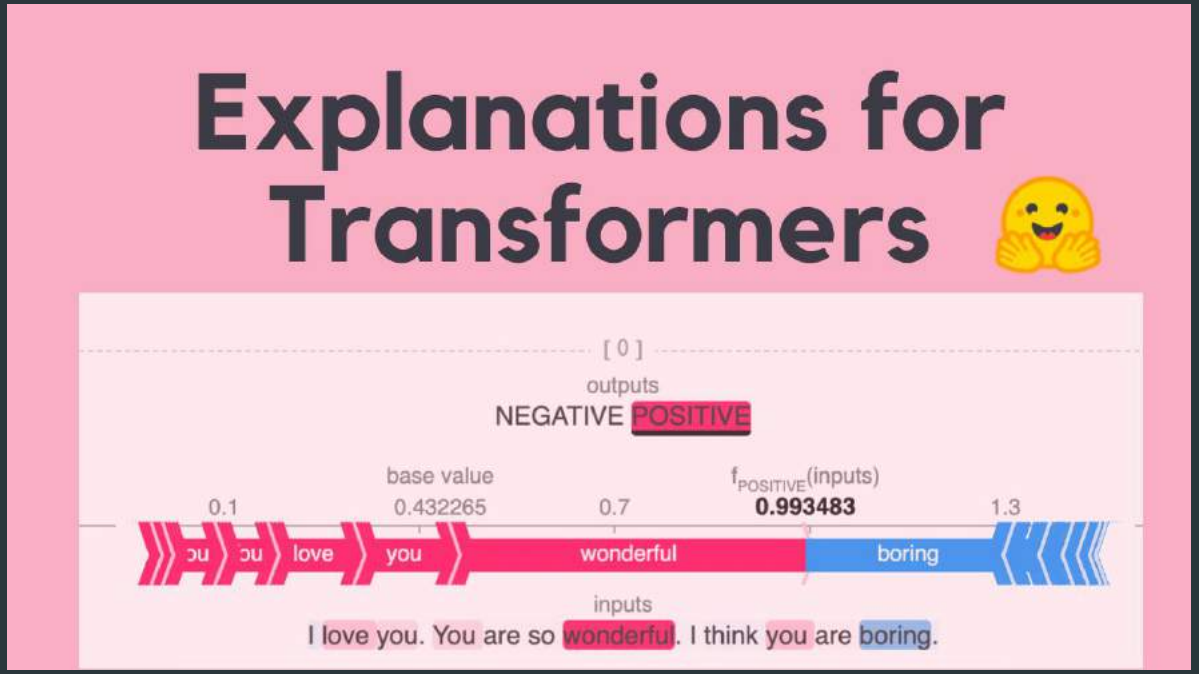
Bias

in Large Scale Pretrained Models

On the Dangers of Stochastic Parrots:
Can Language Models Be Too Big?



Explainability



Modifying
Pre-Train models

What does market mean for your ML model?

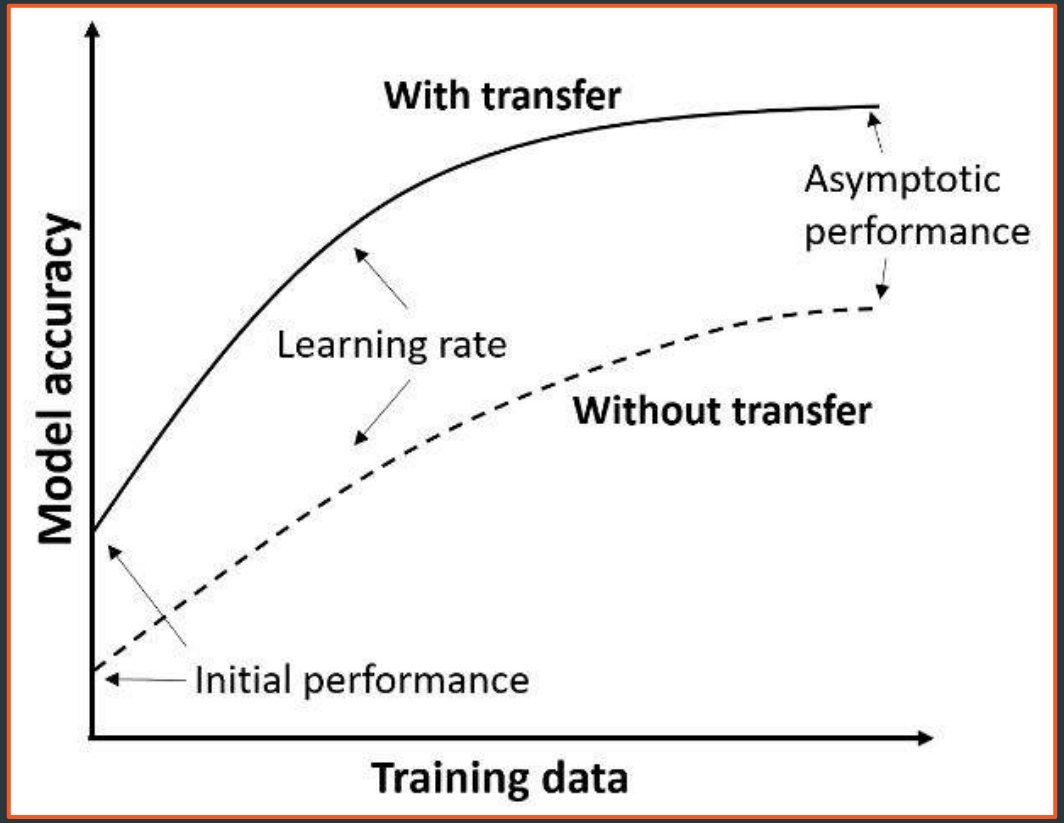
A place where people go to shop?

A place where commercial deals occur?

AI can use Transfer Learning

Model	Large Model	Large Model Retrained
Data	English Wikipedia, BooksCorpus	Task specific dataset
Objective	Predicts Token	Stock Price Sentiment

Transfer Learning



Auditor Stock Sentiment

Model

Finbert

Retrained

Data

Task specific dataset

Objective

Classify stock sentiment

"Operating profit rose to EUR 13.1 mn from EUR 8.7 mn in the corresponding period in 2007 representing 7.7 % of net sales "

POSITIVE

https://huggingface.co/FinanceInc/auditor_sentiment_finetuned

Transforming Enterprise Machine Learning

@rajistics

raj@huggingface.co

Colab Notebook:
https://bit.ly/raj_ai