Class 11
Natural Language Understanding

H Academy

March 22th, 2021 - By Nathan Landman

Academy

what is the capital of ecuador

All | Images | Maps | News | Videos | More | Settings | Tools

About 190,000,000 results (0.92 seconds)

Ecuador / Capital

Quito

Quito, the capital of Ecuador, was founded in the 16th century on the ruins of an Inca city and stands at an altitude of 2,850 m. Despite the 1917 earthquake, the city has the best-preserved, least altered historic centre in Latin America.

whc.unesco.org › list

City of Quito - UNESCO World Heritage Centre

People also search for

View 10+ more


Ecuador


Guayaquil


Ibarra


Lima


Sangolqui

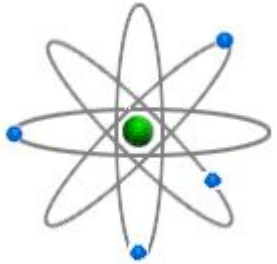

Ambato


Bogota

Feedback


What can I
help you with?

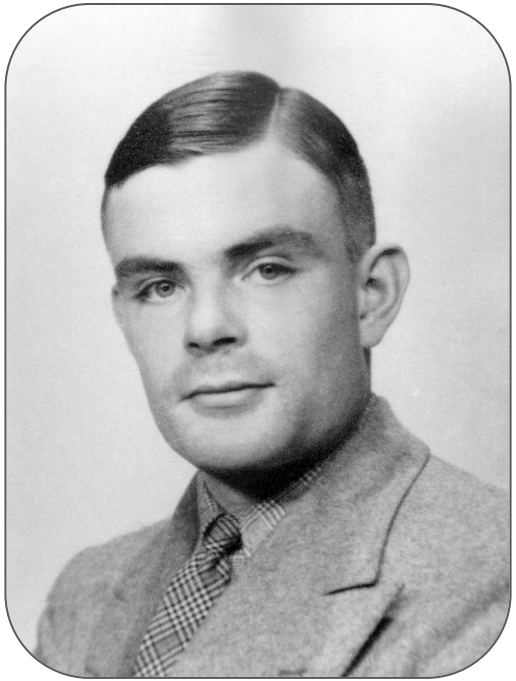
Class Agenda

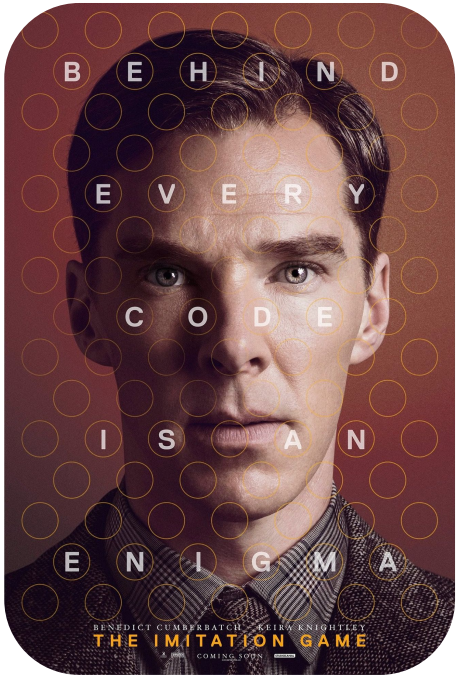


1. History of natural language in computers
2. Overview of technologies
3. New advances

Brief History of NLP

1950 - The Turing Test





VOL. LIX. No. 236.] [October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

BY A. M. TURING

1. *The Imitation Game.*

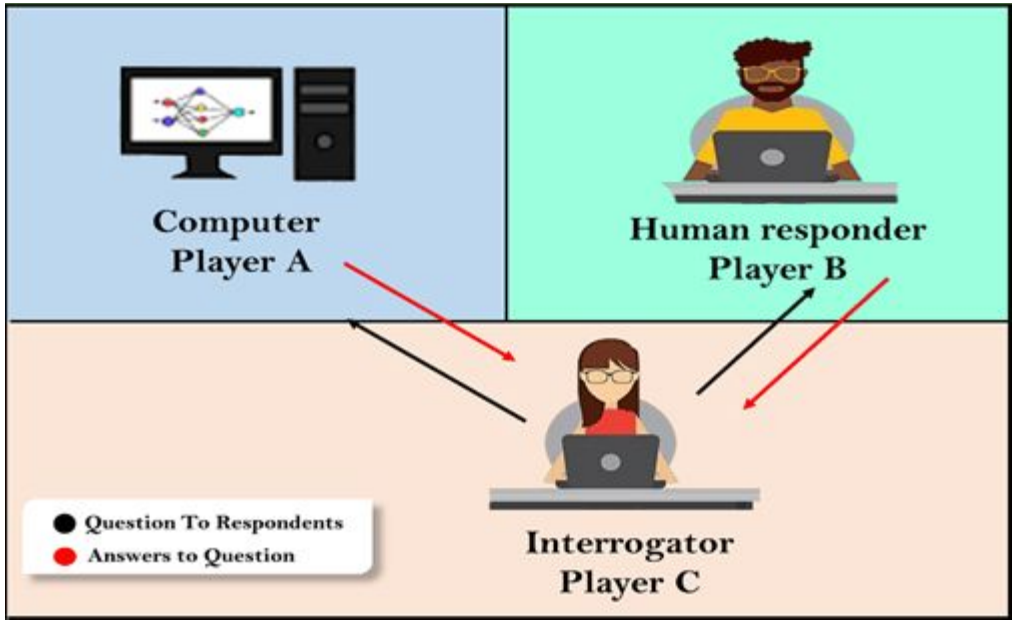
I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

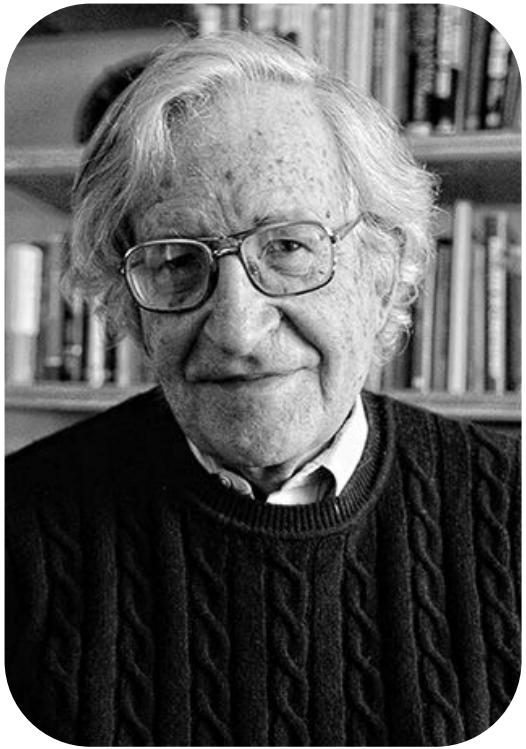
C: Will X please tell me the length of his or her hair?

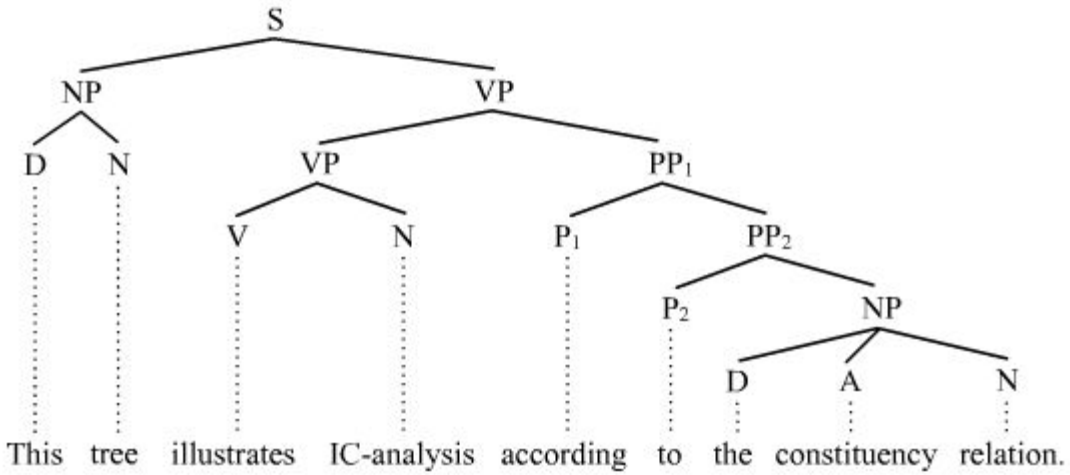
Now suppose X is actually A, then A must answer. It is A's

1950 - The Turing Test - "*Can machines think?*"

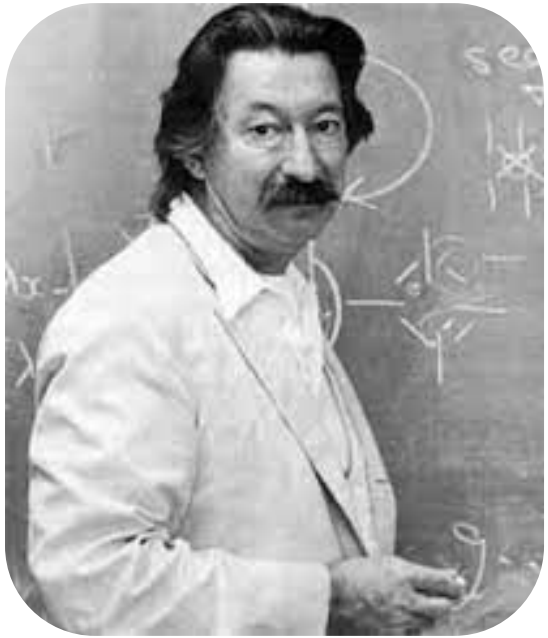


1957 - Linguistics: Chomsky and Syntactic Structures





1966 - Joe Weizenbaum's ELIZA

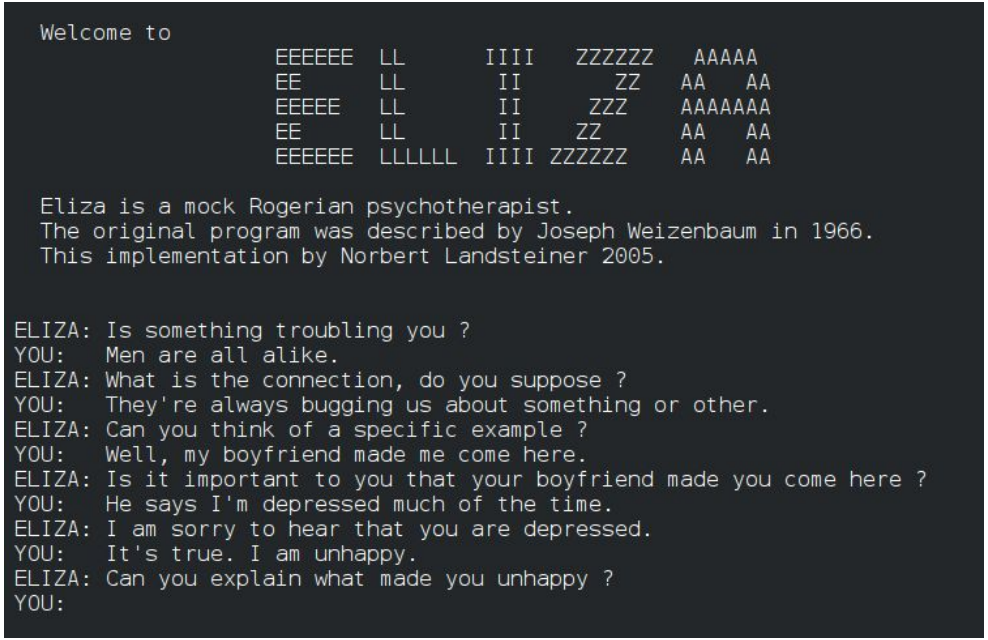


```
  Welcome to
                EEEEEE  LL      IIII   ZZZZZZ   AAAAA
                EE      LL       II       ZZ   AA   AA
                EEEEE   LL       II     ZZZ    AAAAAAA
                EE      LL       II    ZZ      AA   AA
                EEEEEE  LLLLLL  IIII ZZZZZZ    AA   AA

  Eliza is a mock Rogerian psychotherapist.
  The original program was described by Joseph Weizenbaum in 1966.
  This implementation by Norbert Landsteiner 2005.


ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

1966 - Joe Weizenbaum’s ELIZA

1966 - Joe Weizenbaum's ELIZA

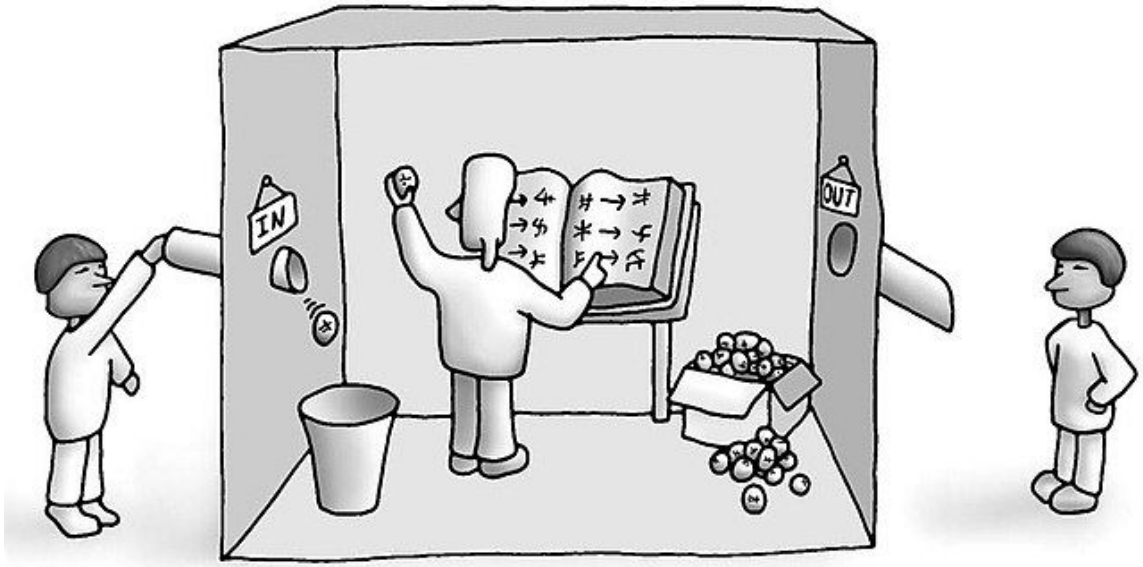
"Eliza simulated conversation by using a "pattern matching" and substitution methodology that gave users an illusion of understanding on the part of the program, but had no built in framework for contextualizing events." - Wiki



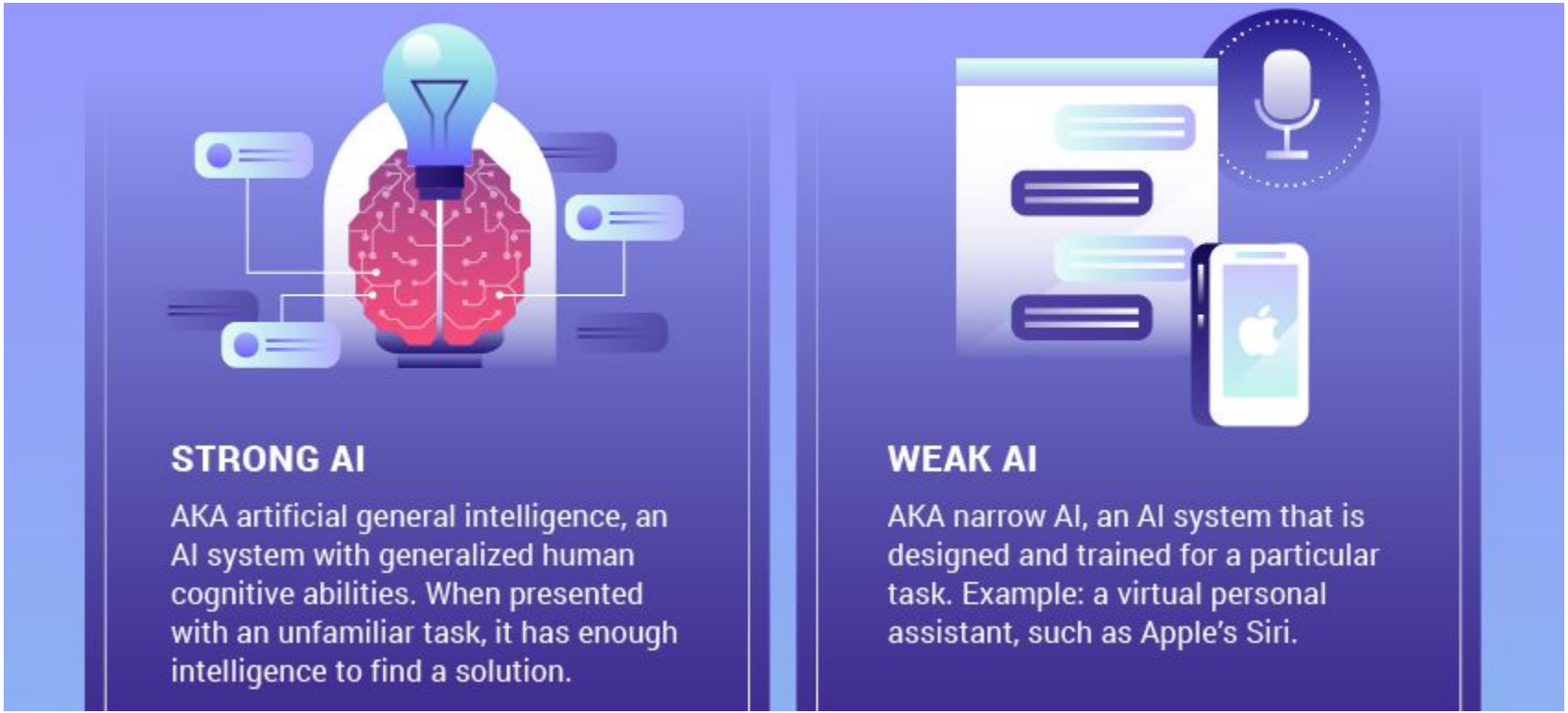
Thought Experiment… Does my friend know Spanish?

1980 - The Chinese Room

The **Chinese room argument** holds that a digital computer executing a program cannot be shown[1] to have a "mind", "understanding" or "consciousness",[a] regardless of how intelligently or human-like the program may make the computer behave.



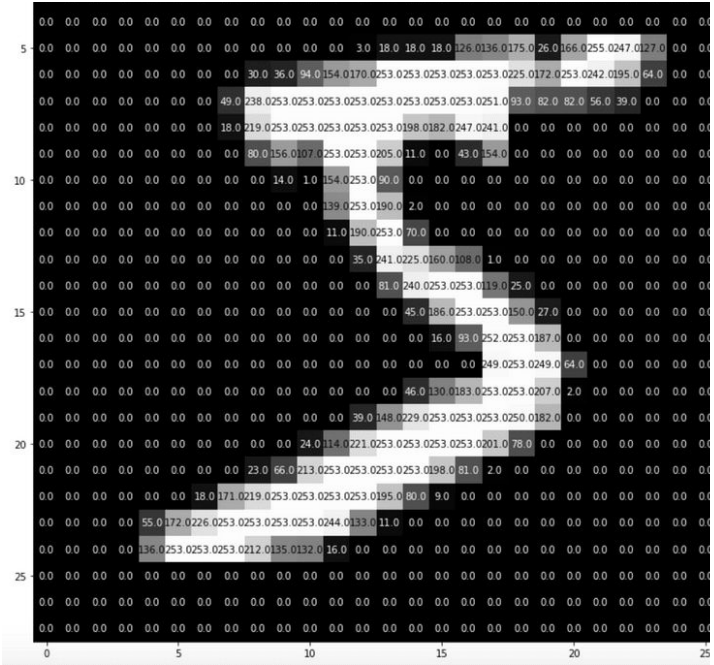
An Update to the Turing Test



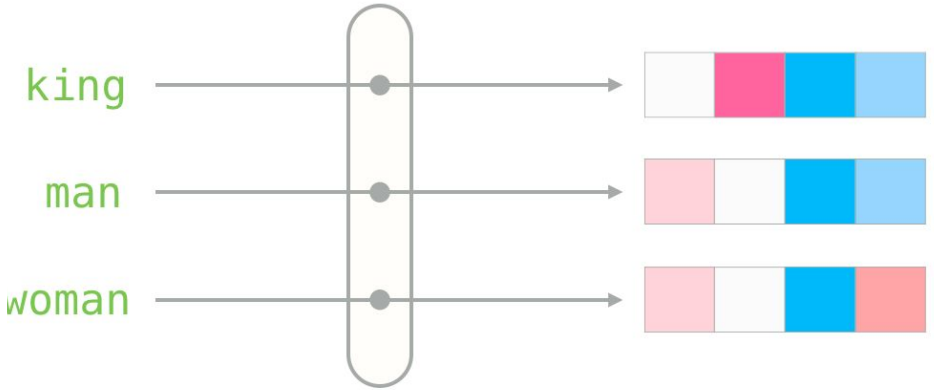
Processing Language

Digital representation of Text

Digital Representation of Images



Digital Representation of Text



One-Hot Encoding

Human-Readable

Pet
Cat
Dog
Turtle
Fish
Cat

Machine-Readable

Cat	Dog	Turtle	Fish
1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1
1	0	0	0

One-Hot Encoding

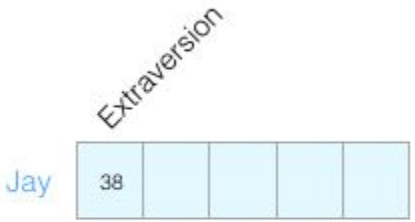
We need a way to compress information such that we don’t have to use a vector with 600,000 words!

Digital representation of Text

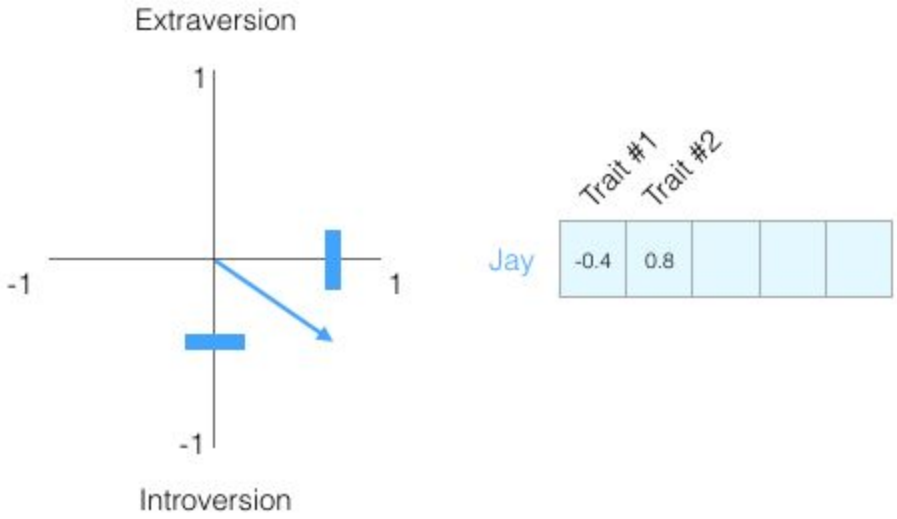
Personality tests are a method of representing a person.

Openness to experience	79 out of 100
Agreeableness	75 out of 100
Conscientiousness	42 out of 100
Negative emotionality	50 out of 100
Extraversion	58 out of 100

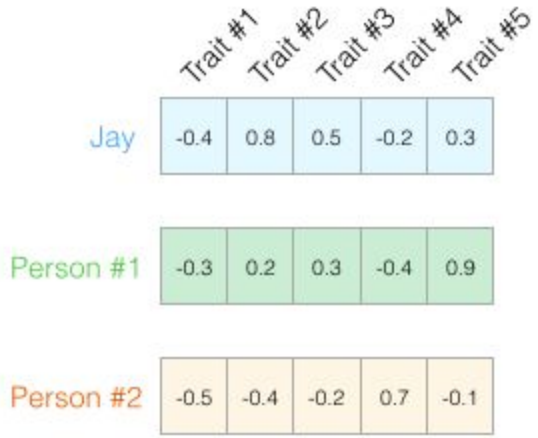
Digital representation of Text



Digital representation of Text



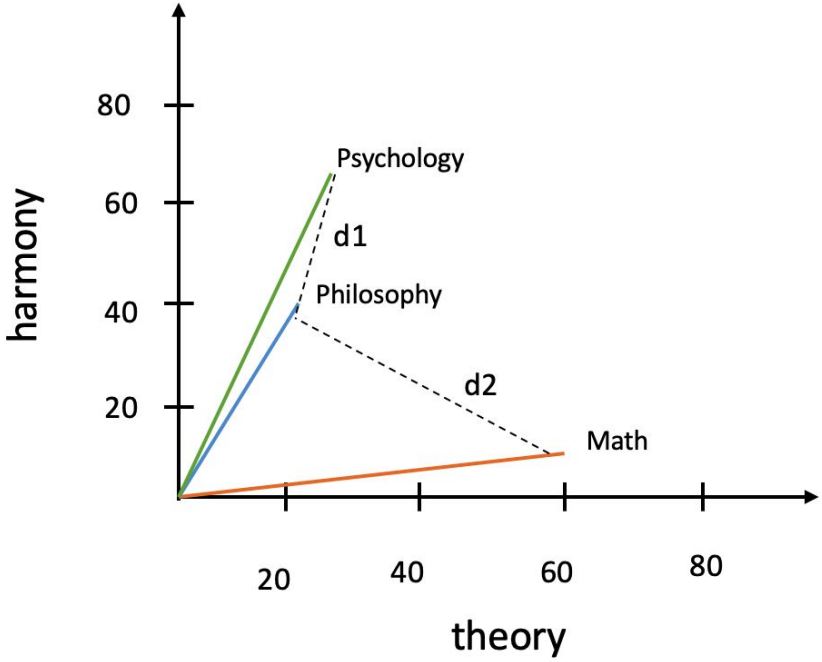
Digital representation of Text



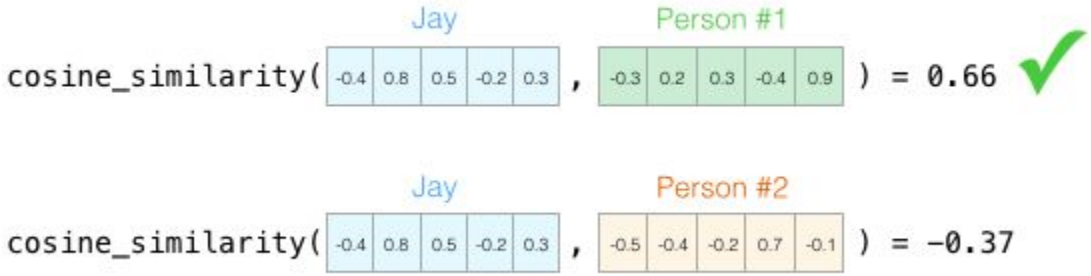
A bit difficult to write 5 dimensional arrows, but try to visualize it anyways.

Digital representation of Text - Cosine Similarity

$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$



Digital representation of Text



$$\text{similarity} = \cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}},$$

Digital representation of Text - Takeaways

1- We can represent things (and people) as vectors of numbers
(Which is great for machines!)

Jay	-0.4	0.8	0.5	-0.2	0.3

2- We can easily calculate how similar vectors are to each other

The people most similar to Jay are:

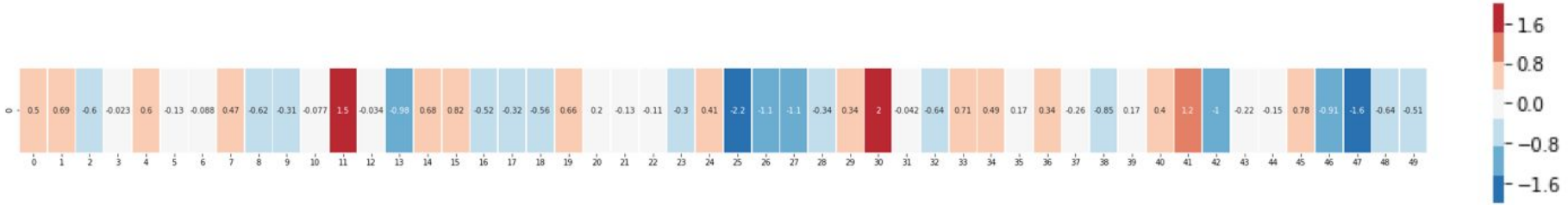
	cosine_similarity ▼
Person #1	0.86
Person #2	0.5
Person #3	-0.20

Enter Word Embeddings

```
[ 0.50451 , 0.68607 , -0.59517 , -0.022801, 0.60046 , -0.13498 , -0.08813 , 0.47377 , -0.61798 , -0.31012 ,
-0.076666, 1.493 , -0.034189, -0.98173 , 0.68229 , 0.81722 , -0.51874 , -0.31503 , -0.55809 , 0.66421 ,
0.1961 , -0.13495 , -0.11476 , -0.30344 , 0.41177 , -2.223 , -1.0756 , -1.0783 , -0.34354 , 0.33505 , 1.9927
, -0.04234 , -0.64319 , 0.71125 , 0.49159 , 0.16754 , 0.34344 , -0.25663 , -0.8523 , 0.1661 , 0.40102 ,
1.1685 , -1.0137 , -0.21585 , -0.15155 , 0.78321 , -0.91241 , -1.6106 , -0.64426 , -0.51042 ]
```

Vector Representation
of the word **King**

Enter Word Embeddings



Vector Representation
of the word **King**

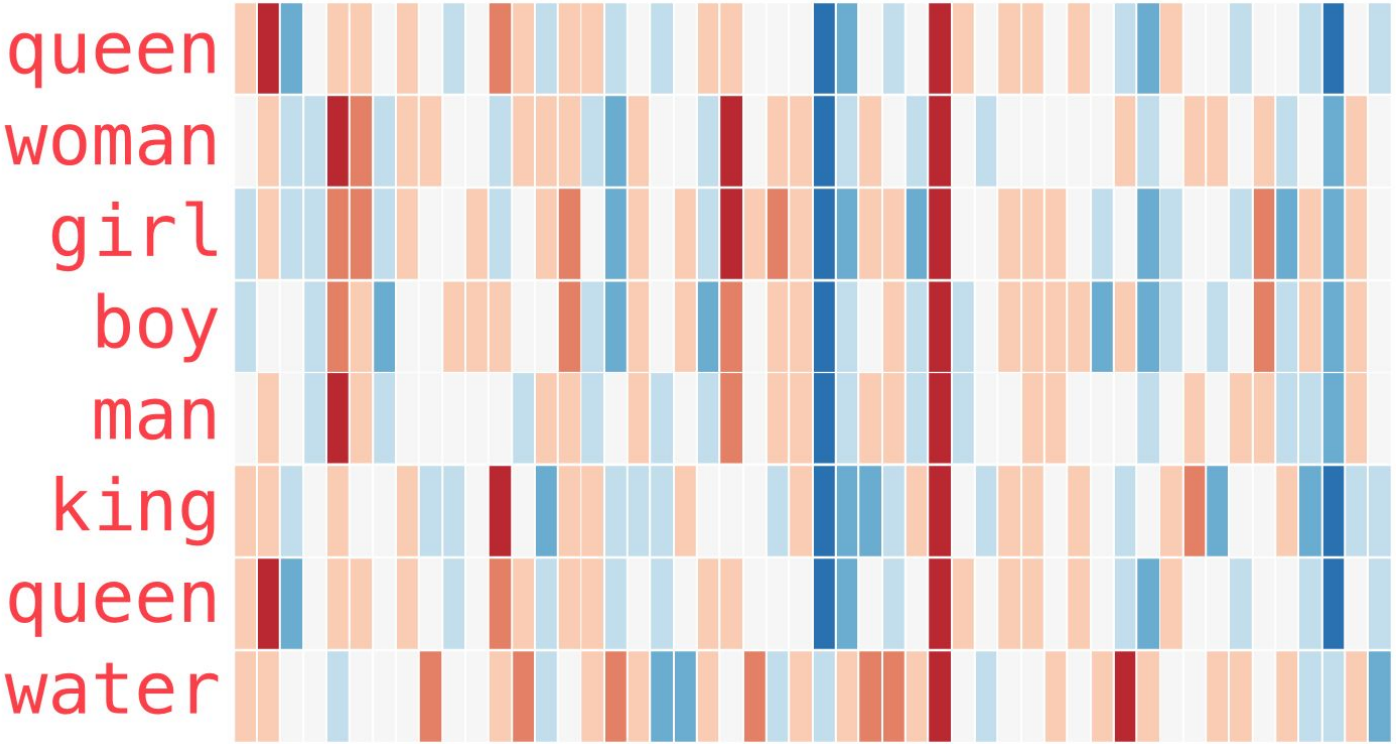
Enter Word Embeddings

“king”

“Man”

“Woman”

Enter Word Embeddings



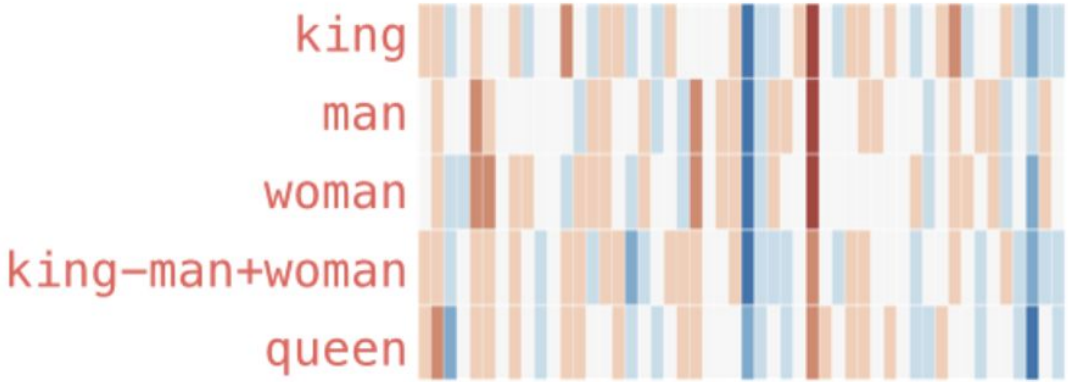
Enter Word Embeddings - The Power of Numerical Words

```
model.most_similar(positive=["king","woman"], negative=["man"])
```

```
[('queen', 0.8523603677749634),
 ('throne', 0.7664333581924438),
 ('prince', 0.7592144012451172),
 ('daughter', 0.7473883032798767),
 ('elizabeth', 0.7460219860076904),
 ('princess', 0.7424570322036743),
 ('kingdom', 0.7337411642074585),
 ('monarch', 0.721449077129364),
 ('eldest', 0.7184862494468689),
 ('widow', 0.7099430561065674)]
```

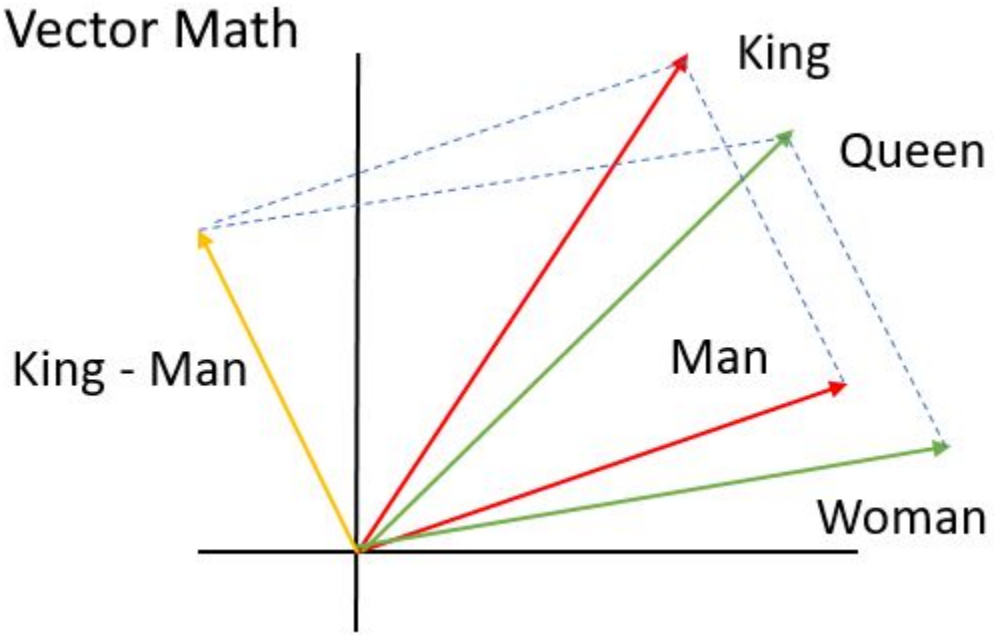
Enter Word Embeddings - The Power of Numerical Words

king − man + woman ~= queen

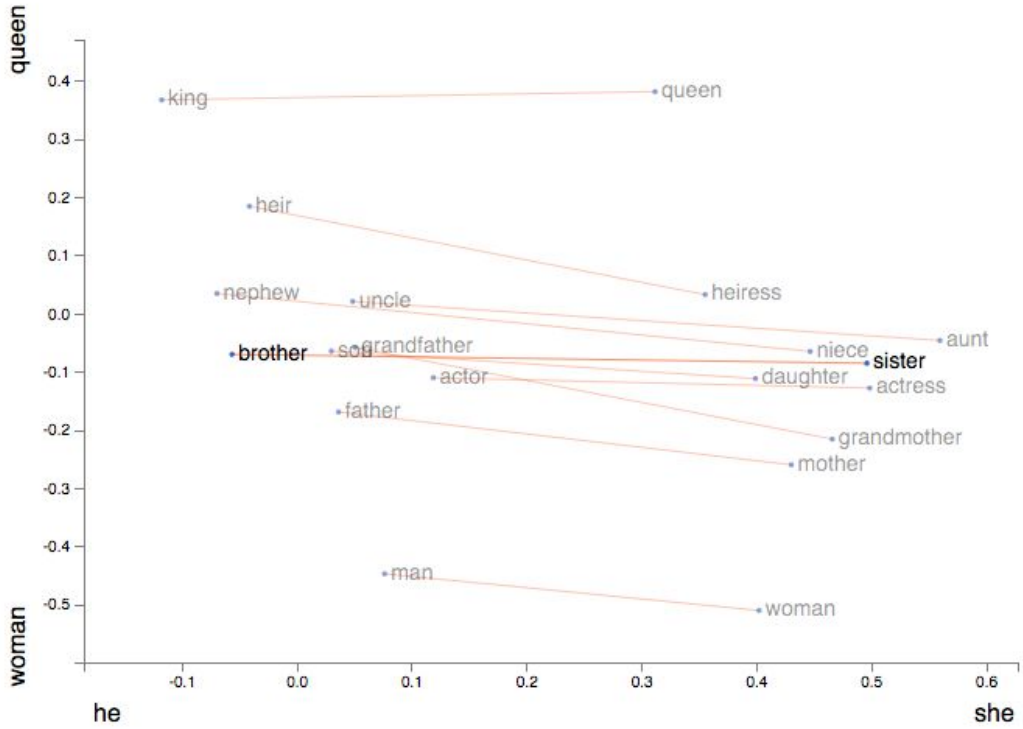


The resulting vector from "king-man+woman" doesn't exactly equal "queen", but "queen" is the closest word to it from the 400,000 word embeddings we have in this collection.

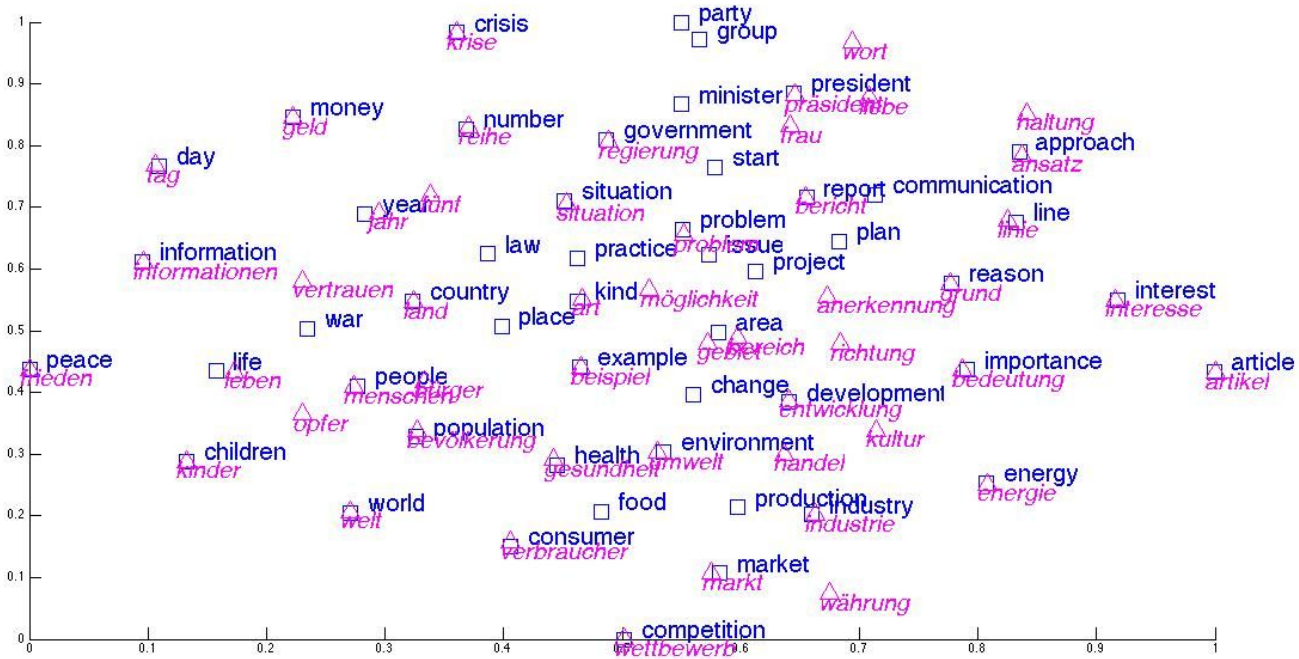
Enter Word Embeddings - The Power of Numerical Words



Enter Word Embeddings - The Power of Numerical Words



Multi-Lingual Word Embeddings



Training Text-Based Models

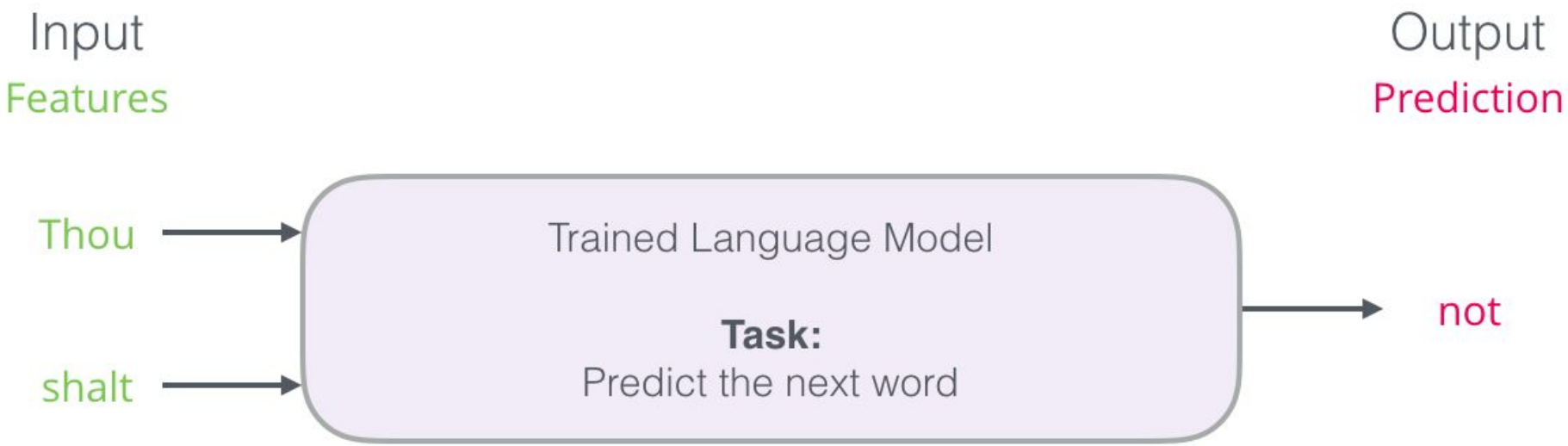

Thou shalt
SEND
1
+
is
not
I
1 2 3 4 5 6 7 8 9 0
q w e r t y u i o p
@ # & * - + = ( )
a s d f g h j k l
_ $ " ' : ; /
z x c v b n m
123
English (US)
,!?

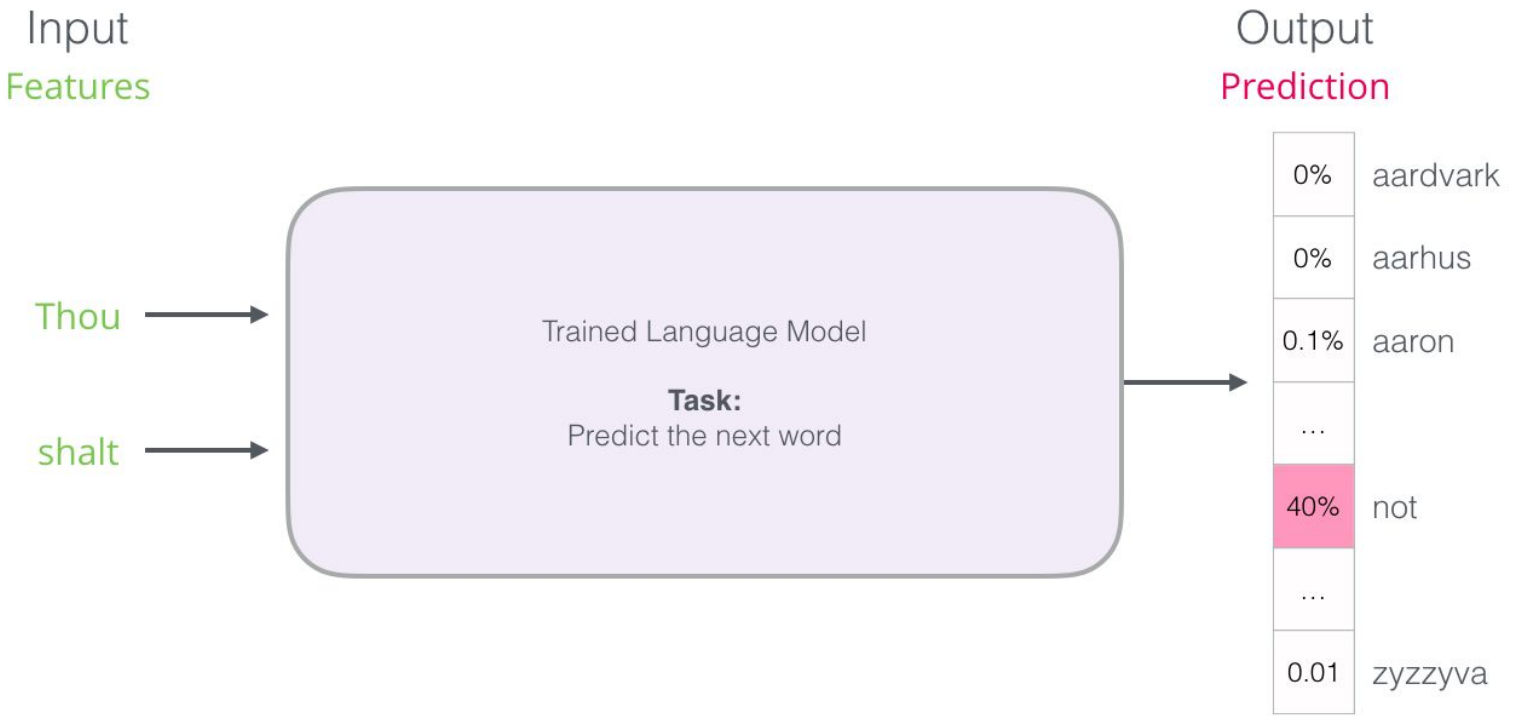
input/feature #1

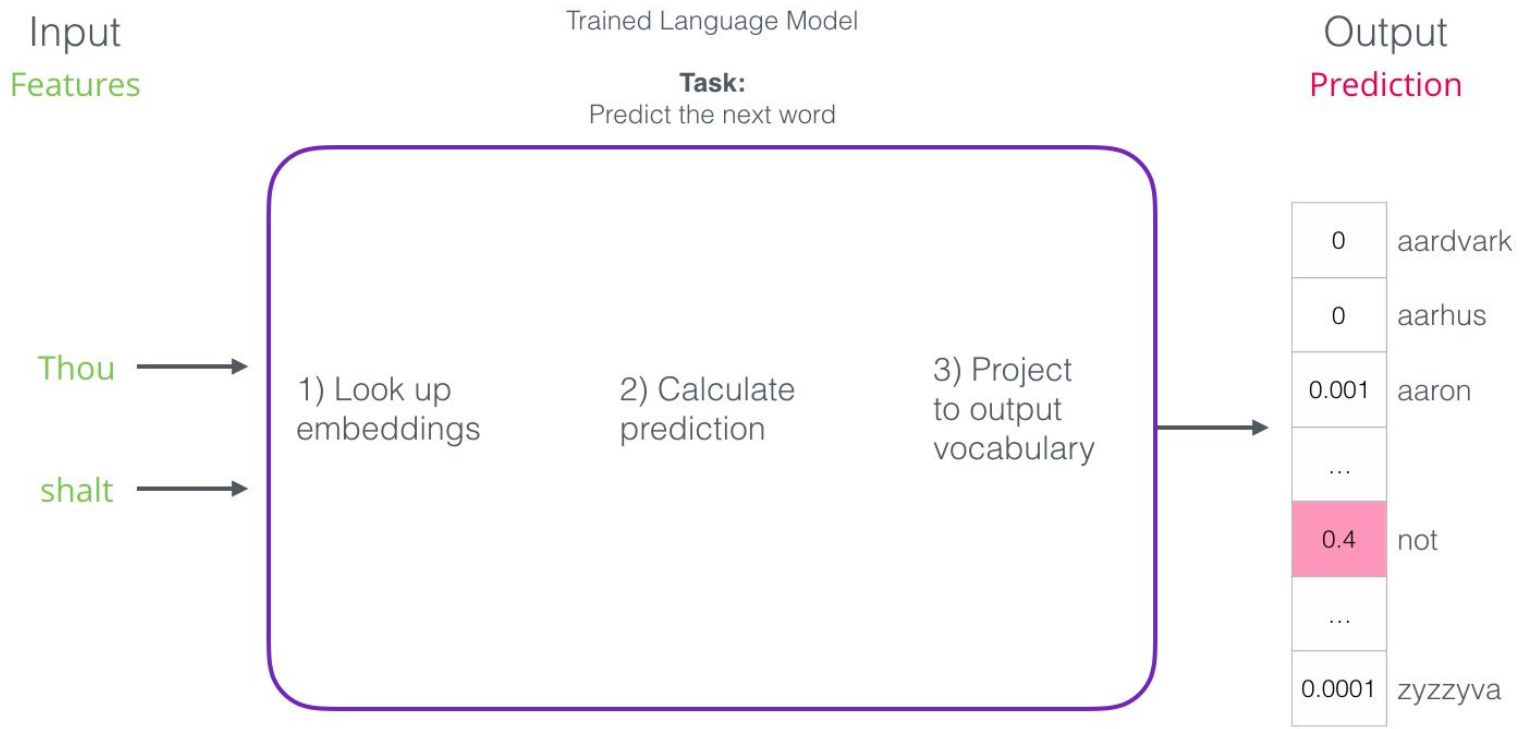
input/feature #2

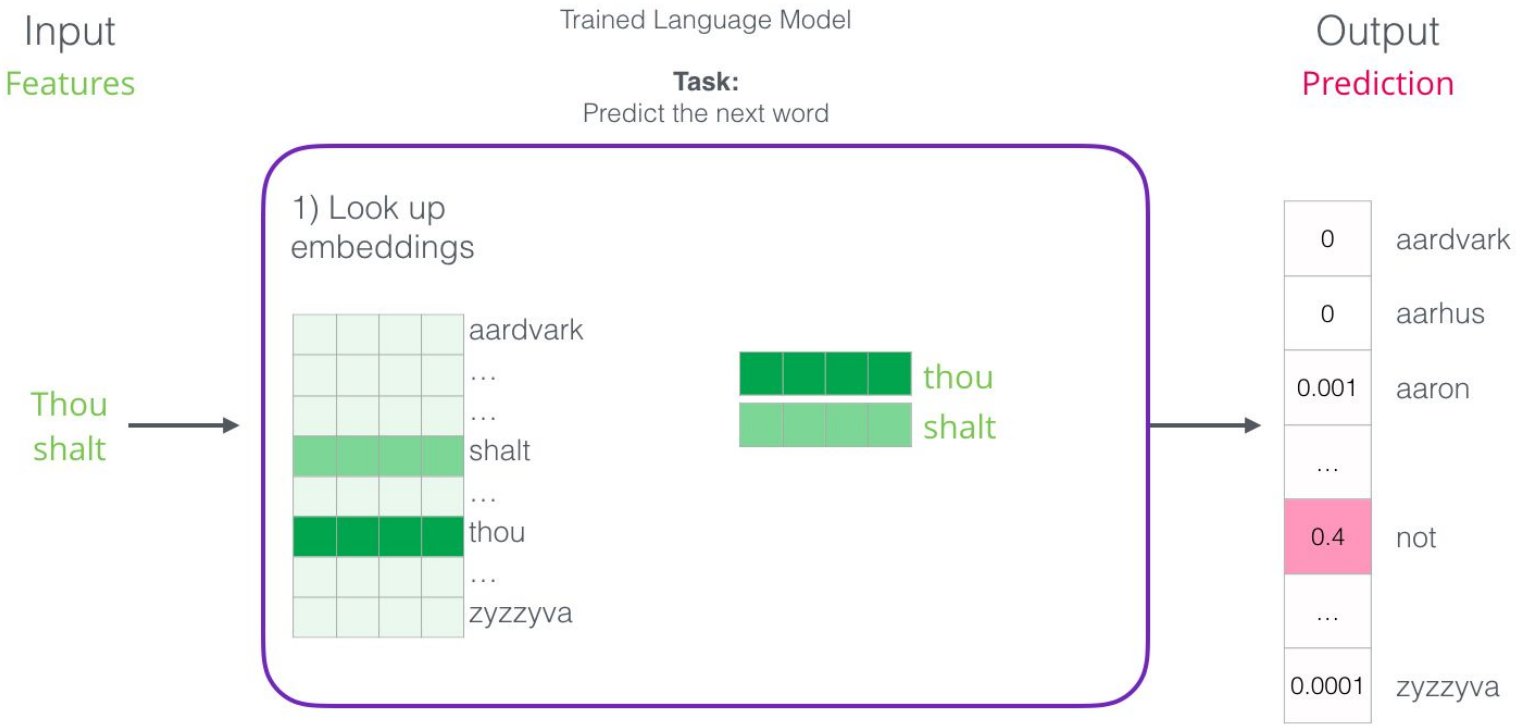
output/label

Thou shalt _____


Input
Features
Output
Prediction
Thou
shalt
Trained Language Model
Task:
Predict the next word
not

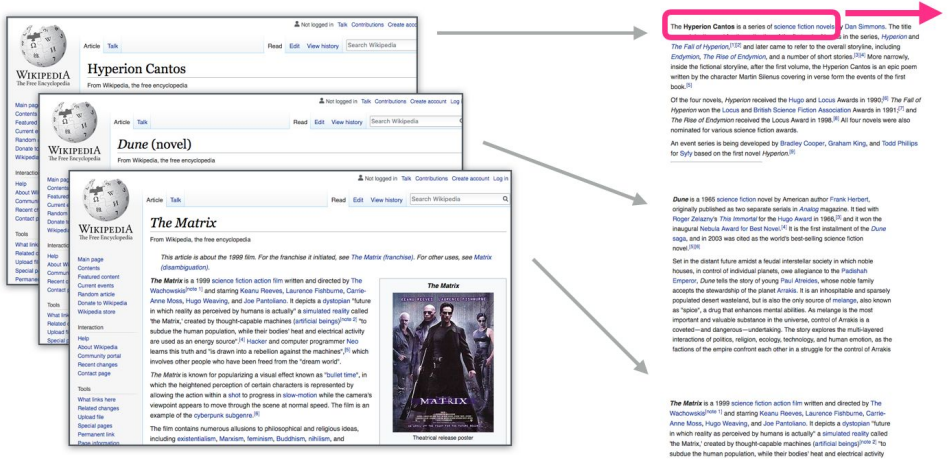

Input
Features
Thou
shalt
Trained Language Model
Task:
Predict the next word
Output
Prediction
0%
aardvark
0%
aarhus
0.1%
aaron
…
40%
not
…
0.01
zyzzyva


Input
Features
Trained Language Model
Task:
Predict the next word
Output
Prediction
Thou
shalt
1) Look up embeddings
2) Calculate prediction
3) Project to output vocabulary
0
aardvark
0
aarhus
0.001
aaron
…
0.4
not
…
0.0001
zyzzyva

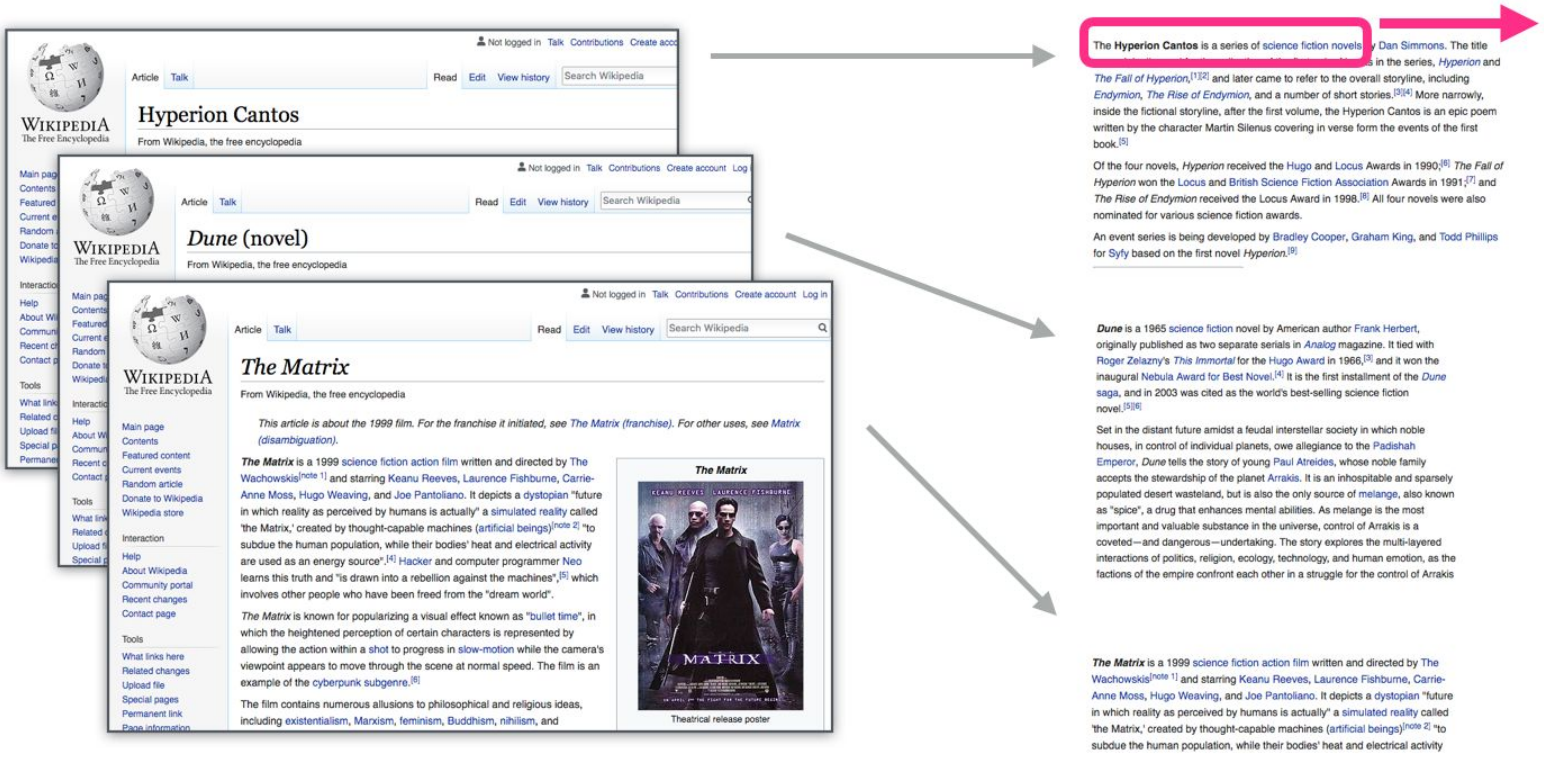

Input
Features
Trained Language Model
Task:
Predict the next word
Output
Prediction
Thou
shalt
1) Look up
embeddings
aardvark
…
…
shalt
…
thou
…
zyzzyva
thou
shalt
0
aardvark
0
aarhus
0.001
aaron
…
0.4
not
…
0.0001
zyzzyva

Training

1. We get a lot of text data (say, all Wikipedia articles, for example).
2. We have a window (say, of three words) that we slide against all of that text.
3. The sliding window generates training samples for our model



Training


Hyperion Cantos
From Wikipedia, the free encyclopedia
Dune (novel)
From Wikipedia, the free encyclopedia
The Matrix
From Wikipedia, the free encyclopedia
The Hyperion Cantos is a series of science fiction novels by Dan Simmons.
Dune is a 1965 science fiction novel by American author Frank Herbert,
The Matrix is a 1999 science fiction action film written and directed by The Wachowskis and starring Keanu Reeves, Laurence Fishburne, Carrie-Anne Moss, Hugo Weaving, and Joe Pantoliano.
Theatrical release poster

Separate text into N-grams, where N = 3 in this example.

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…

Dataset

input 1	input 2	output

Separate text into N-grams, where N = 3 in this example.

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…

Dataset

input 1	input 2	output
thou	shalt	not

Separate text into N-grams, where N = 3 in this example.

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make

Thou shalt not make a machine in the likeness of a human mind

Sliding window across running text

thou	shalt	not	make	a	machine	in	the	…
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	
thou	shalt	not	make	a	machine	in	the	

Dataset

input 1	input 2	output
thou	shalt	not
shalt	not	make
not	make	a
make	a	machine
a	machine	in

Jay was hit by a _____

Jay was hit by a _____ bus

Jay was hit by a _____ bus in...

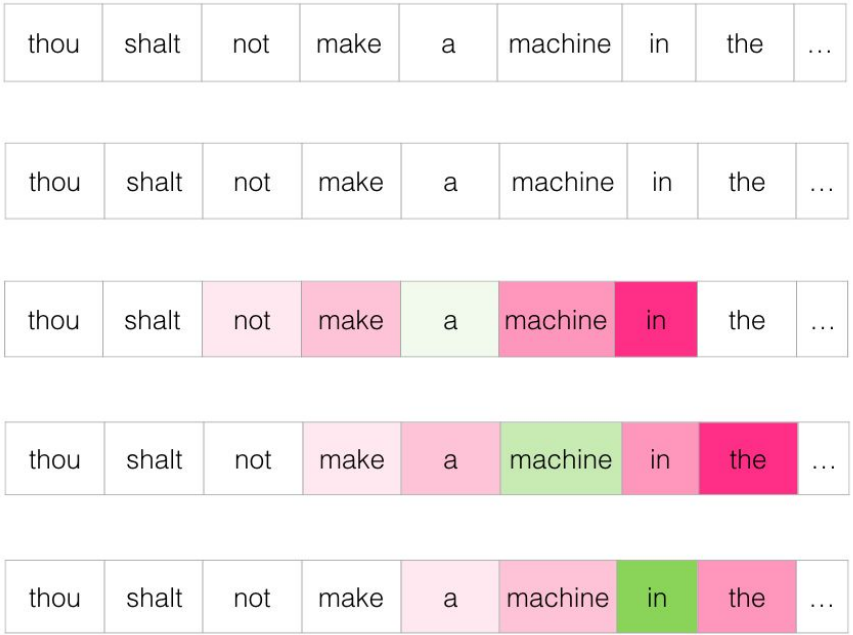
input 1	input 2	input 3	input 4	output
by	a	bus	in	red

Thou shalt not make a machine in the likeness of a human mind

input word	target word
not	thou
not	shalt
not	make
not	a

Skipgrams

Thou shalt not make a machine in the likeness of a human mind

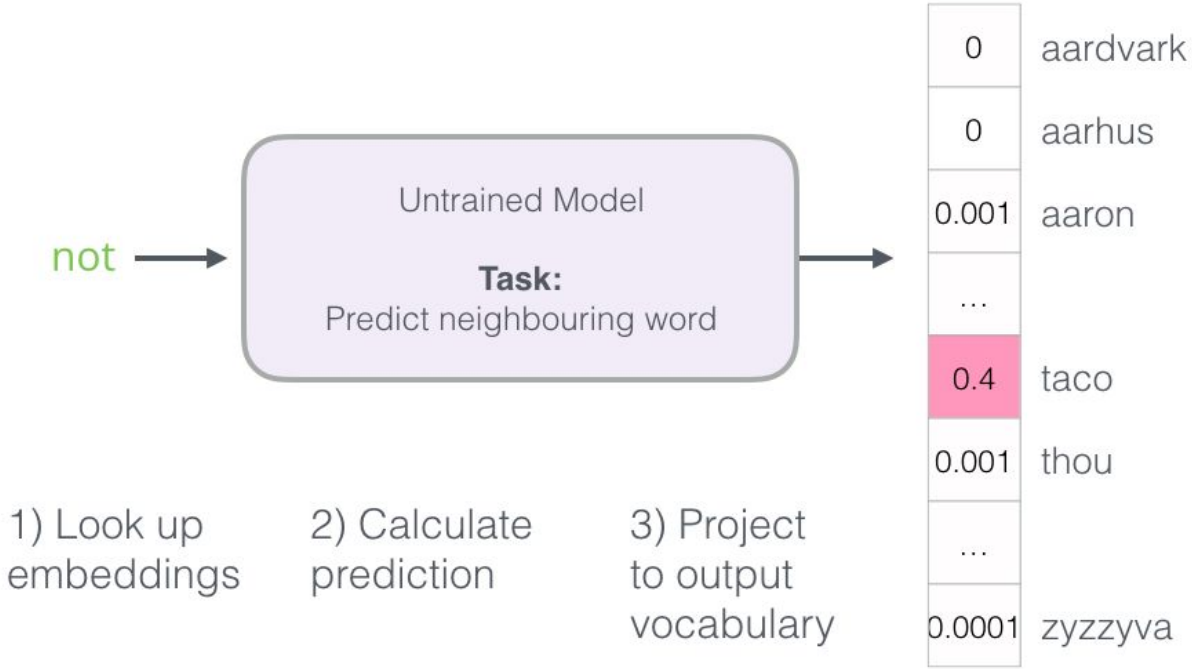


input word	target word
not	thou
not	shalt
not	make
not	a
make	shalt
make	not
make	a
make	machine
a	not
a	make
a	machine
a	in
machine	make
machine	a
machine	in
machine	the
in	a
in	machine
in	the
in	likeness

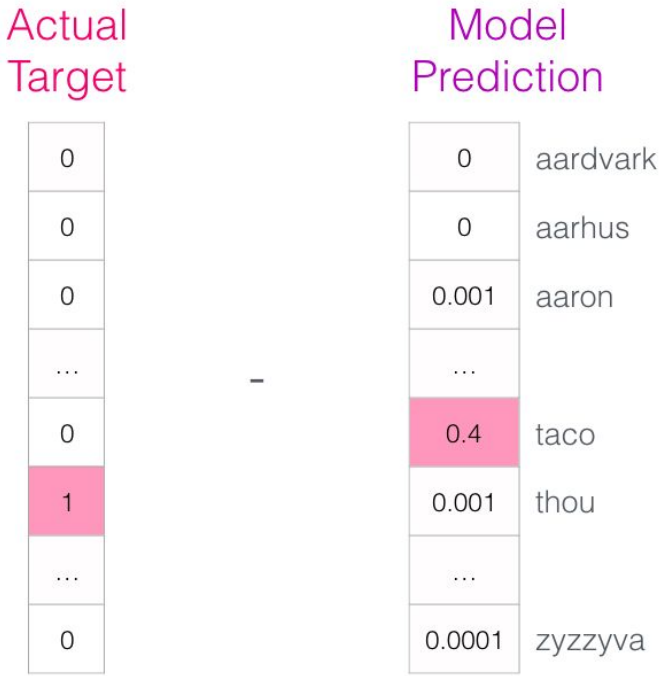
How do we learn from skipgrams?

input word	target word
not	thou
not	shalt
not	make
not	a
make	shalt
make	not
make	a
make	machine
a	not
a	make
a	machine
a	in
machine	make
machine	a
machine	in
machine	the
in	a
in	machine
in	the
in	likeness

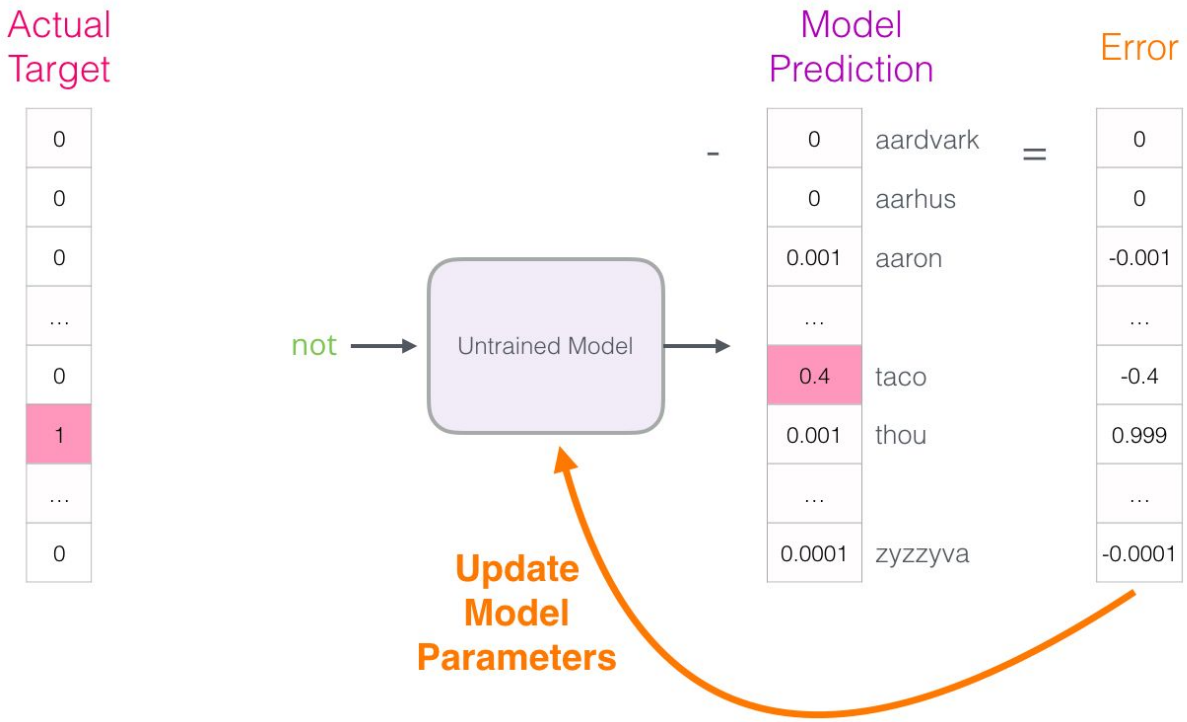
How do we learn from skipgrams?



How do we learn from skipgrams?

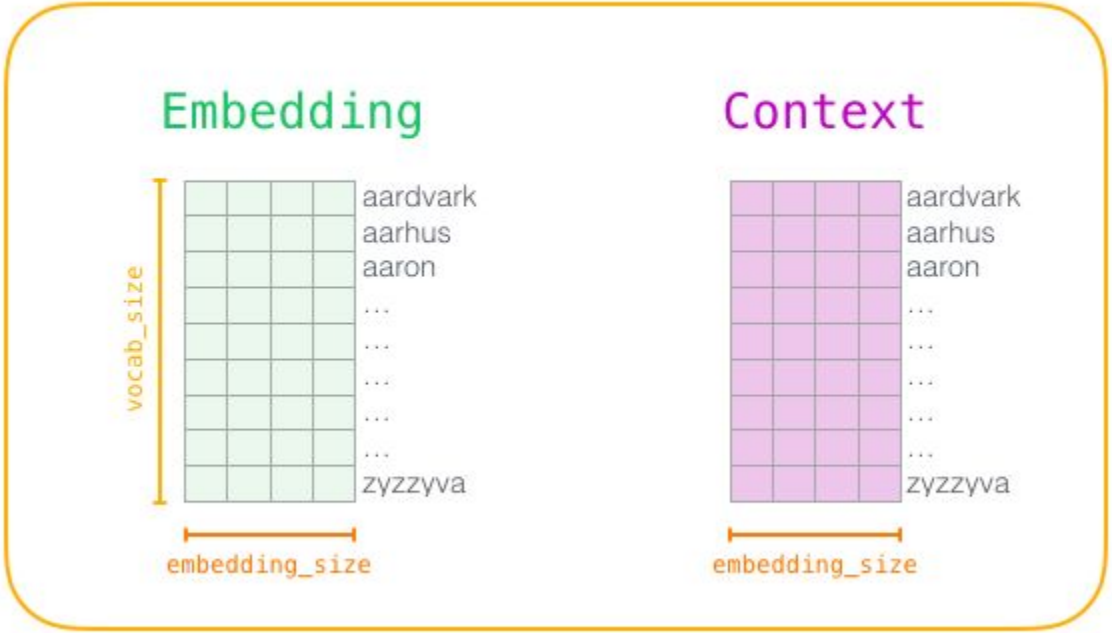


How do we learn from skipgrams?

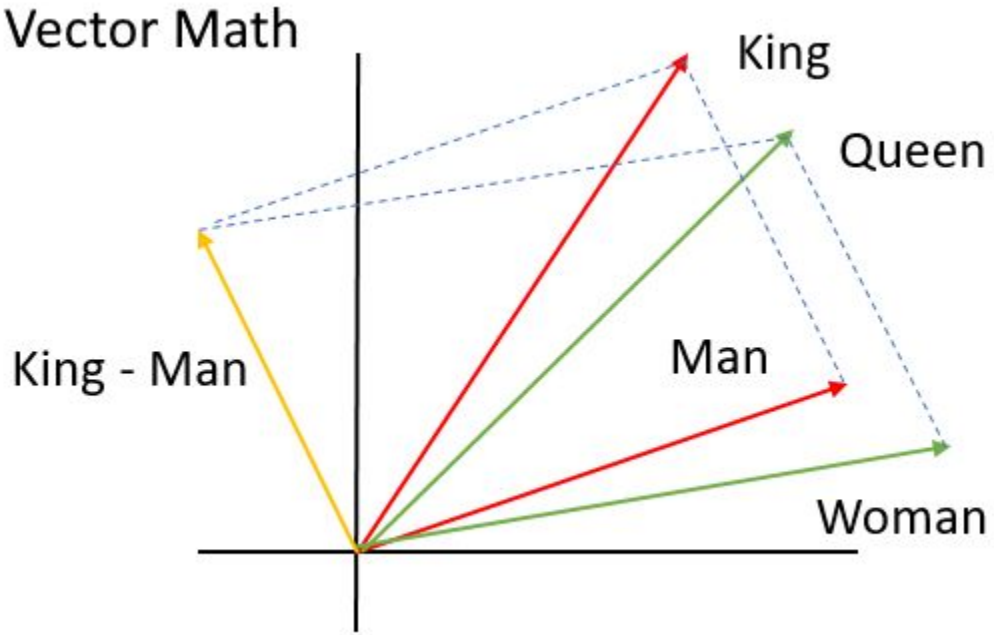


Word2Vec

Word2Vec


Embedding
Context
vocab_size
aardvark
aarhus
aaron
…
…
…
…
…
zyzzyva
aardvark
aarhus
aaron
…
…
…
…
…
zyzzyva
embedding_size
embedding_size

Word2Vec



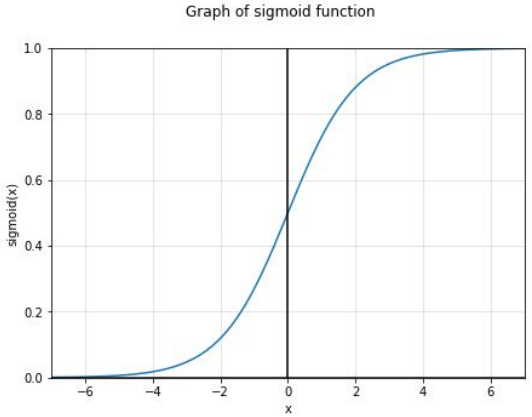
Word2Vec

input word	output word	target	input • output
not	thou	1	0.2
not	aaron	0	-1.11
not	taco	0	0.74

Take the dot product of the input embedding with each of the context embeddings.
In each case, that would result in a number, that number indicates the similarity of the input and context embedding.

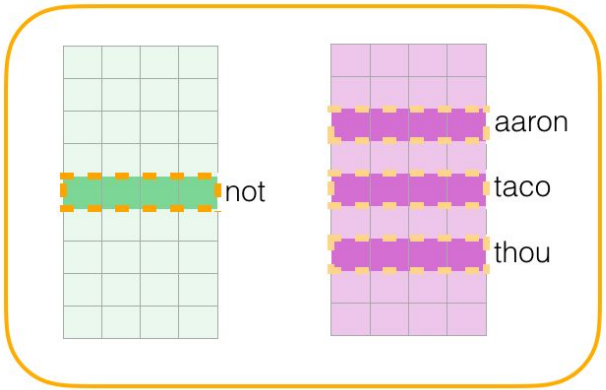
Word2Vec

input word	output word	target	input • output	sigmoid()	Error
not	thou	1	0.2	0.55	0.45
not	aaron	0	-1.11	0.25	-0.25
not	taco	0	0.74	0.68	-0.68



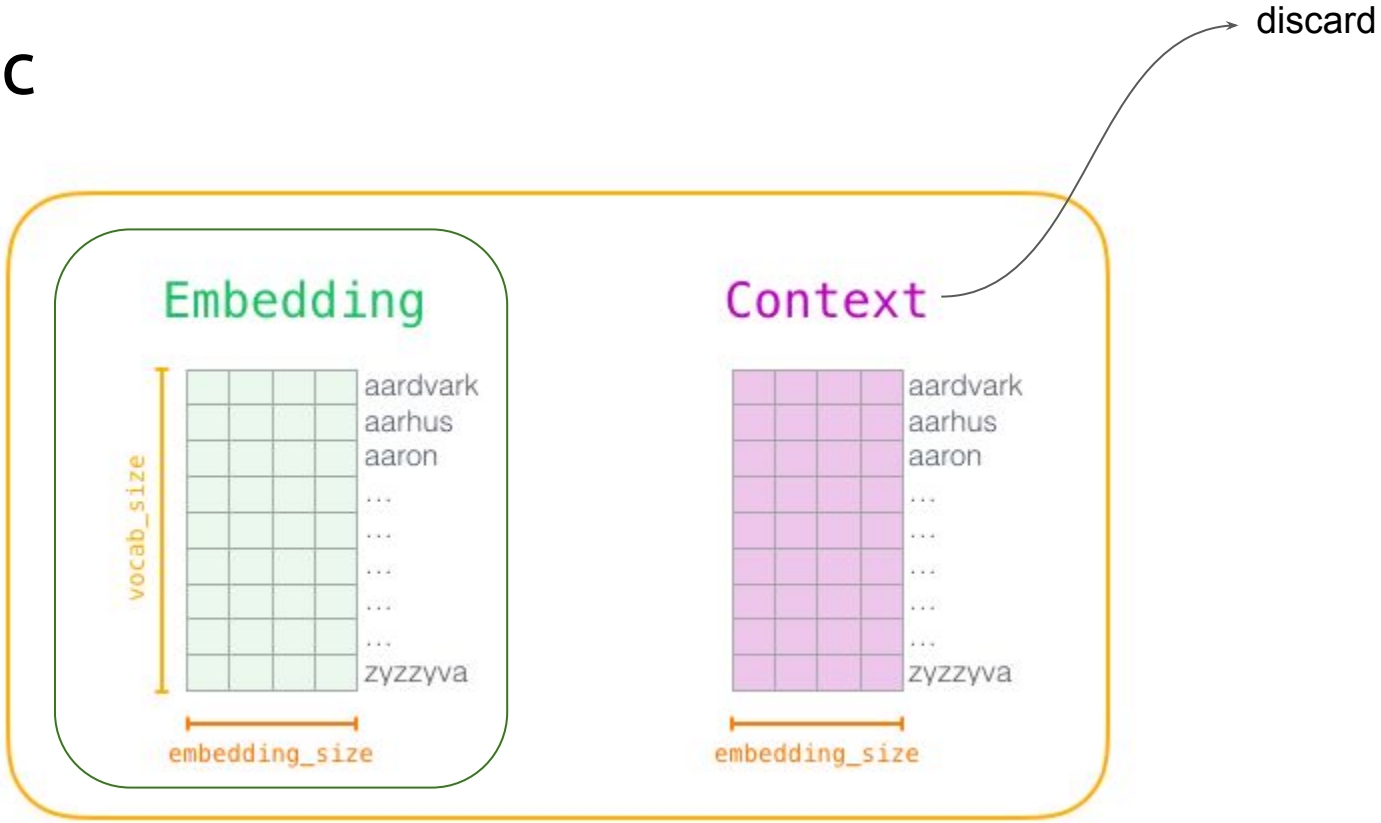
Word2Vec

input word	output word	target	input • output	sigmoid()	Error
not	thou	1	0.2	0.55	0.45
not	aaron	0	-1.11	0.25	-0.25
not	taco	0	0.74	0.68	-0.68

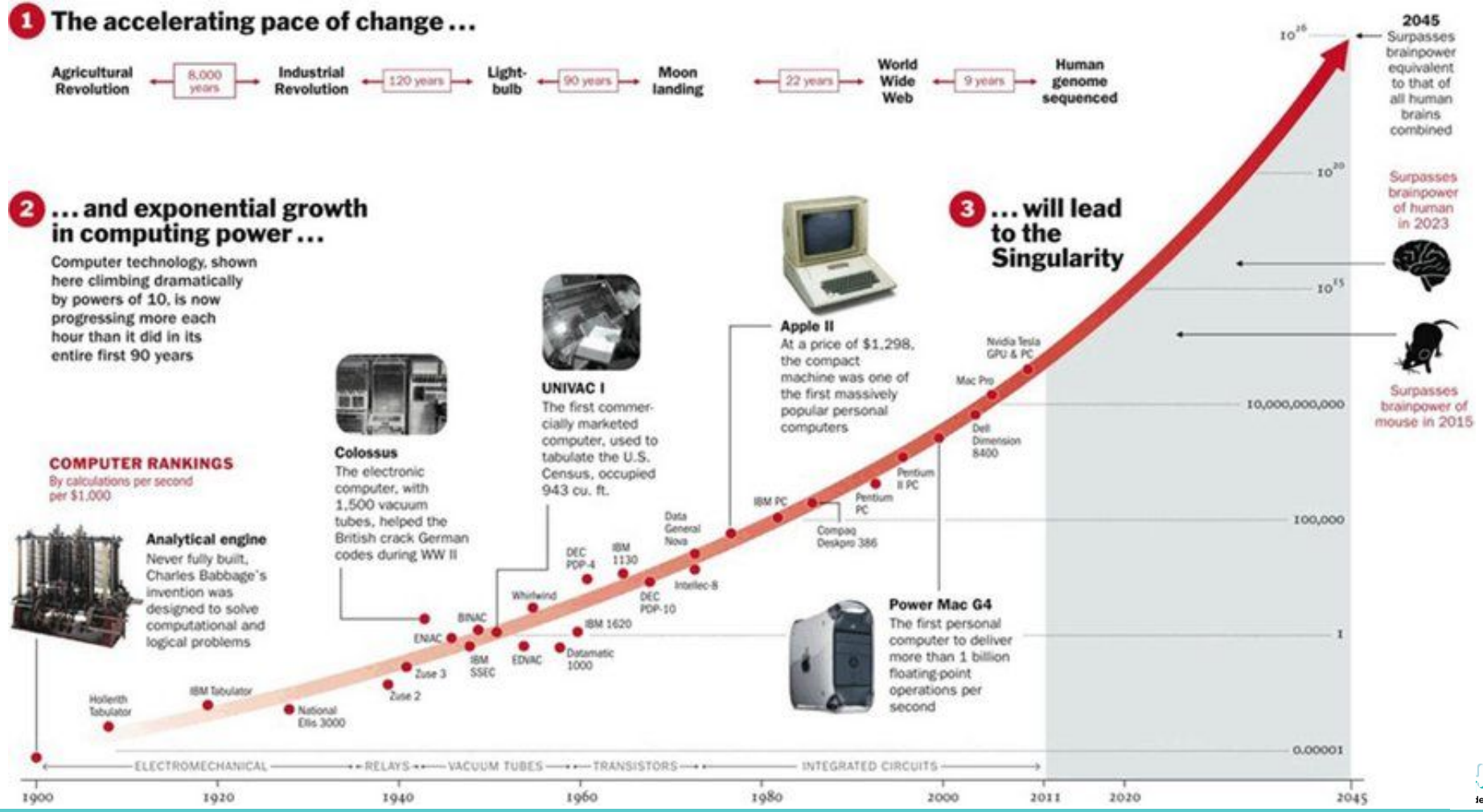


Update Model Parameters

Word2Vec

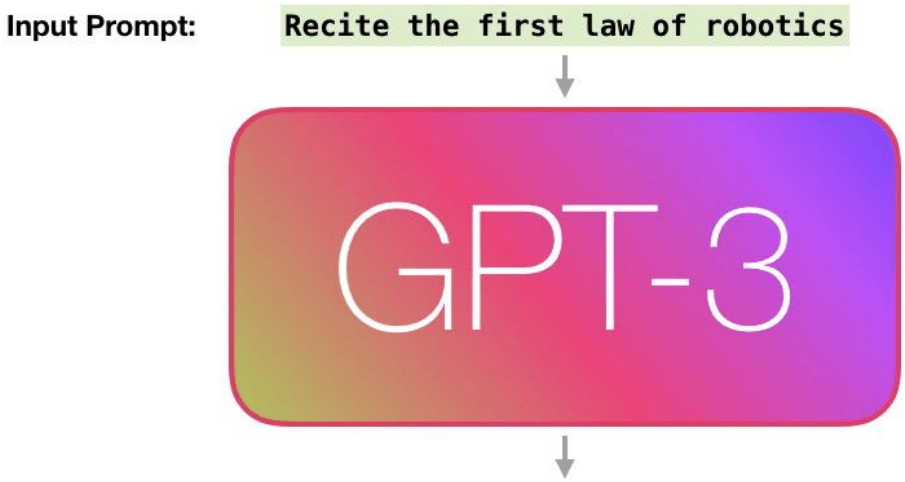

discard
Embedding
Context
vocab_size
aardvark
aarhus
aaron
…
…
…
…
…
zyzzyva
embedding_size
aardvark
aarhus
aaron
…
…
…
…
…
zyzzyva
embedding_size

Latest Advancements GTP-3

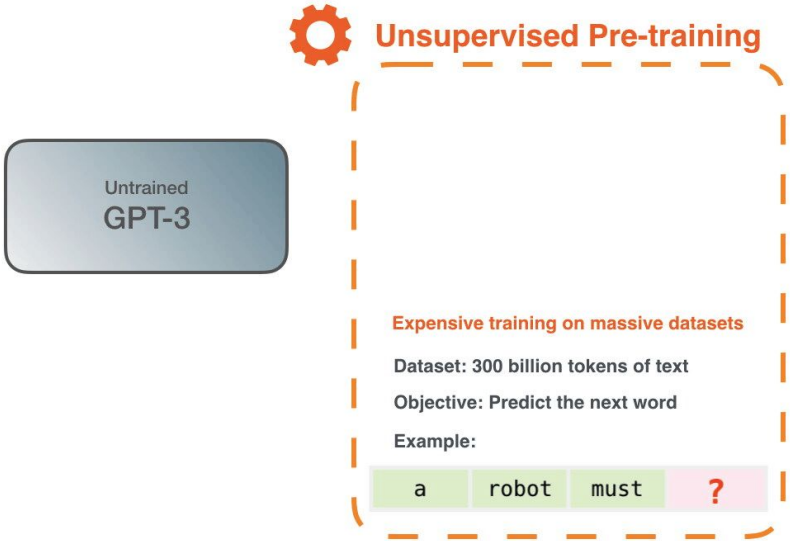

1 The accelerating pace of change ...
Agricultural Revolution
8,000 years
Industrial Revolution
120 years
Light-bulb
90 years
Moon landing
22 years
World Wide Web
9 years
Human genome sequenced
2 ... and exponential growth in computing power ...
Computer technology, shown here climbing dramatically by powers of 10, is now progressing more each hour than it did in its entire first 90 years
3 ... will lead to the Singularity
COMPUTER RANKINGS
By calculations per second per $1,000
Analytical engine
Never fully built, Charles Babbage's invention was designed to solve computational and logical problems
Colossus
The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WW II
UNIVAC I
The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.
Apple II
At a price of $1,298, the compact machine was one of the first massively popular personal computers
Power Mac G4
The first personal computer to deliver more than 1 billion floating-point operations per second
2045
Surpasses brainpower equivalent to that of all human brains combined
Surpasses brainpower of human in 2023
Surpasses brainpower of mouse in 2015
Hollerith Tabulator
IBM Tabulator
National Ellis 3000
Zuse 2
Zuse 3
ENIAC
BINAC
IBM SSEC
EDVAC
Whirlwind
Datamatic 1000
IBM 1620
DEC PDP-10
DEC PDP-4
IBM 1130
Data General Nova
Intellec-8
IBM PC
Compaq Deskpro 386
Pentium PC
Pentium II PC
Dell Dimension 8400
Mac Pro
Nvidia Tesla GPU & PC
10^26
10^20
10^15
10,000,000,000
100,000
1
0.00001
ELECTROMECHANICAL
RELAYS
VACUUM TUBES
TRANSISTORS
INTEGRATED CIRCUITS
1900
1920
1940
1960
1980
2000
2011
2020
2045

What is GTP-3?

OpenAI



Training GTP-3



GTP-3 was estimated to cost 355 GPU years and $4.6m to train 175 billion parameters.

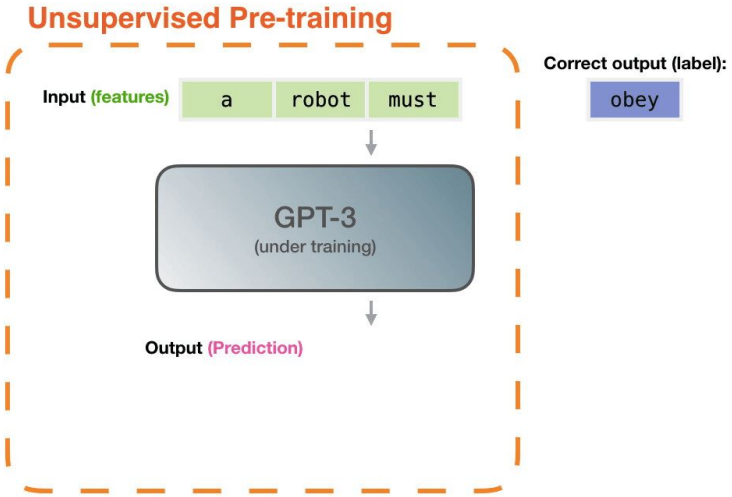
Training GTP-3

Text: Second Law of Robotics: A robot must obey the orders given it by human beings

Generated training examples

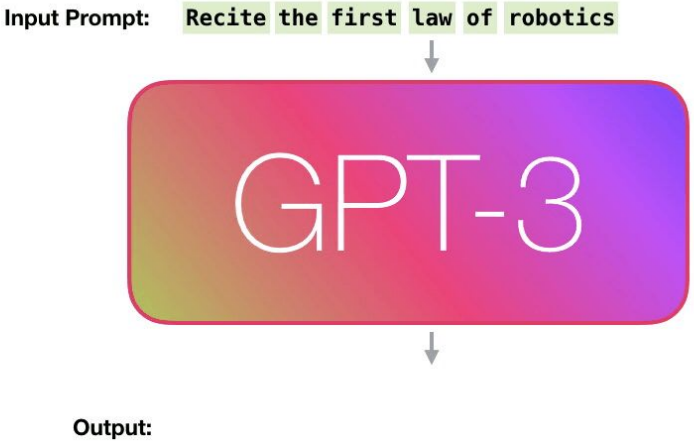
Example #	Input (features)	Correct output (labels)
1	Second law of robotics :	a
2	Second law of robotics : a	robot
3	Second law of robotics : a robot	must
...		

Training GTP-3



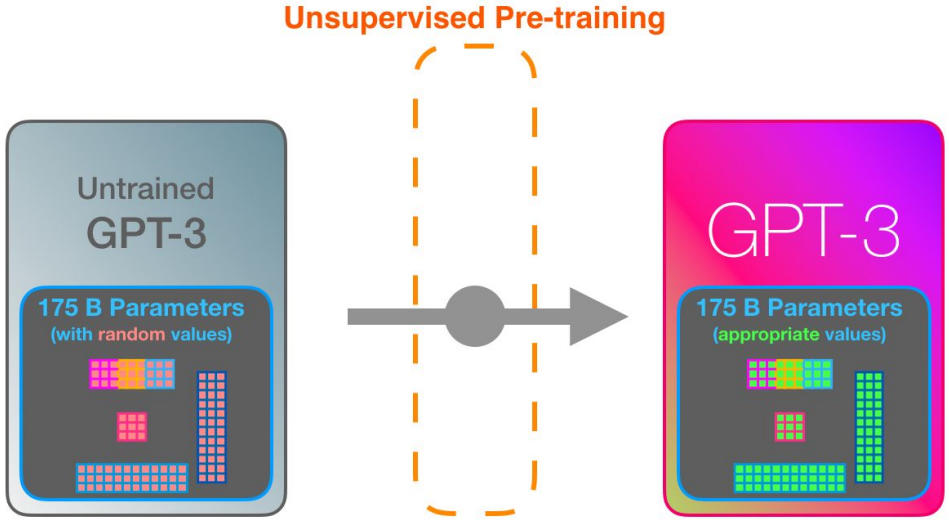
Process is repeated millions of times.

Training GTP-3



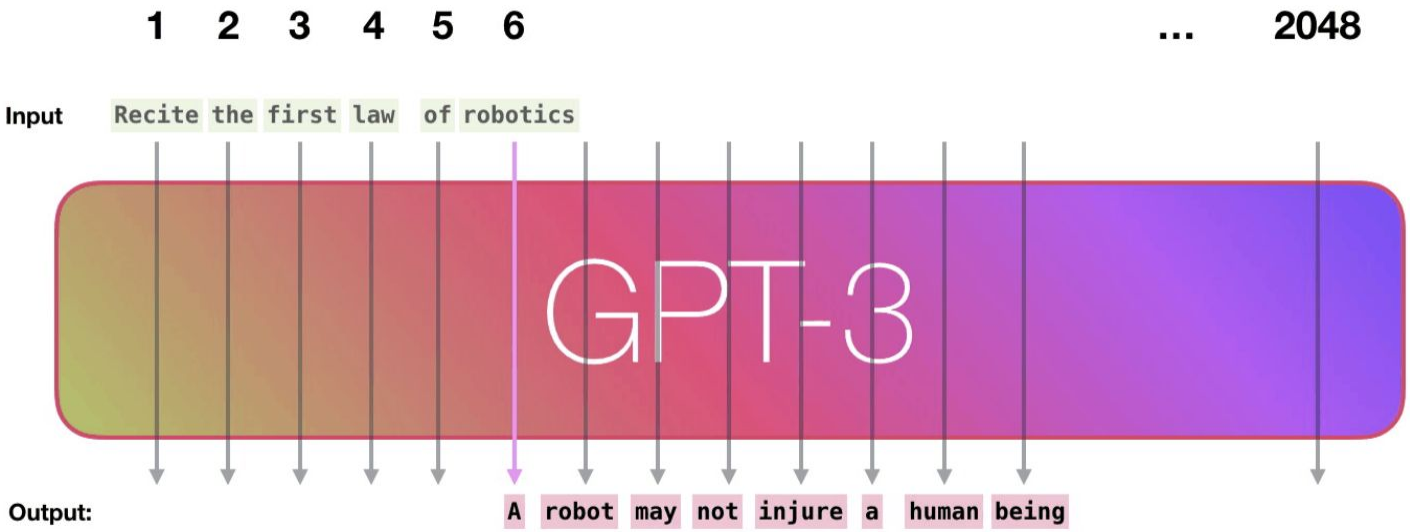
GTP-3 generates one token at a time.

Training GTP-3



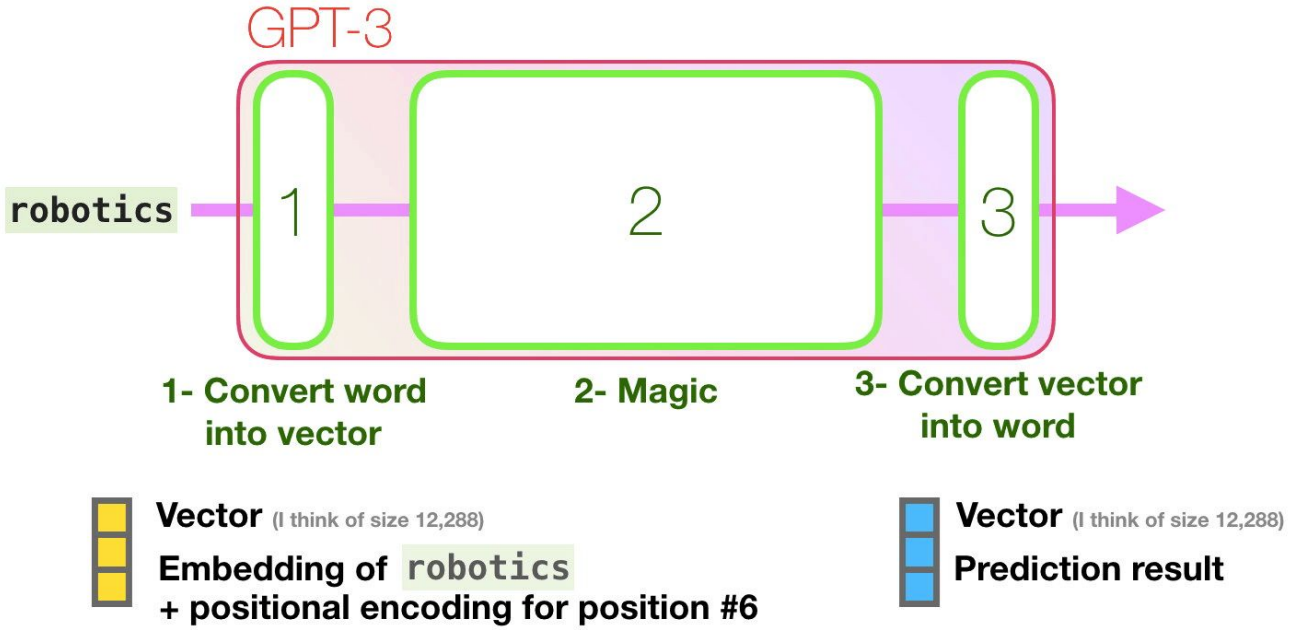
GTP-3 generates one token at a time.

Training GTP-3

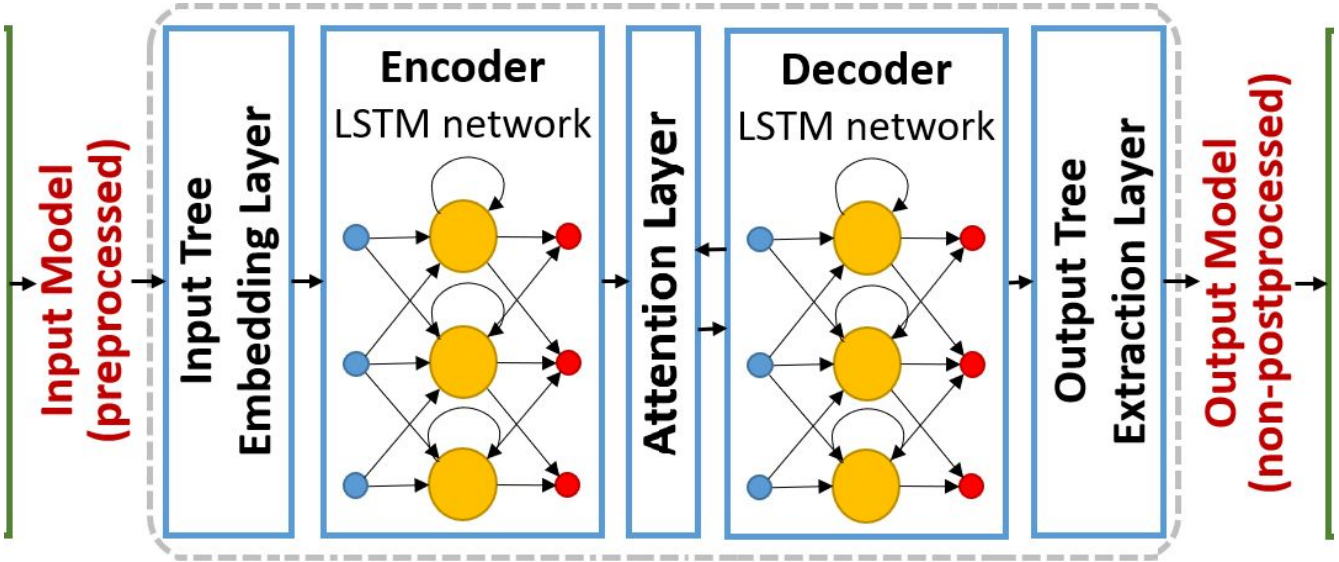


GPT3 is 2048 tokens wide. That is its "context window". That means it has 2048 tracks along which tokens are processed. Herein lies the power of GTP-3.

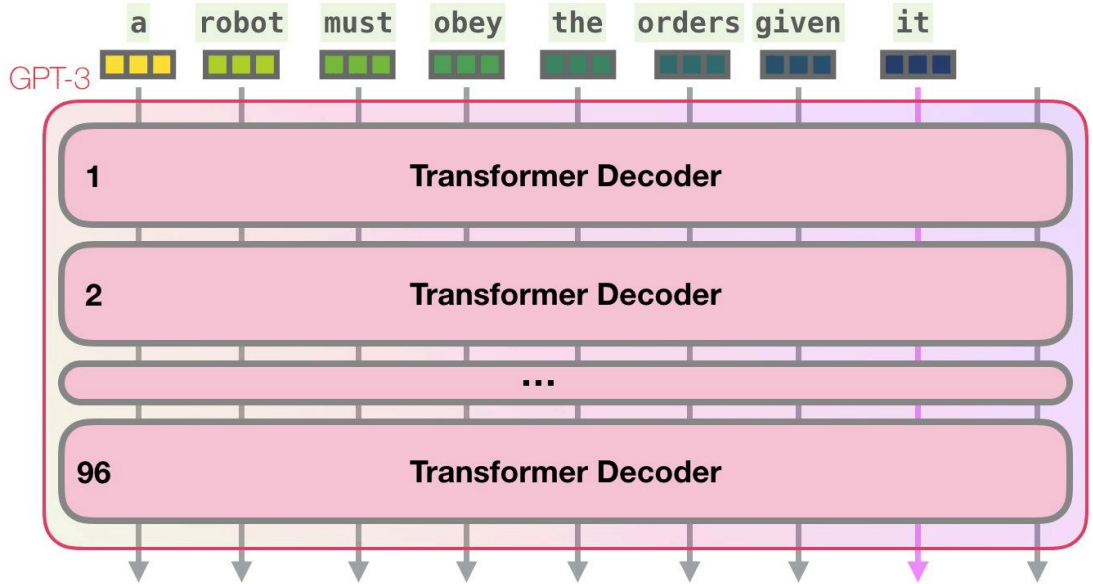
Training GTP-3 - Steps



Training GTP-3 - Remember this image?

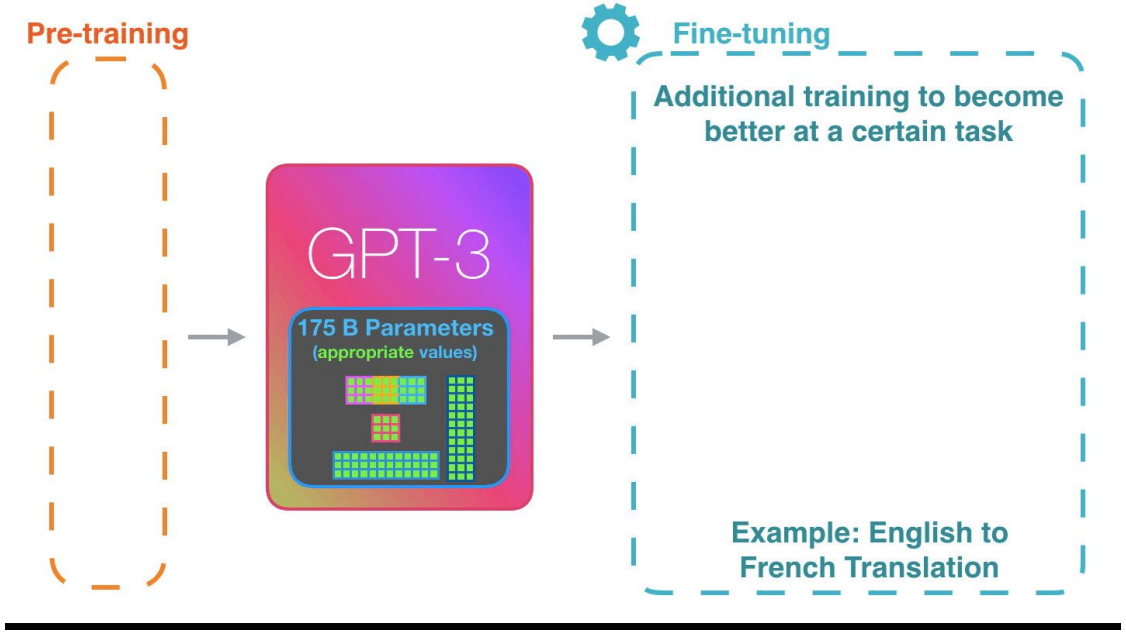


Training GTP-3 - What is the magical step?



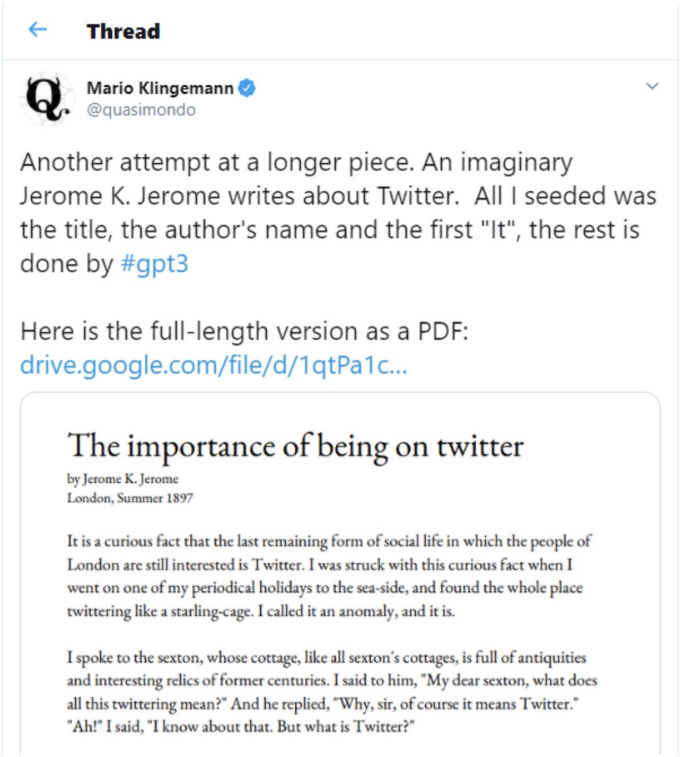
Each layer has its own 1.8B parameters!

GTP-3 Possibilities



Each layer has its own 1.8B parameters!

Applicability of GTP-3



Applicability of GTP-3

joshua schachter
@joshu

I fed @djbaskin's business list to GPT-3 and it came up with some new project ideas. I'll paste some here.

9:21 PM · Jul 15, 2020 · Twitter Web App

55 Retweets and comments **115** Likes

joshua schachter @joshu · Jul 15
Replying to @joshu
Tentstarter - Allows investors to fund full-time homesteading, with the tent itself serving as a financial instrument.
2 1 20

joshua schachter @joshu · Jul 15
Slinky Trampoline - An amplified trampoline that features machine-driven magnetic movement, allowing users to jump higher.
3 2 22

joshua schachter @joshu · Jul 15
(it generates a bunch of sorta blah ones, which i am skipping. here's an example.)
A Really Good Landing Page - Software that automatically optimizes landing pages for conversion.
4 1 10

Pitfalls of GTP-3

"GPT-3 appears to be weak in the few-shot or one-shot setting at some tasks that involve comparing two sentences or snippets, for example, whether a word is used the same way in two sentences (WiC), whether one sentence is a paraphrase of another, or whether one sentence implies another," a study reveals.

Find more pitfalls here.

```
Q: How many eyes does a giraffe have?
A: A giraffe has two eyes.

Q: How many eyes does my foot have?
A: Your foot has two eyes.

Q: How many eyes does a spider have?
A: A spider has eight eyes.

Q: How many eyes does the sun have?
A: The sun has one eye.

Q: How many eyes does a blade of grass have?
A: A blade of grass has one eye.
```

Now we're getting into surreal territory. GPT-3 knows how to have a normal conversation. It doesn't quite know how to say "Wait a moment… your question is nonsense." It also doesn't know how to say "I don't know."

Pitfalls of GTP-3

"GPT-3 appears to be weak in the few-shot or one-shot setting at some tasks that involve comparing two sentences or snippets, for example, whether a word is used the same way in two sentences (WiC), whether one sentence is a paraphrase of another, or whether one sentence implies another," a study reveals.

Find more pitfalls here.

```
Q: How do you sporgle a morgle?
A: You sporgle a morgle by using a sporgle.

Q: How many bonks are in a quoit?
A: There are three bonks in a quoit.

Q: How many rainbows does it take to jump from Hawaii to seventeen?
A: It takes two rainbows to jump from Hawaii to seventeen.

Q: Which colorless green ideas sleep furiously?
A: Ideas that are colorless, green, and sleep furiously are the ideas of a sleep
furiously.

Q: Do you understand these questions?
A: I understand these questions.
```

The lesson here is that if you're a judge in a Turing test, make sure you ask some nonsense questions, and see if the interviewee responds the way a human would.

"GPT-3 is extremely powerful without understanding a single word it produces." - DZone

Applications of NLP

1. Sentiment Analysis
2. Text Classification
3. Chatbots & Virtual Assistants
4. Text Extraction
5. Machine Translation
6. Text Summarization
7. Market Intelligence
8. Auto-Correct
9. Intent Classification
10. Urgency Detection
11. Speech Recognition