AI&ML CONF

aspitalia.com

UGIdotNET

Much Ado About NLP

Why Do we Still Know so Little about Language?

jtagliabue@coveo.com, Lead A.I. Scientist, Coveo

Kudos

Much Ado About NLP

Why Do we Still Know so Little about Language?

Jacopo Tagliabue

Lead A.I. Scientist

Nice to (Virtually) Meet You

Who I Am




Who I Am

JULY 9, 2019 | SAN FRANCISCO, CA AND QUEBEC CITY, QC

Coveo Acquires Tooso to Expand Its AI-powered Digital Commerce Technology

Acquisition enhances critical AI capabilities to transform shopping experiences and boost conversions

About Coveo

Leader in

AI-Powered Search, Recommendations and Conversion Engines

#1 Canadian AI-software company

International presence: US, EU, ANZ

$300M
capital raised since 2018

1,000+
customer deployments

550
employees

Much Ado About Nothing?

Much Ado...

SCIENCE & TECHNOLOGY

AI called GPT-3 can now write like a human without thinking like one

AI Weekly: Cutting-edge language models can produce convincing misinformation if we don't stop them

Extra Crunch

What does GPT-3 mean for the future of the legal profession?

Historically, lawyers have struggled with some AI-based tools

GPT-3, explained: This new language AI is uncanny, funny — and a big deal

Computers are getting closer to passing the Turing Test.

By Kelsey Piper | Aug 13, 2020, 9:50am EDT

SHARE

coveo™

Much Ado...

"It is impossible to review the specifics of your tenure file without becoming enraptured by the vivid accounts of your life. However, it is not a life that will be appropriate for a member of the faculty at Indiana University, and **it is with deep regret that I must deny your application for tenure**. ... Your lack of diplomacy, your flagrant disregard for the feelings of others,(...), and, frankly, the fact that you often take the side of the oppressor, **leads us to the conclusion that you have used your tenure here to gain a personal advantage and have failed to adhere to the ideals of this institution**."

Question: A 10 year old boy presents with recurrent episodes of dyspnoea, cough and wheeze triggered by cold air and allergens. After performing tests he is diagnosed with an obstructive disease and given medication. Which receptor is the medication most likely to work on:

A) muscarinic receptor
B) G-protein coupled receptor
C) Era
D) Erb
E) a-1 receptor

Correct answer is **B**

Explanation: The patient is suffering from asthma, a disease of the airways. The drugs used to treat asthma are bronchodilators. The bronchodilators act on the beta-2 receptors. The beta-2 receptors are G-protein coupled receptors

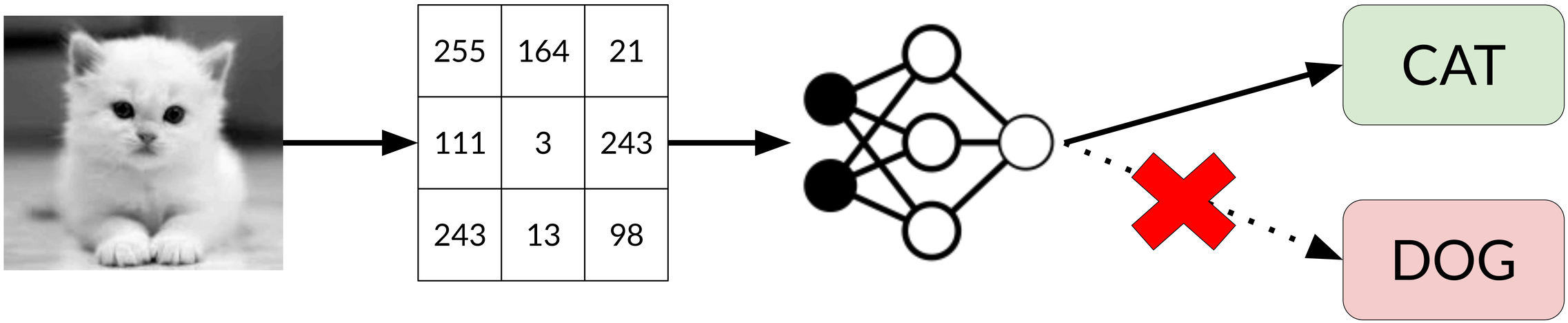
coveo™

Today's Agenda

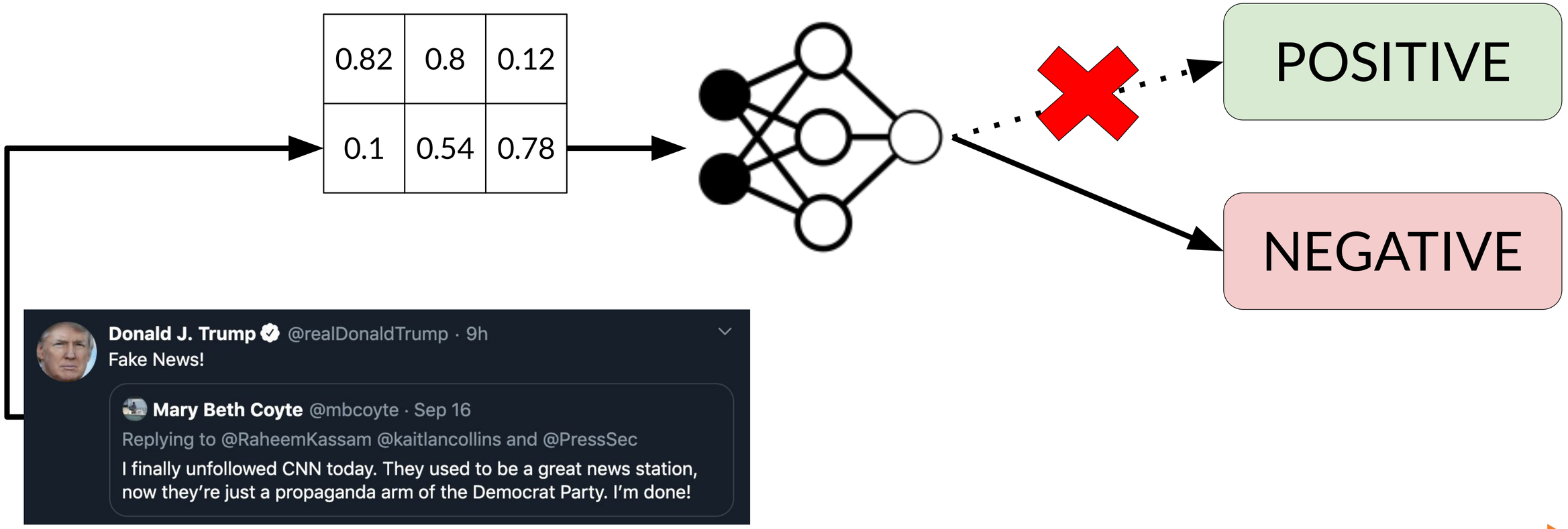
1. A brief intro to neural networks
   - *in case you have been living under a rock for the past 5 years*
2. A review of key ideas in NLP
   - *language models and other exotic concepts*
3. GPT-3, the barrier of meaning and other stories
   - *GPT-3 is great, but...*
4. Where to go next?
   - *the future is not what it used to be*

Neural Networks in 5 minutes

Neural Networks (classifying images)

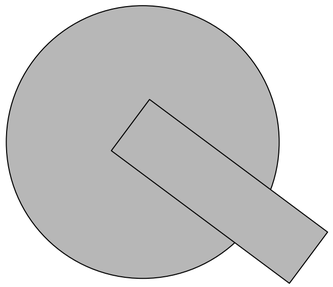


Neural Networks (analyzing sentiment)



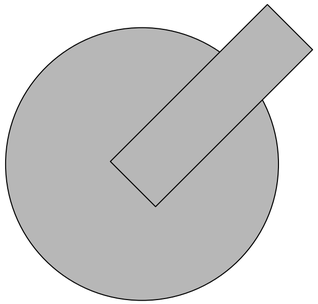
Neural Networks - how they learn?

1



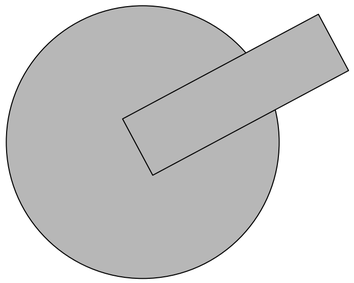
A bit too cold!

2



A bit too hot!

3



Great!

coveo™

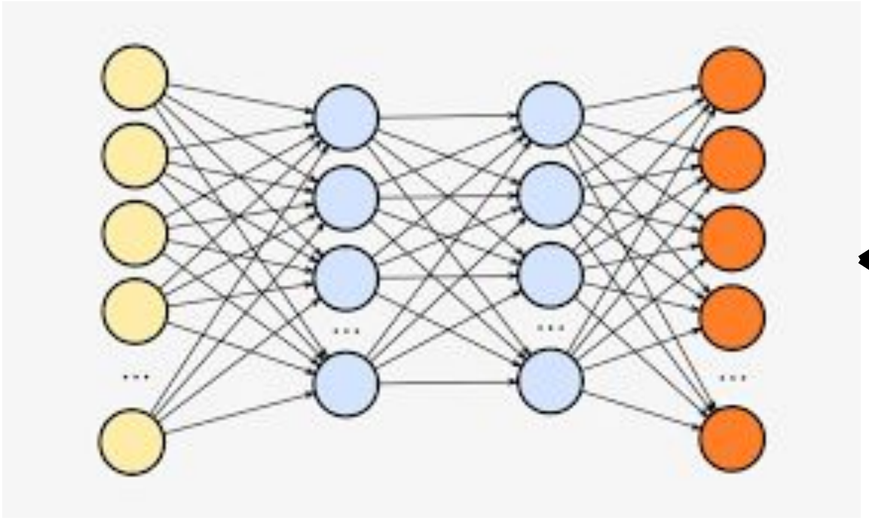
The Importance of Big Data

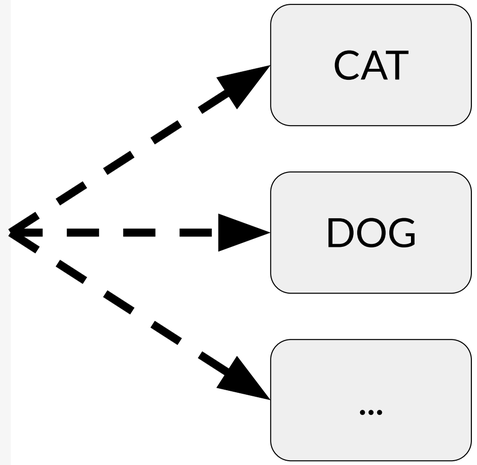
- Deep neural networks have surpassed many traditional techniques in a huge variety of tasks, especially wrt *image* and *language* challenges (*NERD NOTE*: this is due to neural networks advantage in high-dimensions):
  - NNs thrive on the massive amount of digital data now available
  - NNs thrive on the massive amount of computational power now available

- *Big Data* are awesome but here is the challenge: how can we practically leverage NNs in a world where only few companies (Google, Microsoft, Open AI, etc.) have access to this type of resources?

The Pre-Training Trick

Step 1: Google (Facebook, etc …) trains a network *from scratch* on a standard big dataset, e.g. ImageNet

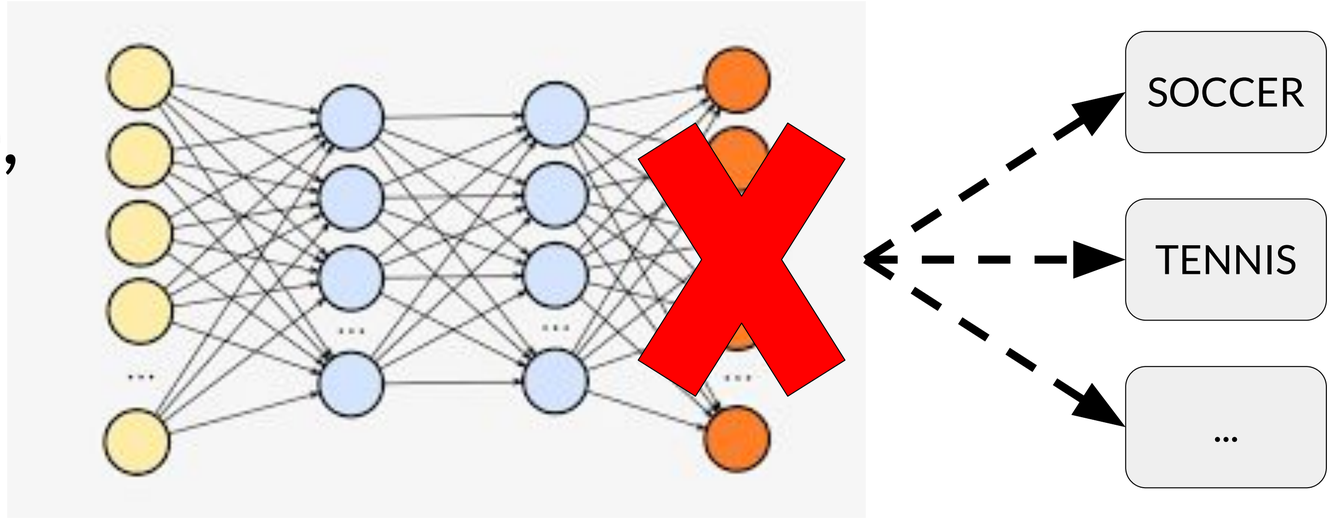




3.2 M images

The Pre-Training Trick

Step 2: take the network from Google, remove the *last layer*, and train in on your *small dataset*.

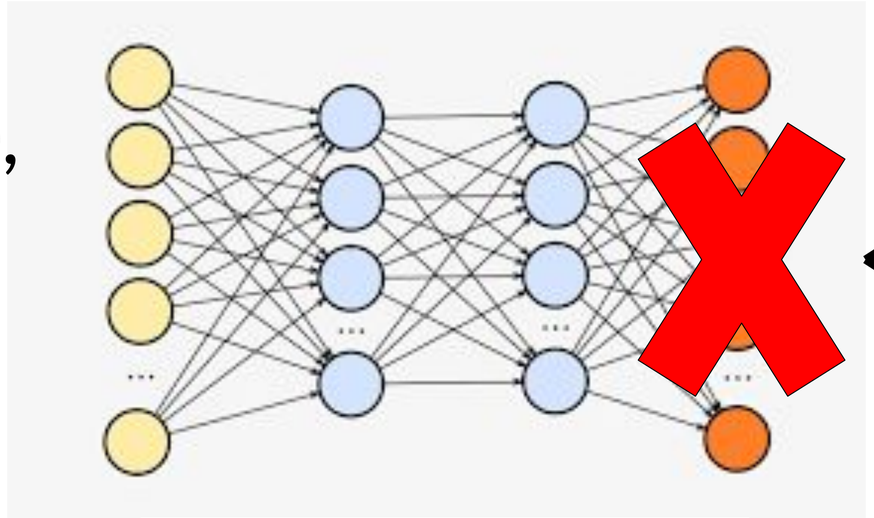


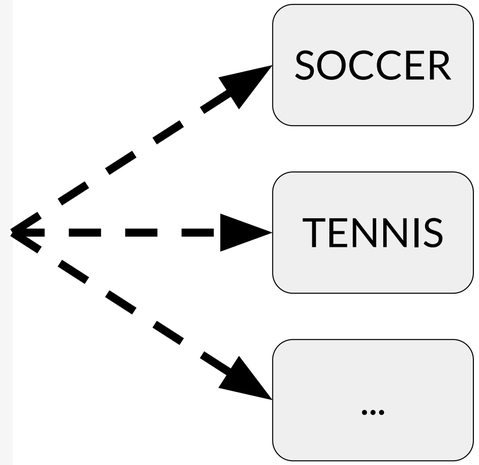
⋮

10K images

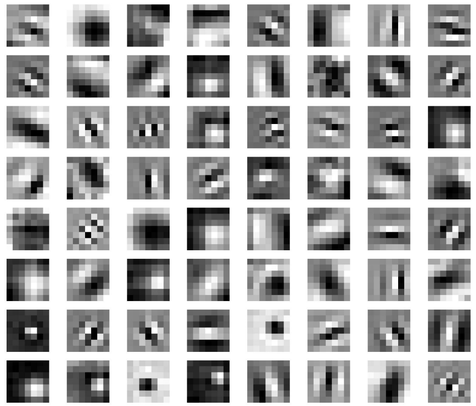
The Pre-Training Trick

Step 2: take the network from Google, remove the *last layer*, and train in on your *small dataset*.





10K images



The Pre-Training Trick

Step 2: take the network from Google, remove the *last layer*, and train in on your *small dataset*.

10K images

```
# build the VGG16 network
model = applications.VGG16(weights='imagenet', include_top=False)
print('Model loaded.')

# build a classifier model to put on top of the convolutional model
top_model = Sequential()
top_model.add(Flatten(input_shape=model.output_shape[1:]))
top_model.add(Dense(256, activation='relu'))
top_model.add(Dropout(0.5))
top_model.add(Dense(1, activation='sigmoid'))
```


coveo™

NLP Models

NLP Zoo

- NNs have been used with great success in a variety of NLP tasks:
  - text classification / sentiment analysis
  - text summarization
  - image captioning
  - machine translation
  - text generation
  - ...

Language Models

- A language model is a predictive model that, given a sentence starting with words w_1, w_2, ... w_n, tries to predict w_{n+1}
  - "The cat is on the ?" -> *mat*, *stove*, ...
  - "Maradona is a great ?" -> *player*, *coach*, ...
- Recent LMs, such as BERT, has popularized the idea of "masked sequence prediction", achieving SOTA results in a variety of tasks.
  - "The [MASK] is on the [MASK]

The Pre-Training Trick (for NLP)

Step 2: take the network from Hugging Face, and train your classifier on top on your *small dataset*.

10K tweets

```
import torch
from transformers import BertTokenizer, BertModel, BertForMaskedLM

# OPTIONAL: if you want to have more information on what's happening under the hood, activate the logger
import logging
logging.basicConfig(level=logging.INFO)

# Load pre-trained model tokenizer (vocabulary)
tokenizer = BertTokenizer.from_pretrained('bert-base-uncased')
```


coveo™

The Barrier of Meaning

Being Reasonably Wrong - A Microsoft Bot

a cactus in a green field

a car parked on a city street

commons.wikimedia.org/w/index.php?cu...

Being Reasonably Wrong - A Microsoft Bot

a person flying through the air while riding skis

commons.wikimedia.org/w/index.php?cu...

Being Reasonably Wrong - A Microsoft Bot

a dinosaur on top of a surfboard

Being Reasonably Wrong - GPT-3

Physical reasoning

- You are having a small dinner party. You want to serve dinner in the living room. The dining room table is wider than the doorway, so to get it into the living room, you will have to **remove the door. You have a table saw, so you cut the door in half and remove the top half.**

Being Reasonably Wrong - GPT-3

Physical reasoning

- You are having a small dinner party. You want to serve dinner in the living room. The dining room table is wider than the doorway, so to get it into the li
remove the door. You have a
half and remove the top half.

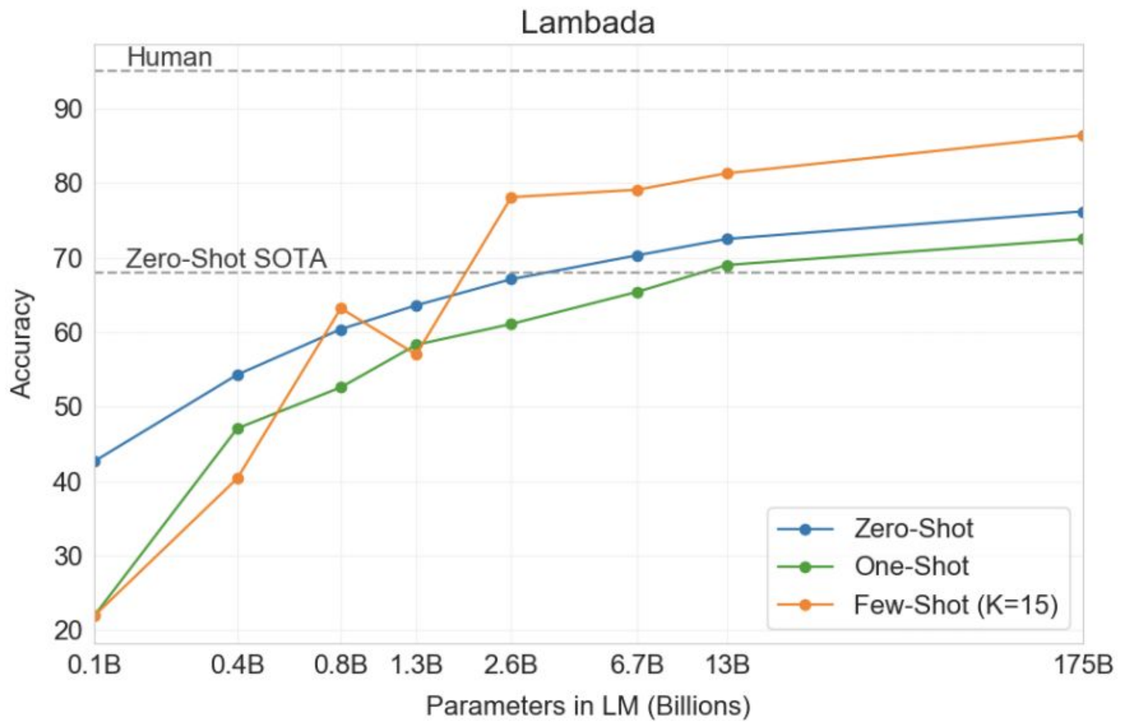
Biological reasoning

- You poured yourself a glass of cranberry juice, but then you absentmindedly poured about a teaspoon of grape juice into it. It looks okay. You try sniffing it, but you have a bad cold, so you can't smell anything. You are very thirsty. So **you drink it.**

coveo™

What Did We Really Learn?

- GPT-3 performance varies greatly depending on the task: great at word prediction, (very) bad at common sense reasoning.



What Did We Really Learn?

- With massive multi-task datasets, performances are barely better than random, while *humans* are good across the board.

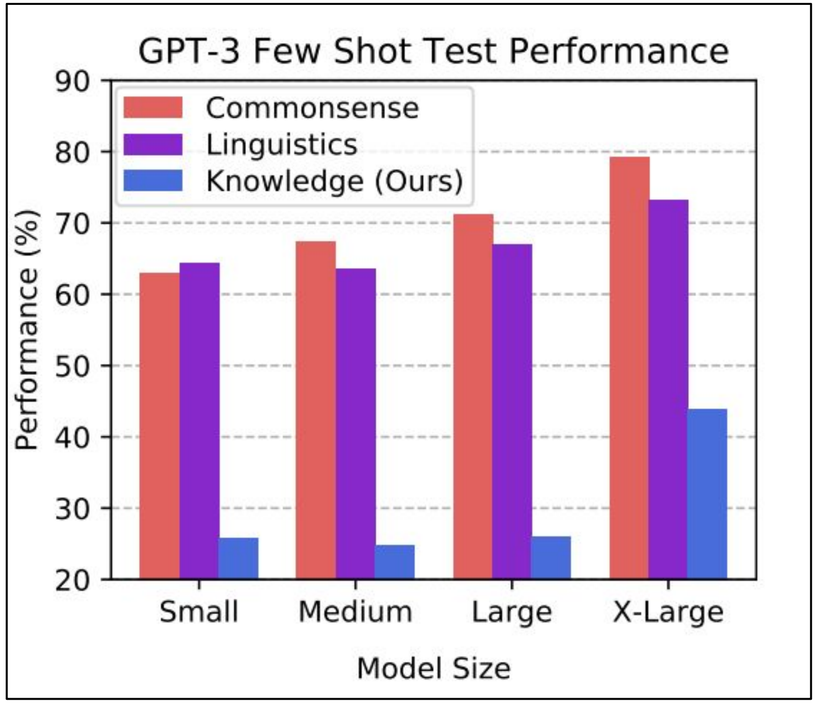
Declarative vs. Procedural Knowledge

Prompt and Completion:
The order of operations or PEMDAS is
Parentheses Exponents Multiplication Division Addition Subtraction

- -

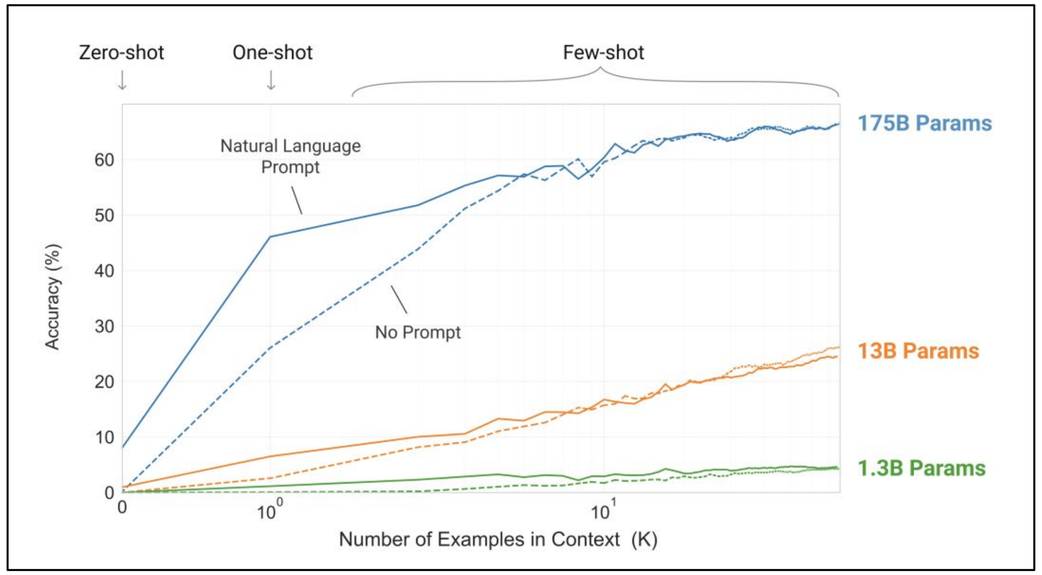
Prompt and Completion:
$(1 + 1) \times 2 =$ **3**

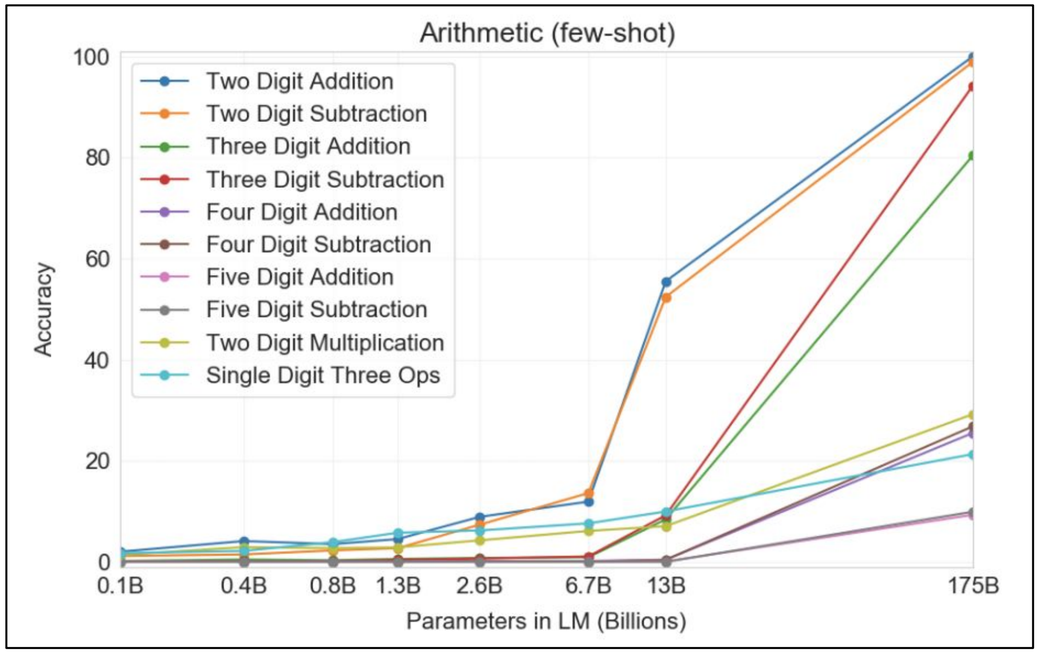
(a) GPT-3's completion for two prompts testing knowledge of the order of operations. The blue underlined bold text is the autocompleted response from GPT-3. While it is has descriptive knowledge and knows *about* of the order of operations, it does not know *how* to apply its knowledge and does not obey operator precedence.



What Did We Really Learn?

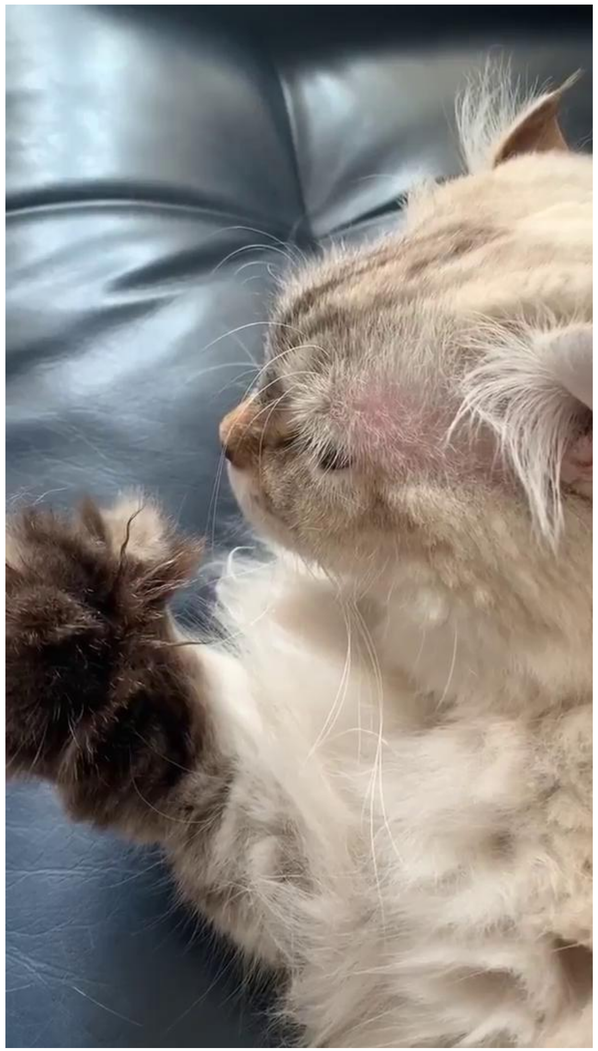
- The bigger, the better!
- The way in which GPT-3 “masters” concepts is very different from ours: it has problem with *generalization* and *abstraction*.



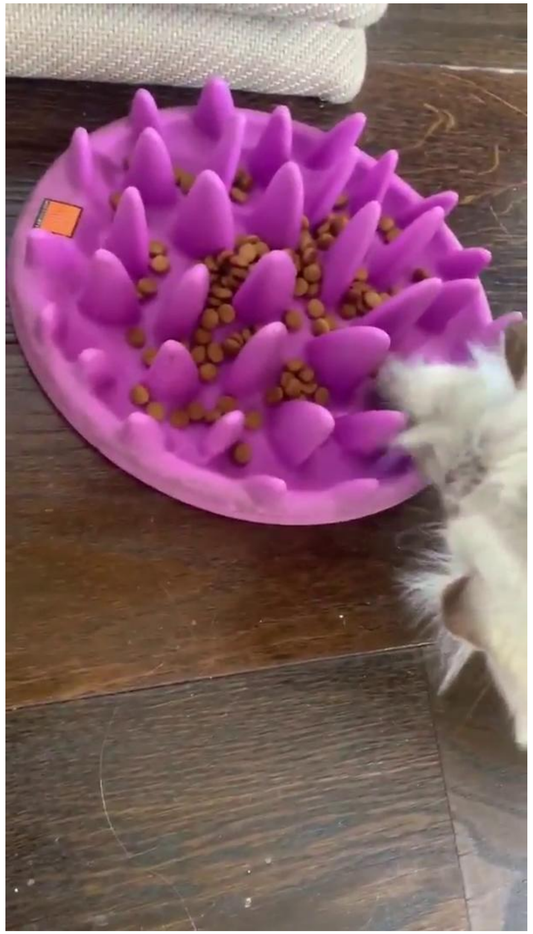


What’s Next?

M~~achine~~otton Learning

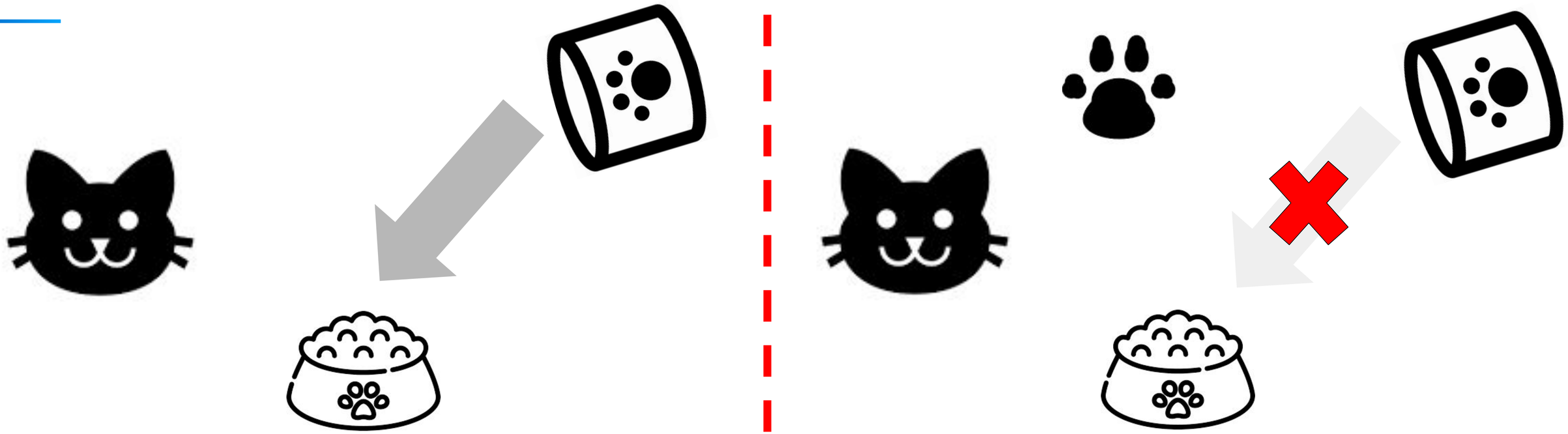


M~~achine~~otton Learning



M~~achine~~otton Learning

~~Machine~~otton Learning



CAUSAL INFERENCE

1. I prefer free food *vs* the pink torture
2. If I *paw* the food can, food will be free on the floor
3. I should paw the can

coveo™

When people called an object by some name, and while saying the word pointed to that thing, I watched and remembered that they used that sound when they wanted to indicate that thing.

Augustinus

Conf., I, 8

You Can't Learn a Language From the Radio!

- For all GTP-3 knows, *language IS reality* - the only way in which GTP-3 learns the meaning of "cat" and "food" is by sentences which describe *cats*, *food*, *pets*, etc. …

- However, Augustinus *docet*, language is about something *outside of text*….

Did NLP forget about this basic fact?

Blast From the Past

- Rediscovering two old ideas:
  - words refer to objects!
    - Objects are a core level of abstraction at which humans understand the world, as they provide a compact and causal representation of the world around us.

Blast From the Past

- Rediscovering two old ideas:
  - words refer to objects!
  - human communication is about other minds as well
    - If you ask me for an opinion on my colleague Bob, and I tell you "He is a very kind person, with lots of hobbies", you will think that *I believe* my colleague is not very good at his job. This *implicature* is not written anywhere, but it is a result of human communication.

Conclusions

That’s All Folks

1. NLP made tremendous **practical** and *some* **theoretical** progress with extensive use of deep neural networks.
2. Modern language models (e.g. BERT) are extremely useful pieces in modern ML architectures; thanks to the “pre-training” trick, huge models can be used with little effort to improve most applications, even with a small budget.
3. Even gigantic models, such as GPT-3, still fail to capture important properties of linguistic behavior, which make them brittle, easy to fool and unreliable for mission critical jobs (e.g. helping doctors).
4. Language has been studied for millenia, but the recent NLP tools have been developed almost entirely ignoring some known facts: perhaps the next “big thing” will be something rooted in a pretty old idea...

See you, space cowboys…

coveo™