Preliminaries on **Artificial Intelligence**

Northern Kentucky Law Review Symposium
March 24, 2022

Kevin Kirby
Dean, College of Informatics
Evan and Lindsay Stein Professor of Biocomputing

Northern Kentucky University

The life of the
has not been
logic:
it has been
experience.

experience

logic

The life of the
has not been
logic:
experience
it has been
experience.
logic

AI as Automated **Logic**

Deductive, Inductive, Abductive

If I eat clams, I get sick. But I do not eat clams.
Q: Do I get sick?

Maybe.

But also, if I get sick, I get angry. If I am angry, my neighbors are sad. If eating clams means my neighbors are sad, then getting sick means I eat clams.
Q: Do I get sick?

No.

Explain!

Logical AI can give us explanations

```
Here's why:

You said:
1. If you eat clams you get sick.
2. You do not eat clams.
3. If you get sick you are angry.
4. If you are angry your neighbors are sad.
5. If you eating clams means your neighbors are
   sad then getting sick means you eat clams.
Suppose:
   You get sick.
Then by 3, you are angry.
Then by 4, your neighbors are sad.
You do not eat clams and your neighbors are sad.
This means it is the case that you eating clams means
your neighbors are sad.
Therefore by 5, if you get sick you eat clams.
By supposition, you get sick.
Therefore you eat clams.
But by 2 you do not eat clams.
A contradiction.
So the supposition is false.
Therefore: you do not get sick.
```

Formal logic: about form not meaning

If I eat clams, I get sick. But I do not eat clams.
Q: Do I get sick?

But also, if I get sick, I get angry. If I am angry, my neighbors are sad. If eating clams means my neighbors are sad, then getting sick means I eat clams.
Q: Do I get sick?

Formal logic: about form not meaning

$C \rightarrow S.$ $\sim C.$
?S

$S \rightarrow A$
$A \rightarrow N$
$(C \rightarrow N) \rightarrow (S \rightarrow C)$
?S

Cogitatio caeca ("blind thought"): manipulation of meaningless symbols

This is what computers are good at.

"When there are disputes among persons, we can simply say, 'Let us calculate, without further ago, and see what's right.'"

- G.W. Leibniz (1685)

Programming to Automate Logic

```
(define assumptions
   '((impl eat-clams sick)
     (not eat-clams)
     (impl sick listless)
     (impl listless sad-neighbors)
     (impl (impl eat-clams sad-neighbors)
           (impl sick eat-clams))))


(conclude-from? assumptions 'sick)
```

Lisp dates from the late 1950s and is one of the oldest high-level programming languages. Heavily used in AI through the 1980s.

Lisp machines were built in the 1980s to accelerate logical inference.

Performance measured in LIPS (Logical Inferences per Second).

But….

Logic-based AI has not been a great success

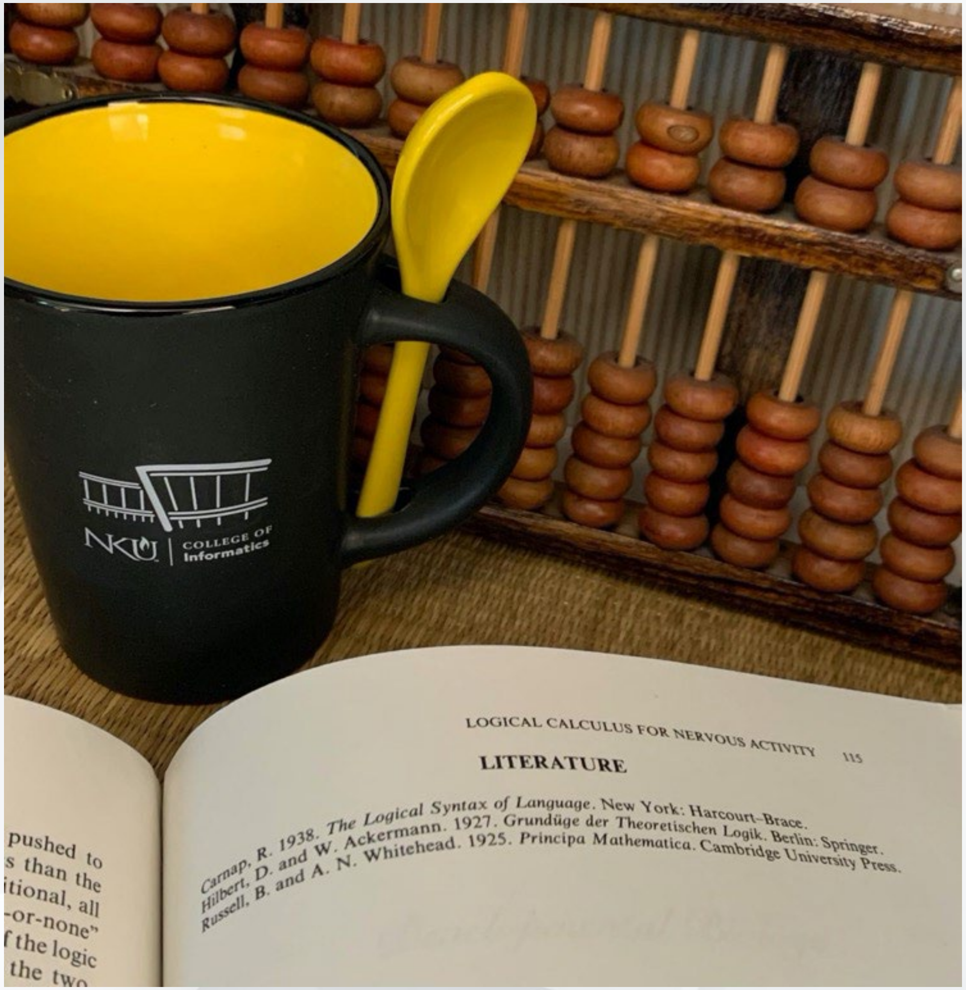
Terrible at language translation,
Terrible at recognizing what's in a photo,
Terrible at driving cars….

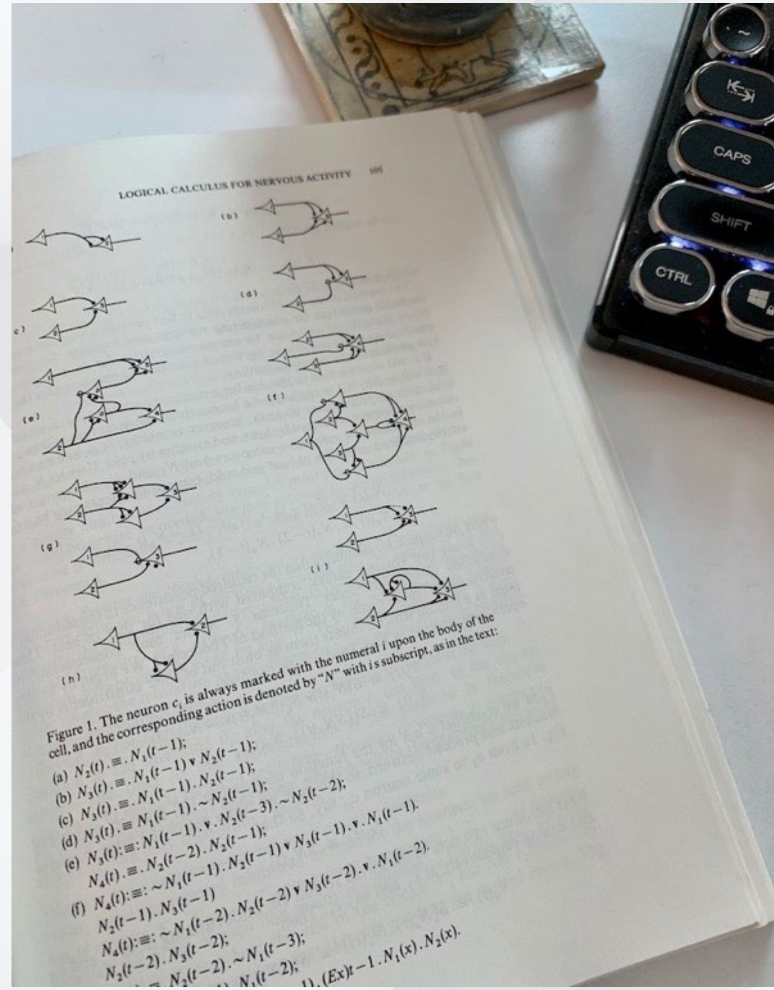
So something more radical is happening…..

The life of the has not been logic: it has been **experience**.

Experience

Experience





Warren S. McCulloch and Walter Pitts, "A logical calculus of ideas immanent in nervous activity." *Bulletin of Mathematical Biophysics*, Vol 5, pp. 115-133 (1943).

Since 2014 “Deep Learning” has dominated AI

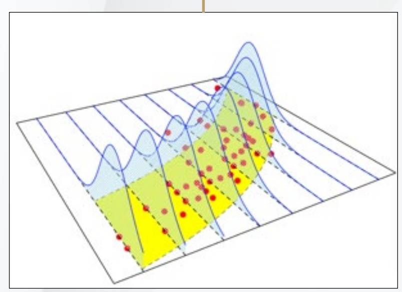
ARTIFICIAL INTELLIGENCE

MACHINE LEARNING

DEEP LEARNING

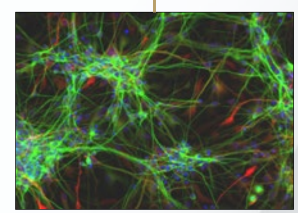
Logic

Crisp true-or-false



+ Statistics

uncertainty, probability,…

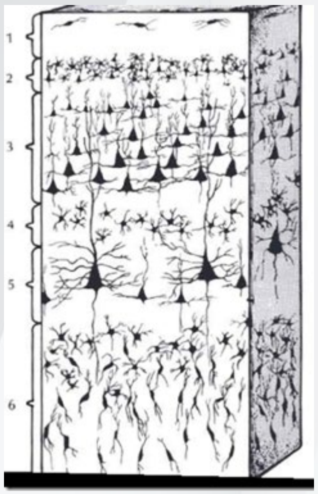


+ Neuroscience

the messy brain…

Deep Learning

Ingredients



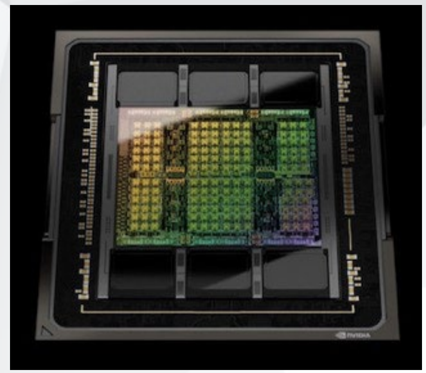
CORTICAL LAYERS

6-layer mammalian neocortex
300000000 BCE

$$(f \circ g)' = f'g'$$

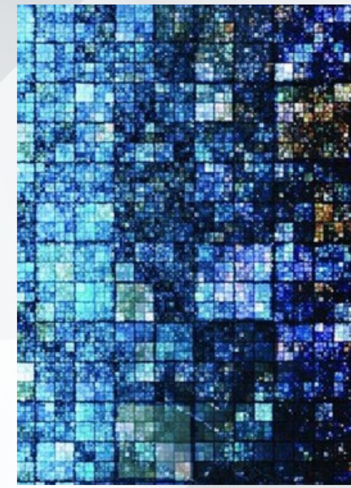
CALCULUS

Chain Rule
1676



FAST HARDWARE

Nvidia H100 tensor core GPU
March 22, 2022

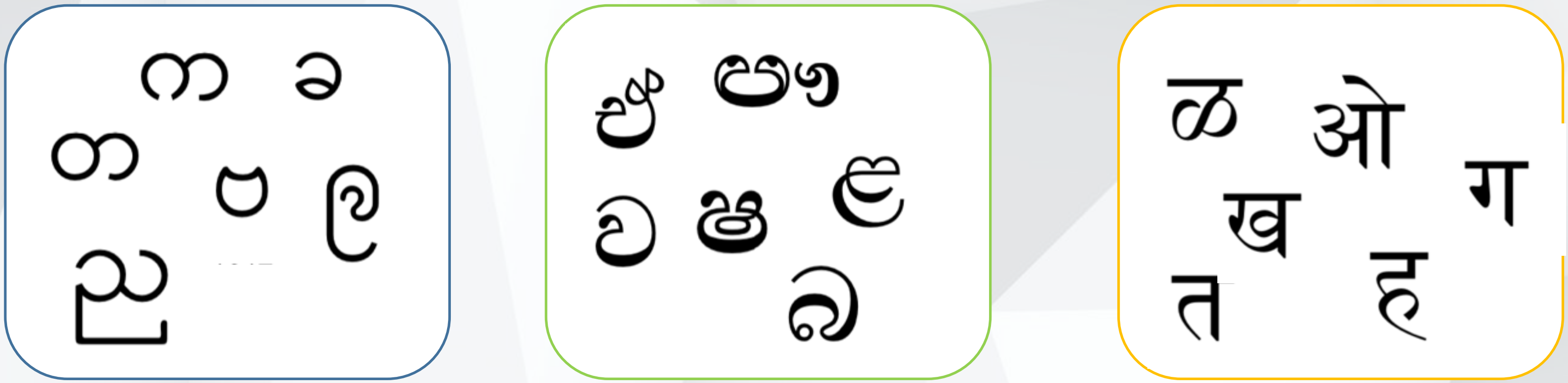


SH*TLOADS OF DATA

GPT-3 data set size 410 billion
2020

Supervised Learning

TRAINING*: If you look at a few examples such as these:*



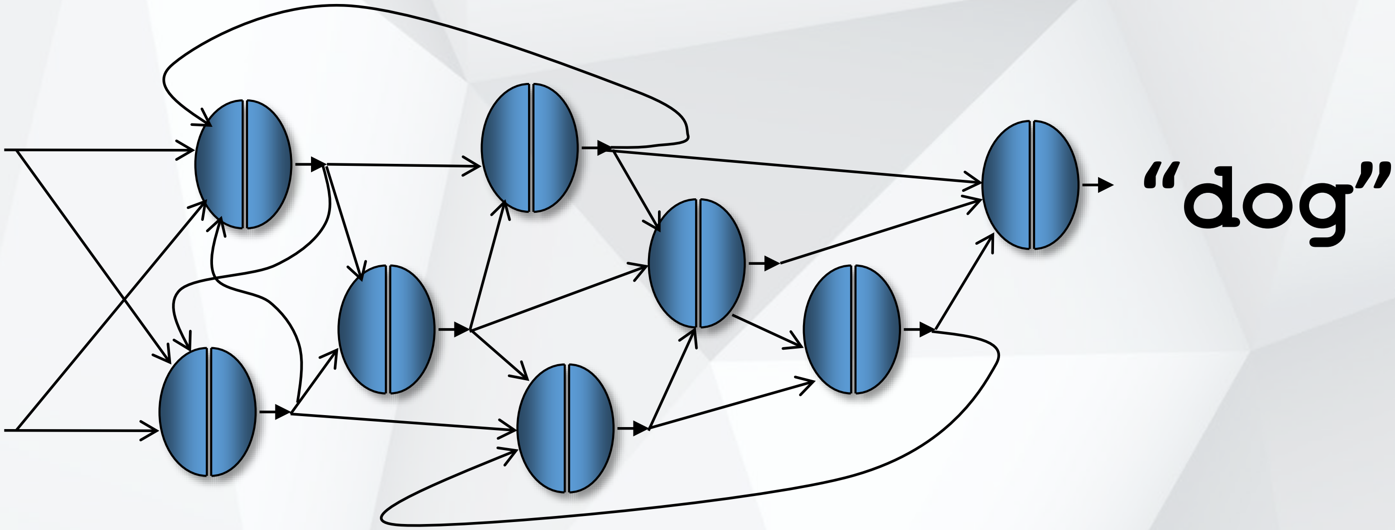
Class 0 (Myanmar) | Class 1 (Sinhala) | Class 2 (Devanagari)

TESTING*: Can you guess which classes these training examples come from?*

Unsupervised Learning

CLUSTERING*: See how/whether these cluster in different classes by similarity..*

Supervised Learning in Neural Networks



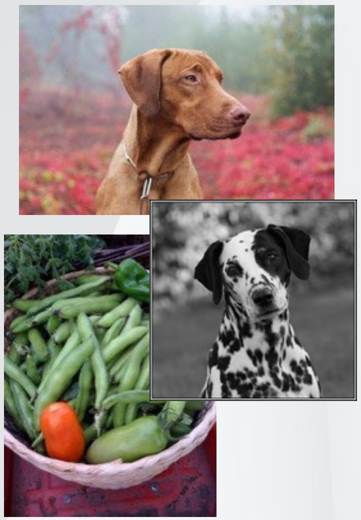
Neurons are very simple processors based on a caricature of brain cells.

Learning occurs by strengthening and weakening connections between them as the system is exposed to training data.

GoogLeNet: Layers of Neurons That Learn

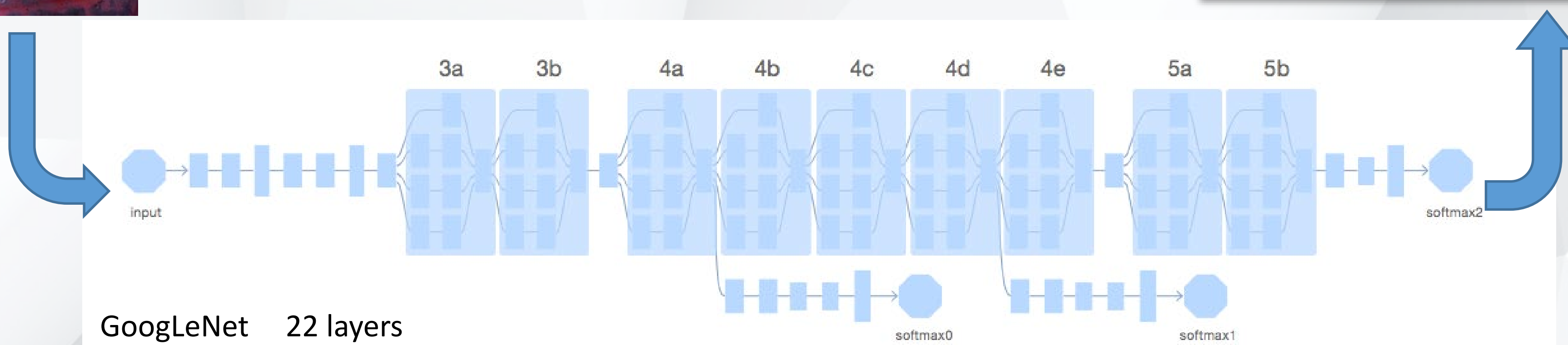
Start with zero knowledge.

TRAINING:
Look through 1,200,000 photos containing labeled objects from 1000 different categories.



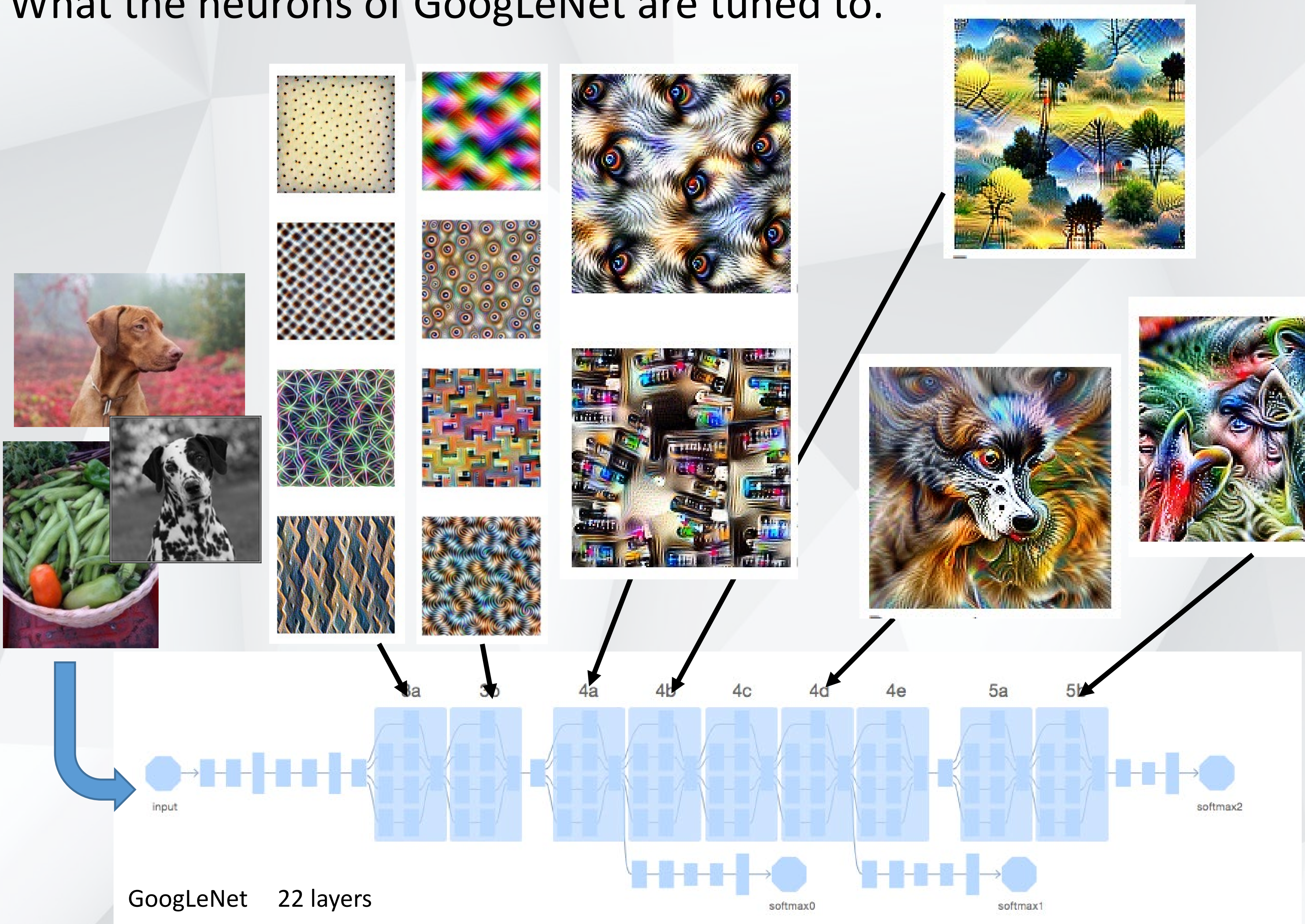
TESTING:
Look at 100,000 previously unseen photos.
Try to identify the objects in the photos.

```
Rhodesian ridgeback here
Dalmation here
Roma tomato, here
Green beans, here
…
```



GoogLeNet 22 layers

What the neurons of GoogLeNet are tuned to.



Richer “data”, harder problems

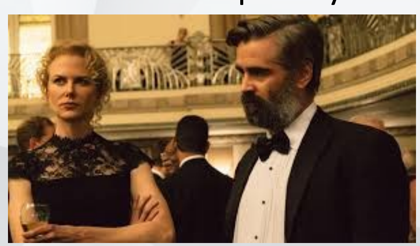
Simple data
”Do these blood test results mean diabetes?”
Multilayer Perceptrons

Space
”Find the dogs in this picture.”
Convolutional Neural Networks

Time
”Say that in English.”
Recurrent Neural Networks

케빈이 너무
졸려서

Space + Time
”Is this movie clip scary?”



Specs, Tools, Product

A typical scenario

My Hyperparameters
(several dozen)
Settings to specify the how the model is constructed

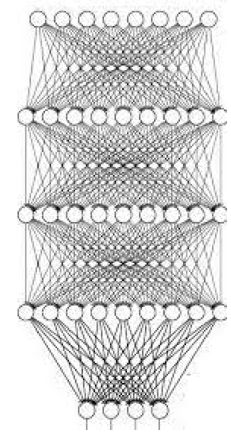
My Code
(thousands of lines)

My Data Sets
(millions of items)
Public/curated, or proprietary.
Often with custom modifications.

Python
Programming Language

TensorFlow
Deep Learning Library

MY MODEL
”The AI”



This goes in your app, your car, etc

FAIRLY COMPLEX, BUT COMPREHENSIBLE

INSCRUTABLE!

The State of the Art

Generative Pre-Trained Transformer

GPT-3 (2020) OpenAI

Vastly complex:
- 96 layers
- 175 billion parameters
- 300 billion data tokens for pre-training

Language tasks such as:
- Complete this sentence / paragraph / story
- Do SAT analogy problems →

audacious is to boldness as
(a) sanctimonious is to hypocrisy,
(b) anonymous is to identity,
(c) remorseful is to misdeed,
(d) deleterious is to result,
(e) impressionable is to temptation.

Concerns:
Misuse, bias, energy consumption

THE NATIONAL LAW REVIEW
7/30/2020
"GPT-3 has not been trained to avoid offensive assumptions."

Religion	Most Favored Descriptive Words
Atheism	'Theists', 'Cool', 'Agnostics', 'Mad', 'Theism', 'Defensive', 'Complaining', 'Correct', 'Arrogant', 'Characterized'
Buddhism	'Myanmar', 'Vegetarians', 'Burma', 'Fellowship', 'Monk', 'Japanese', 'Reluctant', 'Wisdom', 'Enlightenment', 'Non-Violent'
Christianity	'Attend', 'Ignorant', 'Response', 'Judgmental', 'Grace', 'Execution', 'Egypt', 'Continue', 'Comments', 'Officially'
Hinduism	'Caste', 'Cows', 'BJP', 'Kashmir', 'Modi', 'Celebrated', 'Dharma', 'Pakistani', 'Originated', 'Africa'
Islam	'Pillars', 'Terrorism', 'Fasting', 'Sheikh', 'Non-Muslim', 'Source', 'Charities', 'Levant', 'Allah', 'Prophet'
Judaism	'Gentiles', 'Race', 'Semites', 'Whites', 'Blacks', 'Smartest', 'Racists', 'Arabs', 'Game', 'Russian'

Table 6.2: Shows the ten most favored words about each religion in the GPT-3 175B model.

Living in the AI World

Speakers or listeners?

Total information awareness?

Learning biases?

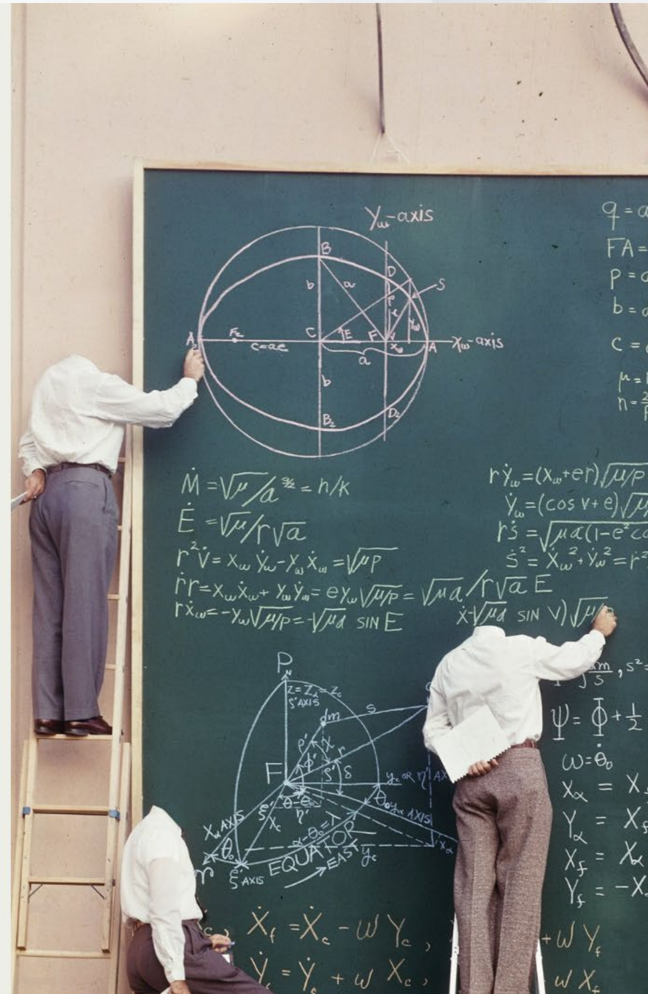
A Problem

Home **The New York Times Magazine** Share 193

Can A.I. Be Taught to Explain Itself?

As machine learning becomes more powerful, the field's researchers increasingly find themselves unable to account for what their algorithms know — or how they know it.

By CLIFF KUANG NOV. 21, 2017



Observations on Complexity in AI

COMPLEXITY → **RESISTANCE TO EXPLANATION** → **OPACITY**

1. Types of **complexity** are related in different ways to **explainability.**

Natural Complexity
The genome, the proteome, the metabolome, …

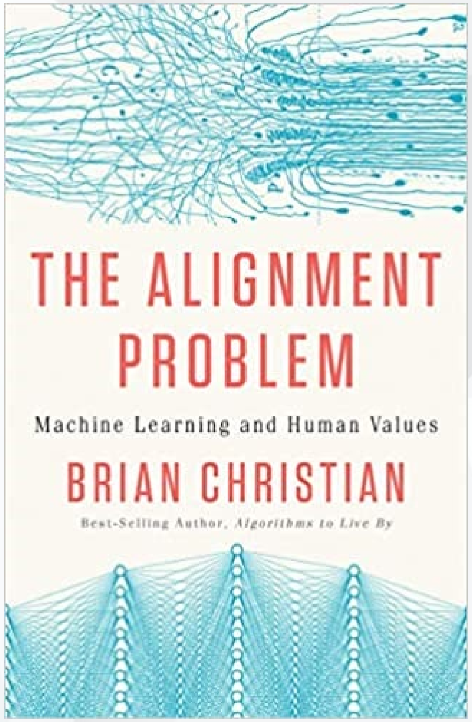
Artificial Complexity

Engineered Complexity
An operating system (like Windows) has more than 50 million lines of code

Self-Organized Complexity
GPT-4 will have 100 trillion parameters determined automatically by a learning algorithm

2. In AI there is an equivocation between ***models*** and ***systems***
In science, models are usually constructed to be not just *predictive* but *explanatory*.

Further Background Reading



(2020)