

CS109: Probability for Computer Scientists

Gather around and let me tell you a story
that we all should know



But first...
who am I?

Professor Chris Piech

Teaching at Stanford

CS106A

Programming
Methodologies

CURRENT

CS106B

Programming
Abstractions

LAST: FALL 2016

8,000+ students over 10 years

CS109

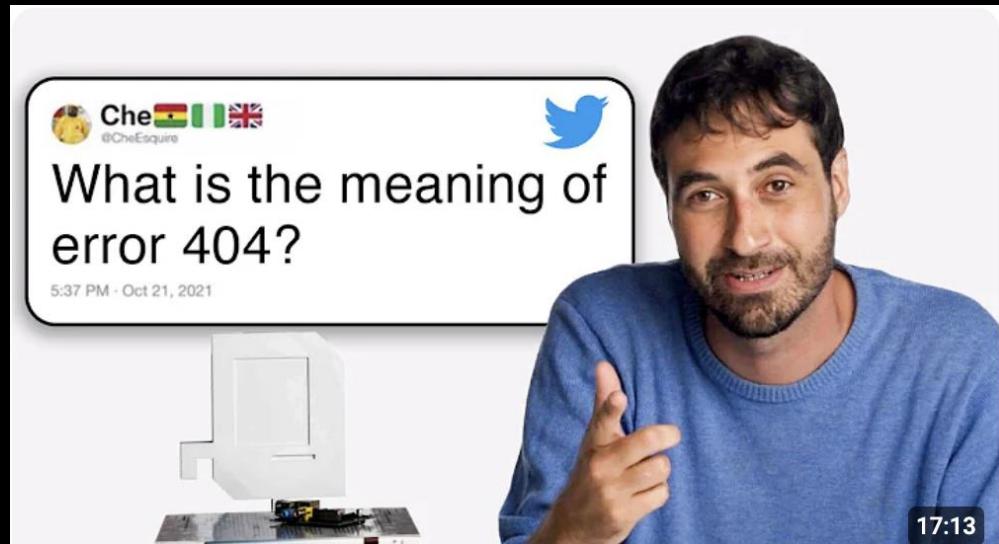
Probability for Computer
Scientists

LAST: FALL 2018

CS221

Intro to Artificial
Intelligence

LAST: SUM 2013



Code in Place: the
course with the most
teachers

From Three Places!



Grew up in Nairobi, Kuala Lumpur before Stanford!



Long History in CS109



I took the first CS109 back when I looked like this



Been teaching it since 2014

Piech, CS109, Stanford University



Stanford University

Computer Science Department

Stanford AI

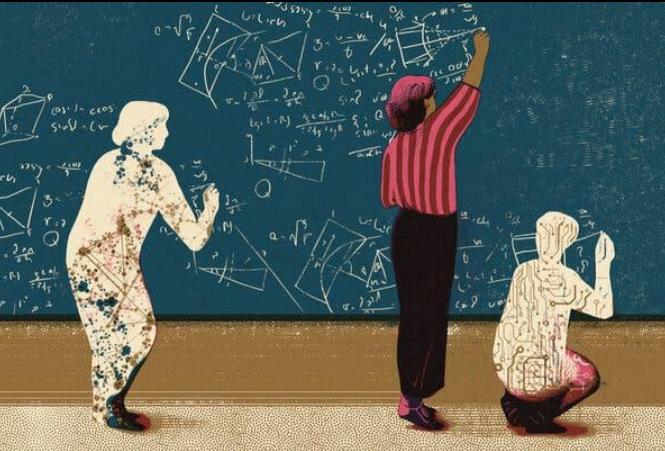
Piech Lab



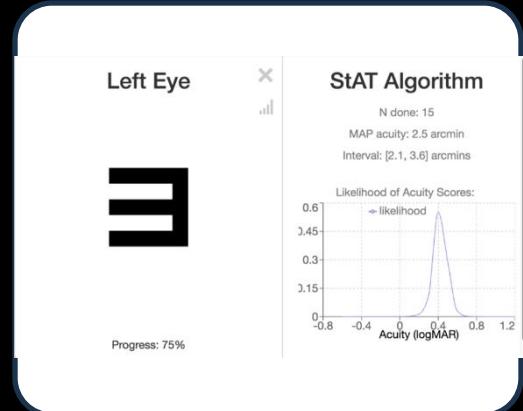
Professor Chris Piech



Invented DKT, the algorithm used in Duolingo's Bird Brain



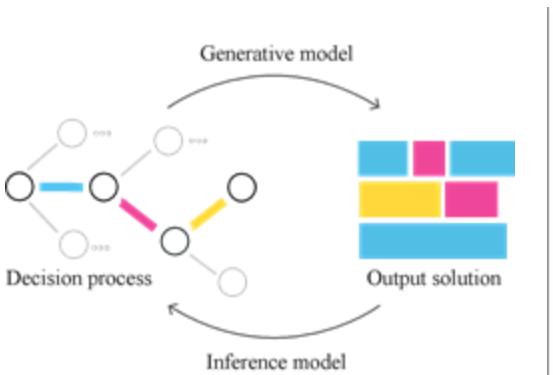
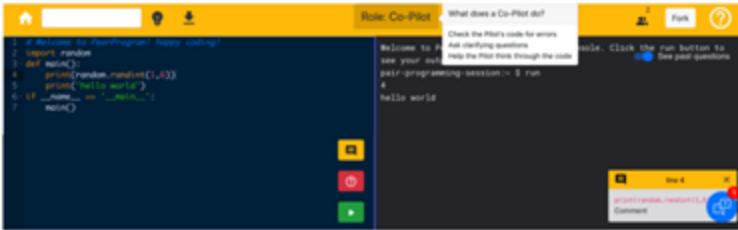
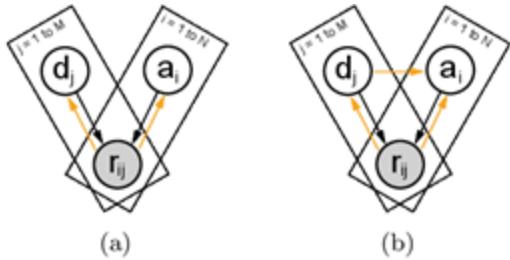
Invented Generative Grading, used in Code.org, featured in the New York Times



Created Stanford Acuity Test, Most Accurate Eye Exam!



Computational Education (and medicine too)



THIS IS MATH

$$\begin{array}{r} 1 \\ \underline{-} 2 \\ \hline 243 \\ - 87 \\ \hline 156 \end{array}$$

THIS IS MATH ON COMMON CORE



AI generated feedback

Students evaluate the feedback

Your Solution

```
def main():
    # write your solution here
    height = float(input("Enter your height in meters: "))
    if height < 0.4:
        print("Below minimum acceptable height")
    if height > 0.4:
        print("Above maximum acceptable height")
    if height == 0.4 and height < 0.4:
        print("Correct height to be an astronaut")
```

Algorithm uses attention to highlight where in the code the error comes from

Syntax error (missing ") here would prevent auto graders from being useful.

Traduza esta frase

esta casa está à venda?

Is this house for sale?



Play to Grade

The screenshot shows a web browser window titled "DreamApp Grading Interface". The URL is "Not Secure | iris-ws-6:3000/grading_display#". The page displays a list of assignments for "demo_student_2" (Not Graded) under the "Assignment: Breakout" category. The assignments and their scores are:

Assignment	Score
Mouse movement	4/4
Brick drawing	5/5 •
Paddle drawing	4/4 •
Ball drawing	4/4 •
Constants	0/2 •
Wall bouncing	3/3 •
Paddle bouncing	2/2 •
Paddle skewering	0/1
Ball does not become skewered on paddle (1)	0/1
New life	4/4 •

Total Score: 37/40

On the right side of the interface, there is a video player with a play button and a progress bar. Below the video player are three buttons: "Video" (highlighted in blue), "Code", and "Demo".

Best of STEM award for AI Tutor with Code.org

Piech, CS109, Stanford University

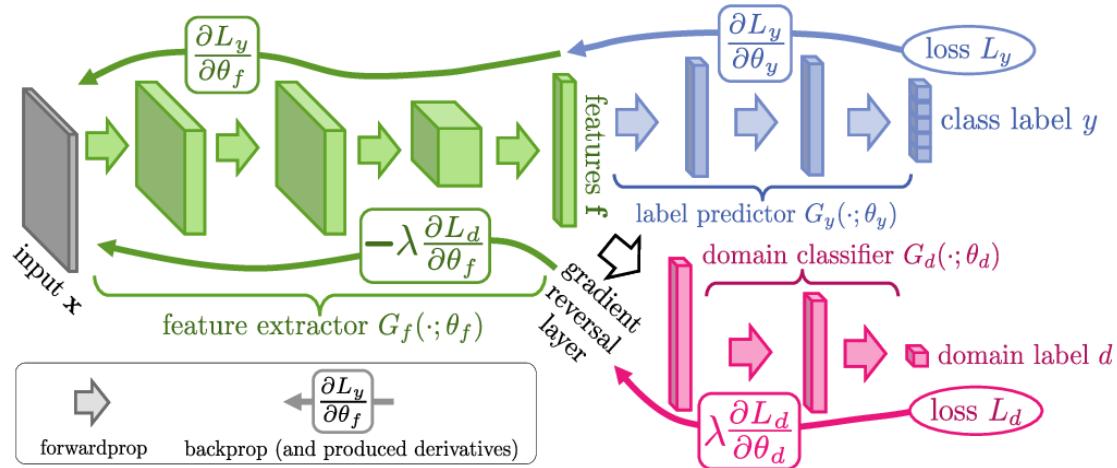


Research by CS109
Students

Fair AI with Adversarial Network

221 Citations

With undergrads Christina Wadsworth and Francesca Vera

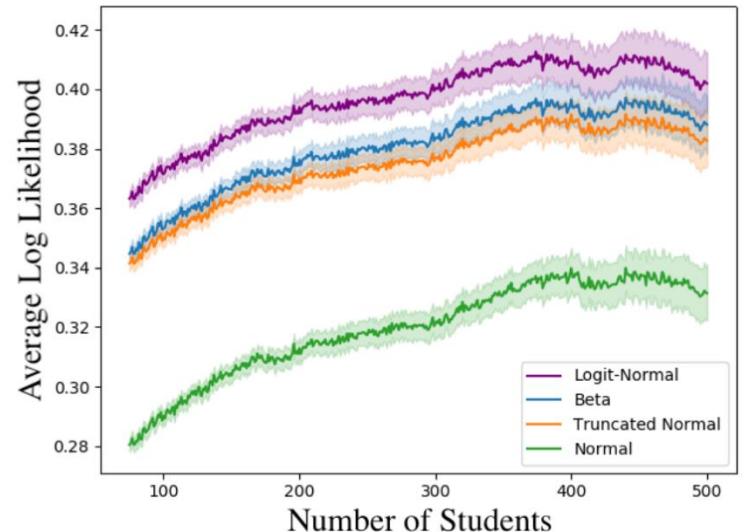
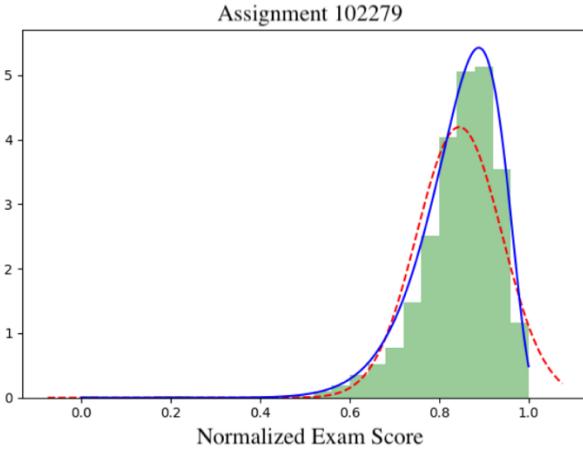
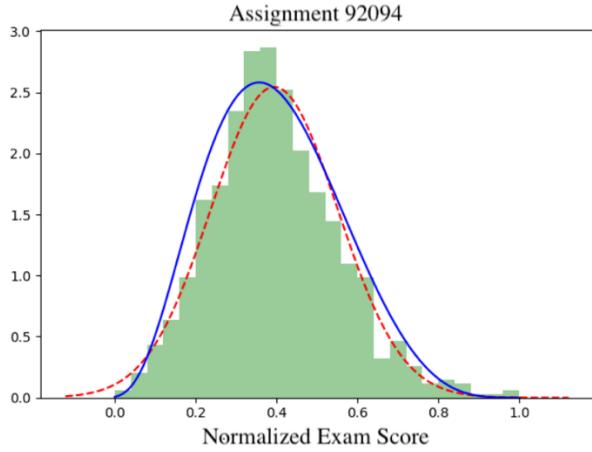
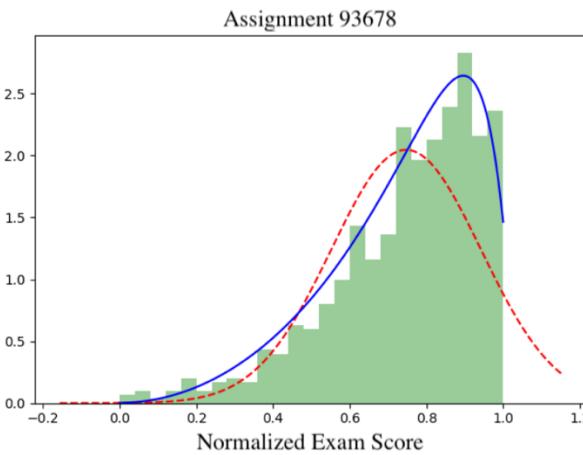
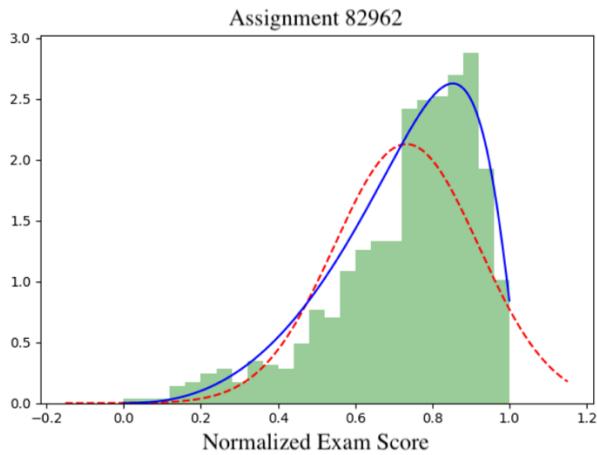


MODEL	ACCURACY	FP GAP	FN GAP
COMPAS SCORES (OUR TEST SET)	0.68	0.17	0.22
OUR RECIDIVISM MODEL	0.70	0.15	0.27
OUR CHOSEN ADVERSARIAL MODEL	0.70	0.01	0.02
BECHAVOD ET AL. AVD PENALIZERS (2017)	0.65	0.02	0.04
BECHAVOD ET AL. SD PENALIZERS (2017)	0.66	0.02	0.03
BECHAVOD ET AL. VANILLA REGULARIZED (2017)	0.67	0.20	0.30
ZAFAR ET AL. (2017)	0.66	0.03	0.11
ZAFAR ET AL. BASELINE (2017)	0.66	0.01	0.09
HARDT ET AL. (2016)	0.65	0.01	0.01

Math question: Can you remove racism from a deep learning predictor?



Grades are Not Normal



Noah Arthurs, CS109 Student

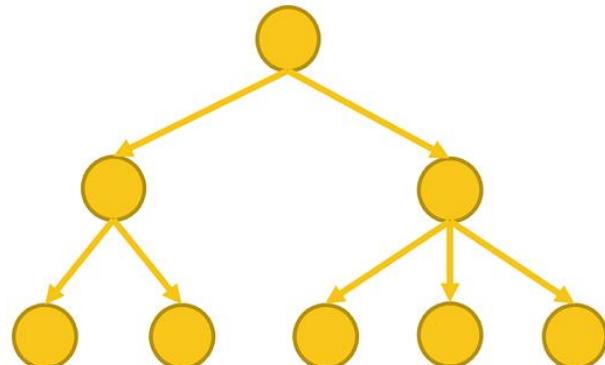


Math question: What is the generative story for grades in a typical classroom assignment?

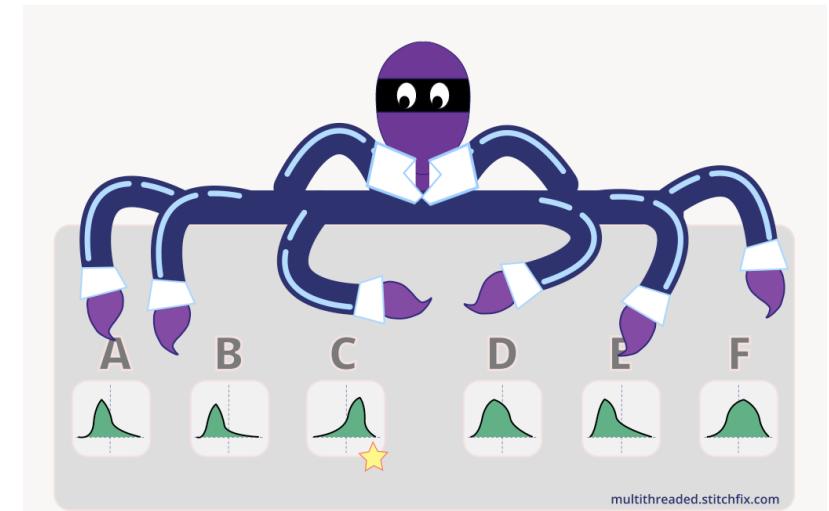
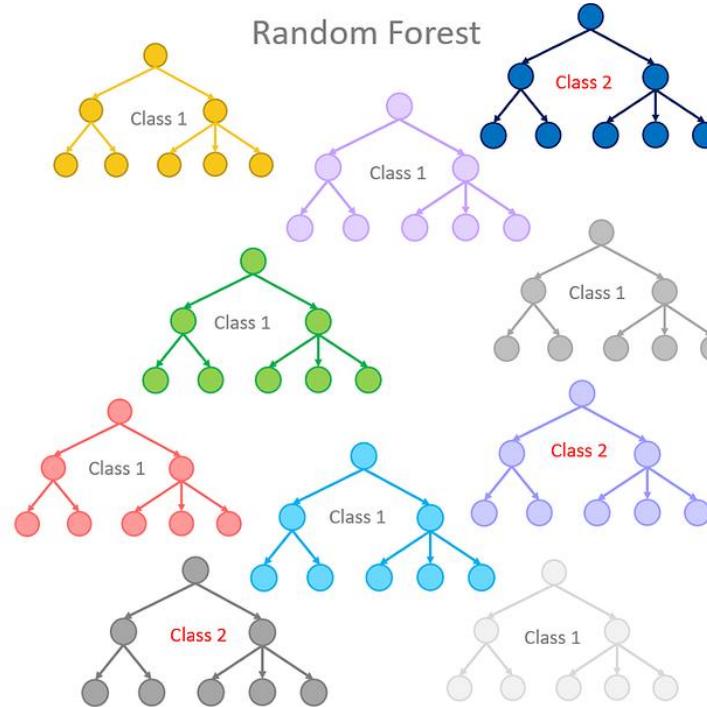


Faster Forest Training Using Multi-Armed Bandits

Single Decision Tree



Random Forest



Math question: Can you speed up classic algorithms if you treat expensive computation as decisions under uncertainty?

```
def example():
    → print("two space")
    → → print('four space')
    → print(two space, no quotation)
```

```
def main():
    → print(antaining)
example_function()
    → print("finishing")
```

```
if __name__ == "main__":
    main()
```

Math question: Can you relate pixel position to intended indentation using probability?

Transcribed handwritten code

```
pythonProject Version control main
Project pythonProject ~/Pycharm
pythonProject venv main.py
...
External Libraries Scratches and Consoles
def example():
    print("two space")
    print('four space')
    print(two_space, no_quotation)
def main():
    print(nothing)
example_function()
    print("finishing")
```

pythonProject > main.py

9:23 LF UTF-8 4 spaces Python 3.10 (pythonProject)



Undergraduate from CS109 Fall 23
Won Best Undergraduate Paper for L@S

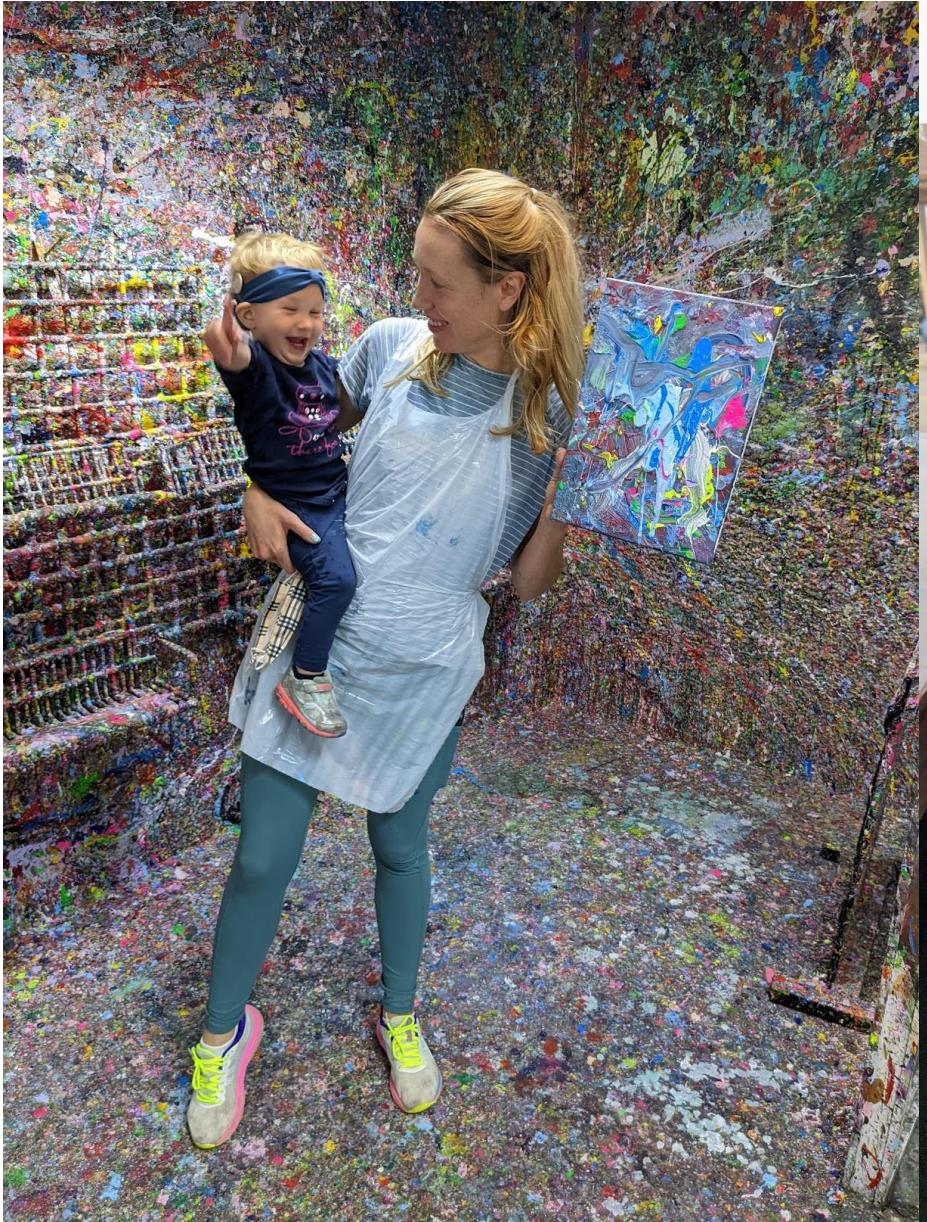
So many things to love in this world



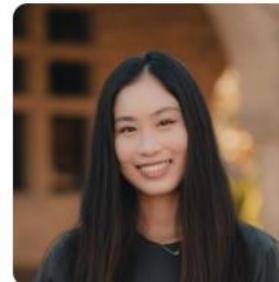
Piech, CS109, Stanford University



Most amazing family

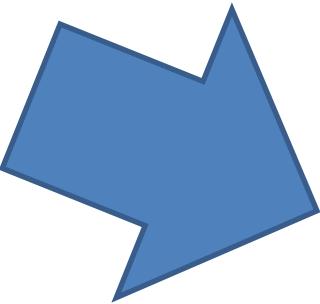


Fantastic Teaching Team

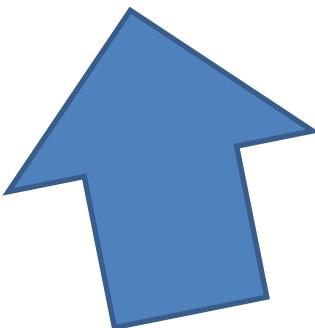
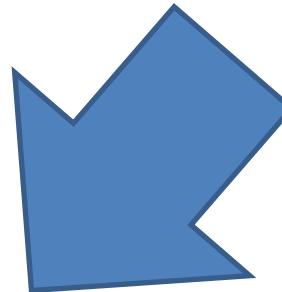


Course mechanics
(this is a light version. Please read the handout
for details).

Essential Information



cs109.stanford.edu



Are you in the right place?

Prerequisites

What you really need:

CS106B/X (important, coreq ok):

- Recursion
- Hash Tables
- Binary Trees
- Programming

CS103 (not necessary):

- Proof techniques (induction)
- Set theory
- Math maturity

Math 51 or CME 100 (important, coreq ok)

- Multivariate differentiation
- Multivariate integration
- Basic facility with linear algebra (vectors)



Coding in CS109

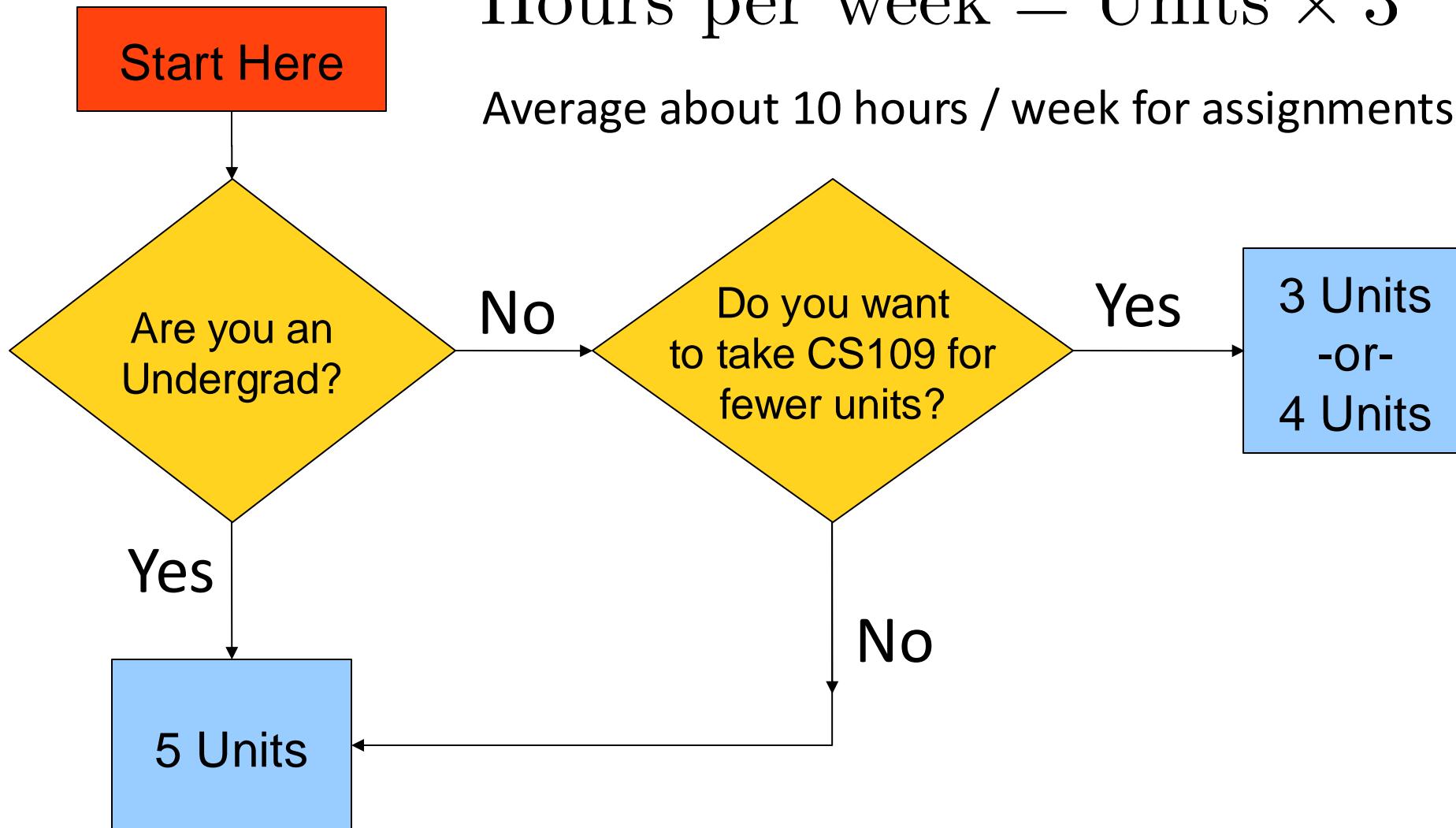


Review session on Friday

Piech, CS109, Stanford University



CS109 Units



Class Breakdown

40% **6 Assignments**

20% **Midterm**
2 hour exam, Oct 29th, 7pm

30% **Final**
3 hour exam, Dec 10th, 8:30am

10% **Participation**
Section, PEP



New this year!

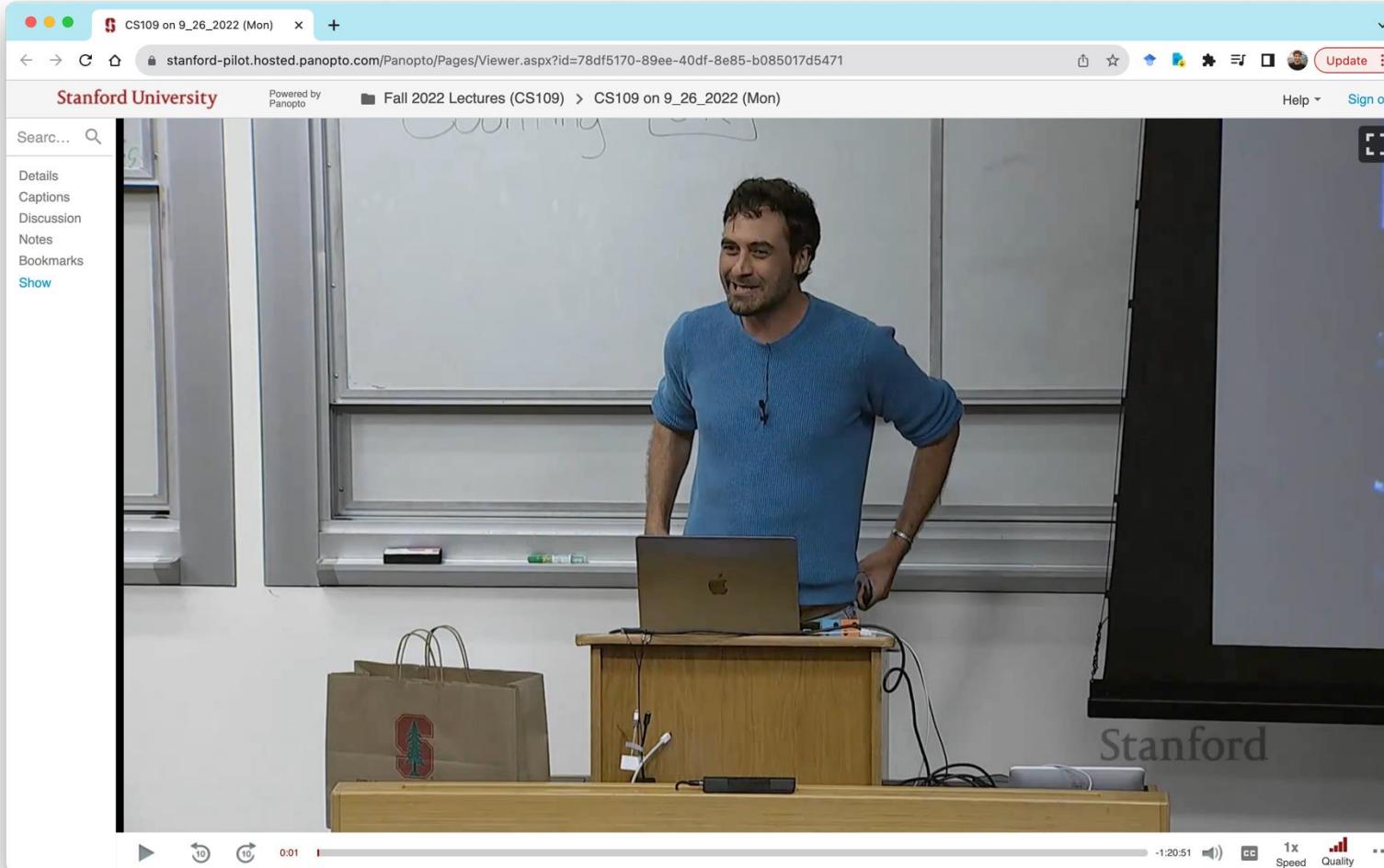
Personal Exam Prep (PEP)

- 15 mins
- Twice in the quarter
- 1:1 with a TA
- Week before each exam
- Participation graded
- Scheduled on Week 4

Song of the Quarter

[coming on Wednesday]

Is Class Online?



TLDR: Yes. Come to live class. It's a good time (and good for you)



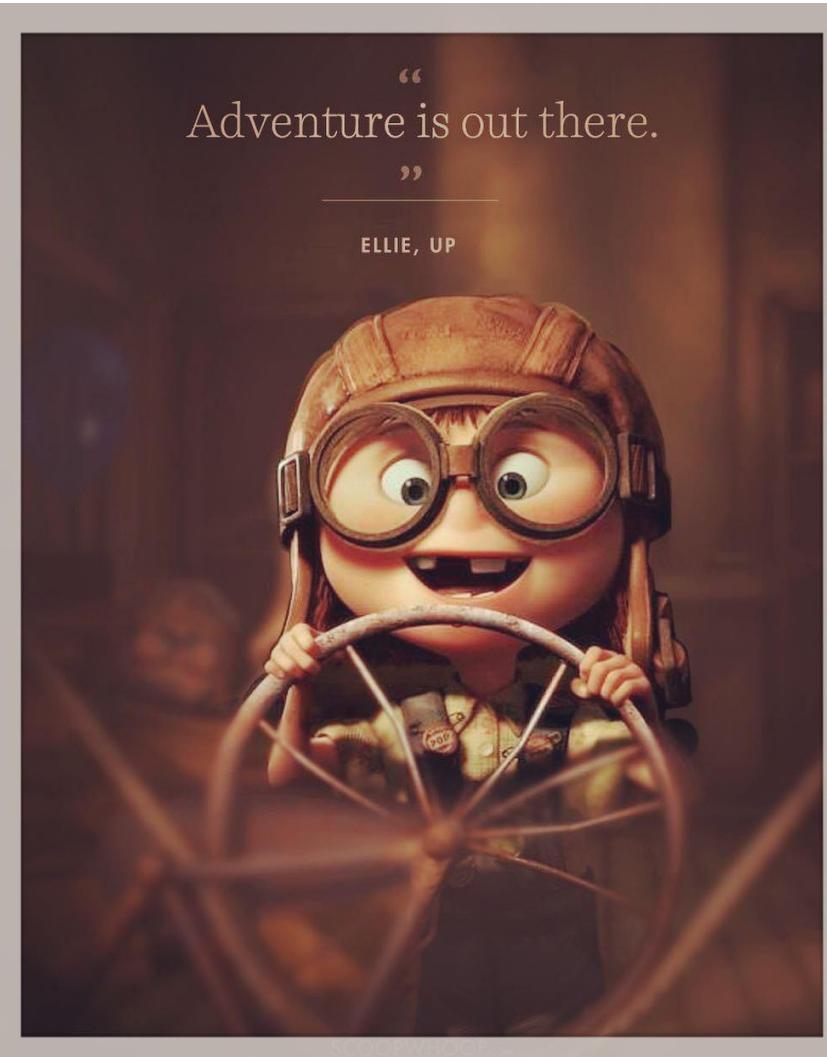
Come to Class

Section assumes you
are caught up to
Monday's lecture

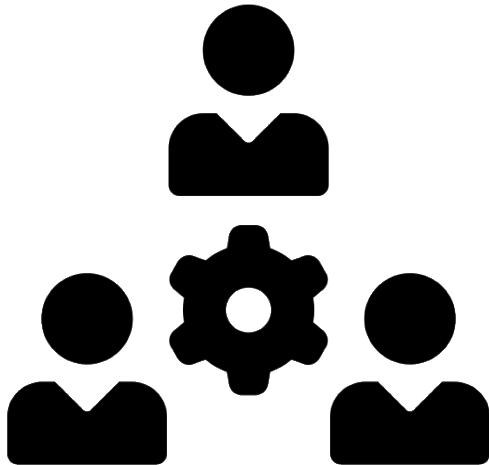
"I met my husband in
your CS109 lectures"
- Student from 3 years
ago

Attending lecture
strongly correlates
with better grades

We have fun.



Ask questions



Q&A forum
All announcements

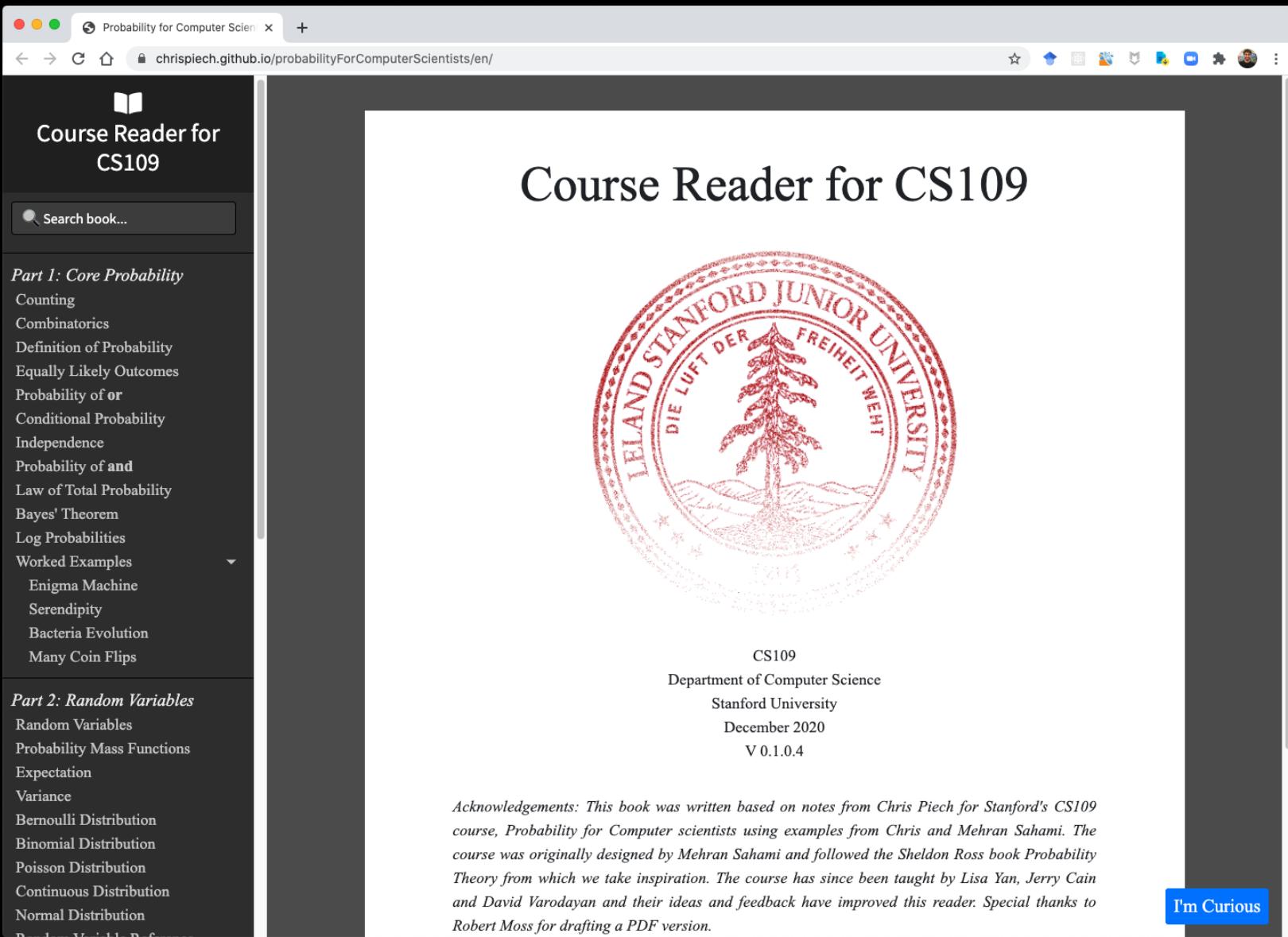
“Working” office hours
start on Saturday

Email cs109@cs.stanford.edu



Chris has 1:1 CS109 office hours

Chris Wrote a Course Reader!



A screenshot of a web browser displaying the "Course Reader for CS109" website. The page title is "Course Reader for CS109". The main content area features the Stanford University seal (Redwood tree) and text: "Course Reader for CS109", "CS109", "Department of Computer Science", "Stanford University", "December 2020", and "V 0.1.0.4". A sidebar on the left lists course topics under "Part 1: Core Probability" and "Part 2: Random Variables". A blue button at the bottom right says "I'm Curious".

Course Reader for CS109

CS109

Department of Computer Science

Stanford University

December 2020

V 0.1.0.4

Acknowledgements: This book was written based on notes from Chris Piech for Stanford's CS109 course, Probability for Computer scientists using examples from Chris and Mehran Sahami. The course was originally designed by Mehran Sahami and followed the Sheldon Ross book Probability Theory from which we take inspiration. The course has since been taught by Lisa Yan, Jerry Cain and David Varodayan and their ideas and feedback have improved this reader. Special thanks to Robert Moss for drafting a PDF version.

I'm Curious

Course Reader for CS109

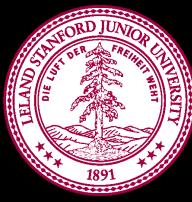
Search book...

Part 1: Core Probability

- Counting
- Combinatorics
- Definition of Probability
- Equally Likely Outcomes
- Probability of or
- Conditional Probability
- Independence
- Probability of and
- Law of Total Probability
- Bayes' Theorem
- Log Probabilities
- Worked Examples
 - Enigma Machine
 - Serendipity
 - Bacteria Evolution
 - Many Coin Flips

Part 2: Random Variables

- Random Variables
- Probability Mass Functions
- Expectation
- Variance
- Bernoulli Distribution
- Binomial Distribution
- Poisson Distribution
- Continuous Distribution
- Normal Distribution



CS109 | Syllabus

localhost:8001/handouts/syllabus.html

Syllabus

Honor Code

Office Hours

Lecture Videos

Syllabus

...ED IN 12 MINUTES

If you have any questions after reading this Syllabus, post on our [discussion forum](#).

Teaching Team



Professor: Chris Piech
✉ piech @ cs
🏡 Durand 305

We are lucky to have a phenomenal group of Course Assistants:



Schedule

Teaching Team

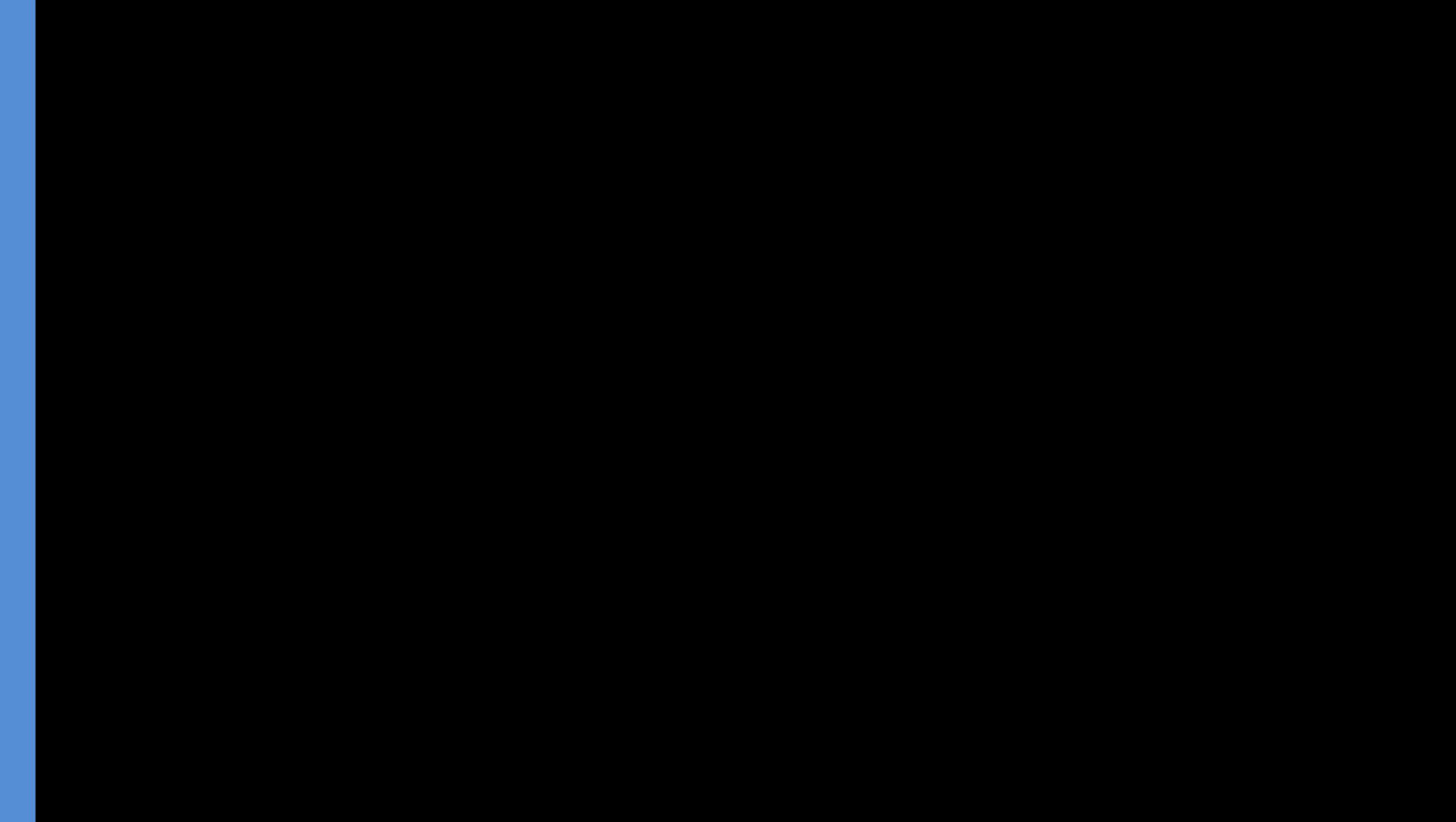
I. Course Overview

II. Course Structure

III. Course Resources

IV. Honor Code

Looking Forward to a Great Quarter



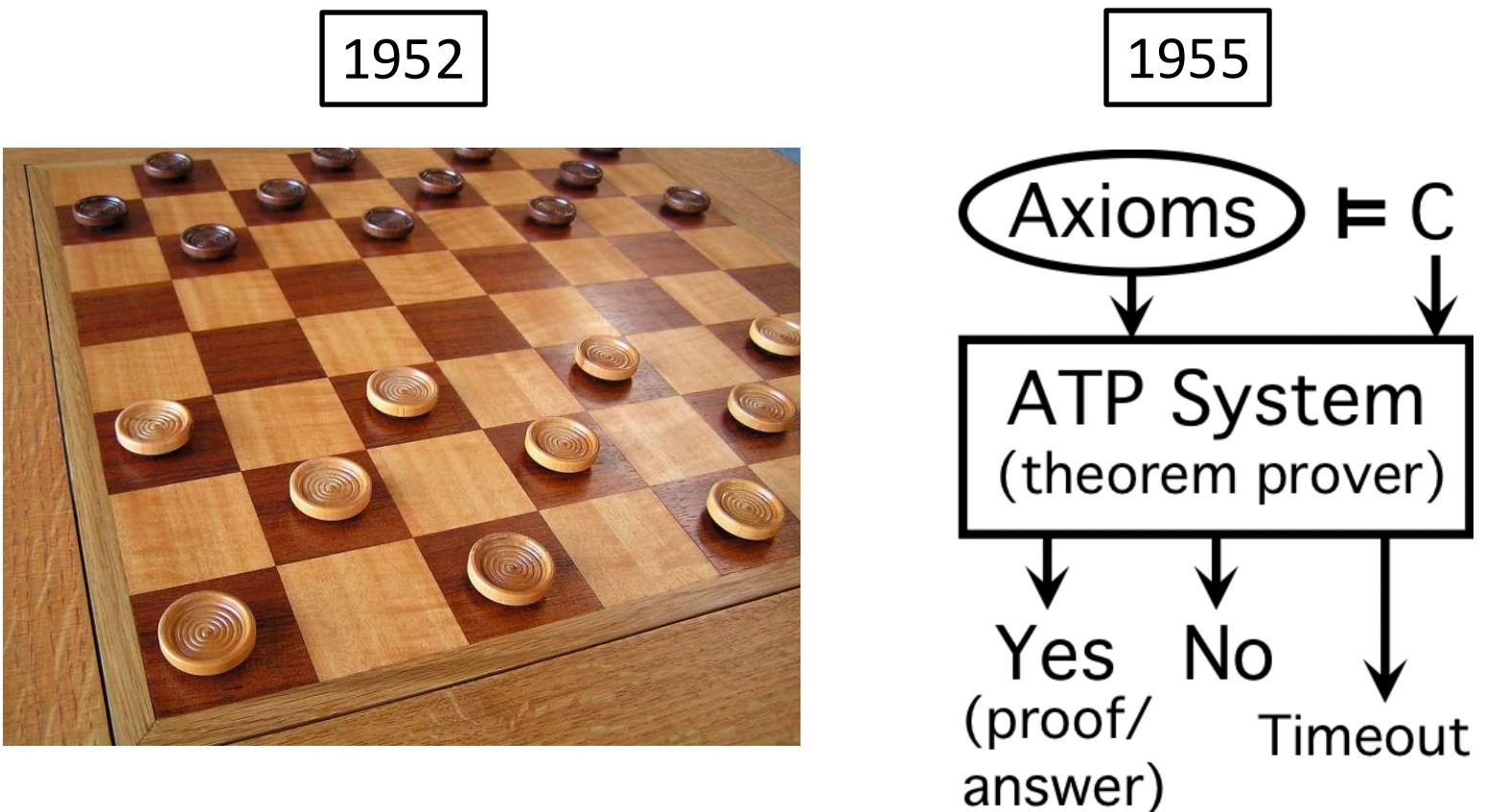
Story of Modern AI

Modern AI
or, How we learned to combine
probability and programming

Brief History



Early Optimism 1950s



Early Optimism 1950s

“Machines will be capable,
within twenty years, of doing
any work a man can do.”
–Herbert Simon, 1952



Underwhelming Results 1950s to 1980s

The spirit is willing but the flesh is weak.



(Russian)



The vodka is good but the meat is rotten.

The world is too complex



BRACE YOURSELVES



WINTER IS COMING



Something is going on in the world of AI

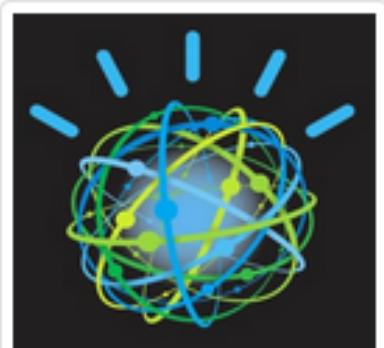
Big Milestones Part 1



1997 Deep Blue

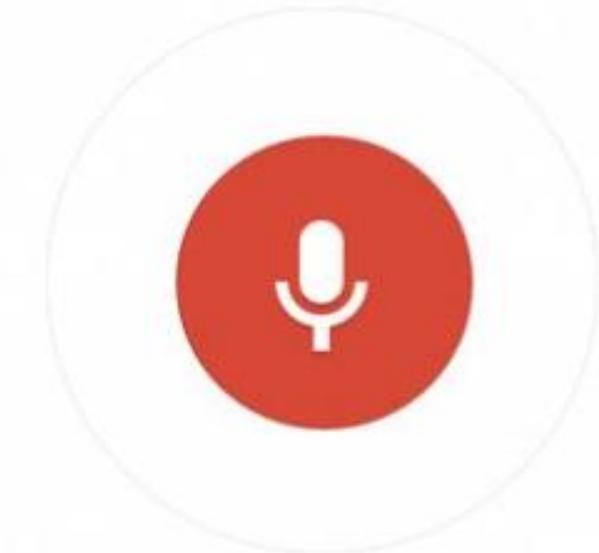


2005 Stanley



2011 Watson

I was told speech was 30 years out



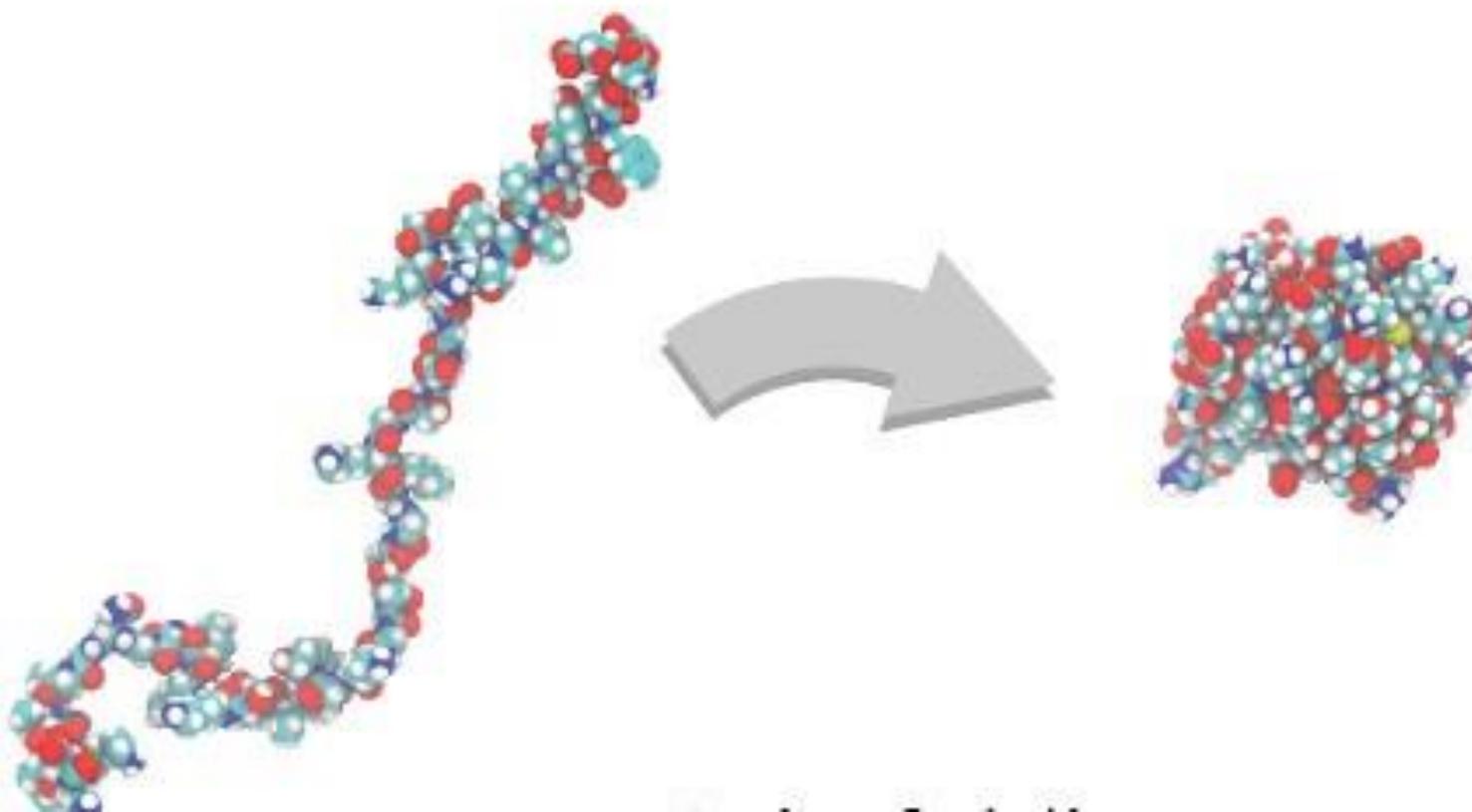
Almost perfect...



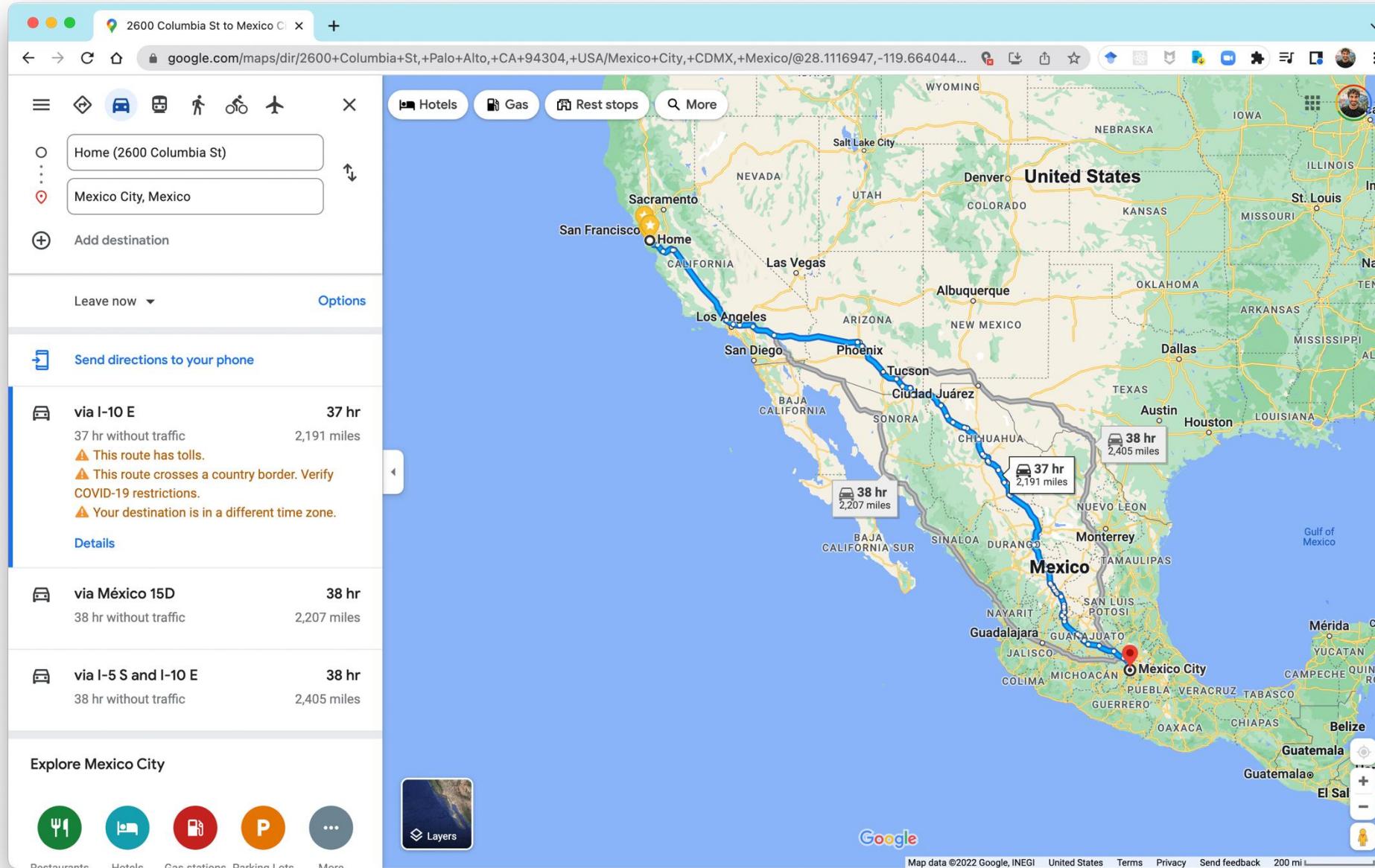
The last remaining board game



Protein Folding



Directions From A to B



Piech, CS109, Stanford University



3:31



Google Translate



English



Ukrainian



ENGLISH



Please translate this into Ukrainian.
Thank you



Camera



Conversation



Transcribe

UKRAINIAN



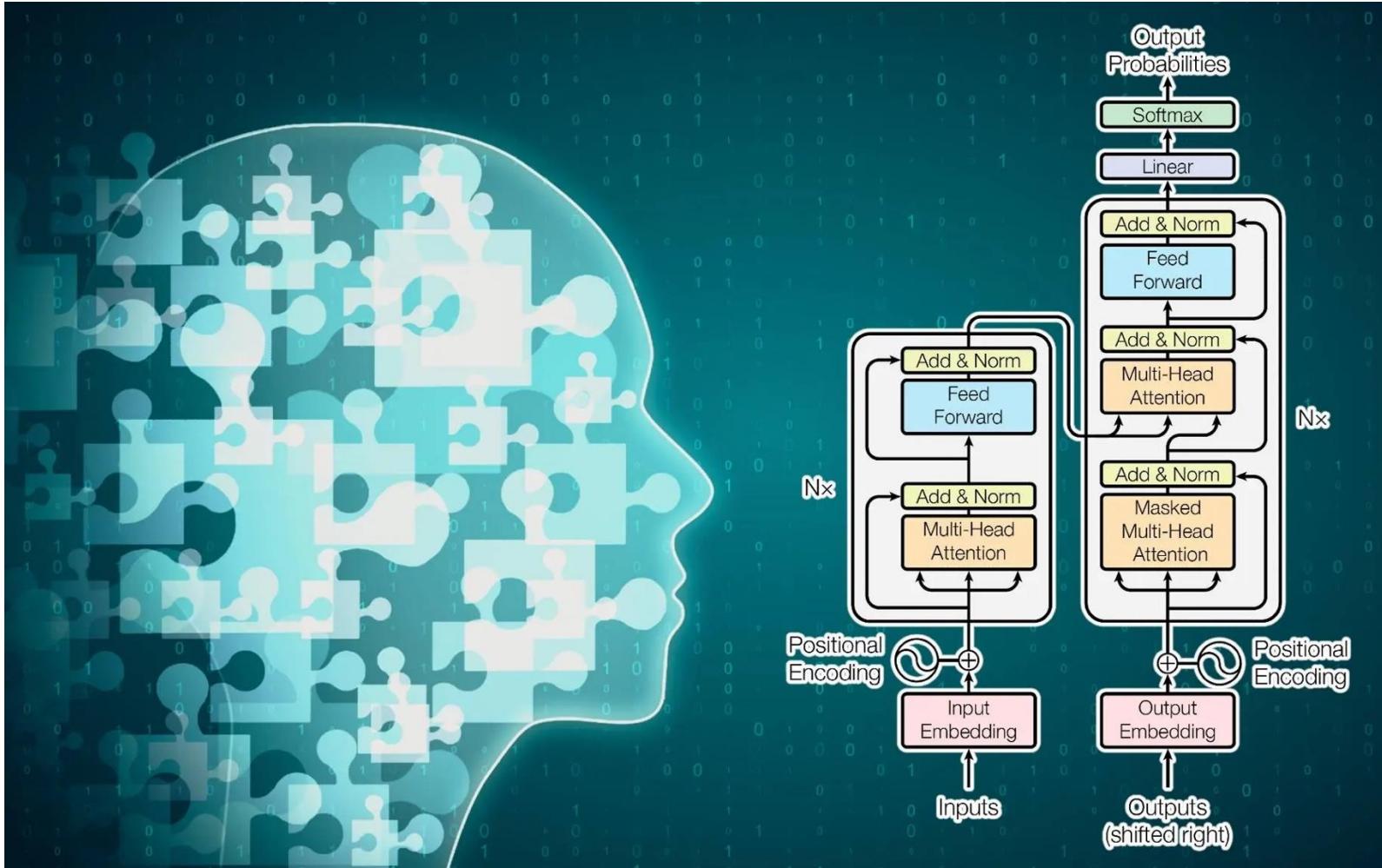
Будь ласка, переведіть це

Self Driving Cars

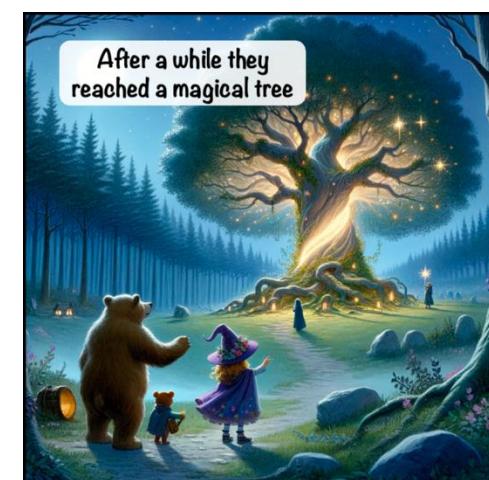
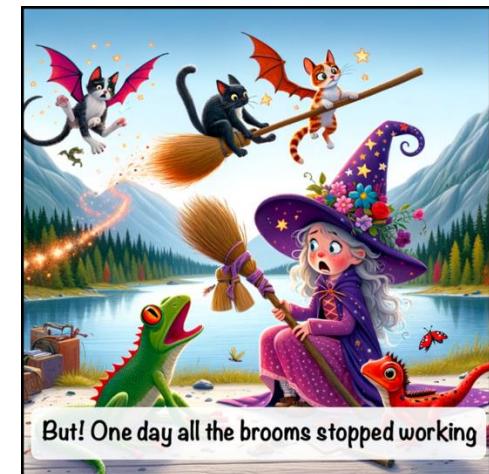
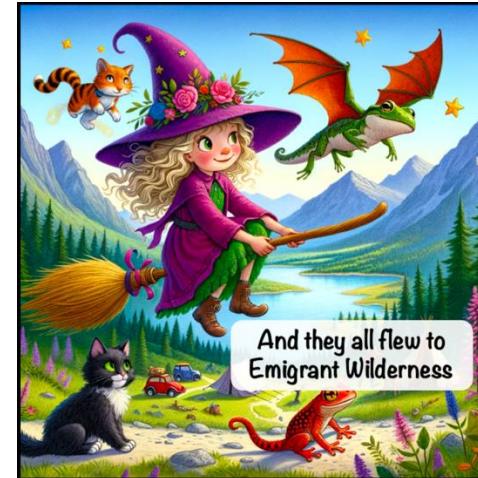


And then last year, everything changed, again

AI that (seems) to understand language



DallE Images made by Freya (2 years old)



What is going on?

[suspense]

Focus on one problem

Computer Vision



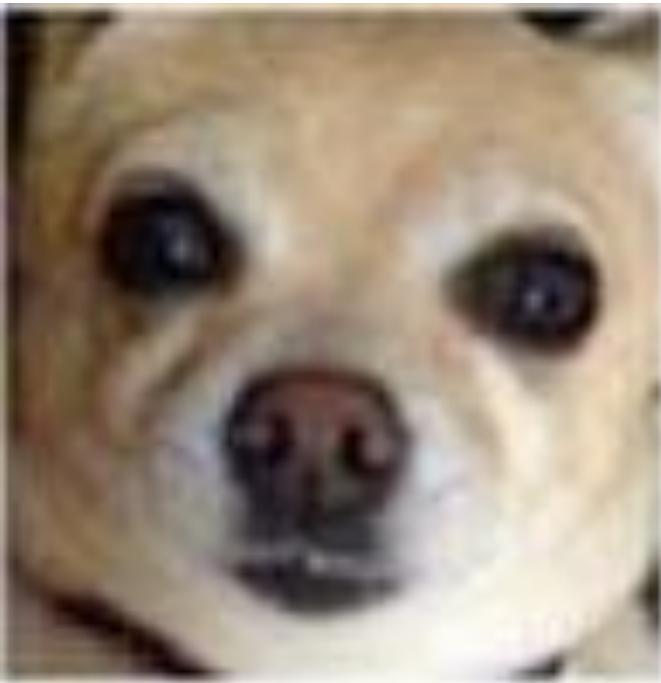
Chihuahua or muffin?

Piech, CS109, Stanford University



Can you do it?

Chihuahua or Muffin?



Piech, CS109, Stanford University



Chihuahua or Muffin?



Piech, CS109, Stanford University



How about now?

What a computer sees

0	0	1	0	1	0	1	0	0	0	1	1	1	1	0	1
1	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0
1	1	1	0	1	0	0	1	1	0	0	1	0	1	0	0
1	1	1	1	1	0	0	0	0	0	0	1	1	0	1	1
0	0	0	1	1	0	0	1	0	0	0	1	1	1	1	0
1	0	0	1	1	0	0	0	1	0	0	1	0	0	0	0
1	1	0	1	1	0	0	1	1	0	0	1	0	1	0	0
1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0
0	0	0	0	1	0	1	0	1	0	1	1	1	1	0	0
0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	0
0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0
0	1	1	1	0	1	0	0	0	1	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1
0	0	1	1	1	0	1	0	1	1	1	1	1	1	1	1



What a human sees



Why is it easy for Humans?



About 30% of your cortex is used from vision
3% is used to process hearing







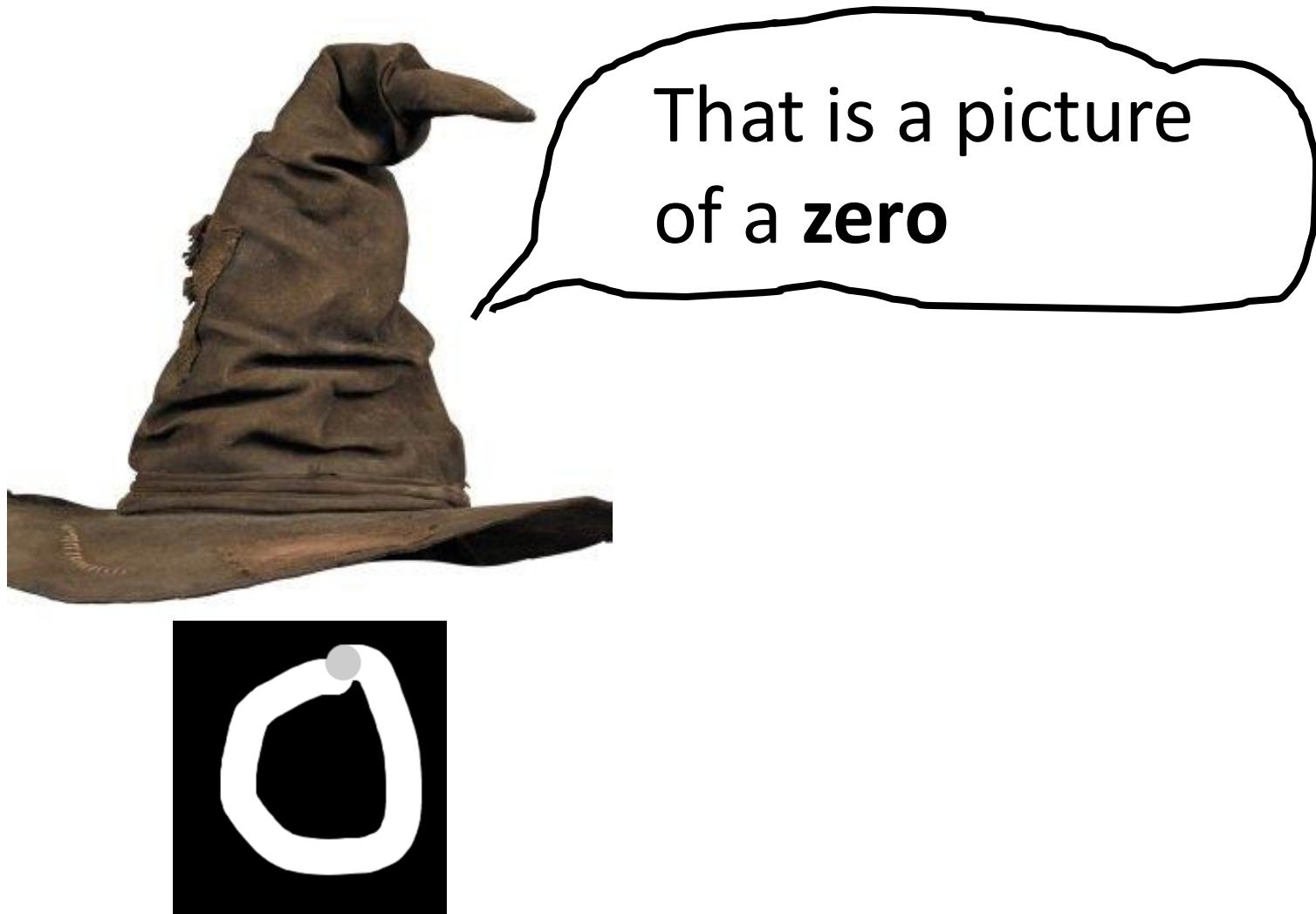
Make a Harry Potter Sorting Hat



Classification



Classification



Classification



* It doesn't have to be
correct all of the time



How about now?

What a computer sees

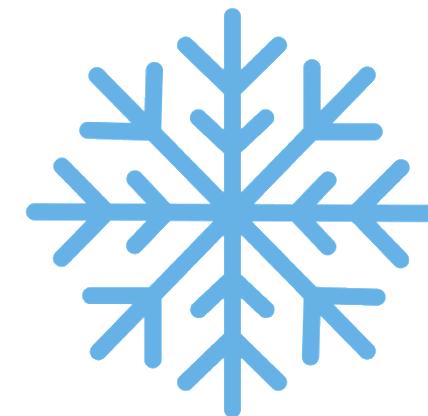
0	0	1	0	1	0	1	0	0	0	1	1	1	1	0	1
1	0	0	1	0	1	1	1	0	1	0	0	0	0	0	0
1	1	1	0	1	0	0	1	1	0	0	1	0	1	0	0
1	1	1	1	1	0	0	0	0	0	1	1	0	1	1	1
0	0	0	1	1	0	0	0	1	0	0	1	1	0	1	1
1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0
1	1	0	1	1	0	0	1	1	0	0	1	1	0	1	0
1	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0
0	0	0	0	1	0	1	0	1	0	1	0	1	1	1	1
0	1	1	0	0	0	0	0	0	1	1	1	1	1	1	1
0	0	1	0	1	1	1	0	0	0	0	0	0	0	0	0
0	1	1	1	0	1	0	0	0	1	0	0	1	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	1	0	1	1	1	1	1
0	0	1	1	1	0	1	0	1	1	0	1	1	1	1	1



What a human sees



Very hard to Program



```
class DigitDetector:  
  
    def detect(raw_image):  
        # Return a 0 or 1 based on the pixels of the image  
  
        # TODO
```



Perhaps there is an insight?

Two Great Ideas

1. Artificial Neurons

2. Learn by Example

Two Great Ideas

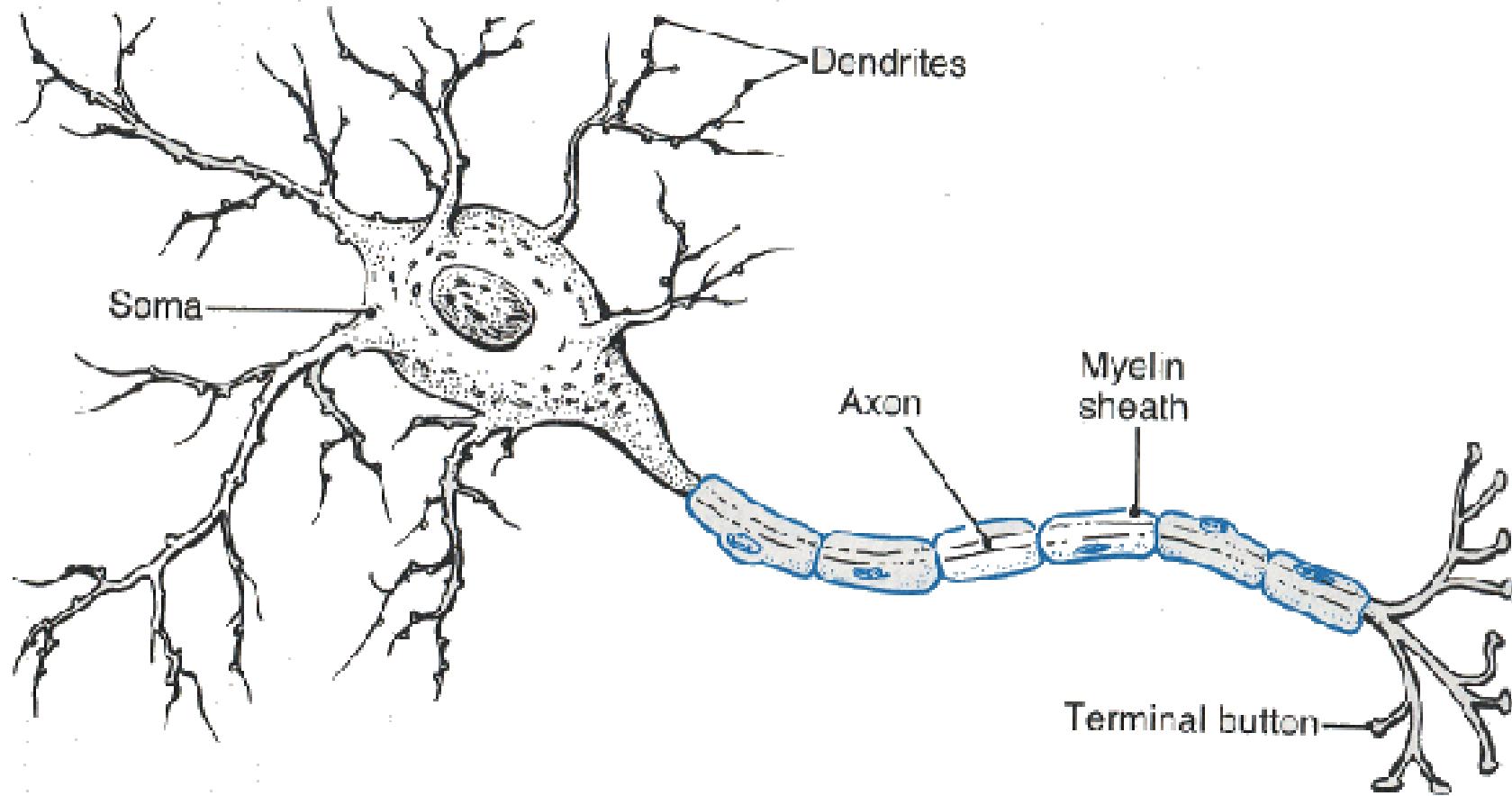
1. Artificial Neurons

2. Learn by Example

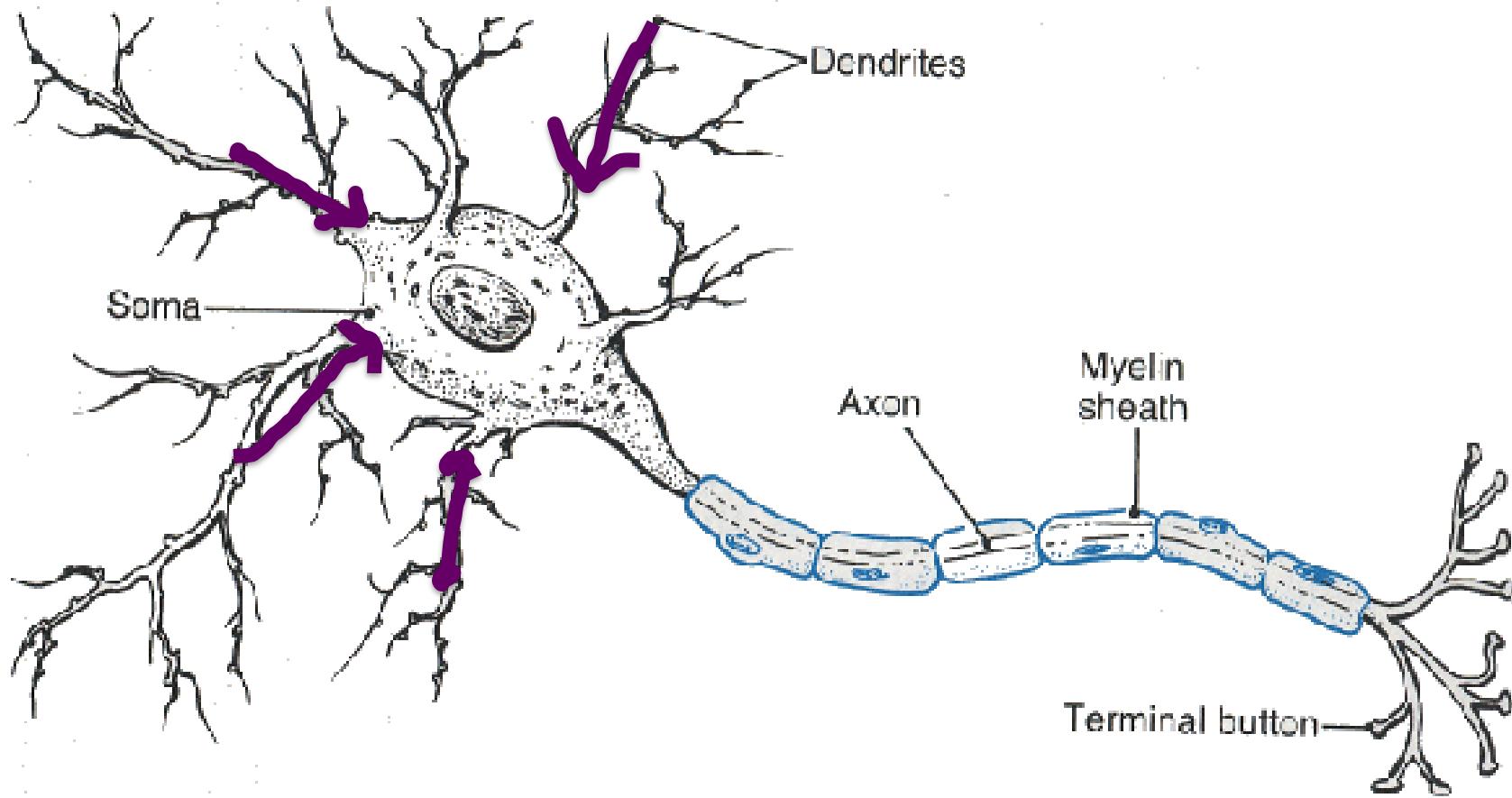
1. Artificial Neurons



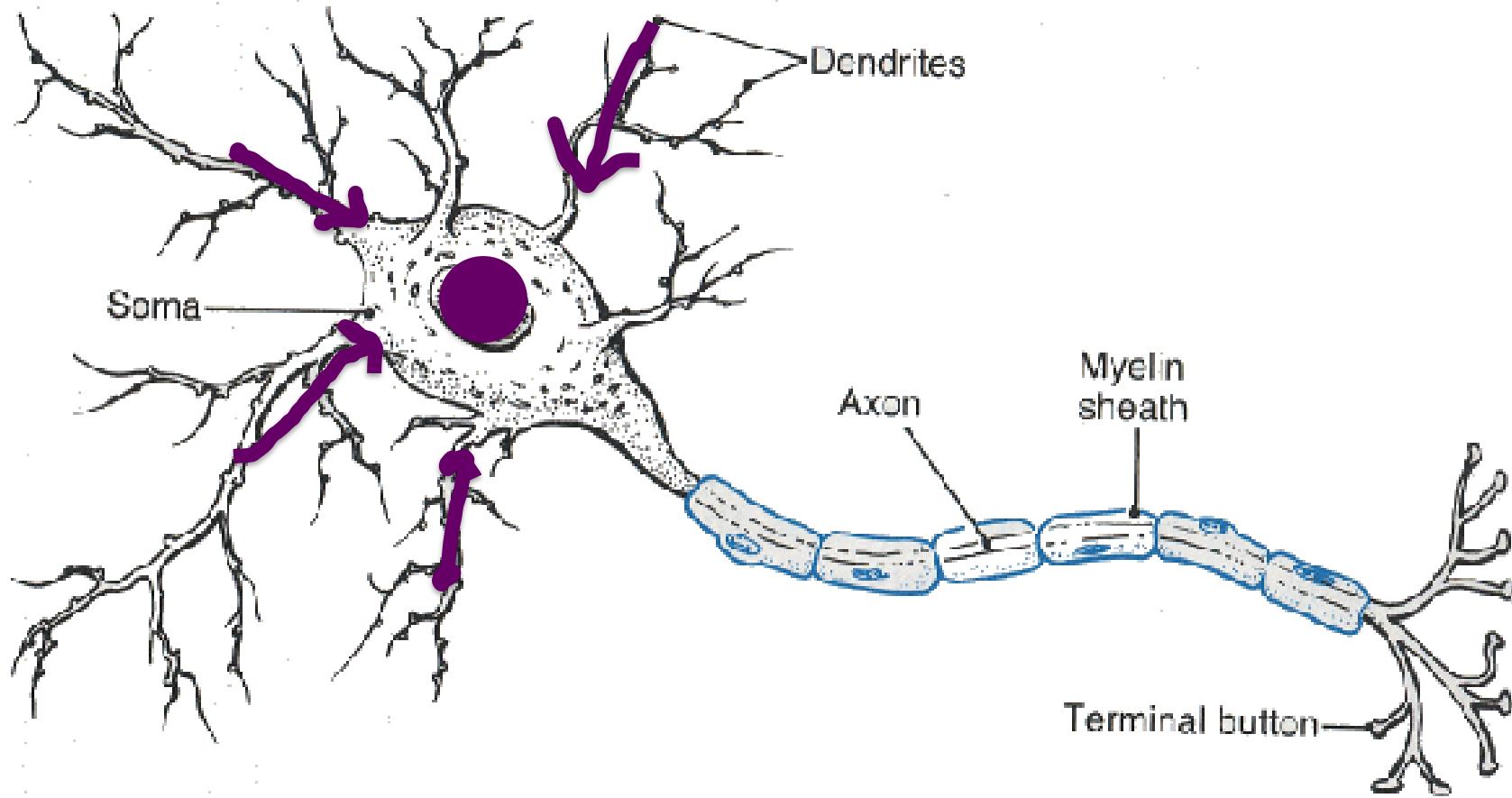
Neuron



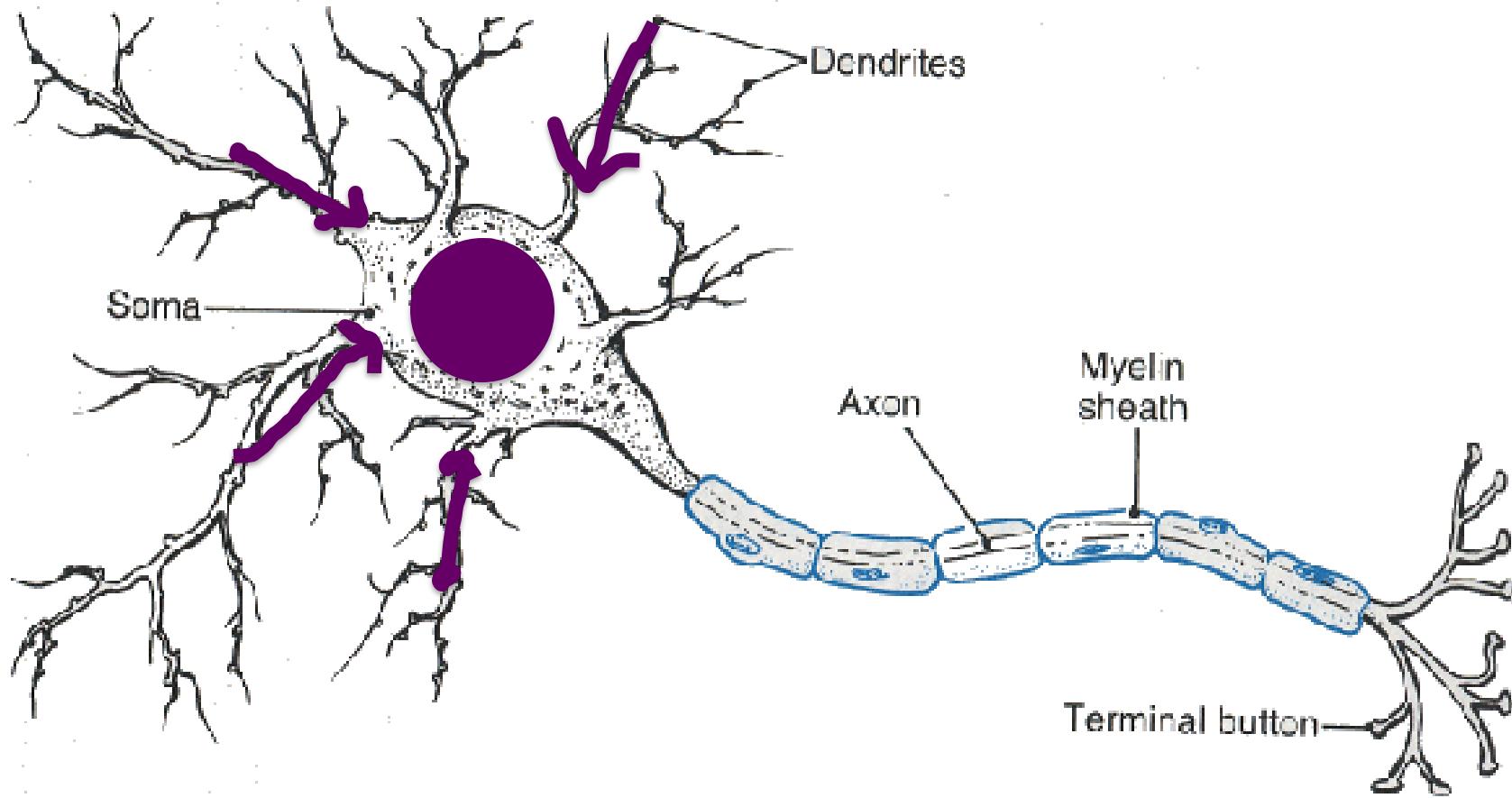
Neuron



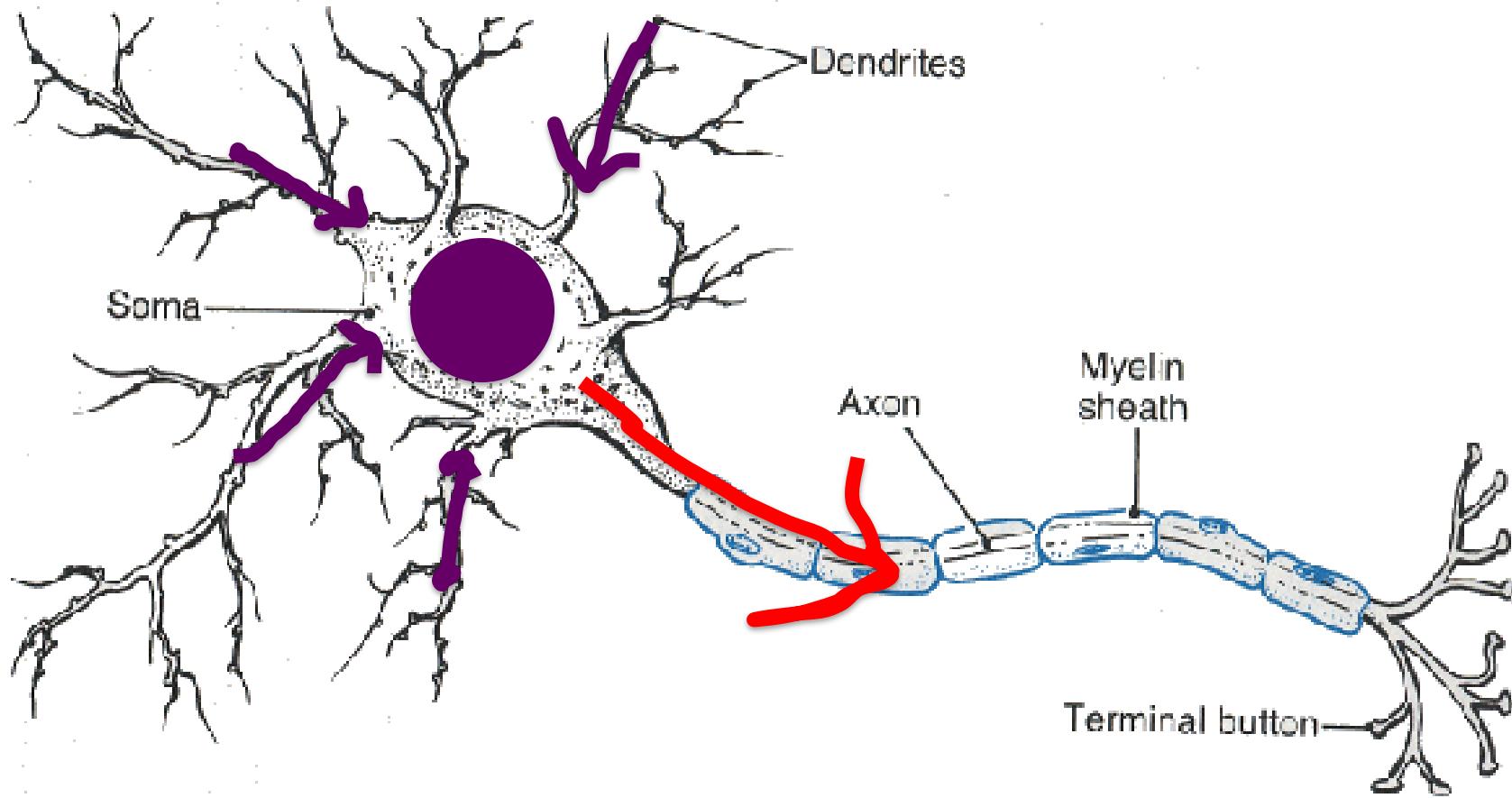
Neuron



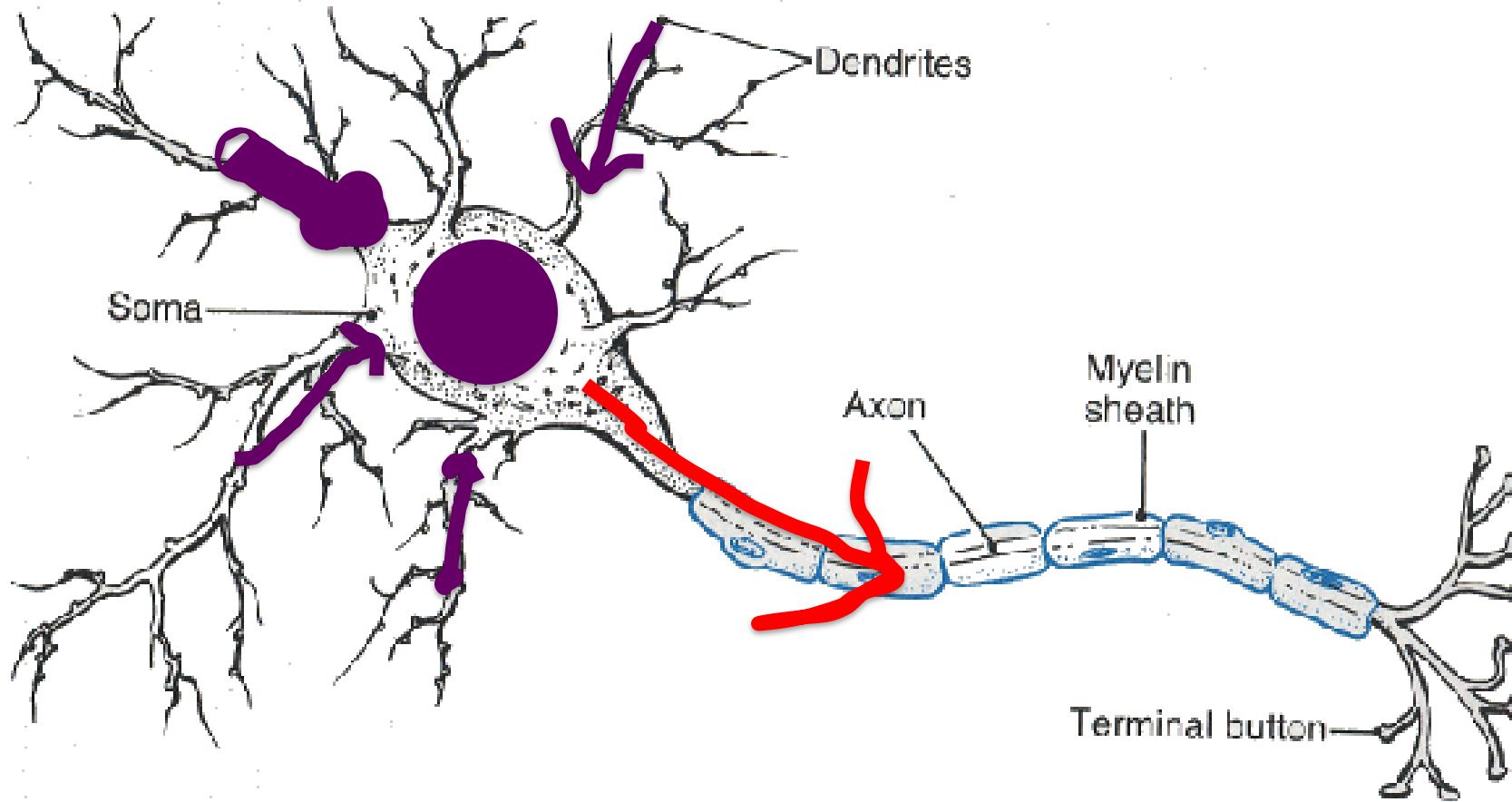
Neuron



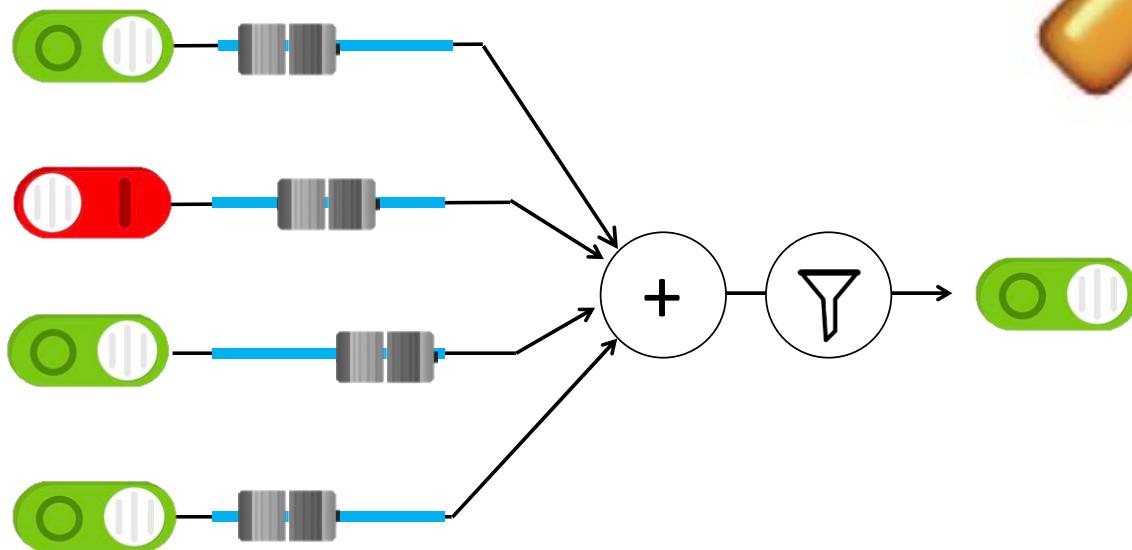
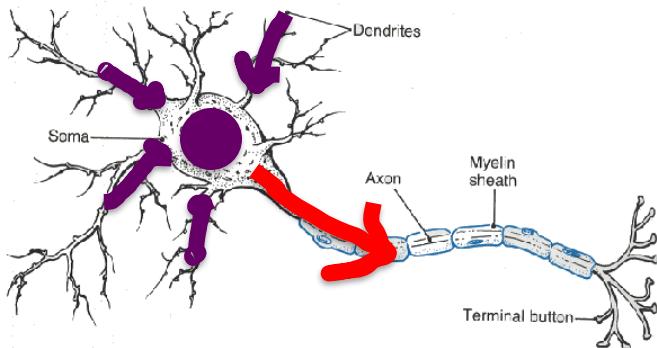
Neuron



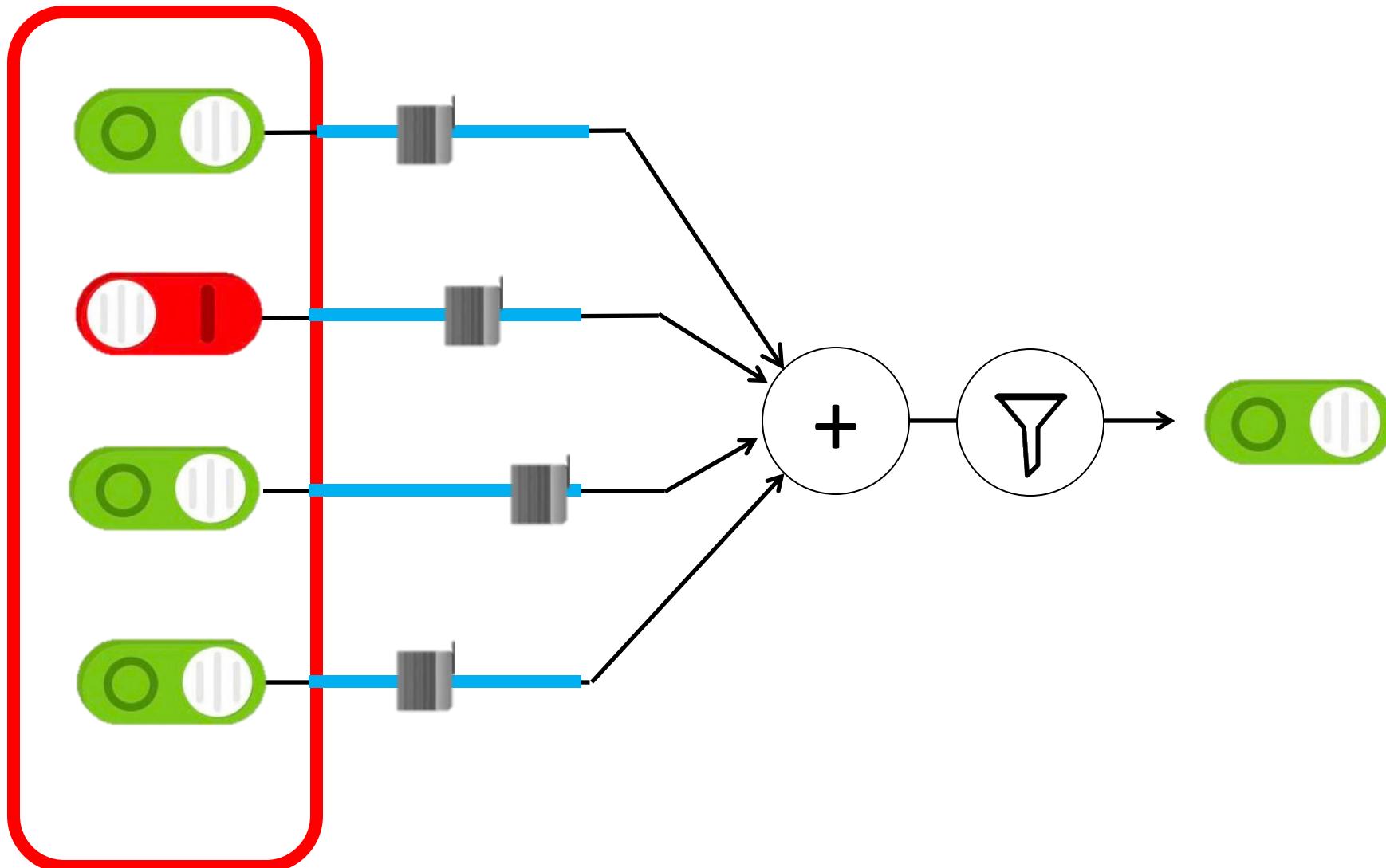
Some Inputs are More Important



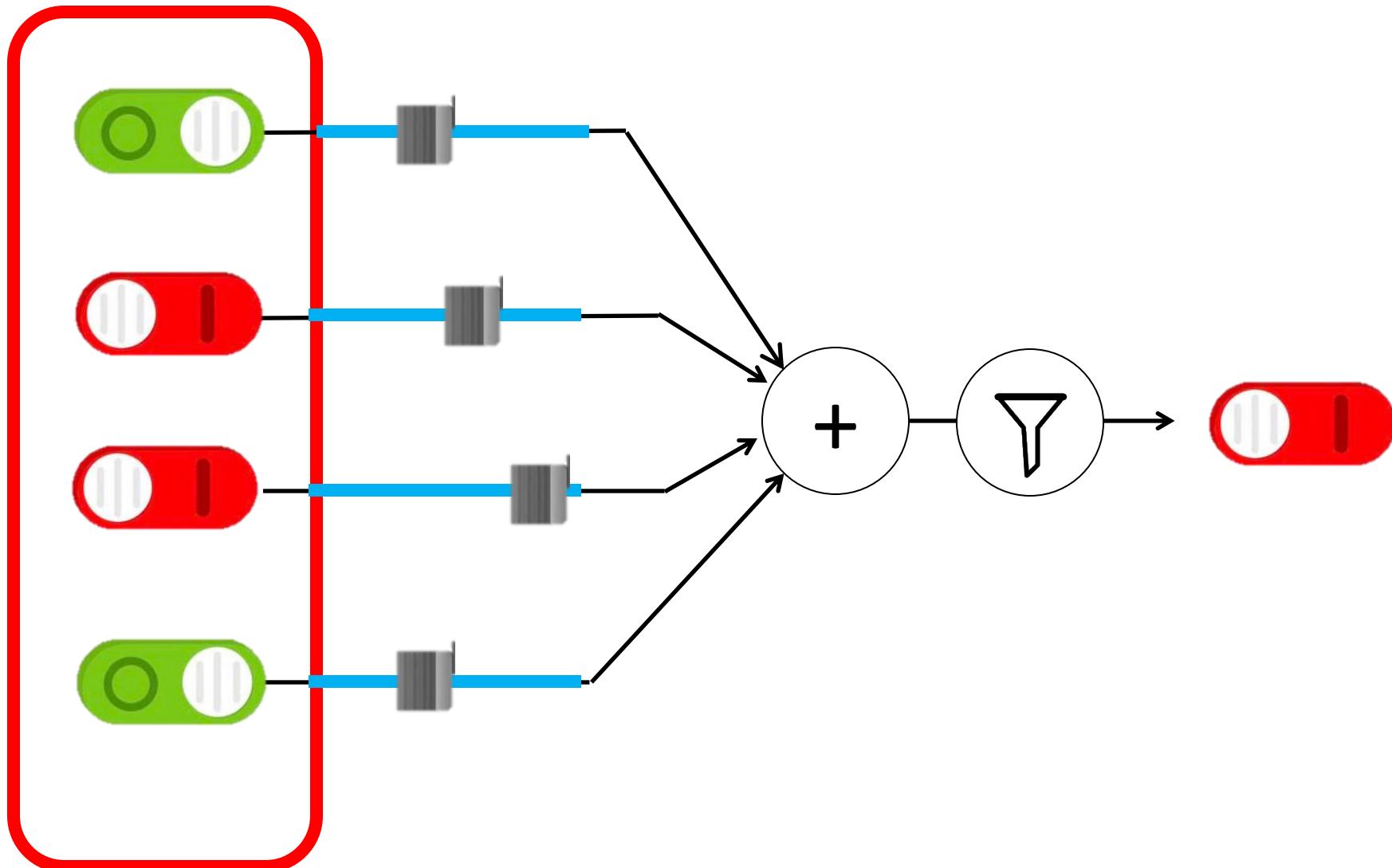
Artificial Neuron



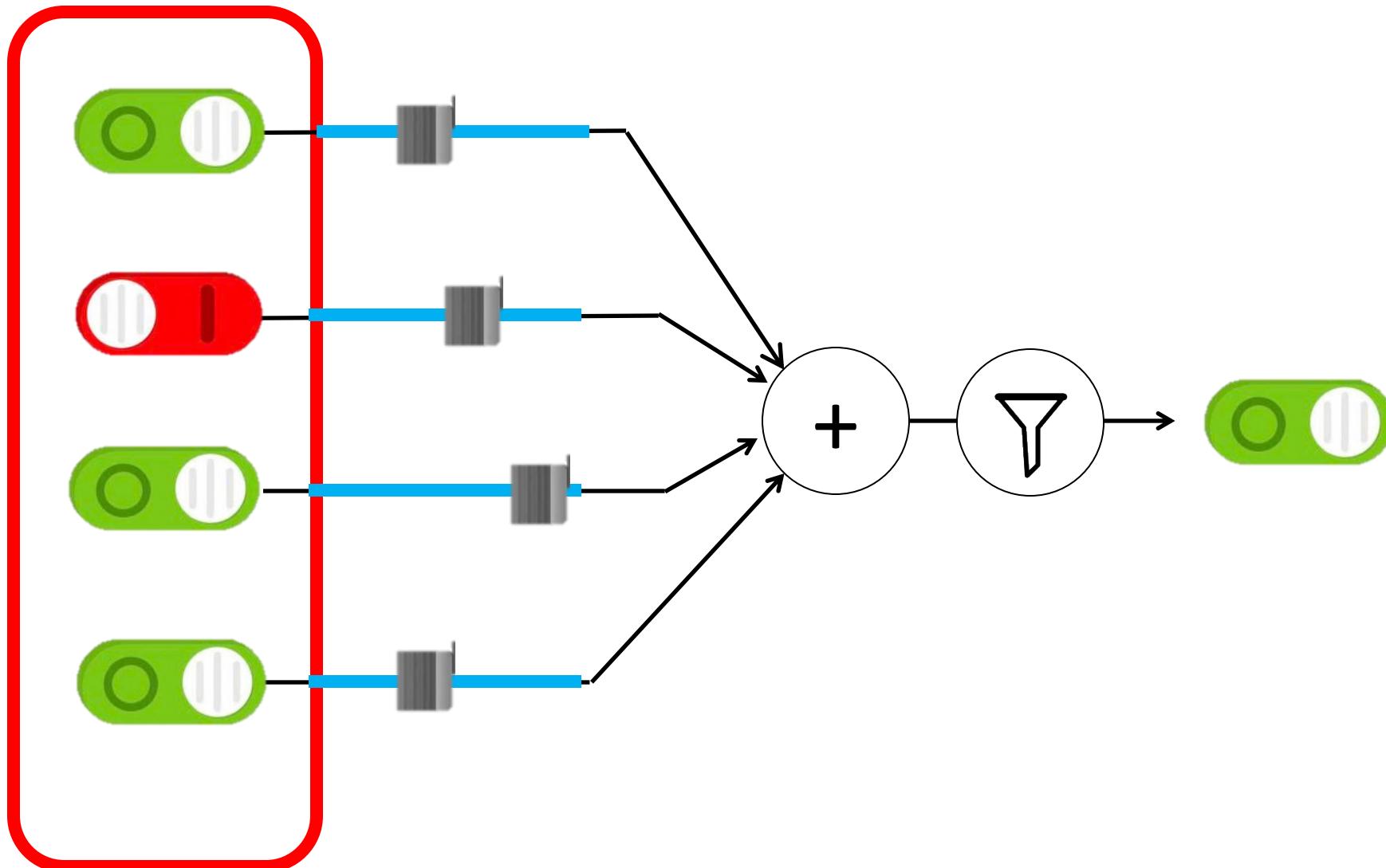
Inputs



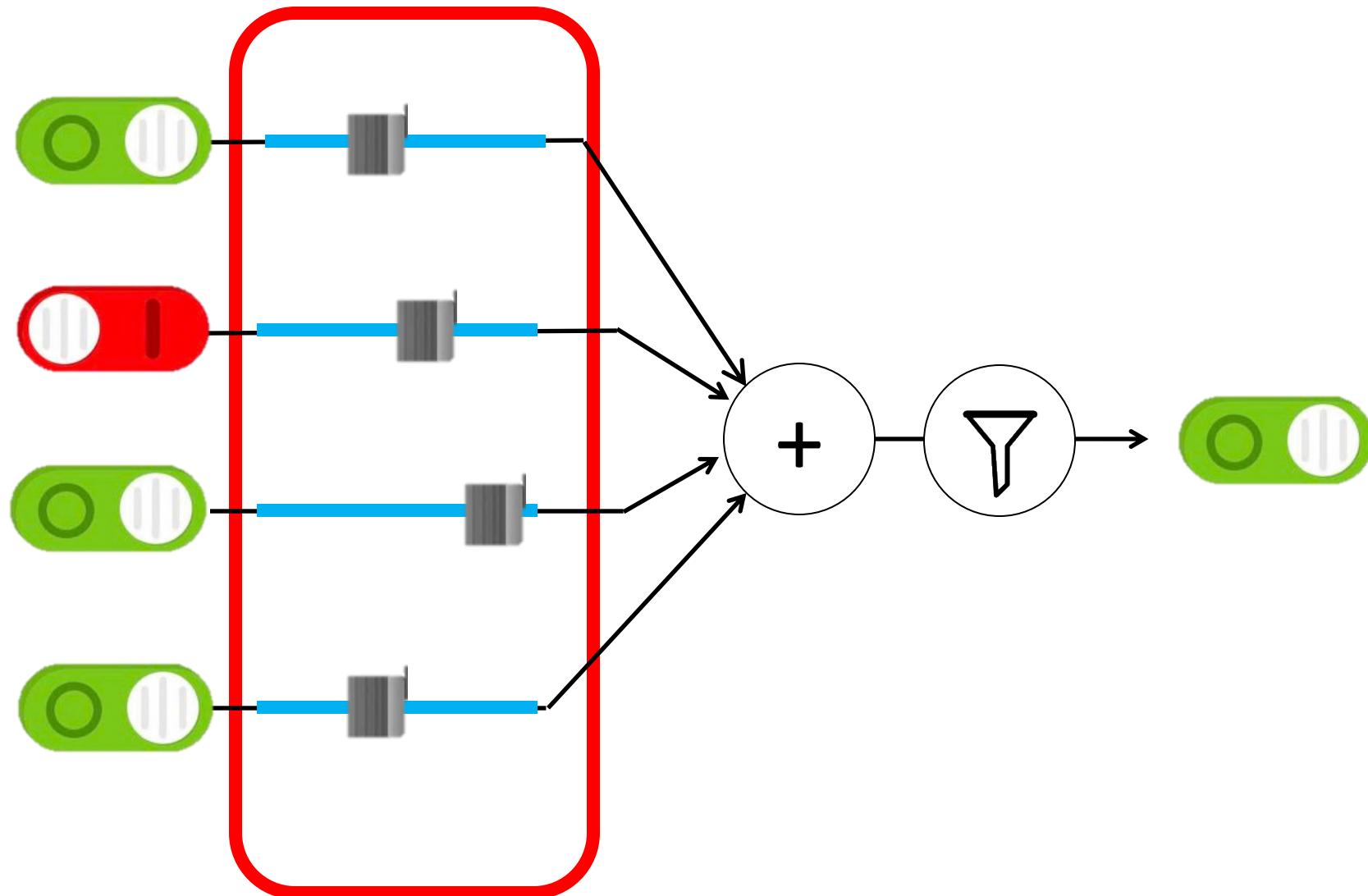
Inputs



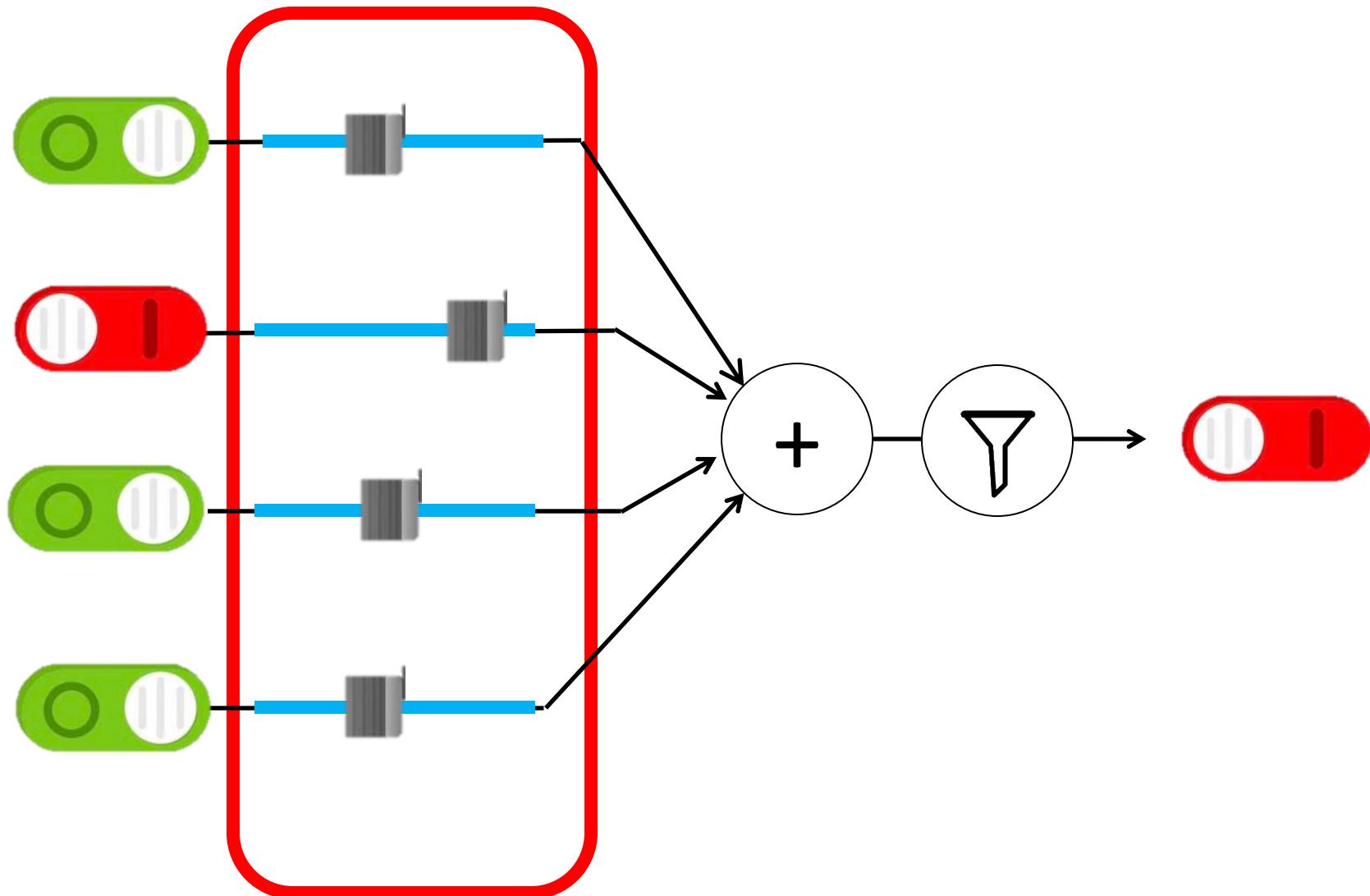
Inputs



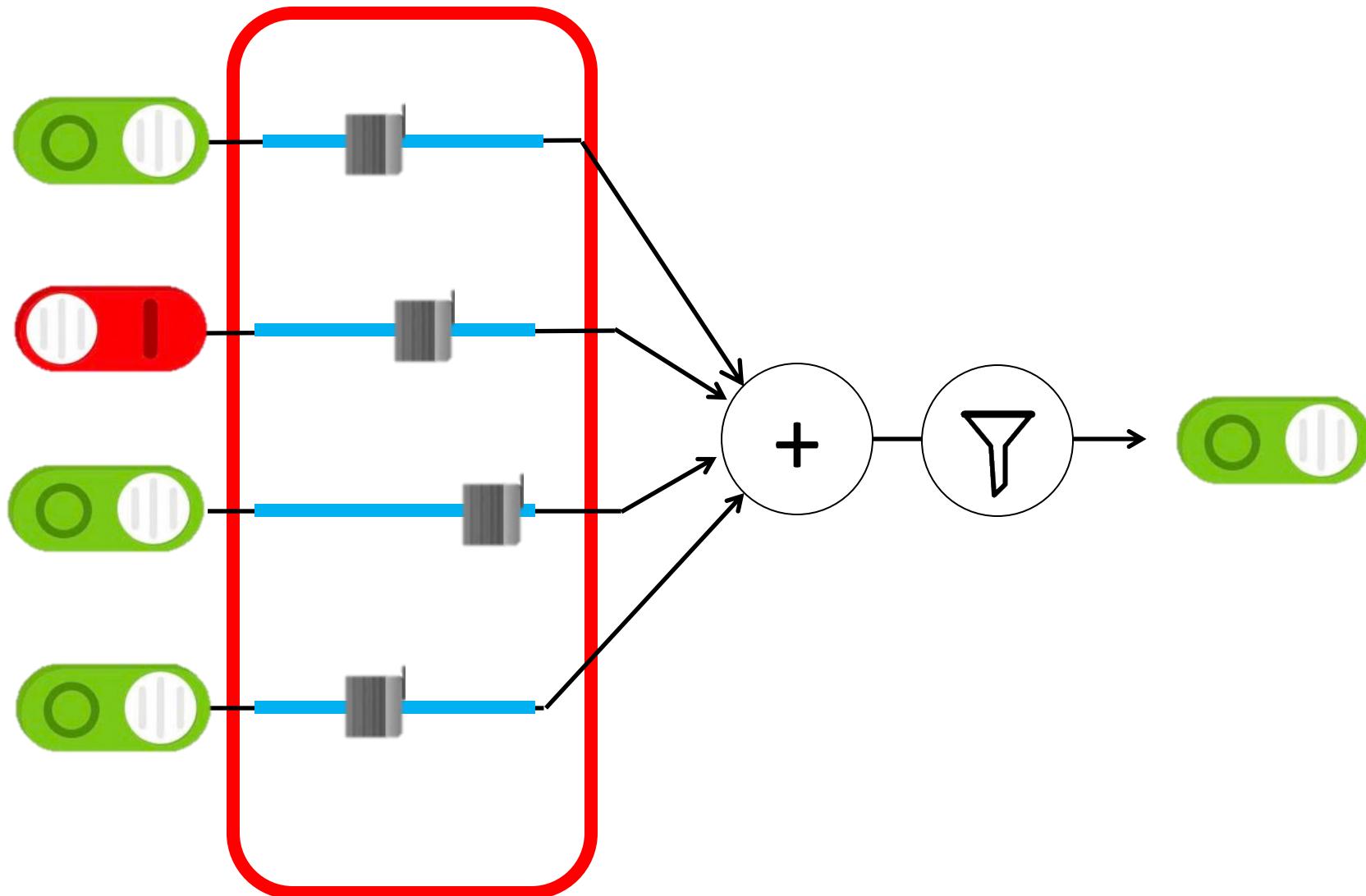
Weights



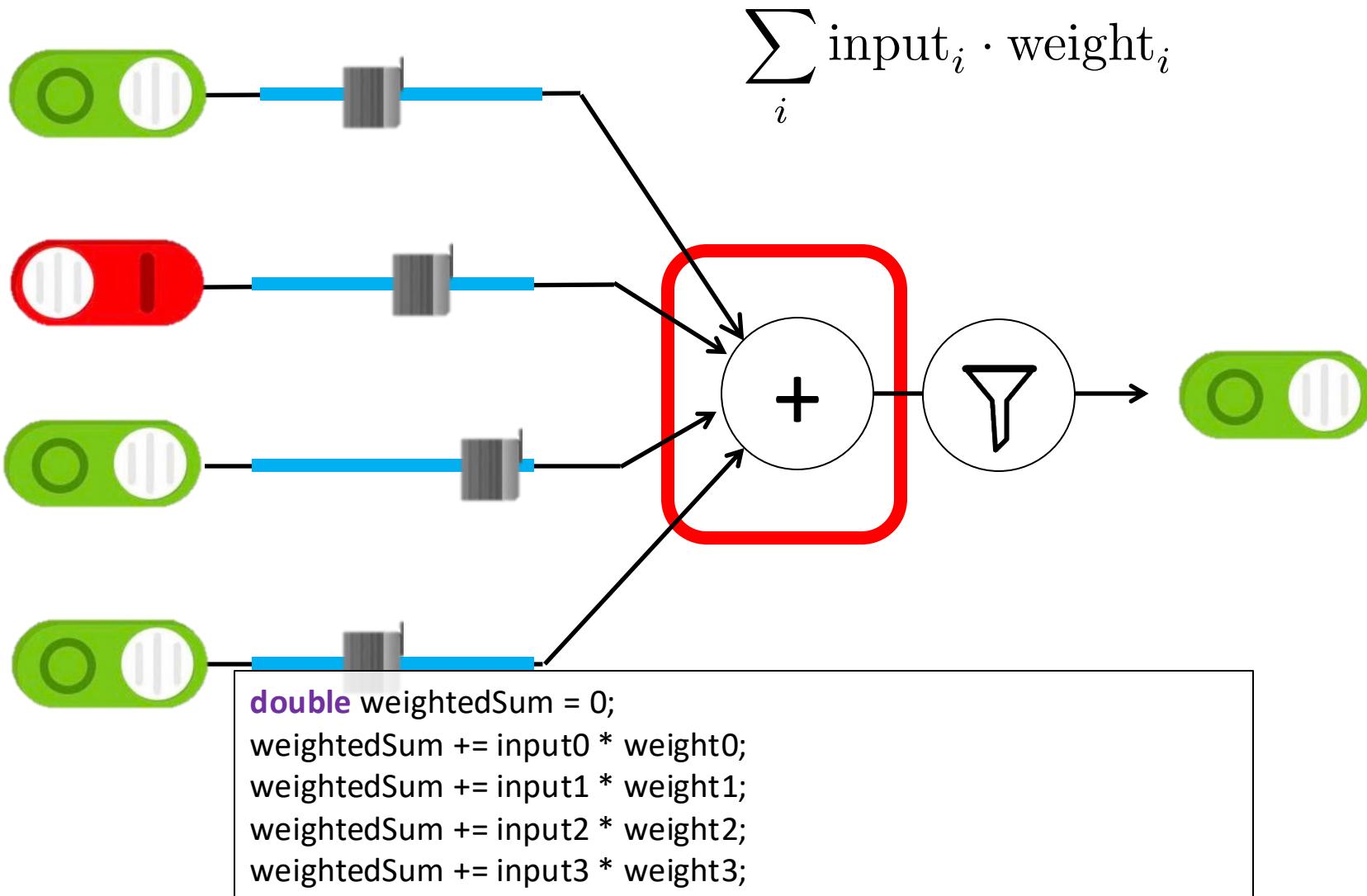
Weights



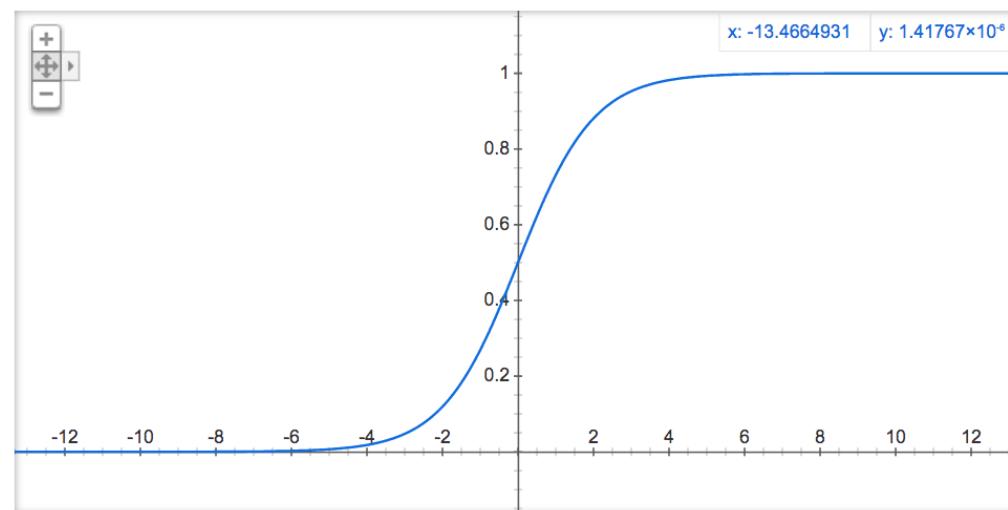
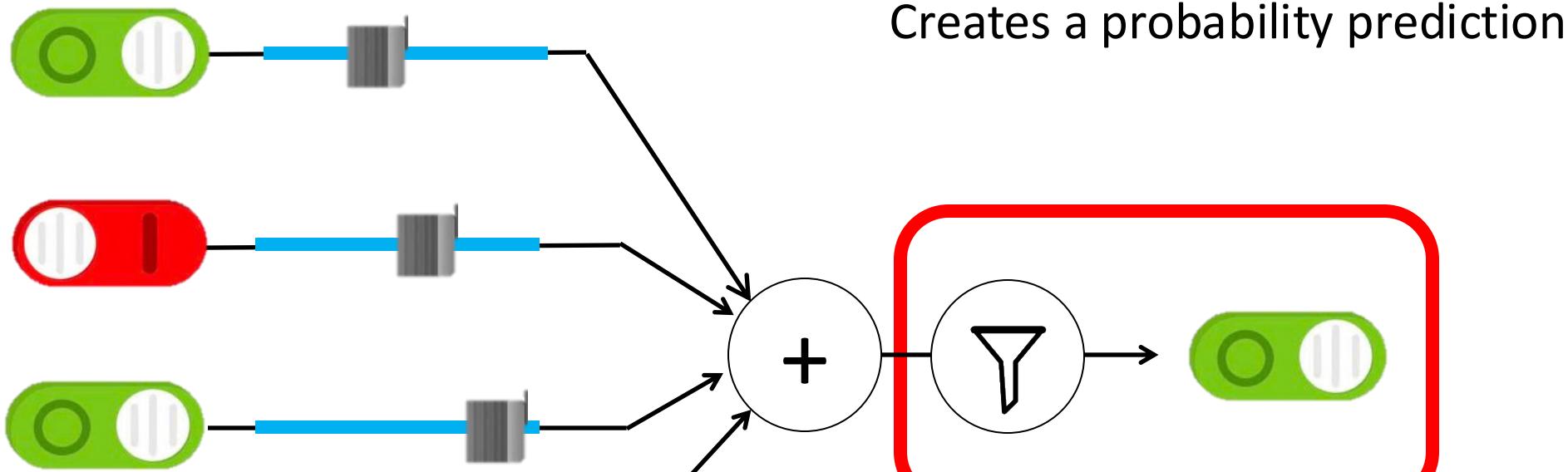
Weights



Weighted Sum

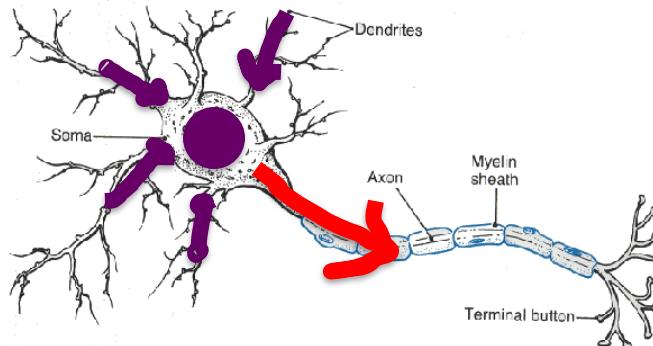


Filter and Output

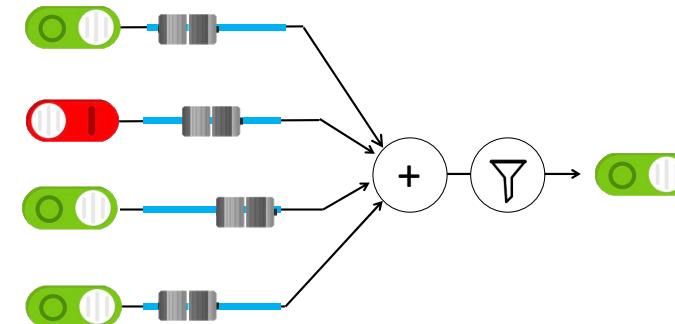
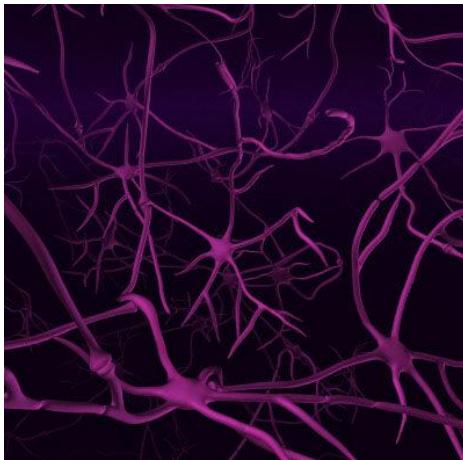


Biological Basis for Neural Networks

- A neuron



- Your brain



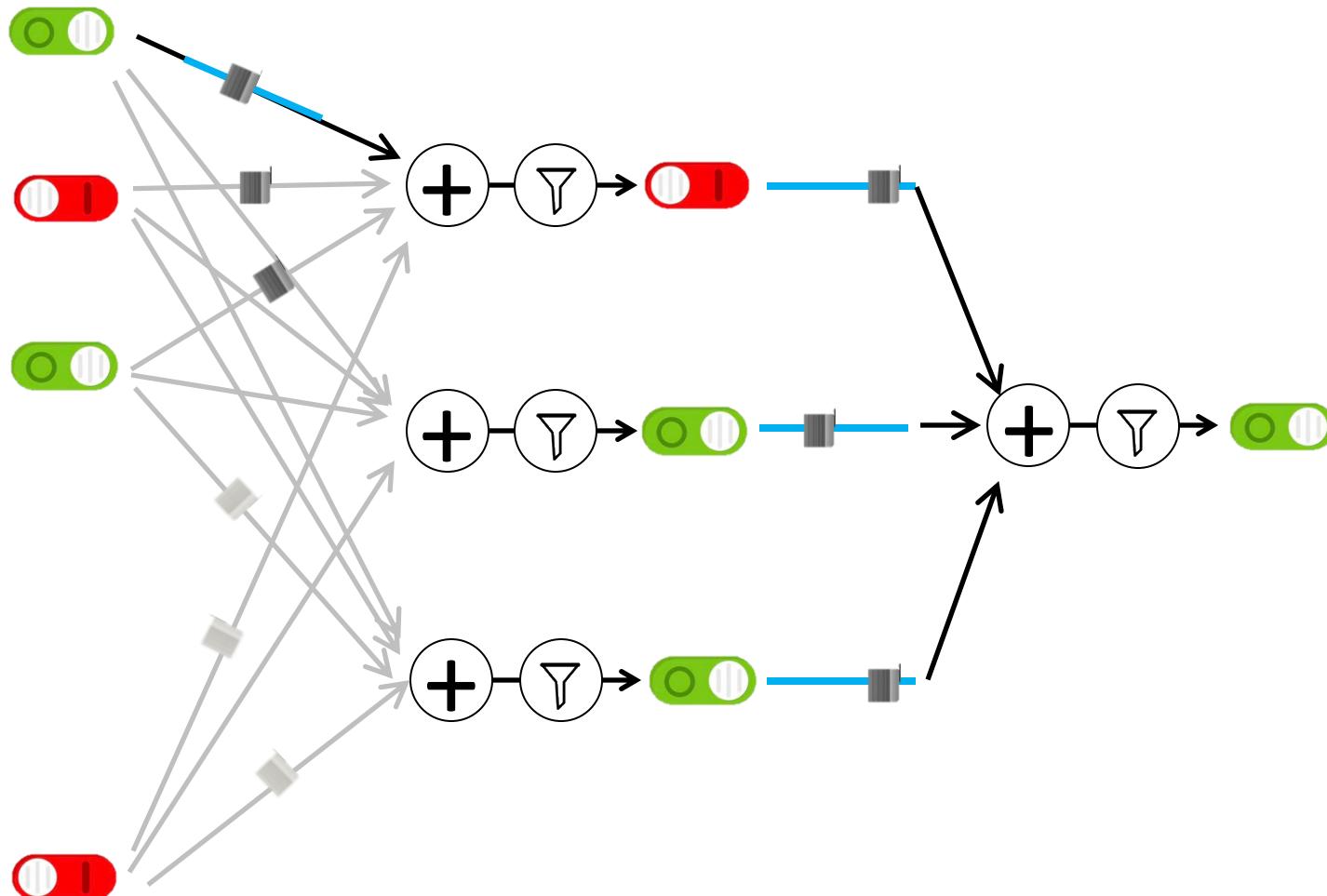
???

Actually, it's probably someone else's brain

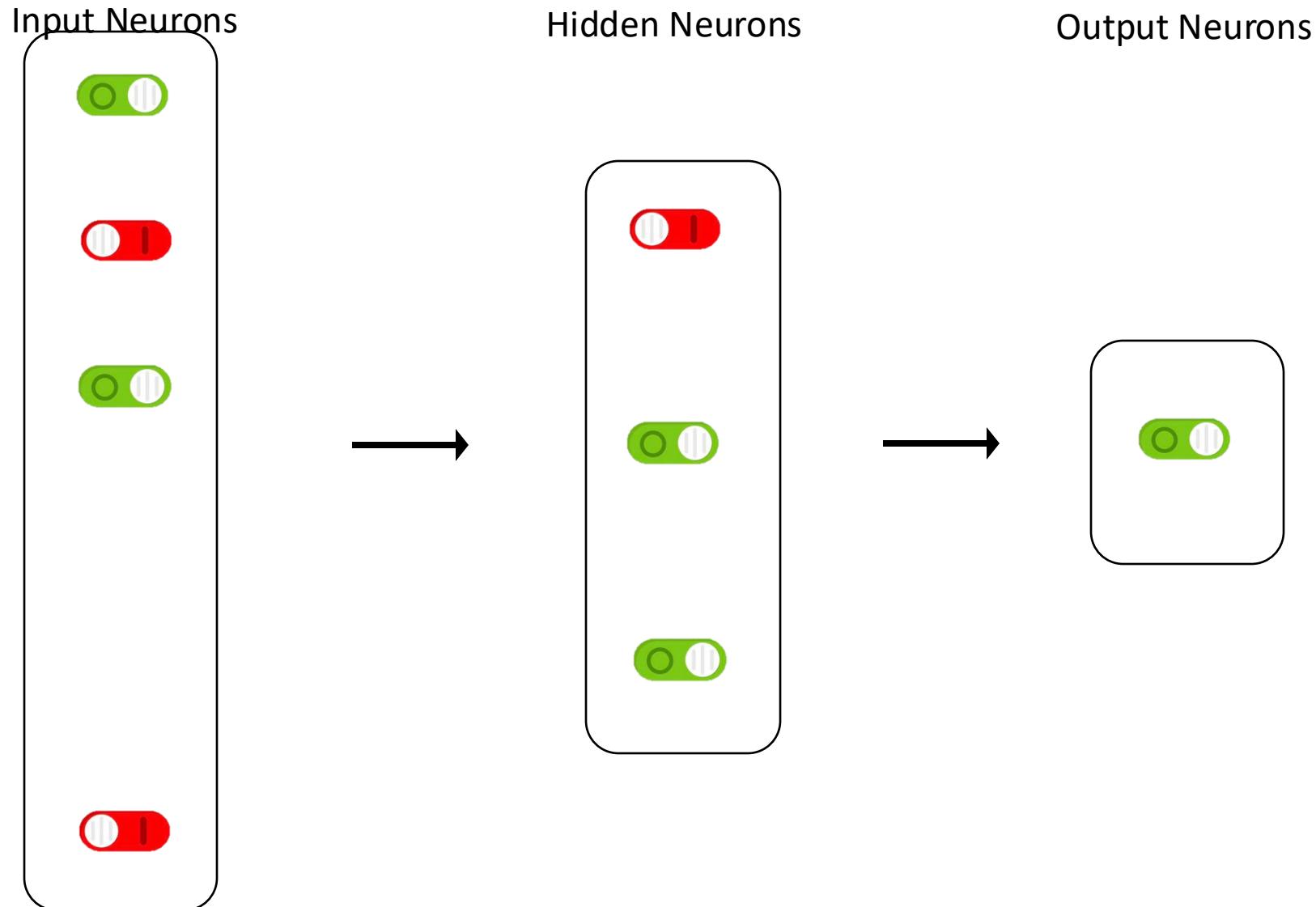


Piech

Put Many Together

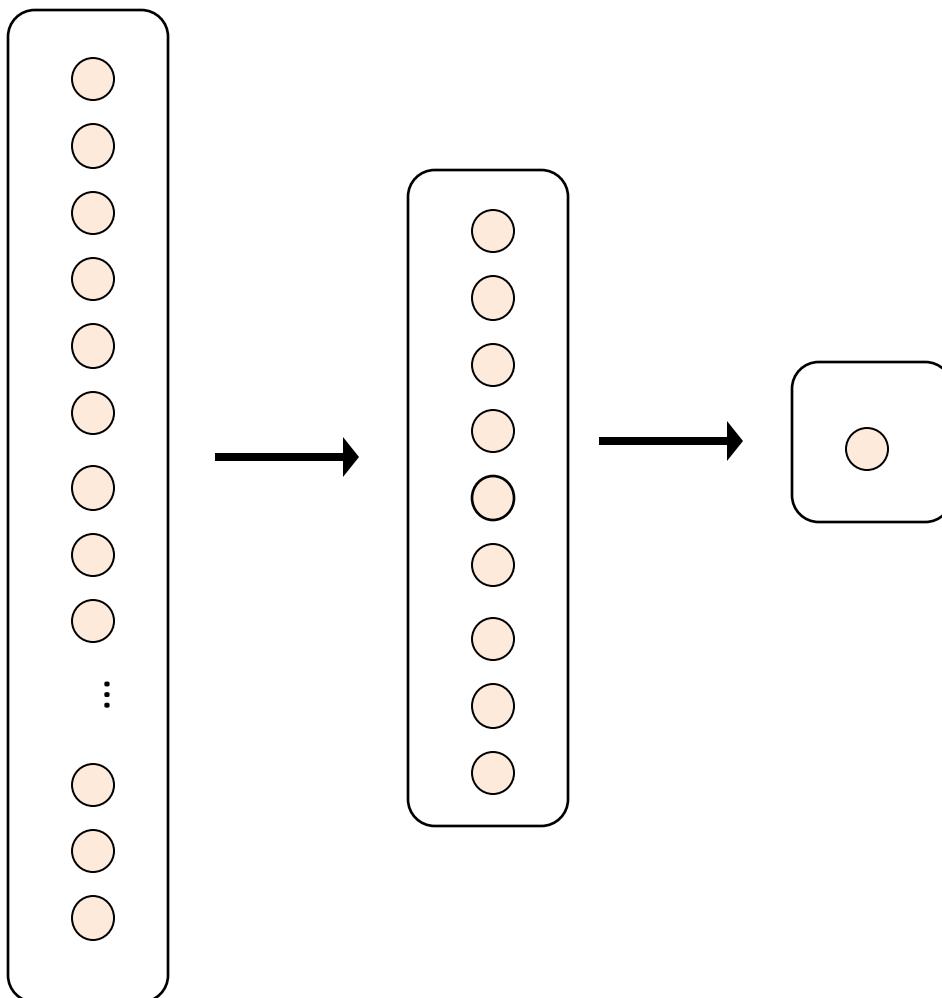


Put Many Together

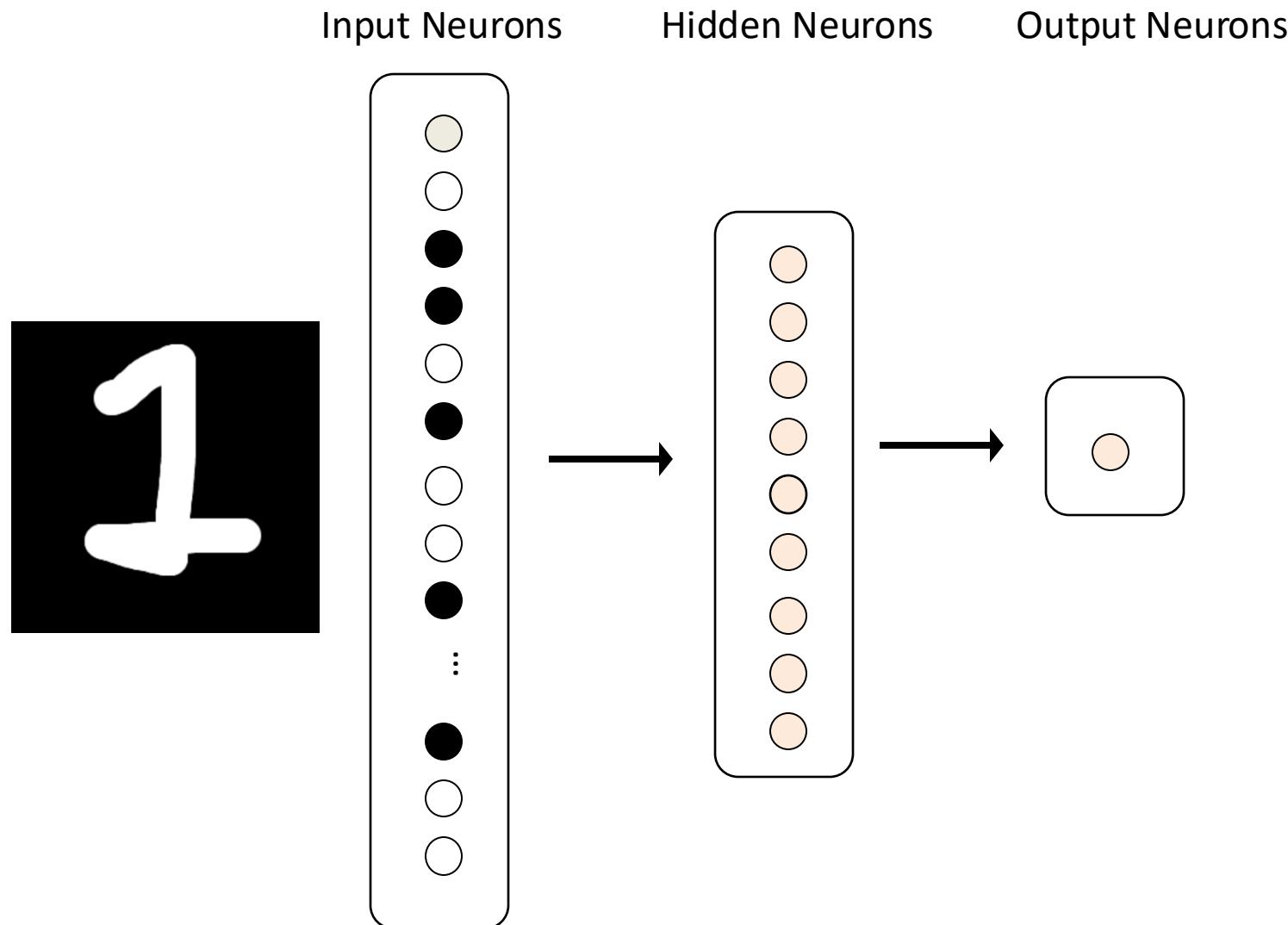


Making a Prediction

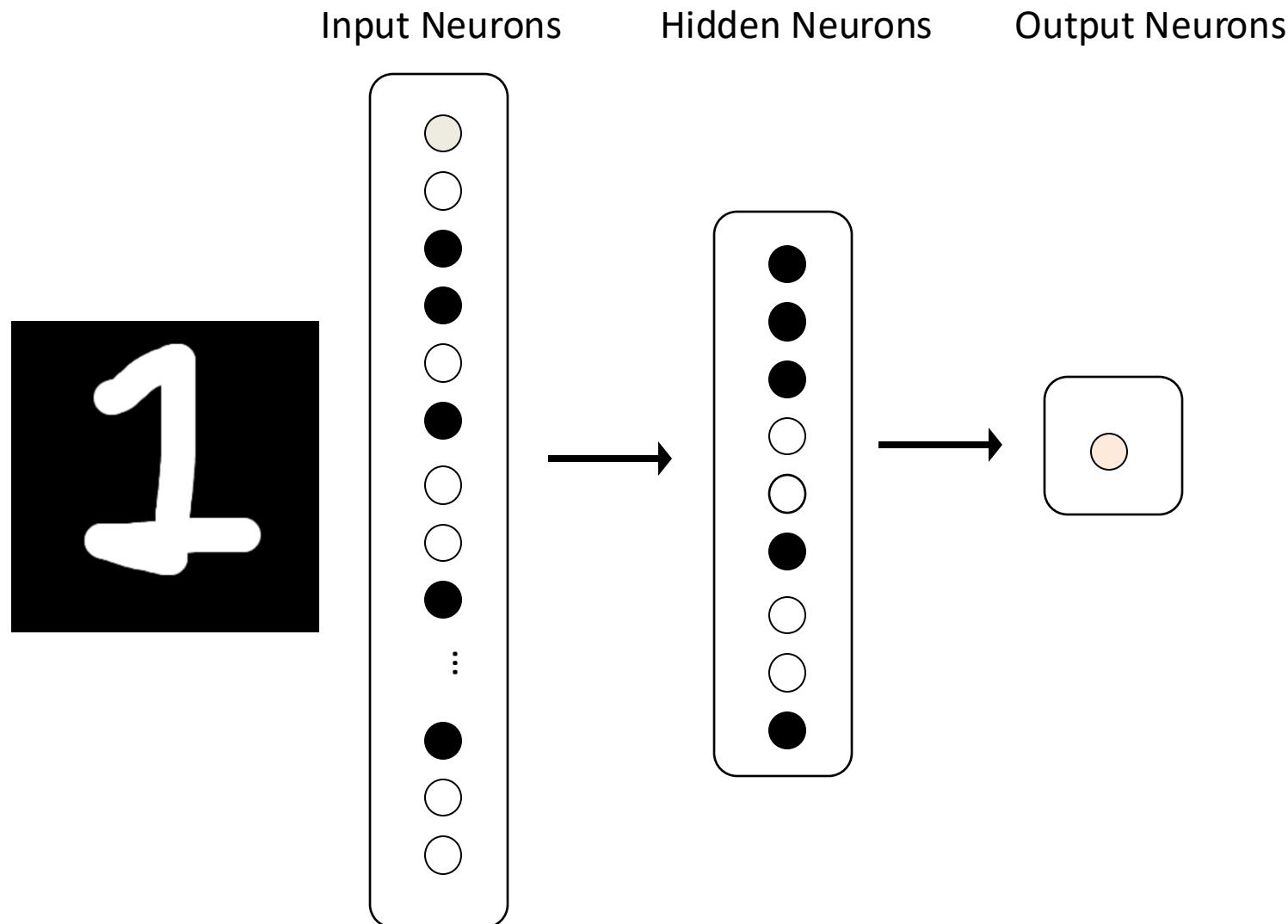
Input Neurons Hidden Neurons Output Neurons



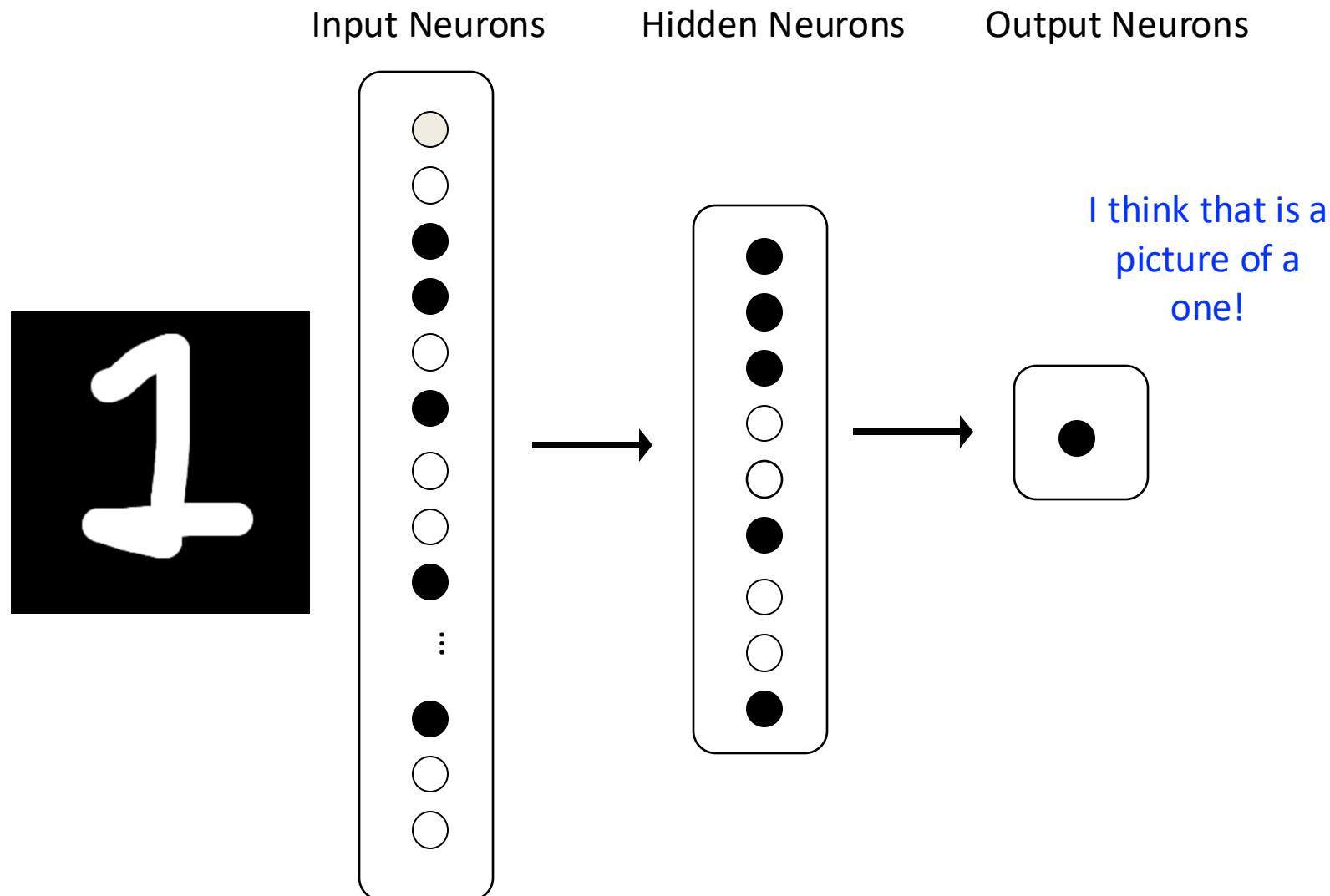
Making a Prediction



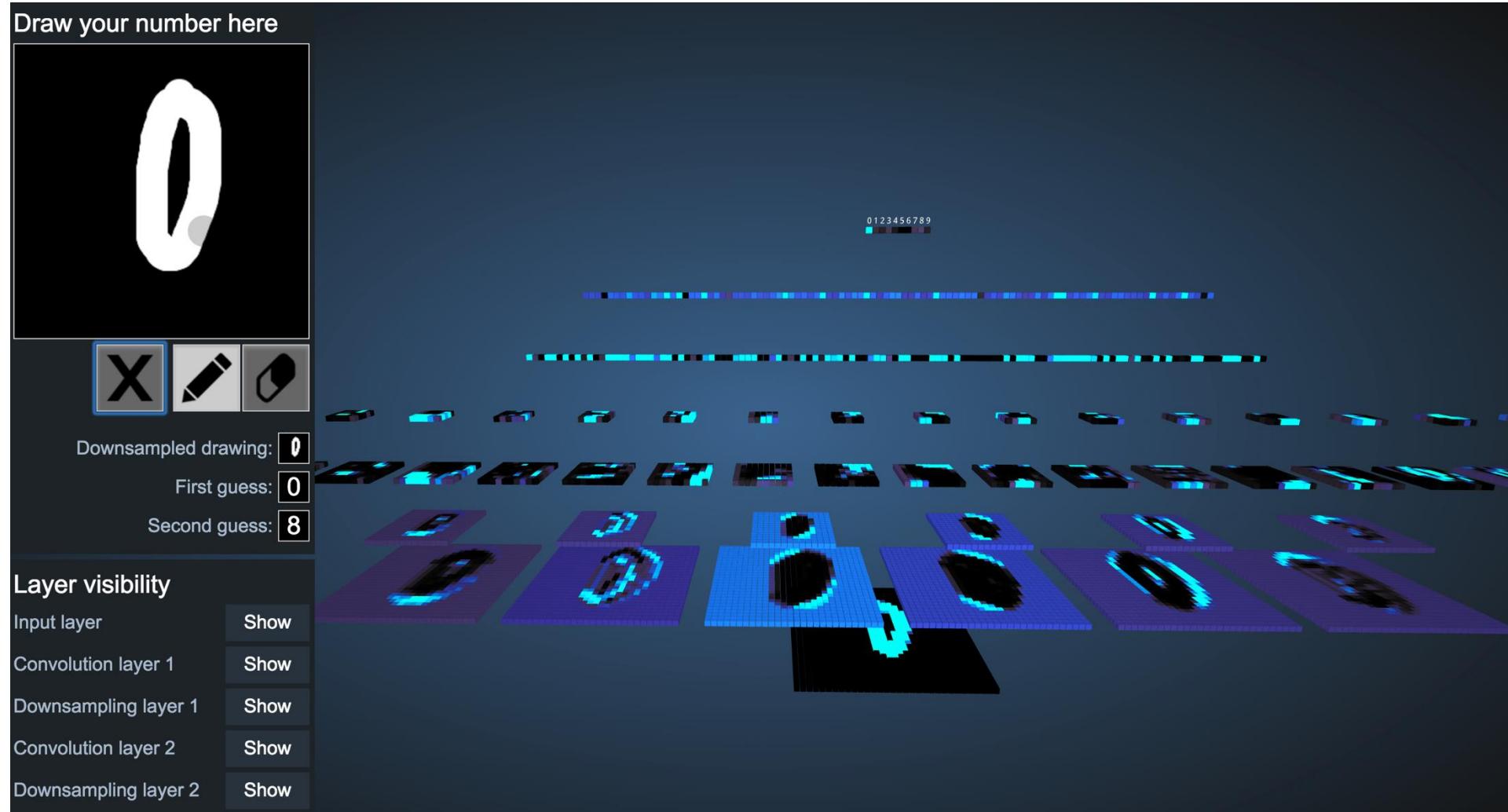
Making a Prediction



Making a Prediction



Demonstration

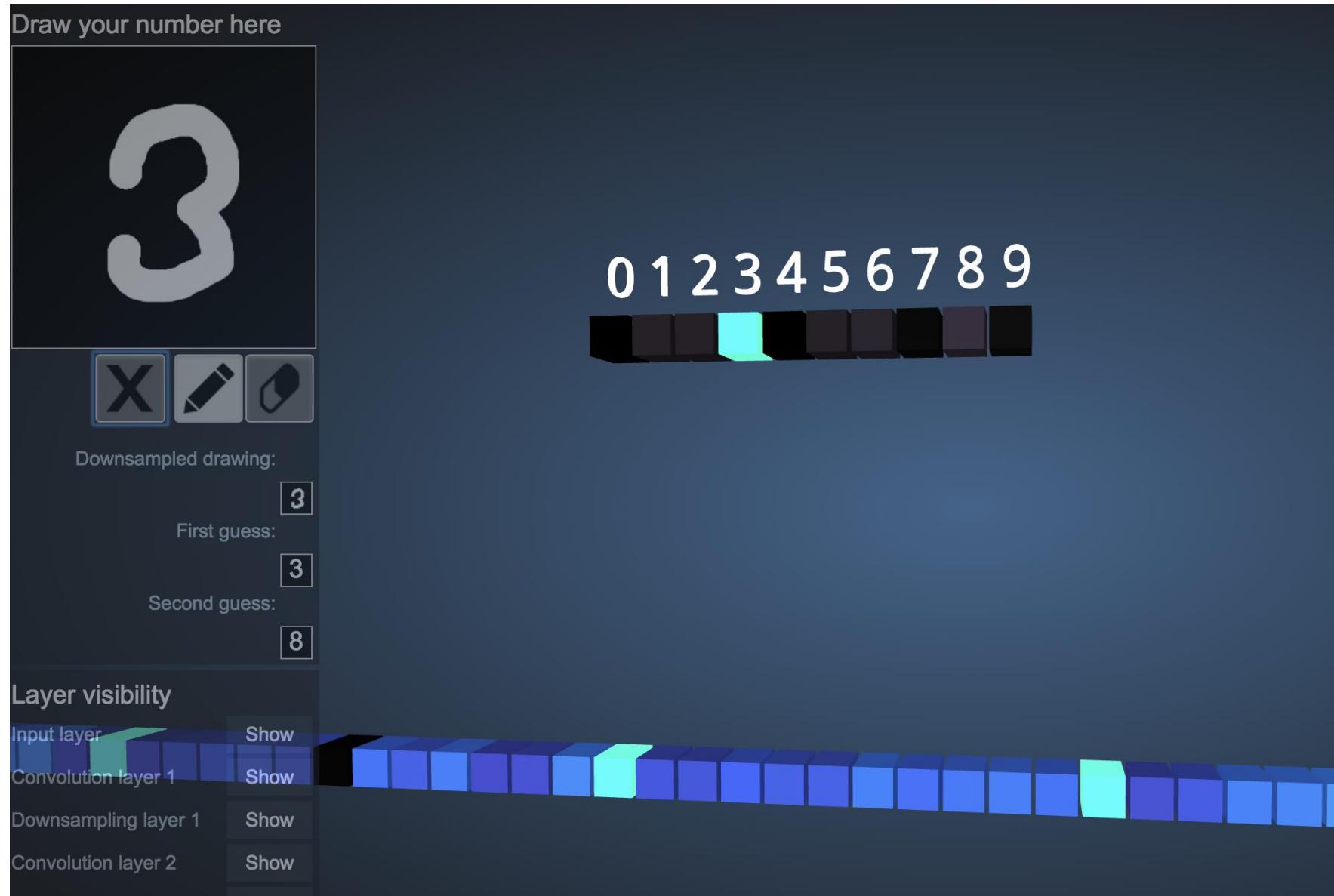


<https://web.archive.org/web/20211117115916/https://www.cs.ryerson.ca/~aharley/vis/conv/>

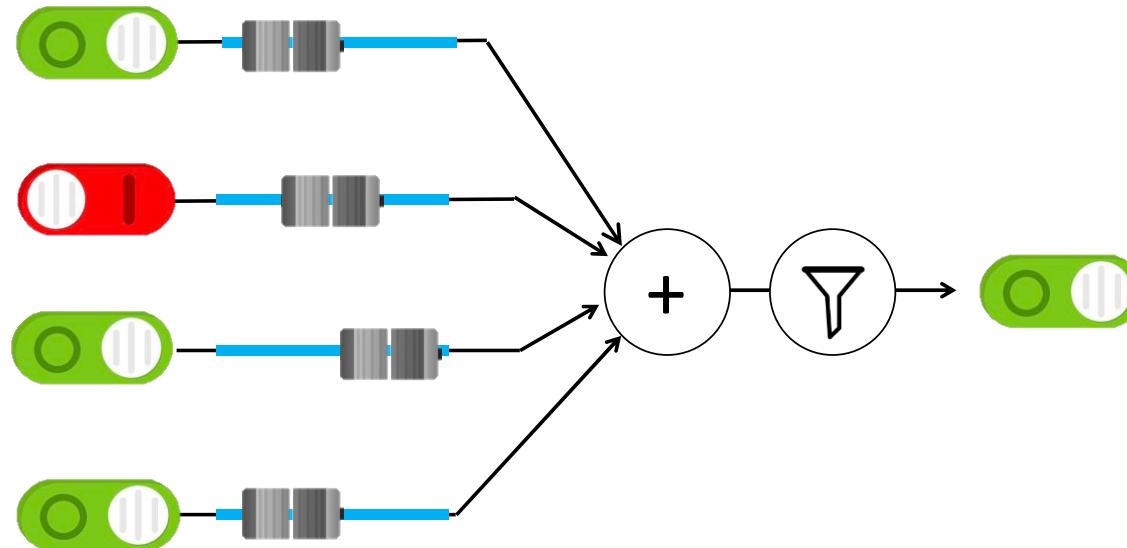
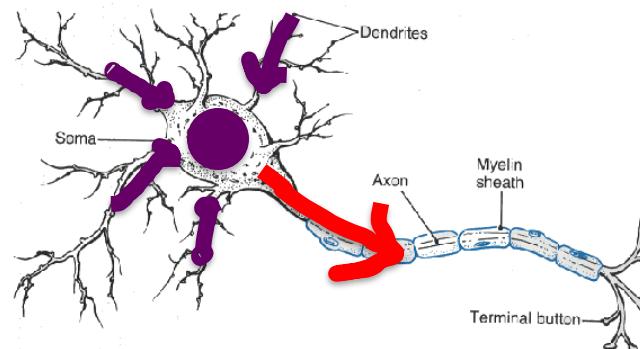


Piech

Interpret the Output as Prediction



Great Idea: Artificial Neurons



Where do Artificial
Neural Networks
get their
intelligence from?



Neural Networks get their intelligence from their sliders (parameters)



Two Great Ideas

1. Artificial Neurons

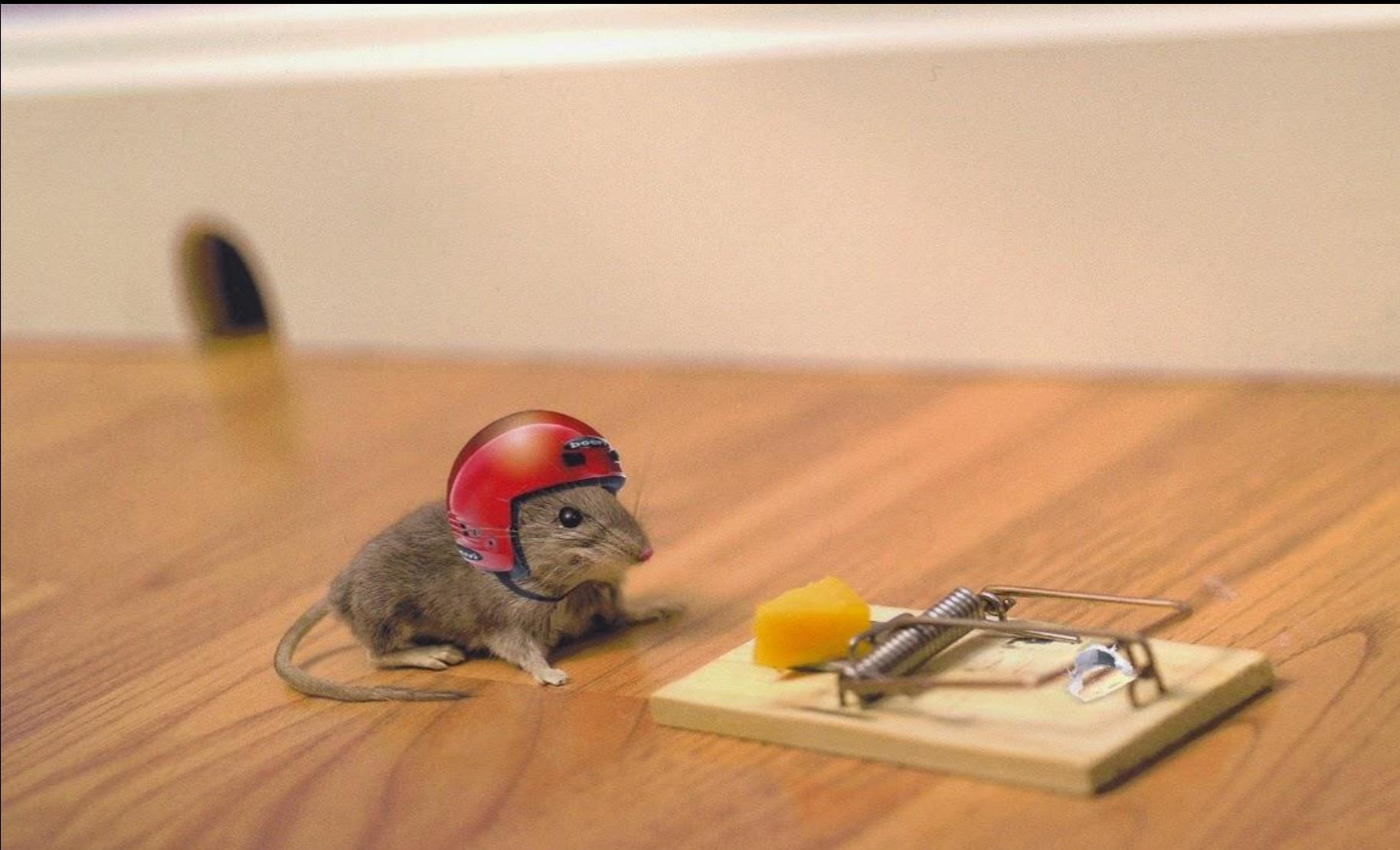
2. Learn by Example

Two Great Ideas

1. Artificial Neurons

2. Learn by Example

2. Learn From Experience



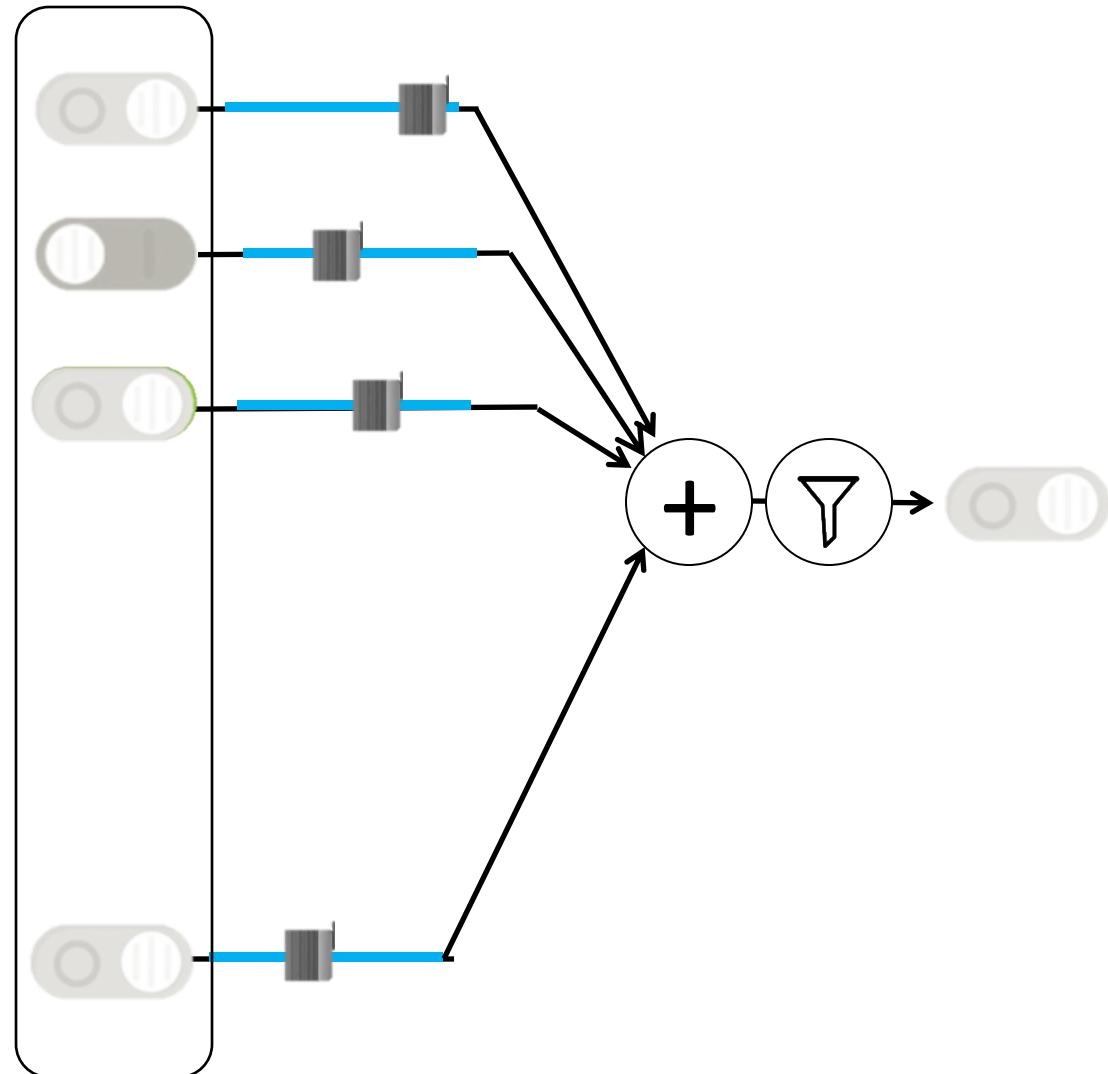
Piech

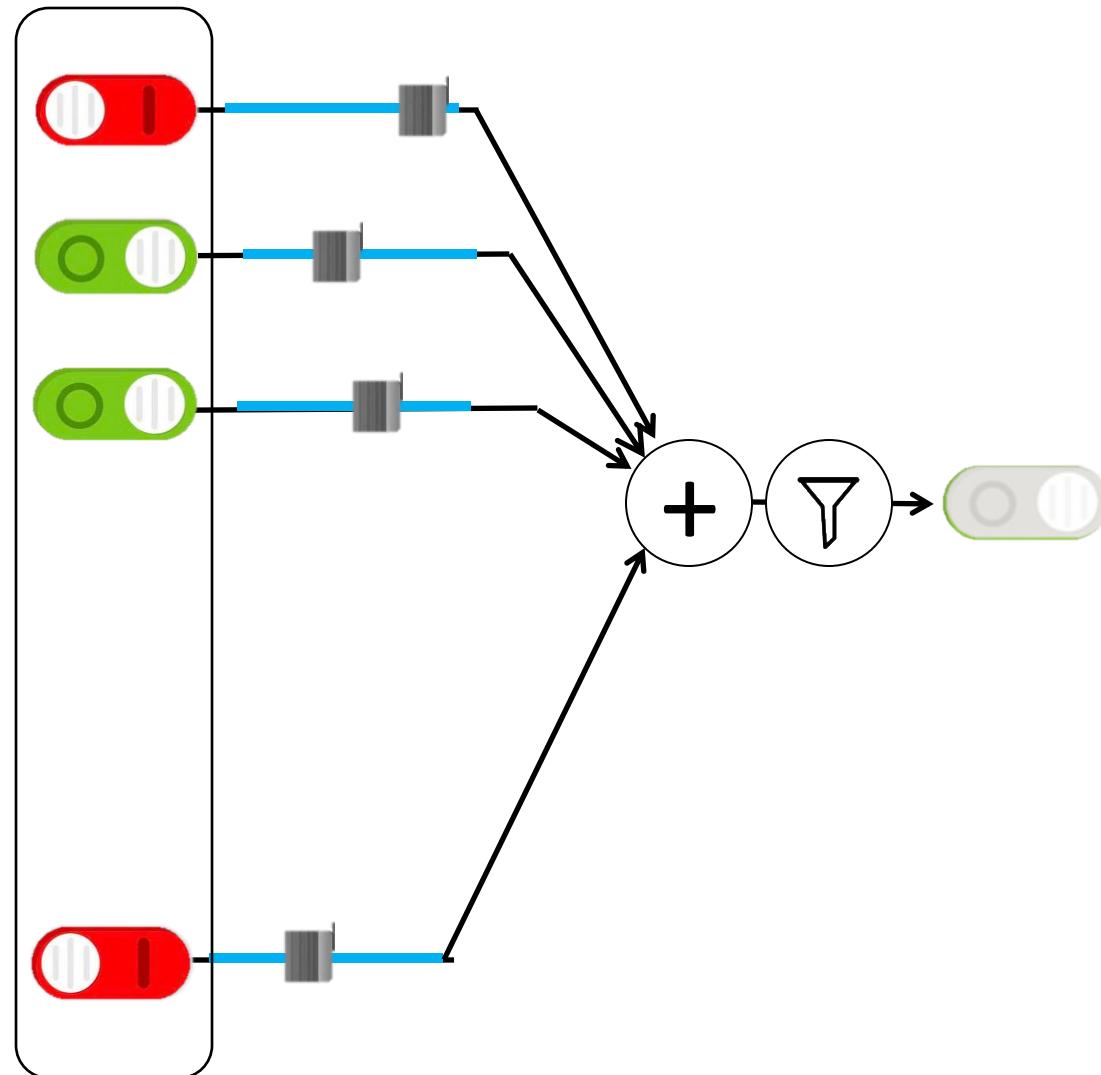
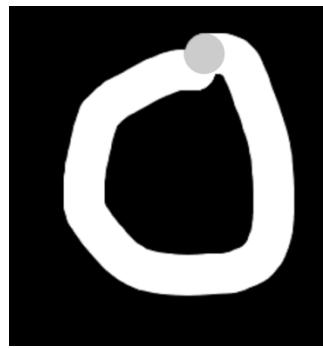


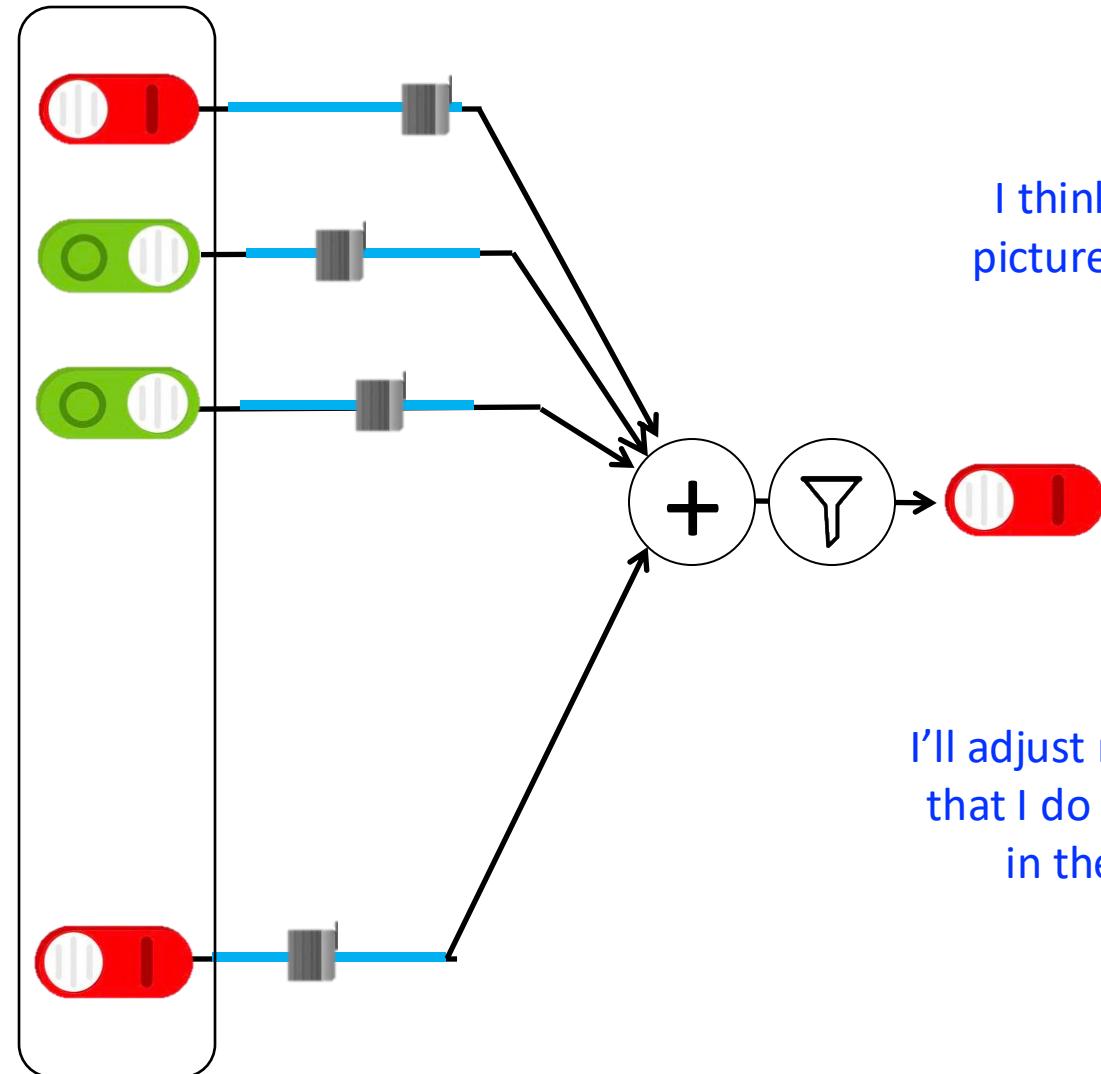
Learn by Example

0000000000000000
1111111111111111
2222222222222222
3333333333333333
4444444444444444
5555555555555555
6666666666666666
7777777777777777
8888888888888888
9999999999999999







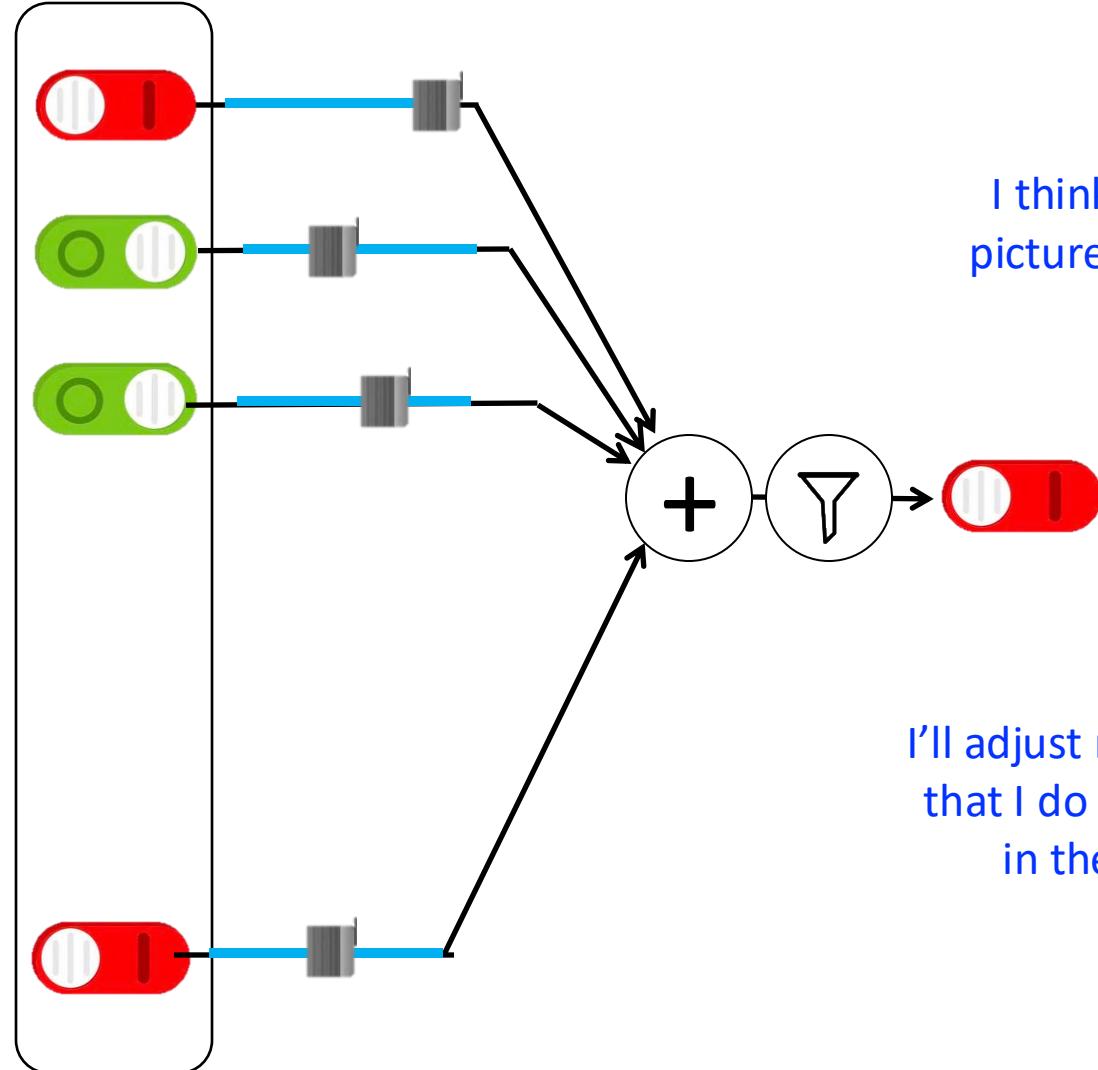


I think that is a picture of a **One!**

What do you mean it's actually a **Zero?**

I'll adjust my sliders so that I do a better job in the future



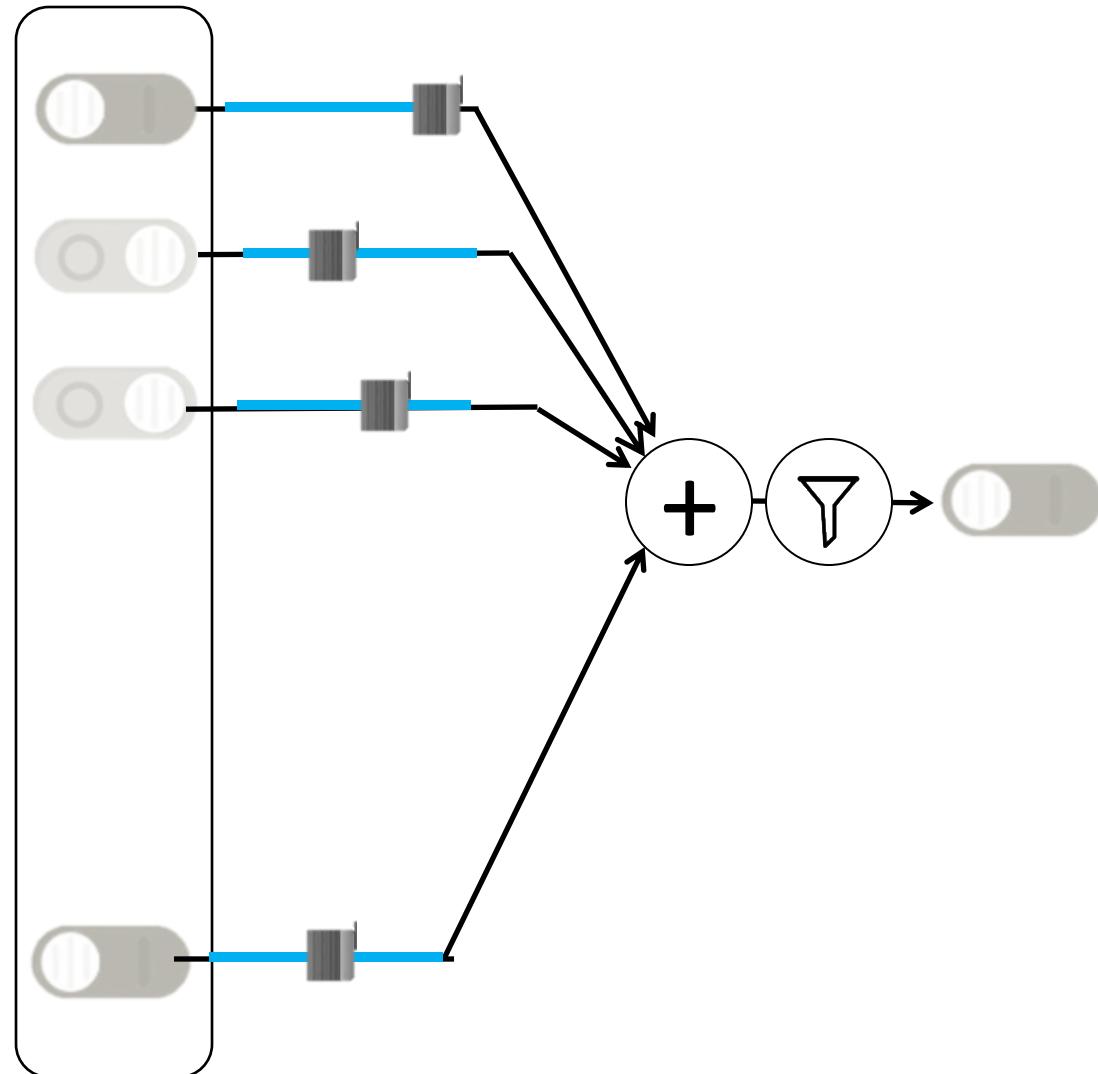


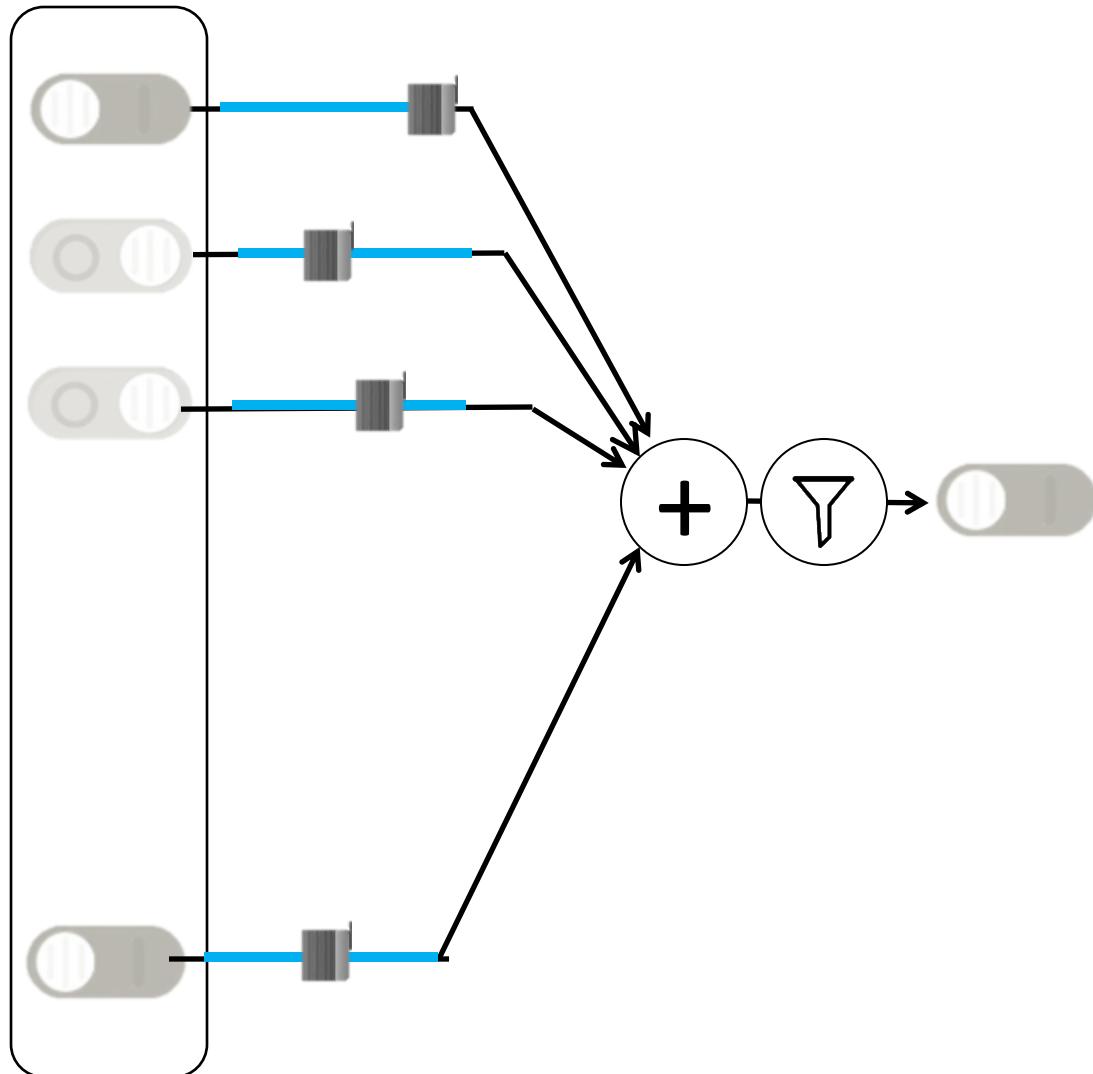
I think that is a
picture of a **One!**

What do you
mean it's actually
a **Zero?**

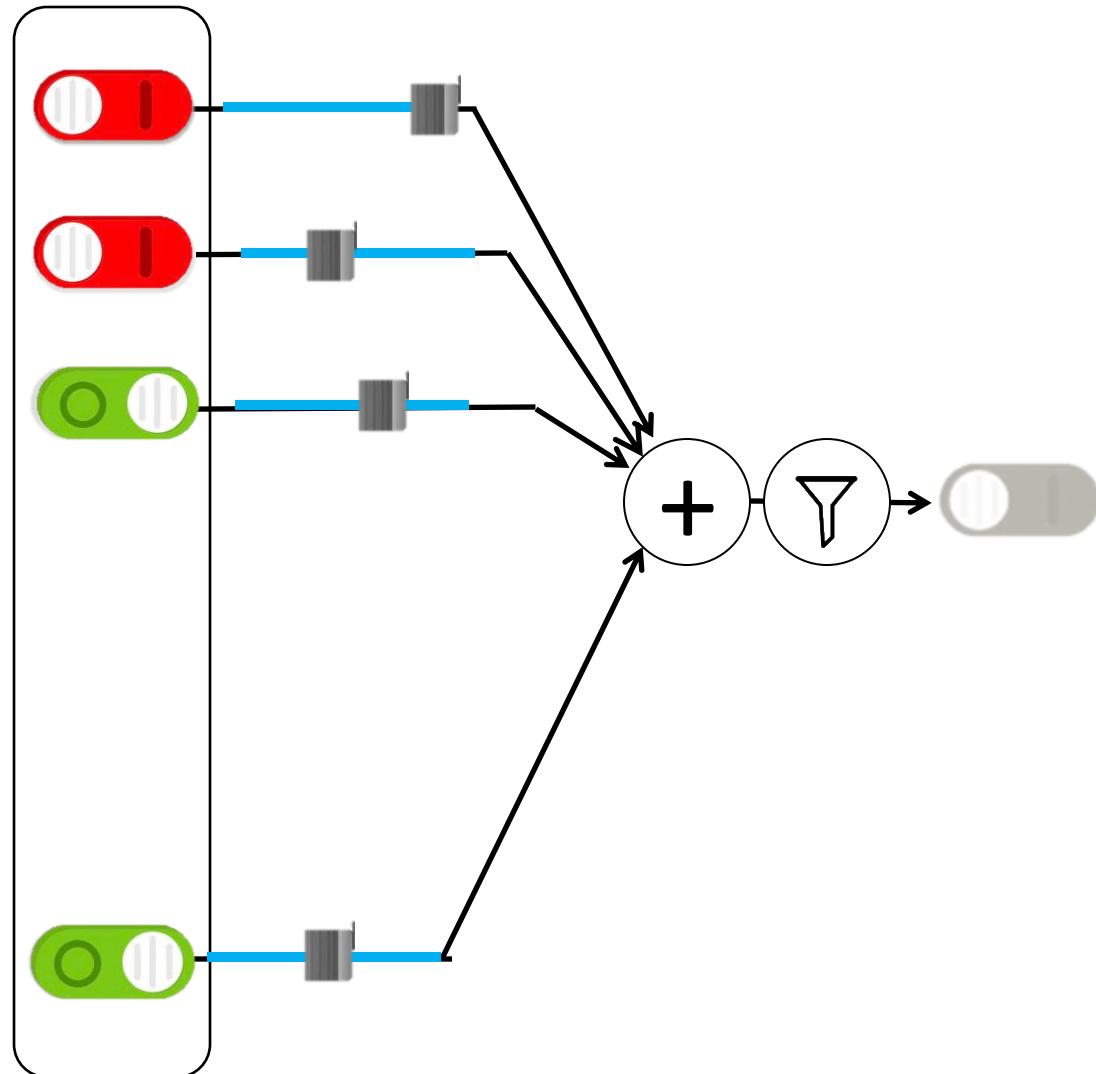
I'll adjust my sliders so
that I do a better job
in the future

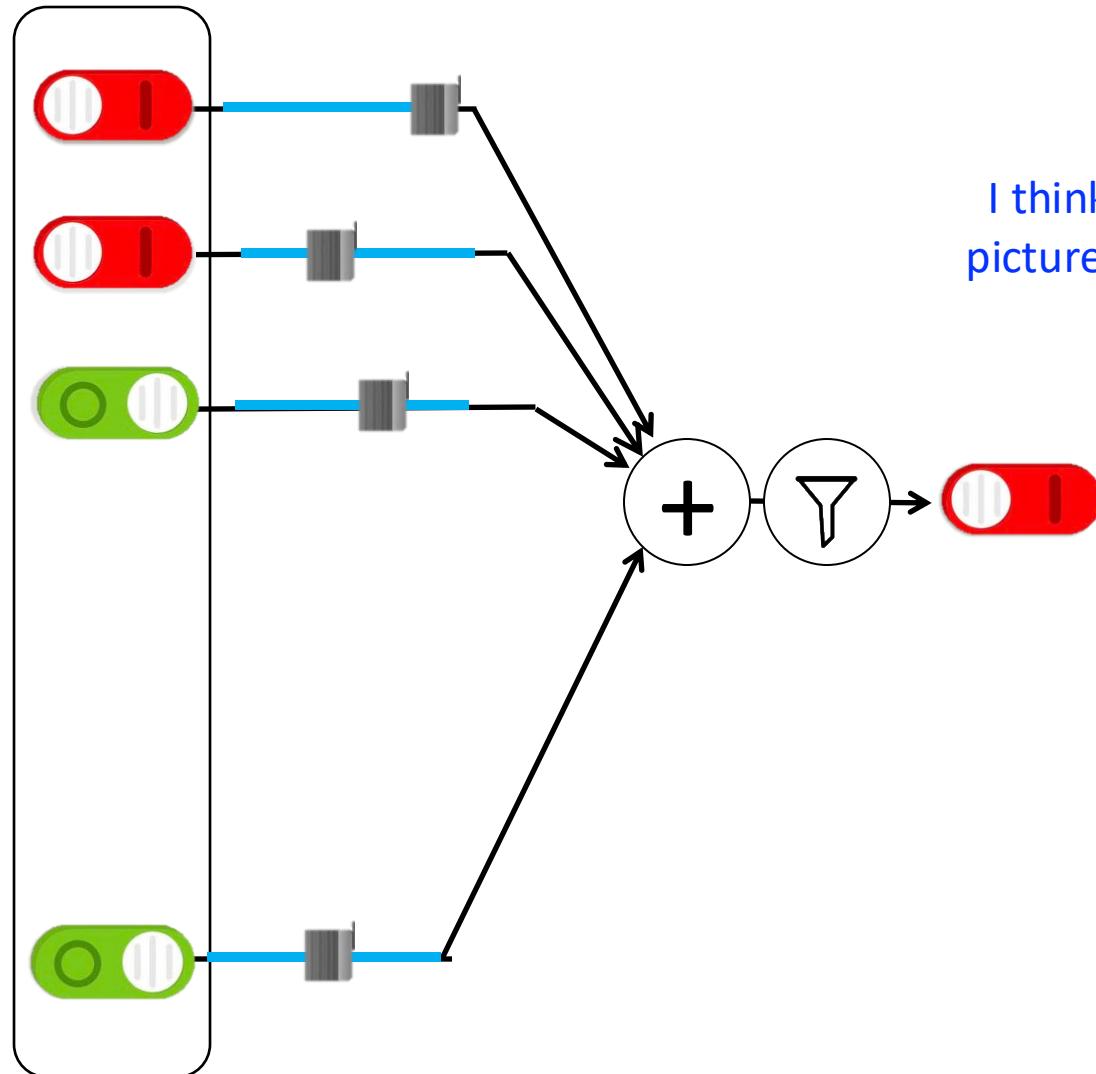






1

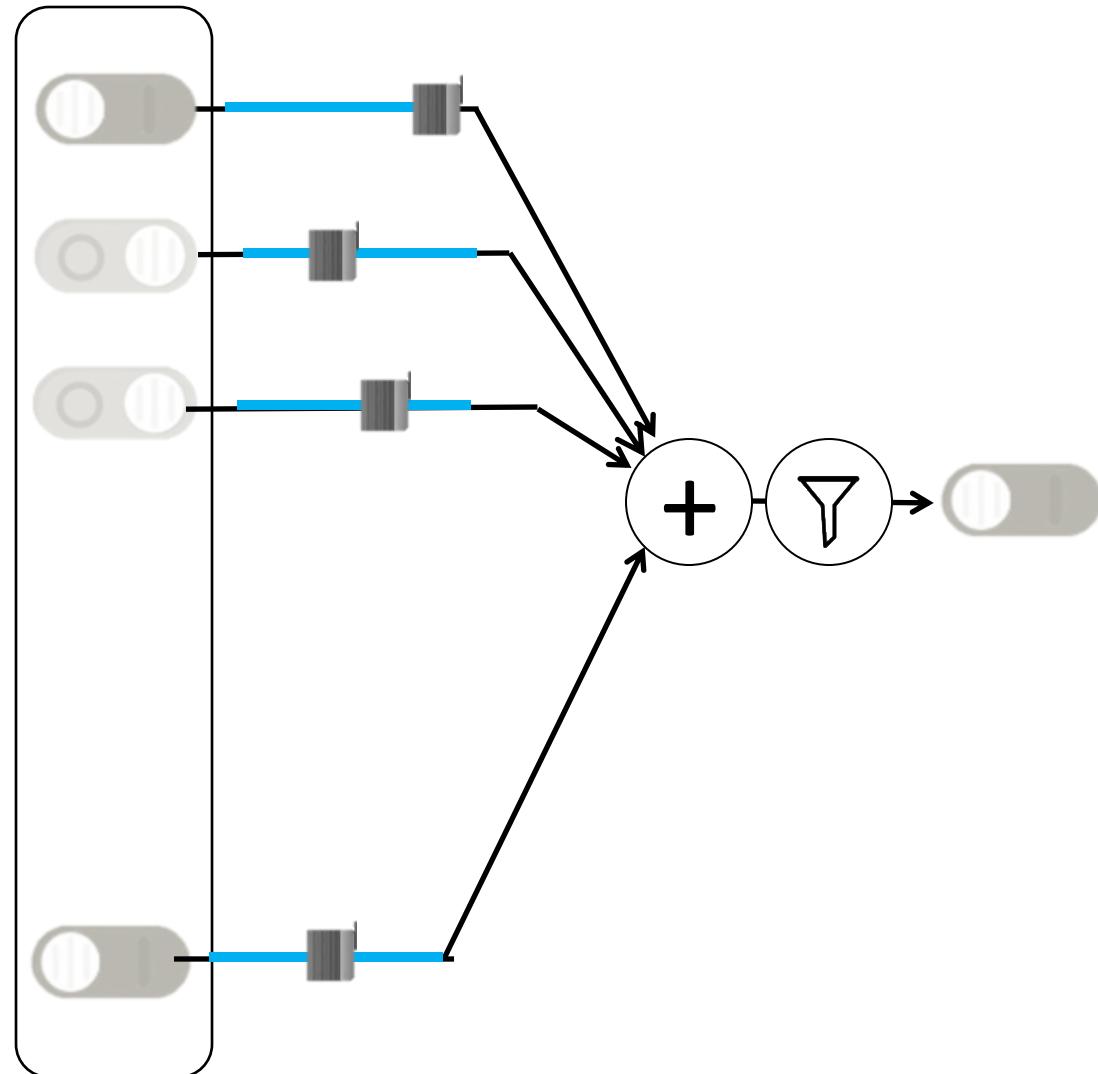




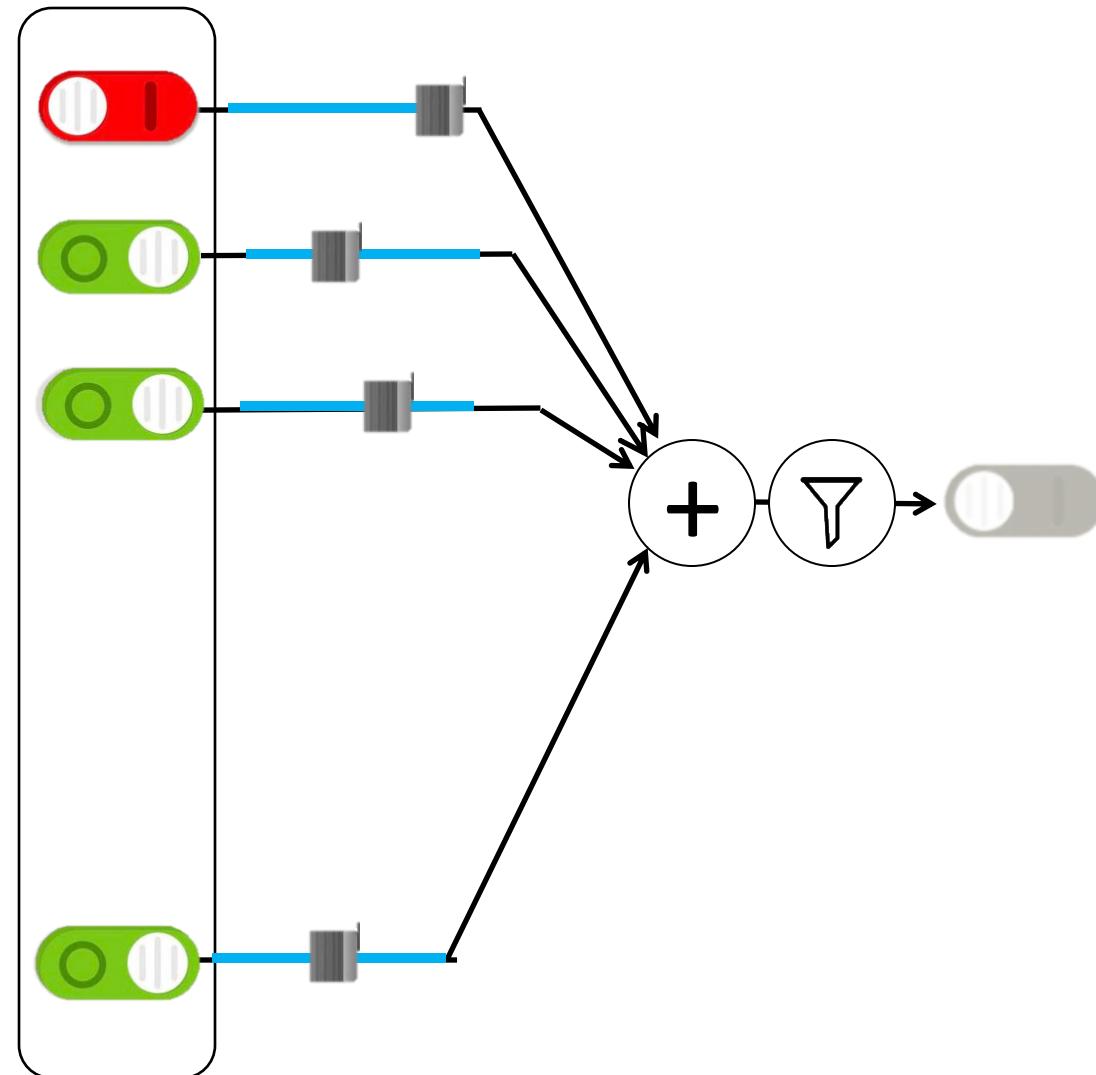
I think that is a
picture of a **One!**

Wahoo I got it
right!

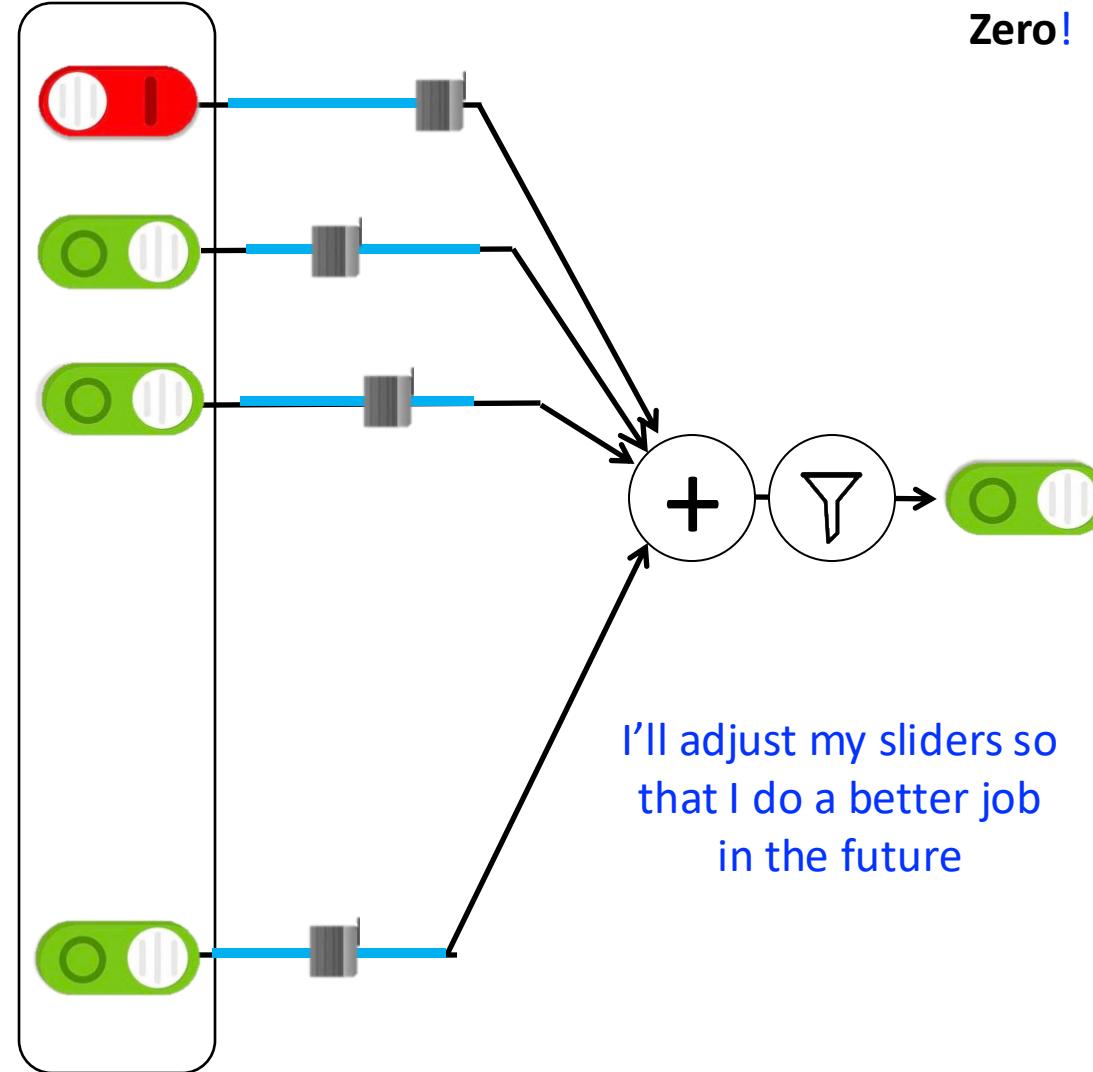




1



1



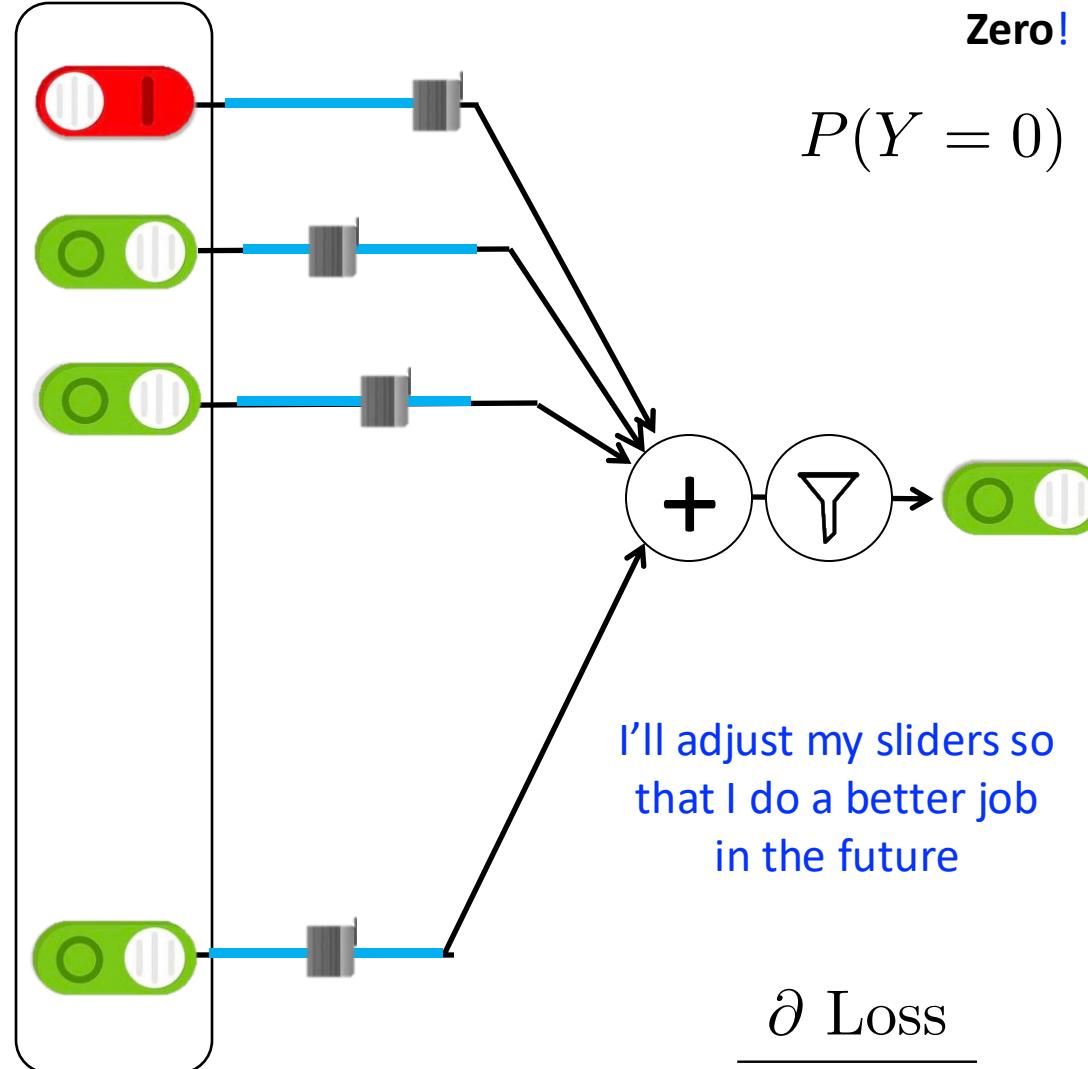
I think that is a picture of a
Zero!

But it is
actually a Zero

I'll adjust my sliders so
that I do a better job
in the future



1



I think that is a picture of a
Zero!

$$P(Y = 0) = 0.9$$

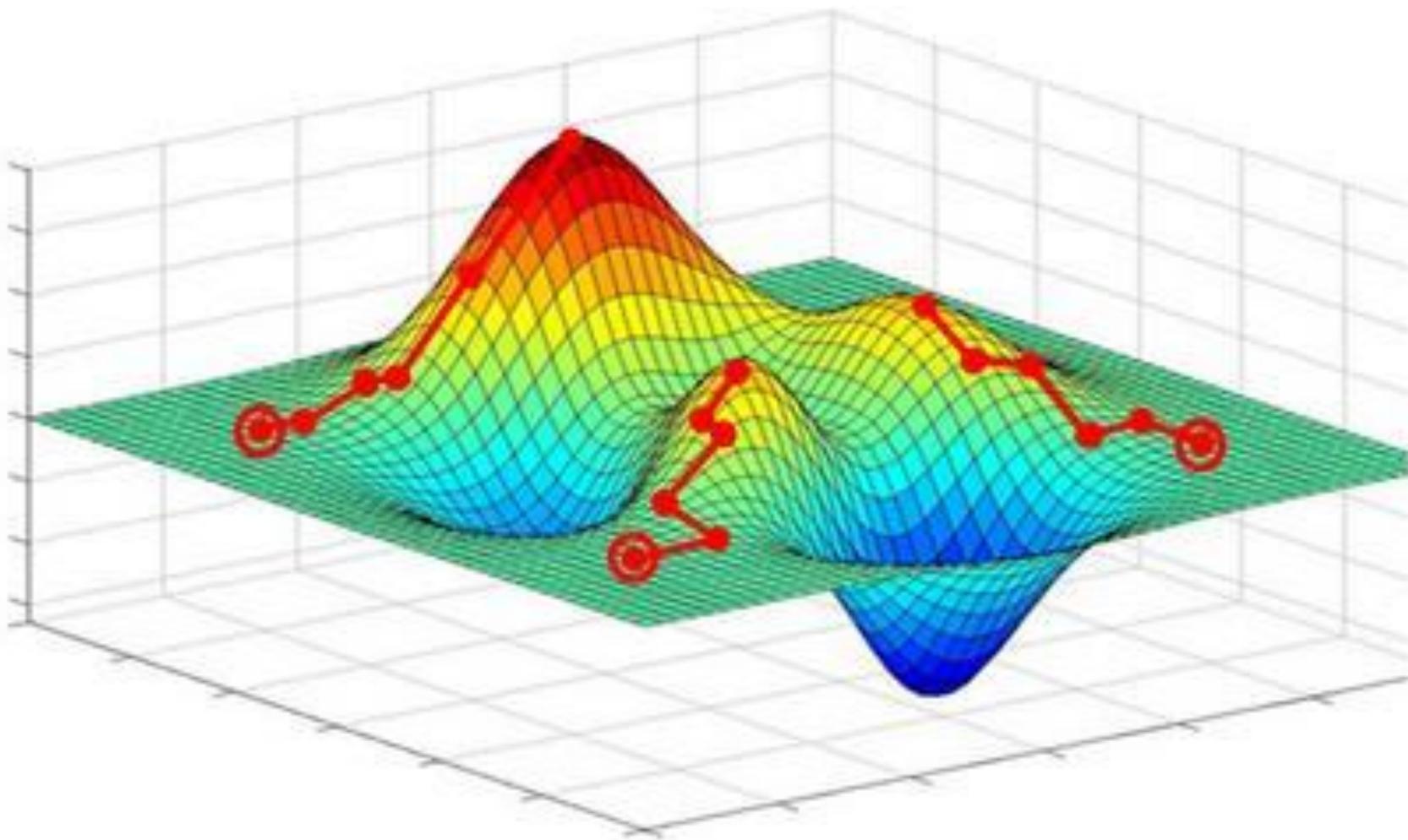
But it is
actually a Zero
Loss = 1

I'll adjust my sliders so
that I do a better job
in the future

$$\frac{\partial \text{Loss}}{\partial \text{Slider}_i}$$



Gradient Descent



Walk uphill and you will find a local maxima
(if your step size is small enough)
Piech, CS109, Stanford University



Gradient of Probability

$$\frac{\partial L}{\partial \theta_i^{(\hat{y})}} = \frac{\partial L}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial \theta_i^{(\hat{y})}}$$

$$\hat{y} = \sigma \left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})} \right)$$

$$\frac{\partial \hat{y}}{\partial \theta_i^{(\hat{y})}} = \sigma \left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})} \right) \left[1 - \sigma \left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})} \right) \right] \cdot \frac{\partial}{\partial \theta_i^{(\hat{y})}} \sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}$$

$$= \hat{y}[1 - \hat{y}] \cdot \frac{\partial}{\partial \theta_i^{(\hat{y})}} \sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})}$$

$$= \hat{y}[1 - \hat{y}] \cdot h_i$$

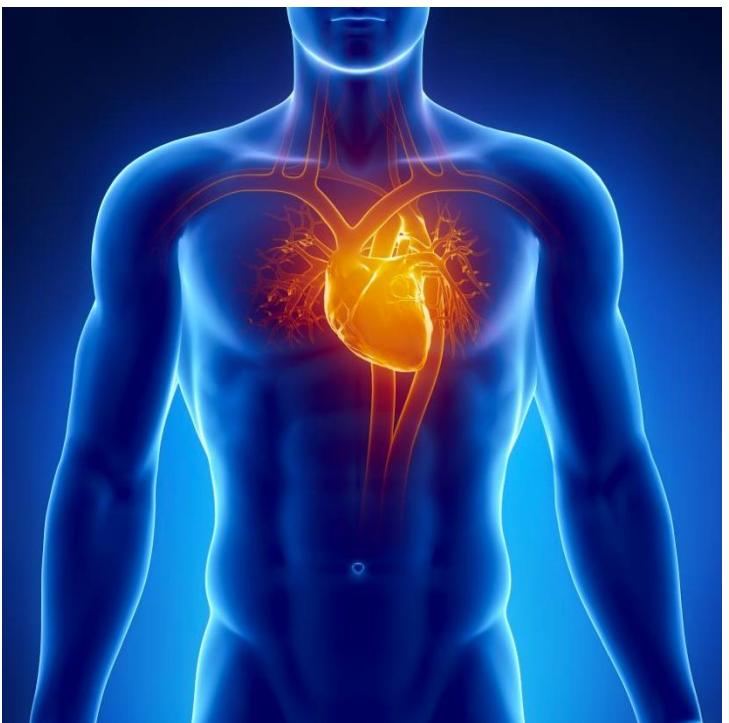
You will be able to do this.



Where you will be by the end of class

CS109: Theory Class focused on Applications

Heart



Ancestry



Netflix



Where is this Useful?



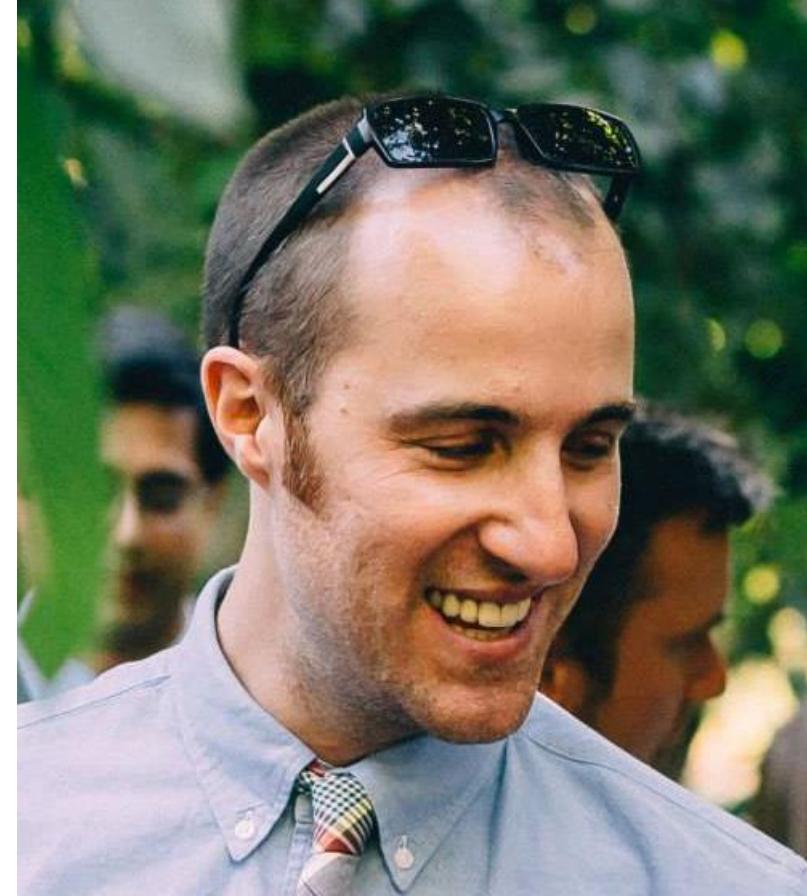
A machine learning algorithm performs **better than** the best dermatologists.

Developed recently, at Stanford.

Esteva, Andre, et al. "Dermatologist-level classification of skin cancer with deep neural networks." *Nature* 542.7639 (2017): 115-118.

What about Generative AI?

Who Invented Generative AI for Images?



Deep Unsupervised Learning using Nonequilibrium Thermodynamics by Jascha Sohl Dickstein

Dalle2. Prompt “a large lecture class at stanford learning probability for computer scientists in the style of vangough”



Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein
Stanford University

JASCHA@STANFORD.EDU

Eric A. Weiss
University of California, Berkeley

EWEISS@BERKELEY.EDU

Niru Maheswaranathan
Stanford University

NIRUM@STANFORD.EDU

Surya Ganguli
Stanford University

SGANGULI@STANFORD.EDU

Abstract

A central problem in machine learning involves modeling complex data-sets using highly flexible families of probability distributions in which learning, sampling, inference, and evaluation are still analytically or computationally tractable. Here, we develop an approach that simultaneously achieves both flexibility and tractability. The essential idea, inspired by non-equilibrium statistical physics, is to systematically and slowly destroy structure in a data distribution through an iterative forward diffusion process. We then learn a reverse diffusion process that restores structure in data, yielding a highly flexible and tractable generative model of the data. This approach allows us to rapidly learn, sample from, and evaluate probabilities in deep generative models with thousands of layers or time steps, as well as to compute conditional and posterior probabilities under the learned model. We additionally release an open source reference implementation of the algorithm.

1. Introduction

Historically, probabilistic models suffer from a tradeoff between two conflicting objectives: *tractability* and *flexibility*. Models that are *tractable* can be analytically evaluated and easily fit to data (e.g. a Gaussian or Laplace). However,

these models are unable to aptly describe structure in rich datasets. On the other hand, models that are *flexible* can be molded to fit structure in arbitrary data. For example, we can define models in terms of any (non-negative) function $\phi(\mathbf{x})$ yielding the flexible distribution $p(\mathbf{x}) = \frac{\phi(\mathbf{x})}{Z}$, where Z is a normalization constant. However, computing this normalization constant is generally intractable. Evaluating, training, or drawing samples from such flexible models typically requires a very expensive Monte Carlo process.

A variety of analytic approximations exist which ameliorate, but do not remove, this tradeoff—for instance mean field theory and its expansions (T, 1982; Tanaka, 1998), variational Bayes (Jordan et al., 1999), contrastive divergence (Welling & Hinton, 2002; Hinton, 2002), minimum probability flow (Sohl-Dickstein et al., 2011b;a), minimum KL contraction (Lyu, 2011), proper scoring rules (Gneiting & Raftery, 2007; Parry et al., 2012), score matching (Hyvärinen, 2005), pseudolikelihood (Besag, 1975), loopy belief propagation (Murphy et al., 1999), and many, many more. Non-parametric methods (Gershman & Blei, 2012) can also be very effective¹.

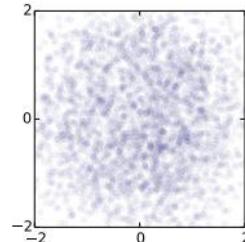
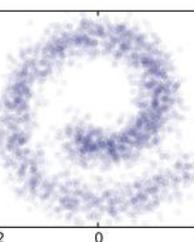
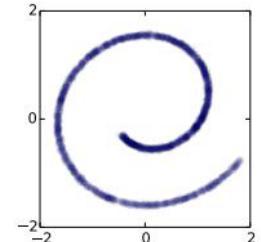
1.1. Diffusion probabilistic models

We present a novel way to define probabilistic models that allows:

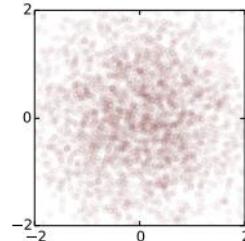
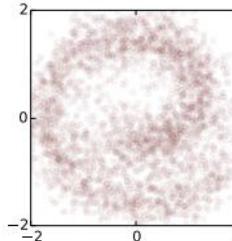
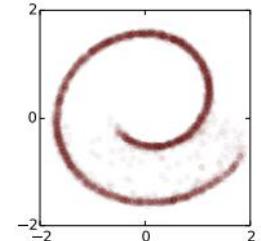
1. extreme flexibility in model structure,
2. exact sampling,

¹Non-parametric methods can be seen as transitioning smoothly between tractable and flexible models. For instance, a non-parametric Gaussian mixture model will represent a small number of components as a Gaussian, and a large number of components as a uniform distribution.

$$q(\mathbf{x}^{(0 \dots T)})$$



$$p(\mathbf{x}^{(0 \dots T)})$$



Will teach this at the end
of the quarter

AI has constantly been revolutionized by people
who understood probability theory.

End of Story

Except it isn't the end of the story...

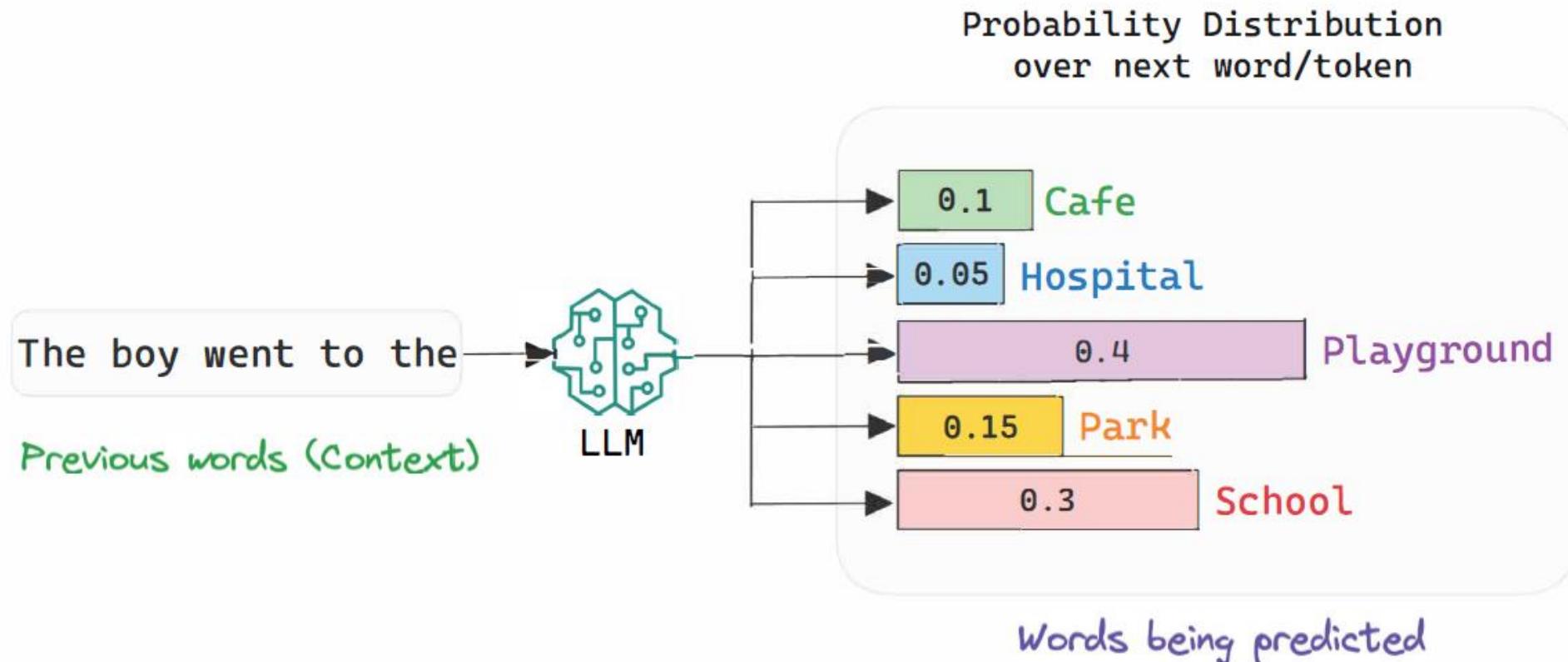
Probability is more than just machine learning

Abundance of important problems



Many Open Problems

Why is beam search (a **probabilistic algorithm** that generates text) for LLMs (think **GPT4**) more effective with a small beam size?



Many Open Problems

Why is beam search (a **probabilistic algorithm** that generates text) for LLMs (think **GPT4**) more effective with a small beam size?

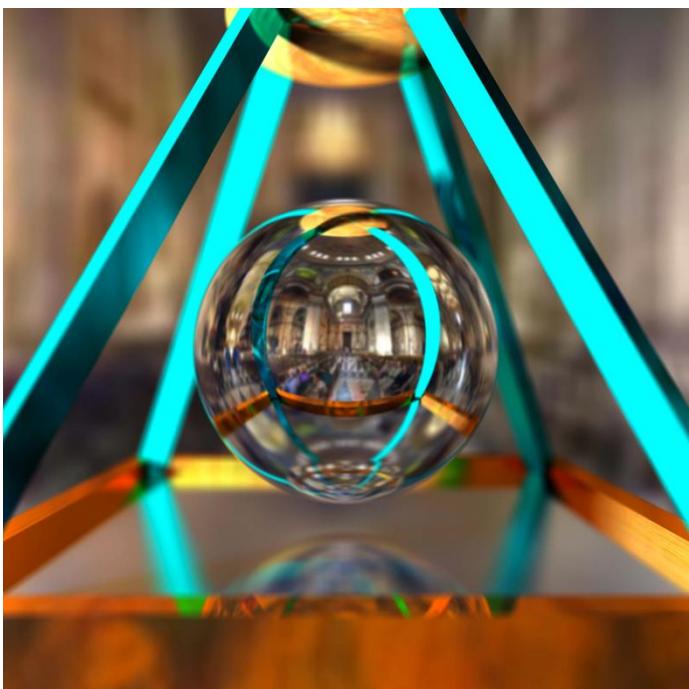
0.3 0.2
The boy went to the **school playground yesterday**
 0.8

- | | |
|-------------------|--|
| High beam size: | Most likely sentence given the context |
| Small beam size: | Greedy next word selection |
| Medium beam size: | ??? |

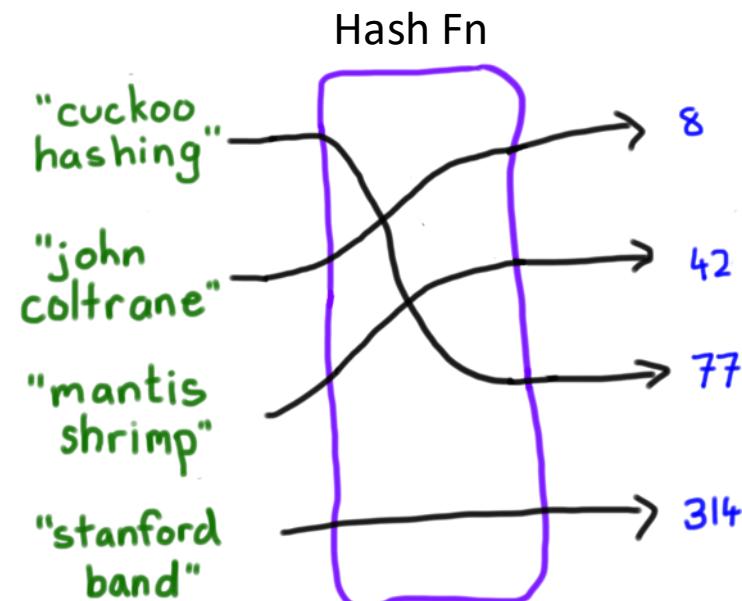


Algorithms and Probability

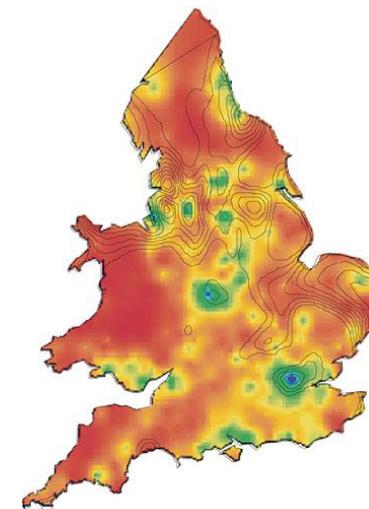
Eg Raytracing



Eg HashMaps



Understanding the world and building tools



Recommender Systems

amazon.com

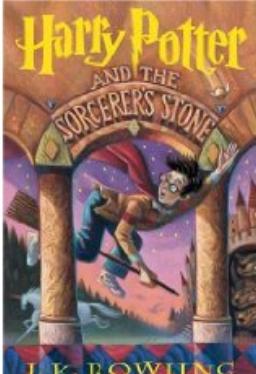
Hello. Sign in to get personalized recommendations. New customer? Start here.

FREE 2-Day Shipping, No Minimum Purchase

Your Account | Help

Shop All Departments Search Books Cart Your Lists

Books Advanced Search Browse Subjects Hot New Releases Bestsellers The New York Times® Best Sellers Libros En Español Bargain Books Textbooks


Harry Potter and the Sorcerer's Stone (Book 1) (Hardcover)
by J.K. Rowling (Author), Mary GrandPré (Illustrator)
 (5,471 customer reviews)

List Price: \$24.99
Price: **\$15.92** & eligible for **FREE Super Saver Shipping** on orders over \$25.
[Details](#)
You Save: \$9.07 (36%)

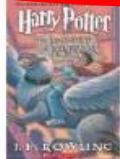
In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.

Quantity: 1
or
[Sign in to turn on 1-Click ordering.](#)
or

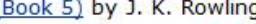
Amazon Prime Free Trial required. Sign up when you check out. [Learn More](#)

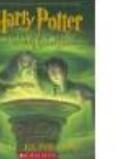
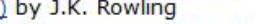
Customers Who Bought This Item Also Bought

Page 1 of 20


[Harry Potter and the Prisoner of Azkaban \(Book 3\) by J.K. Rowling](#)
 \$16.49


[Harry Potter and the Goblet of Fire \(Book 4\) by J.K. Rowling](#)
 \$19.79


[Harry Potter and the Order of the Phoenix \(Book 5\) by J. K. Rowling](#)
 \$10.18

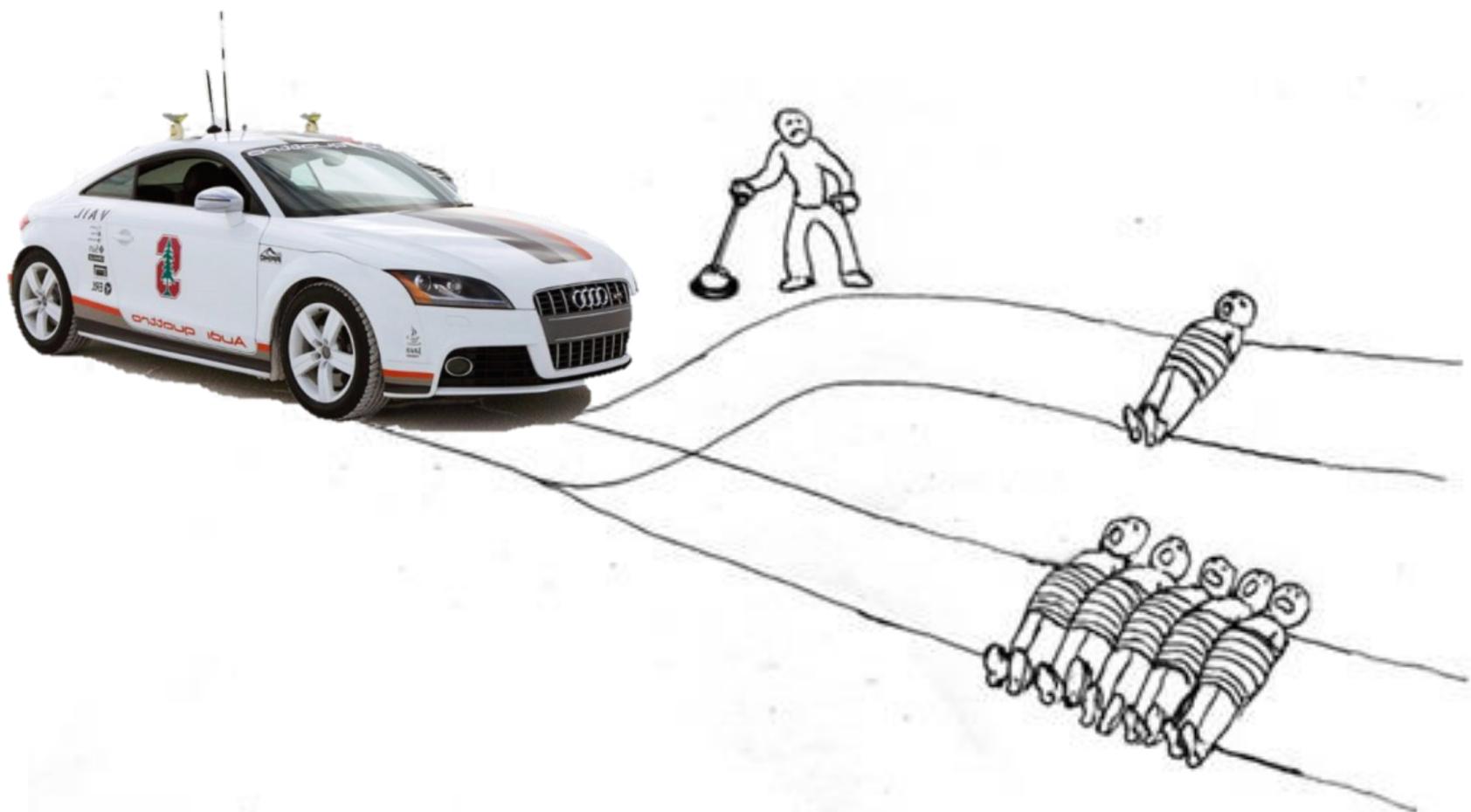

[Harry Potter and the Half-Blood Prince \(Book 6\) by J.K. Rowling](#)
 \$10.18


[The Tales of Beedle the Bard, Collector's Ed... by J. K. Rowling](#)
 (176)





Philosophy and Ethics



Most Desired Skill in Industry

Forbes Billionaires Innovation Leadership Money Consumer

30,575 views | Jan 29, 2018, 02:47pm

Data Scientist Is the Best Job In America According Glassdoor's 2018 Rankings

TWEET THIS

 Data Scientist has been named the best job in America for three years running, with a median base salary of \$110,000 and 4,524 job openings.

 DevOps Engineer is the second-best job in 2018, paying a median base salary of \$105,000 and 3,369 job openings.

f   in 



Job Score is based on:

- Earning potential
- Number of jobs
- Job satisfaction rating

“Data science and machine learning are generating more jobs than candidates right now, making these two areas the *fastest growing employment areas*.”

9.8 times more jobs than five years ago.

[LinkedIn's 2017 U.S. Emerging Jobs Report](#)



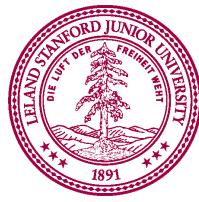
Most Desired Skill in Academia

Most CS PhD students list their highest desiderata upon graduation as:

“Better understanding of probability”



Learn Real Skills in CS109



Spring 2017



Patient sees a series of letters of different font size, and for each, answers correct or incorrect

You decide that the vision tests given by eye doctors could have more precise results if we used an approach inspired by logistic regression. In a vision test a user looks at a letter with a particular font size and either correctly guesses the letter or incorrectly guesses the letter.

You assume that the probability that a particular patient is able to guess a letter correctly is:

$$p = \sigma(\theta - f)$$

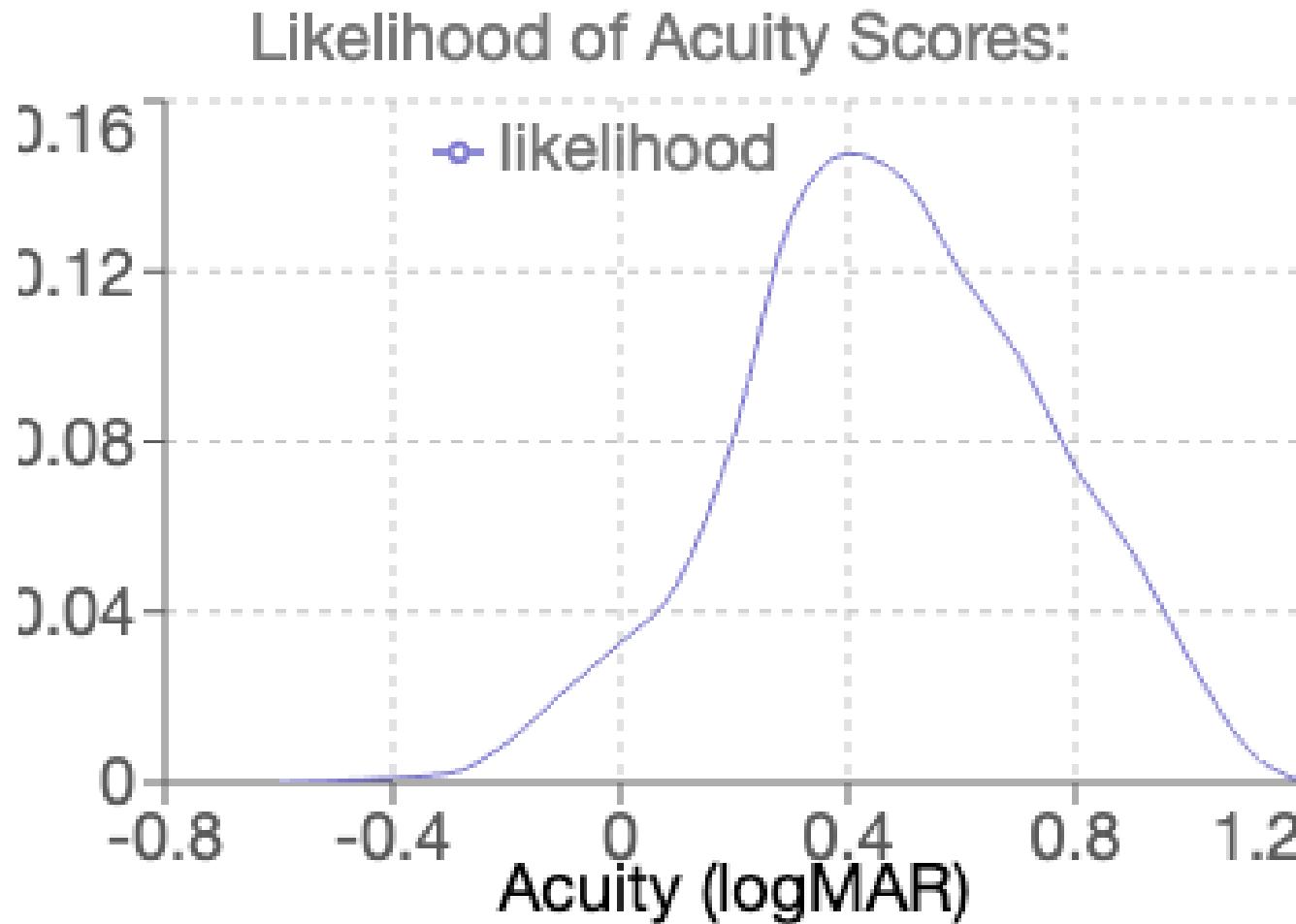
Where θ is the user's vision score and f is the font size of the letter.

Explain how you could estimate a user's vision score (θ) based on their 20 responses $(f^{(1)}, y^{(1)}) \dots (f^{(20)}, y^{(20)})$, where $y^{(i)}$ is an indicator variable for whether the user correctly identified the i th letter and $f^{(i)}$ is the font size of the i th letter. Solve for any and all partial derivatives required by your answer.

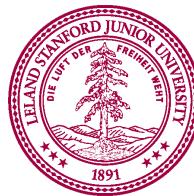
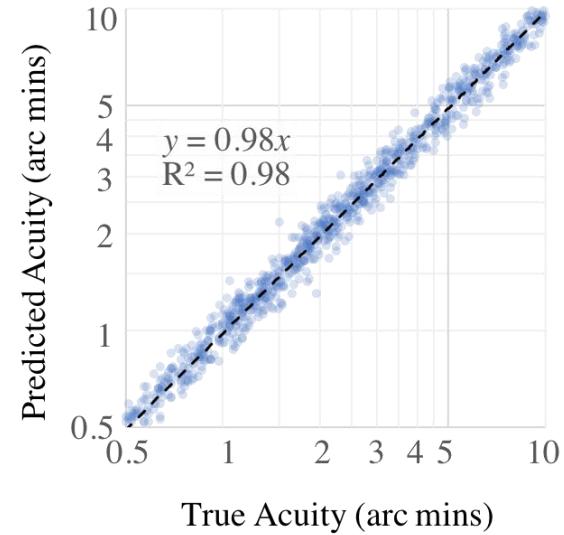
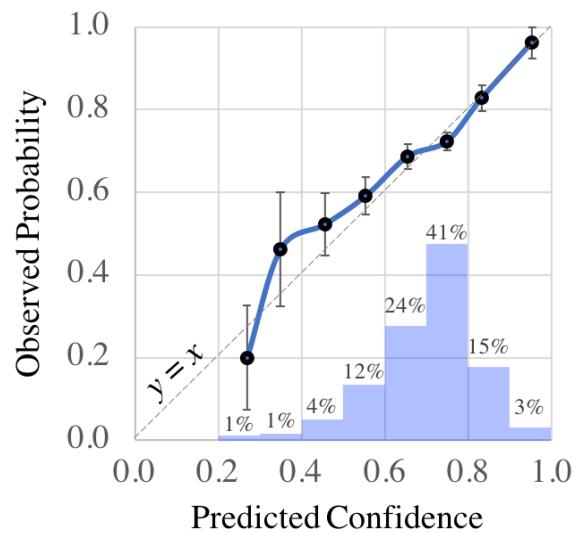
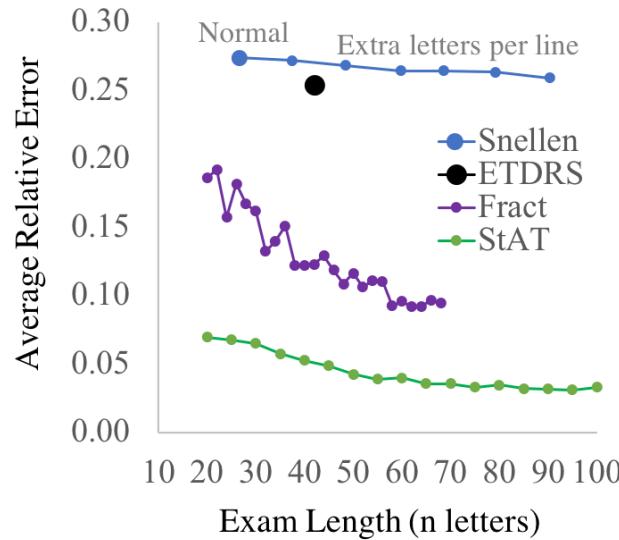


Learn Real Skills in CS109

A patient has answered 20 “letter sizes” and got a few correct. What is your belief in how well they can see?



Now state of the art for eye exam theory



Learn Real Skills in CS109

The Stanford Acuity Test: A Precise Vision Test Using Bayesian Techniques and a Discovery in Human Visual Response

Chris Piech,^{*1} Ali Malik,^{*1} Laura M Scott, Robert T Chang,² Charles Lin²

¹Department of Computer Science, Stanford University

² Department of Ophthalmology, Stanford University

{piech, malikali}@cs.stanford.edu, {rchang3, lincc}@stanford.edu

Abstract

Chart-based visual acuity measurements are used by billions of people to diagnose and guide treatment of vision impairment. However, the ubiquitous eye exam has no mechanism for reasoning about uncertainty and as such, suffers from a well-documented reproducibility problem. In this paper we make two core contributions. First, we uncover a new parametric probabilistic model of visual acuity response based on detailed measurements of patients with eye disease. Then, we present an adaptive, digital eye exam using modern artificial intelligence techniques which substantially reduces acuity exam error over existing approaches, while also introducing the novel ability to model its own uncertainty and incorporate prior beliefs. Using standard evaluation metrics, we estimate a 74% reduction in prediction error compared to the ubiquitous chart-based eye exam and up to 67% reduction compared to the previous best digital exam. For patients with eye disease, the novel ability to finely measure acuity from home could be a crucial part in early diagnosis. We provide a web implementation of our algorithm for anyone in the world to use. The insights in this paper also provide interesting implications for the field of psychometric Item Response Theory.

1 Introduction

Reliably measuring a person's visual ability is an essential component in the detection and treatment of eye diseases around the world. However, quantifying how well an individual can distinguish visual information is a surprisingly difficult task—without invasive techniques, physicians rely on chart-based eye exams where patients are asked visual questions and their responses observed.

Historically, vision has been evaluated by measuring a patient's *visual acuity*: a measure of the font size at which a patient can correctly identify letters shown a fixed distance away. Snellen, this statistic by asking the patient to identify the size of letters correct. This

treatment of patients; yet, it suffers from some notable shortcomings. Acuity exams such as these exhibit high variance in their results due to the large role that chance plays in the final diagnosis, and the approximation error incurred by the need to discretize letter sizes on a chart. On the other hand, digital exams can show letter of any size and can *adaptively* make decisions based on intelligent probabilistic models. As such they have potential to address the shortcomings of analog charts.

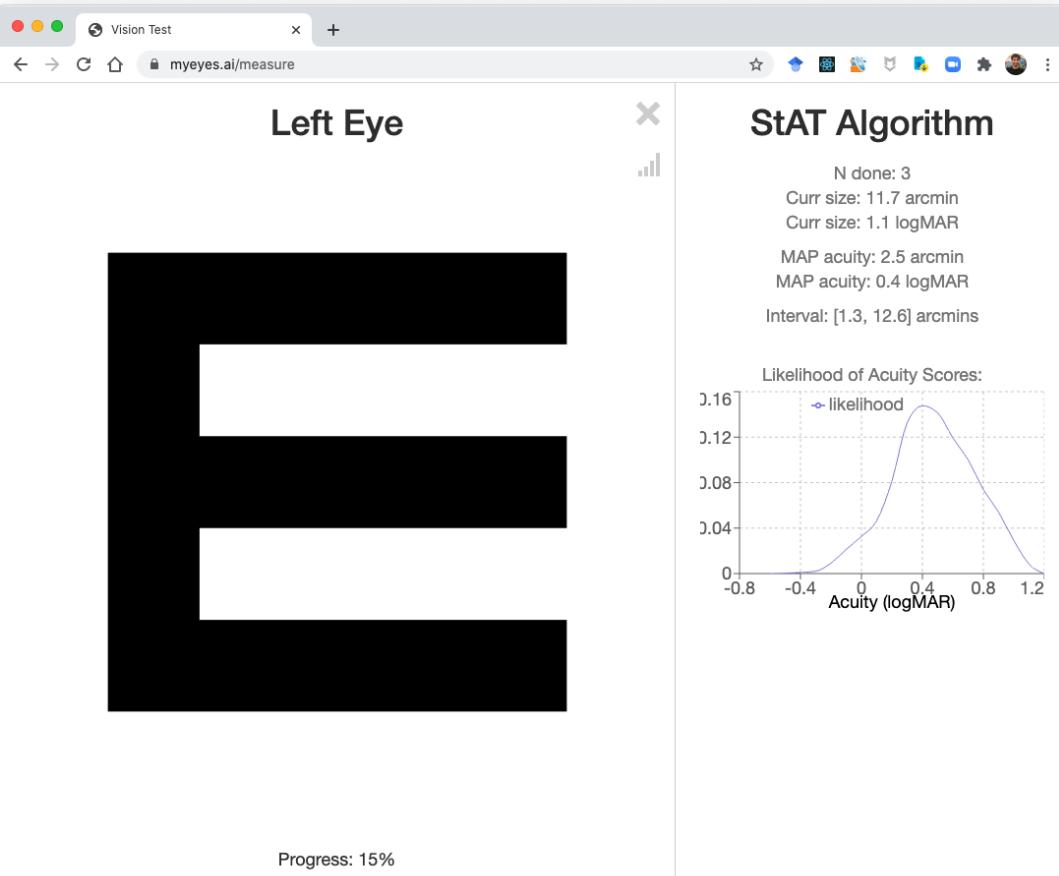
While promising, contemporary digital exams have yet to dramatically improve accuracy over traditional chart-based approaches. The current best digital exam uses a psychometric Item Response Theory (IRT) algorithm for both selecting the next letter size to query and for making a final prediction of acuity. Under simulation analysis, this digital exam results in a 19% reduction in error over traditional chart-based approaches. The separate fields of reinforcement learning and psychometric IRT have independently explored how to effectively make decisions under uncertainty. By merging the good ideas from both disciplines we can develop a much better visual acuity test.

In this paper we make two main contributions. First, we revisit the human Visual Response Function—a function relating the size of a letter to the probability of a person identifying it correctly—and discover that it follows an interpretable parametric form that fits real patient data. Second, we present an algorithm to measure a person's acuity which uses several Bayesian techniques common in modern artificial intelligence. The algorithm, called the Stanford Acuity Test (StACT)¹, has the following novel features:

1. Uses the new parametric form of the human Visual Response Function.
2. Returns a soft inference prediction of the patient's acuity, confidence in the final

ing algorithm to adapt to a user. This reflects acuity belief.

StACT was named after us. We continue in this



Science

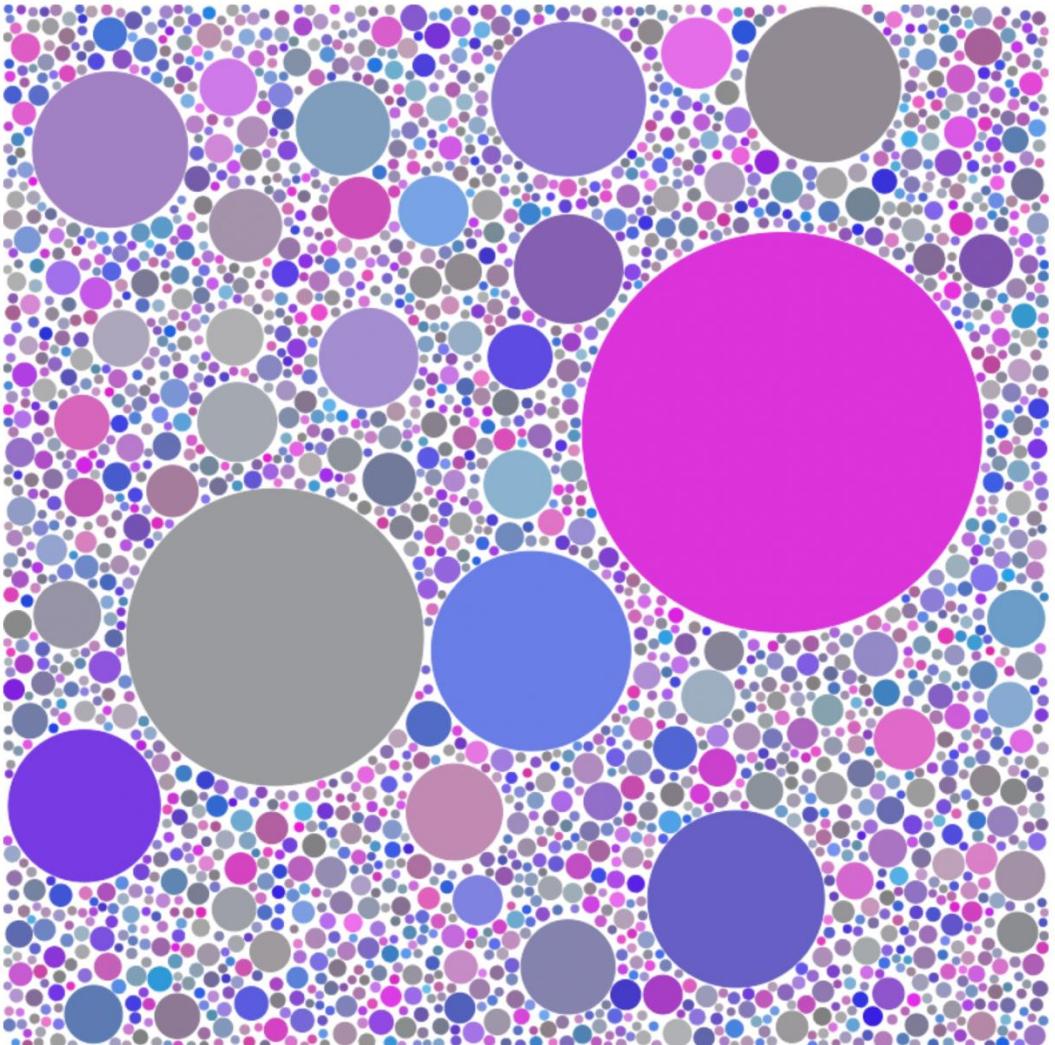
*Equal contrit
Copyright © 2021
Intelligence (www

THE LANCET

Piech, CS109, Stanford University



What is on a typical final?



Regenerate

RIECH, CSUS, STANFORD University

1. Algorithmic Art
2. Lucky Events
3. Supply Chain Decision Making
4. P-Hacking
5. Chess.com Puzzle Ability
6. ML Calibration

https://chrispiech.github.io/probabilityForComputerScientists/en/examples/algorithmic_art/



Foundation for your future

But its not always intuitive

But Its not Always Intuitive



A patient has a
positive Zika test.

What is the probability they have zika?

-
- *0.8% of people have zika*
 - *Test has 90% positive rate for people with zika*
 - *Test has 7% positive rate for people without zika*

The right answer is 9%

Probability = Important + Needs Study

Delayed gratification

CS109 View of Probability

Teach you how to write programs
that most people are not able to write.

CS109 View of Probability

Teach you the theory you need to do the math that
most people are not able to do.

CS109

AI

Uncertainty Theory

Single Random
Variables

Probabilistic Models

Counting

Probability Fundamentals

Lets dive in...

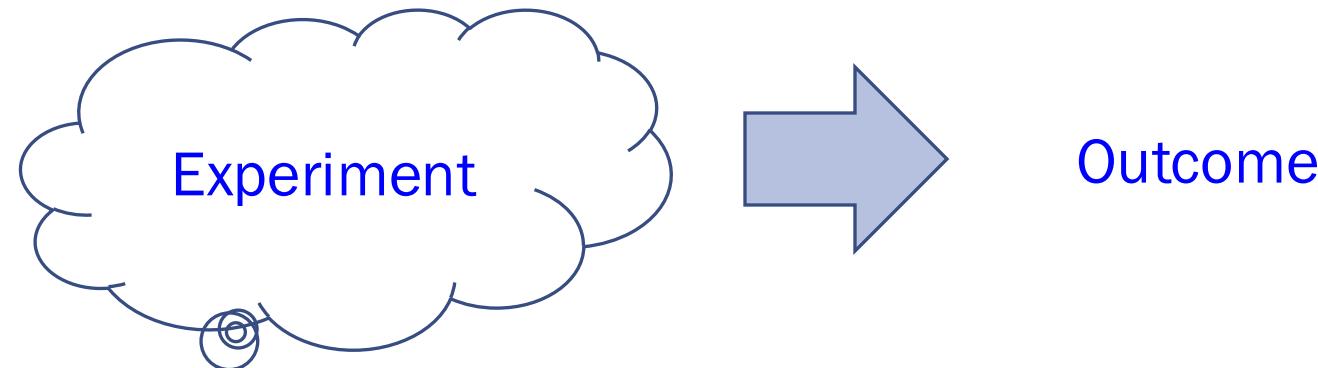
2 min pedagogic pause.

Counting I



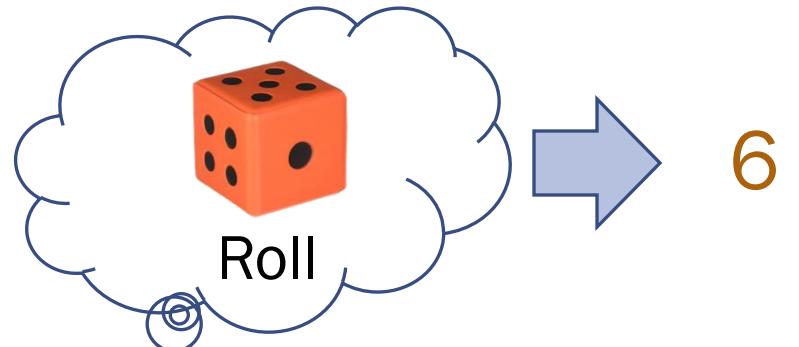
What is Counting?

An experiment
in probability:

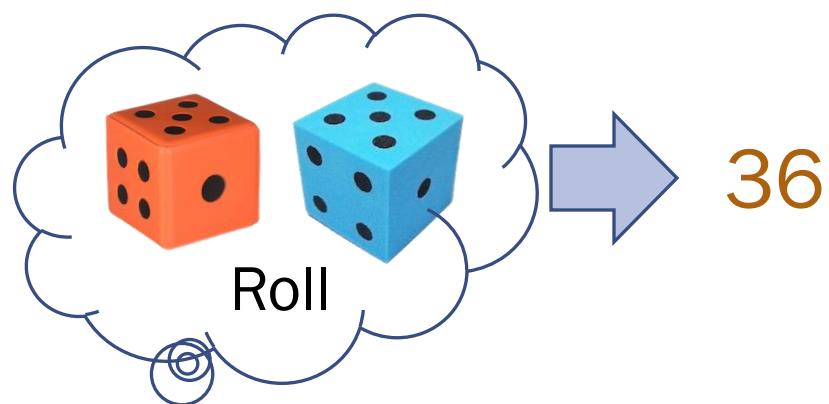
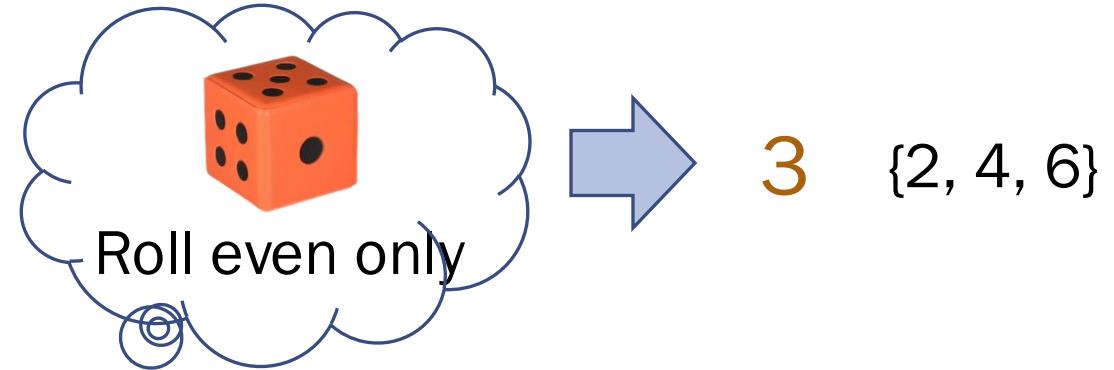


Counting: How many possible **outcomes** satisfy some **event**?

What is Counting?



6
 $\{1, 2, 3, 4, 5, 6\}$



36
 $\{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), (2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6), (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6), (4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6), (5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6), (6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}$

Think about a generative story...

Step Rule of Counting (aka Product Rule of Counting)

If an experiment has two steps, where

The first step's outcomes are from Set A , where $|A| = m$,
and the second step's outcomes are from Set B , where $|B| = n$,
and $|B|$ is unaffected by outcome of first step.

Then the number of outcomes of the experiment is

$$|A||B| = mn.$$

Two-step experiment

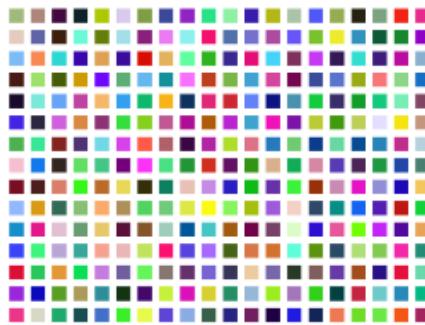


How Many Unique Images?

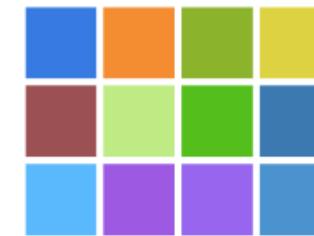
Each pixel can be one of 17 million distinct colors



(a) 12 million pixels



(b) 300 pixels



(c) 12 pixels

$$(17 \text{ million})^n$$

A silhouette of a person standing on a dark, wavy horizon, holding a long stick or staff. They are looking up at a vast, colorful sky filled with stars. The sky transitions from deep blue at the top to vibrant orange and red near the horizon, suggesting a sunset or sunrise. A bright, central light source, possibly the sun, creates a lens flare effect.

10^{80}

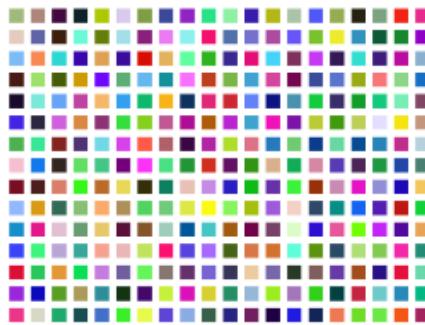
How Many Unique Images?

Each pixel can be one of 17 million distinct colors



(a) 12 million pixels

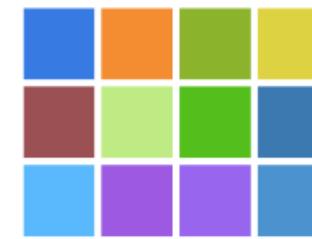
$$\approx 10^{86696638}$$



(b) 300 pixels

$$\approx 10^{2167}$$

$$(17 \text{ million})^n$$



(c) 12 pixels

$$\approx 10^{86}$$

Follow Along Online

The screenshot shows a web browser window with two tabs open. The active tab is titled "Lecture 1 - Welcome to CS109" and displays the "CS109 | Welcome to CS109" page from web.stanford.edu/class/cs109/lectures/1>Welcome/. The page content includes:

- Lecture 1: Welcome to CS109** (Section 1)
- Date: SEPT 27TH, 2023
- Location: HEWLETT 200, 3:30P
- Lecture Materials**:
 - [Slides PDF](#) (represented by a document icon)
 - [Lecture Questions](#) (represented by a blue question mark icon)A red circle highlights the "Lecture Questions" link.
- Learning Goals**:

You should feel welcomed to CS109 and have an idea of why probability is exciting to learn. Then we will jump into content. By the end of class you should know the fundamentals of counting, especially the step rule (aka product rule) and the "or" rule (aka inclusion/exclusion).
- Reading**:

[Syllabus, Counting](#)

Follow Along Online

Sum Rule of Counting

If the outcome of an experiment can be either from

Set A , where $|A| = m$,

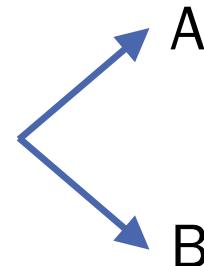
or Set B , where $|B| = n$,

where $A \cap B = \emptyset$,

Then the number of outcomes of the experiment is

$$|A| + |B| = m + n.$$

One experiment



How many toys?

Question: All of Freya's toys are either Balls **OR** Plush Animals. She has 2 Balls and 3 Plush Animals. How many toys does she have?



Answer: $2 + 3$



iversity

How Many Bit Strings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01

010000
010001
010010
010011
010100
010101
010110
010111
011000
011001
011010
011011
011100
011101
011110
011111

Set *A*

2^4 end with 10

000010
000110
001010
001110
010010
010110
011010
011110
100010
100110
101010
101110
110010
110110
111010
111110

Set *B*

How Many Bit Strings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01

010000
010001
010010
010011
010100
010101
010110
010111

Set *A*

2^4 end with 10

000010
000110
001010
001110
010010
010110
011010
011110

Set *B*

How Many Bit Strings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

2^4 start with 01

010000
010001
010010
010011
010100
010101
010110
010111
011000
011001
011010
011011
011100
011101
011110
011111

Set *A*

2^4 end with 10

000010
000110
001010
001110
010010
010110
011010
011110
100010
100110
101010
101110
110010
110110
111010
111110

Set *B*

How Many Bit Strings?

Problem: A 6-bit string is sent over a network. The valid set of strings recognized by the receiver must either start with "01" or end with "10". How many such strings are there?

Answer

$$\begin{aligned}N &= |A| + |B| - |A \text{ and } B| \\&= 16 + 16 - 4 \\&= 28\end{aligned}$$

2^4 start with 01

010000
010001
010010
010011
010100
010101
010110
010111
011000
011001
011010
011011
011100
011101
011110
011111

Set *A*

2^4 end with 10

000010
000110
001010
001110
010010
010110
011010
011110
100010
100110
101010
101110
110010
110110
111010
111110

Set *B*

Or Rule of Counting (aka Inclusion/ Exclusion)

If the outcome of an experiment can be either from

Set A , where $|A| = m$,

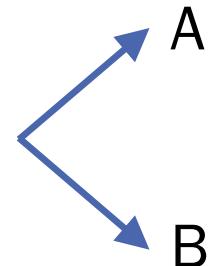
or Set B , where $|B| = n$,

where $A \cap B$ *may not be empty*,

Then the number of outcomes of the experiment is

$$N = |A| + |B| - |A \cap B|.$$

One experiment



Core Counting

Counting with steps

Definition: Step Rule of Counting (aka Product Rule of Counting)

If an experiment has two parts, where the first part can result in one of m outcomes and the second part can result in one of n outcomes regardless of the outcome of the first part, then the total number of outcomes for the experiment is $m \cdot n$.

Counting with “or”

Definition: Inclusion Exclusion Counting

If the outcome of an experiment can either be drawn from set A or set B , and sets A and B may potentially overlap (i.e., it is not the case that A and B are mutually exclusive), then the number of outcomes of the experiment is $|A \text{ or } B| = |A| + |B| - |A \text{ and } B|$.

Challenge Problem

1. Strings

- How many *different* orderings of letters are possible for the string BOBA?

BOBA, ABOB, OBBA...



Incredible time. Incredible
school at which to study
probability!
Exciting.