



CS109: Probability for Computer Scientists

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



But first...
who am I?

Chris Piech

Teaching at Stanford

8,000+ students over 10 years

CS106A

Programming
Methodologies

CURRENT

CS106B

Programming
Abstractions

LAST: FALL 2016

CS109

Probability for Computer
Scientists

LAST: FALL 2018

CS221

Intro to Artificial
Intelligence

LAST: SUM 2013



Created a research lab in:
AI for Social Good (esp Education)



Grew up in Nairobi, Kuala Lumpur before Stanford!



Long History in CS109



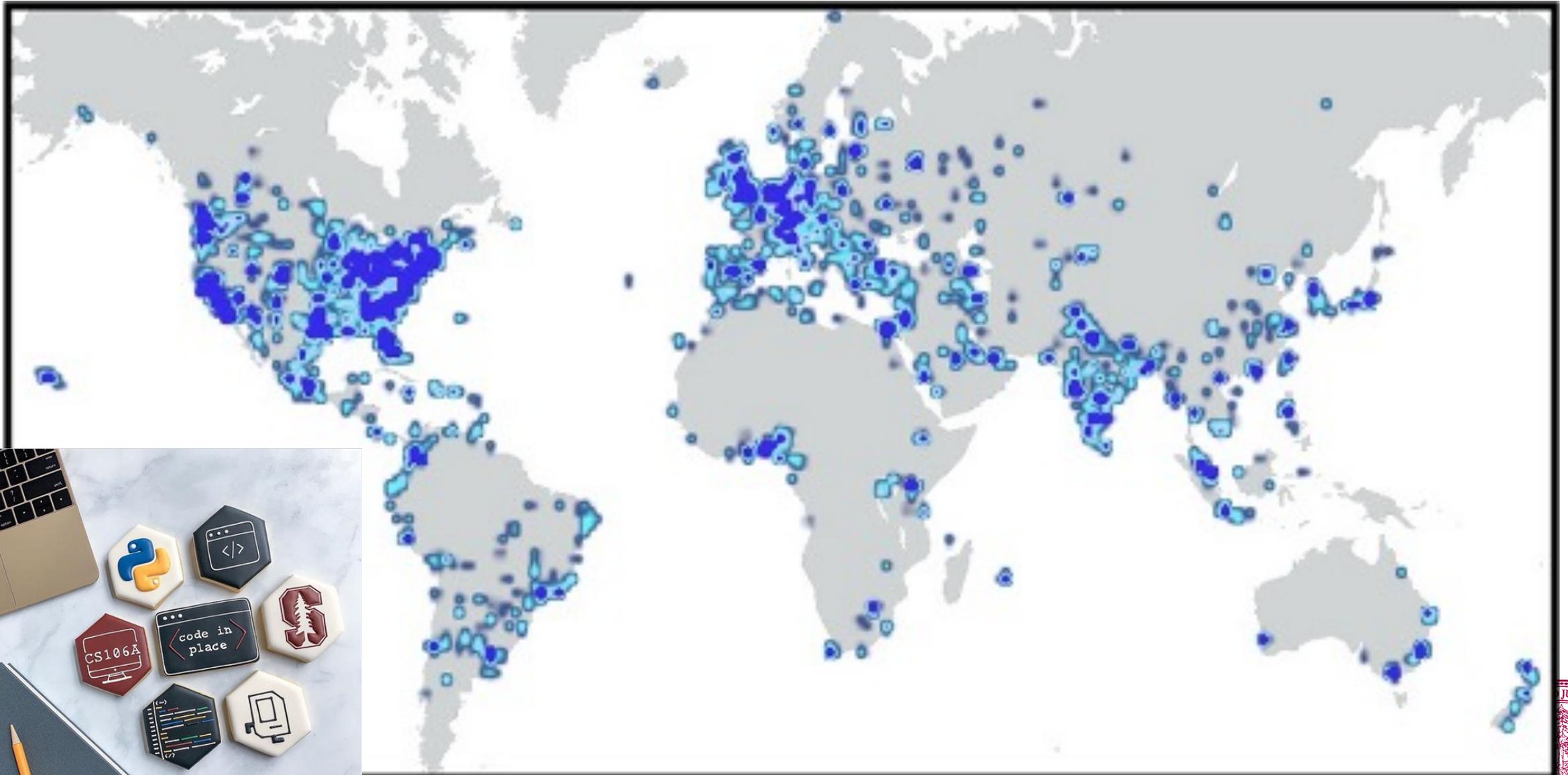
I took the first CS109 back when I looked like this



Been teaching it since 2014



What Sort of Thing Does Chris Research?



Piech, CS109, Stanford University



Deployed to grade 16,000 submissions in Code.org

The screenshot shows a web browser window titled "Code in Place Feedback" with the URL "codeinplace.stanford.edu/diagnostic/feedback". The interface includes navigation tabs for "Overview", "Question 1", "Question 2", "Question 3", "Question 4", "Question 5", and "Wrap-Up". The "Question 1" tab is active. On the left, there are "Back", "Feedback", and "Next" buttons. The "Feedback" section contains the following text:

GETTING INPUT FROM USER
This question requires you to get input from the user, convert it to a number, and save it as a variable. Did you correctly do all of these steps?

Close. There is a minor error with your logic to get input from user. This could be something like forgetting to convert user input to a float

Do you agree with the feedback in the purple box?

Below this text are thumbs-up and thumbs-down icons, and a text input field labeled "Please explain (optional):".

The "Your Solution" section on the right contains the following Python code:

```
def main():  
    # TODO write your solution here  
    height=input("Enter your height in meters: ")  
    if height < 1.6:  
        print("Below minimum astronaut height")  
    if height > 1.9:  
        print("Above maximum astronaut height")  
    if height >= 1.6 and height <= 1.9:  
        print("Correct height to be an astronaut")  
  
if __name__ == "__main__":  
    main()
```

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

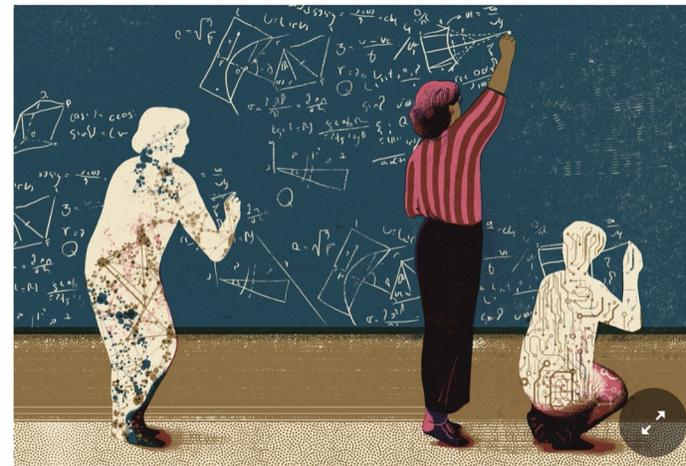
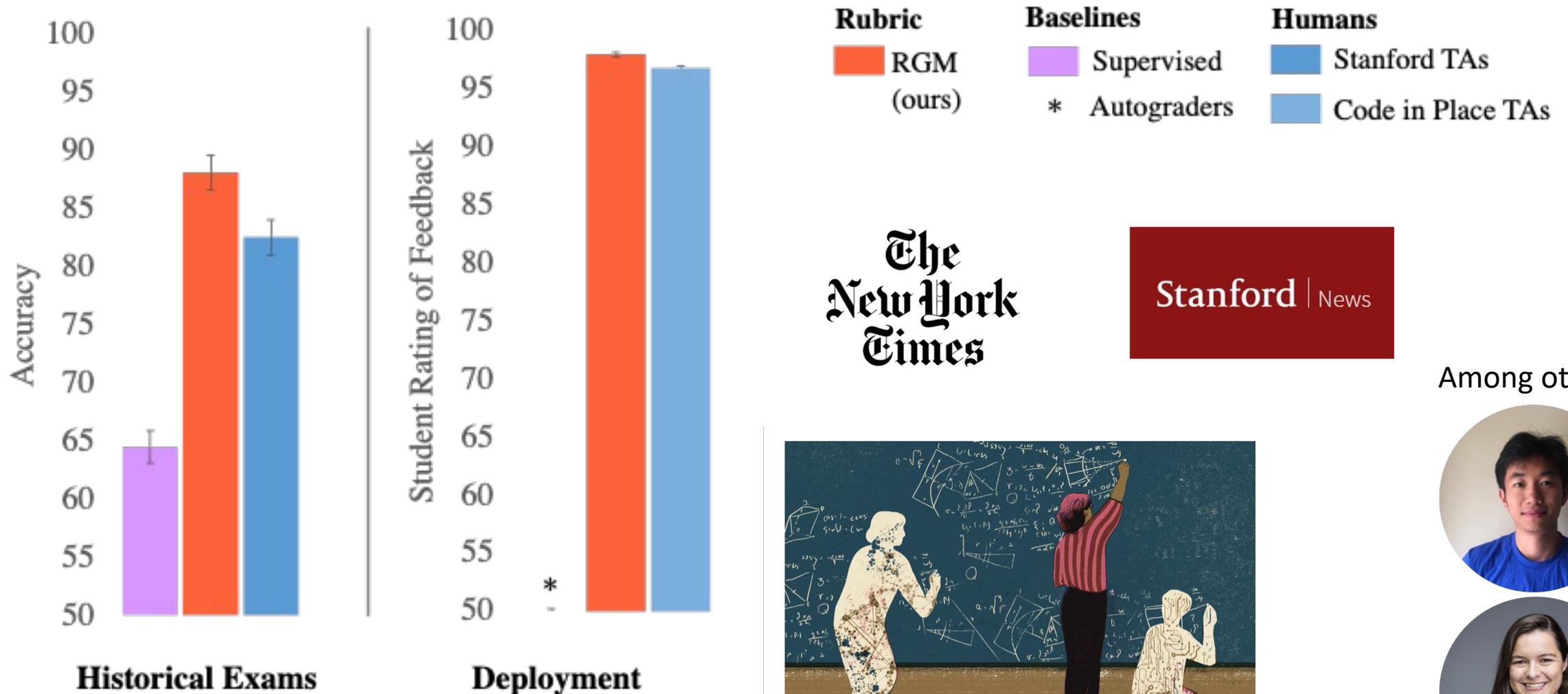
Syntax error (missing ") here would prevent auto graders from being useful.

AI generated feedback

Students evaluate the feedback



Impact and Fairness Analysis (work in progress)



AI Teacher Training

1

Record Class



2

**Transcribe & Anonymize
Recording**



3

**Analyze
Transcripts**



4

**Generate
Insights**



Causal impact of AI teacher Training

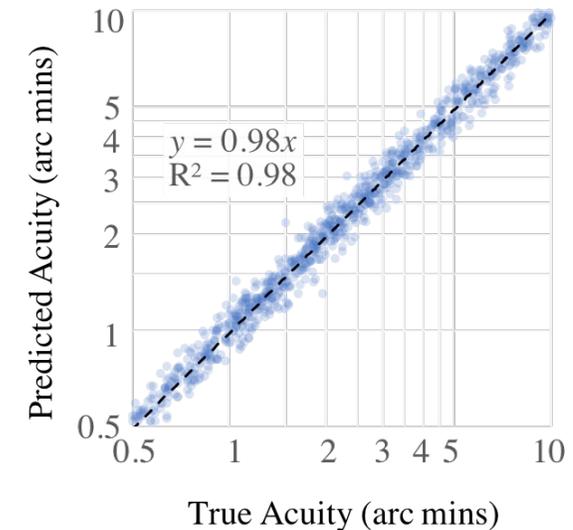
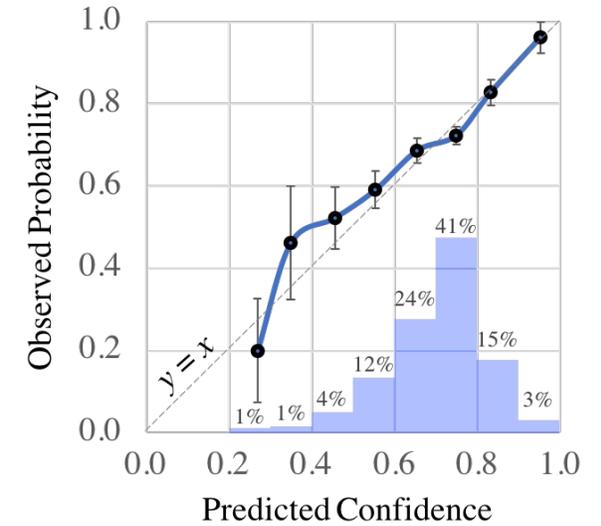
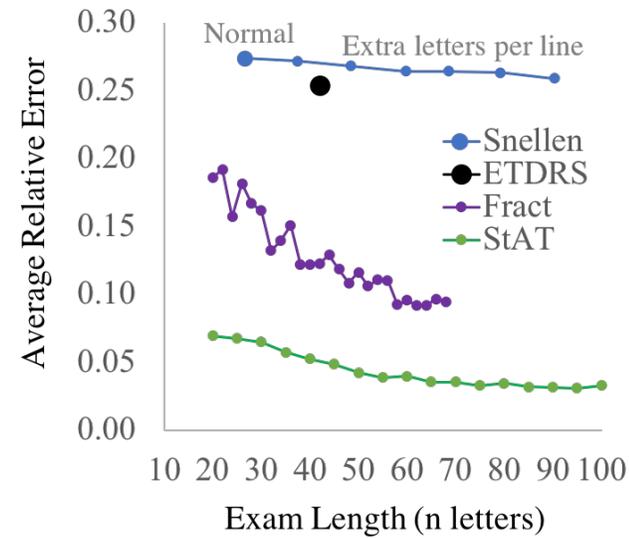
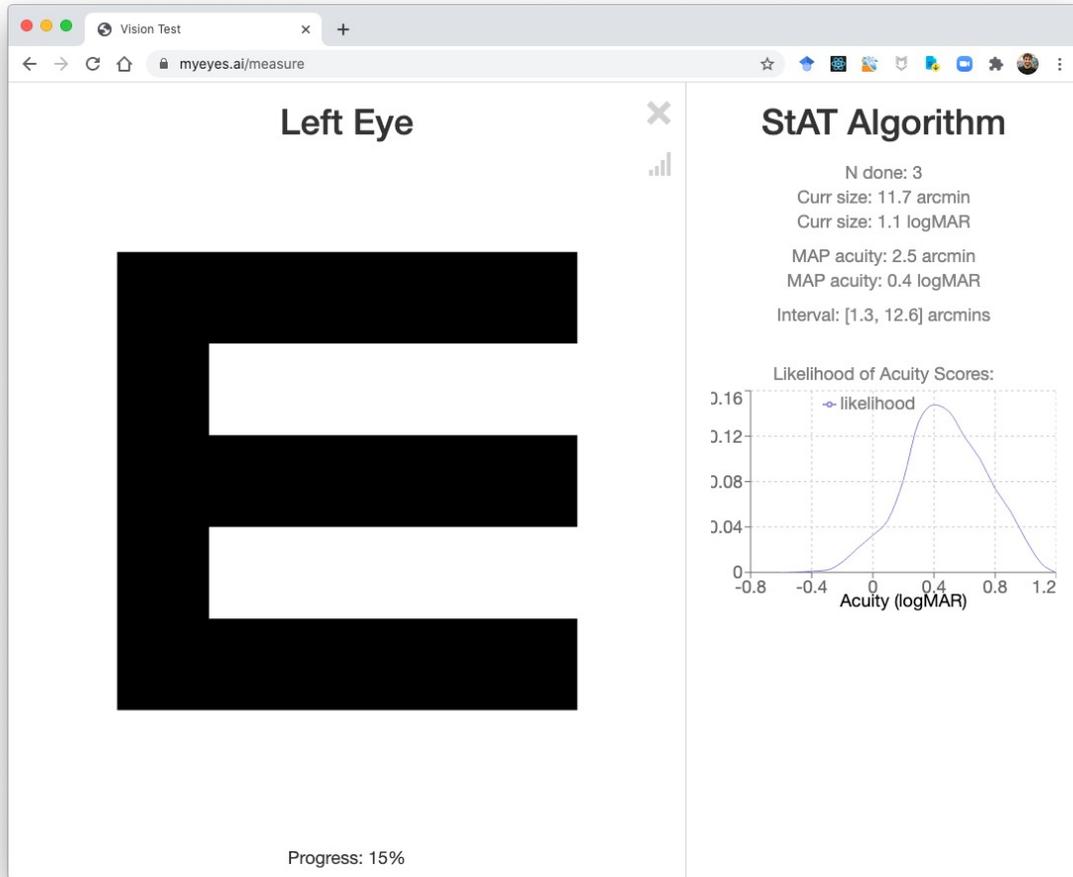
Within 4 lessons, as a result of training

1. Teachers asked 10% **more questions**
2. Teachers “**took up**” student ideas 10% more
3. Students were significantly more likely to:
recommend the class ($p < 0.05$),
find sections helpful ($p < 0.05$)



Probability in my Research: Better Eye Exam

Jan 2020, With a former CS109 student, Ali Malik

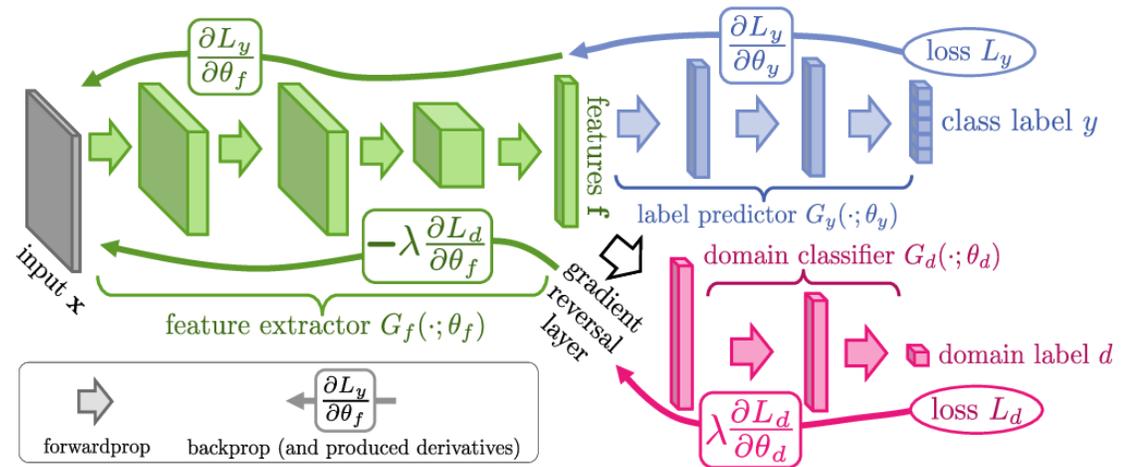


Math question: Estimate a continuous valued number.
Get to run noisy experiments of your choosing.



Fair AI with Adversarial Network

2018, 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
BEHAVOD ET AL. AVD PENALIZERS (2017)	0.65	0.02	0.04
BEHAVOD ET AL. SD PENALIZERS (2017)	0.66	0.02	0.03
BEHAVOD 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?



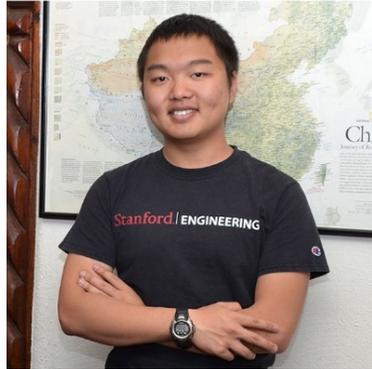
So many things to love in this world



Most amazing family



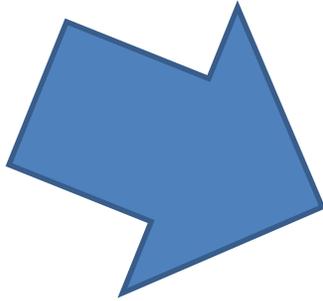
Amazing Teaching Team



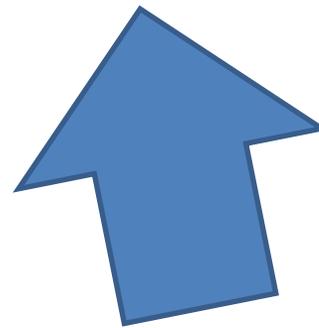
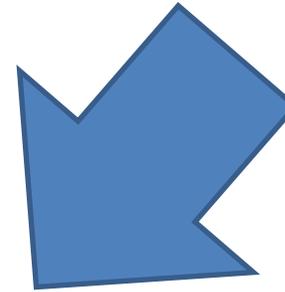
Course mechanics

(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):

- 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



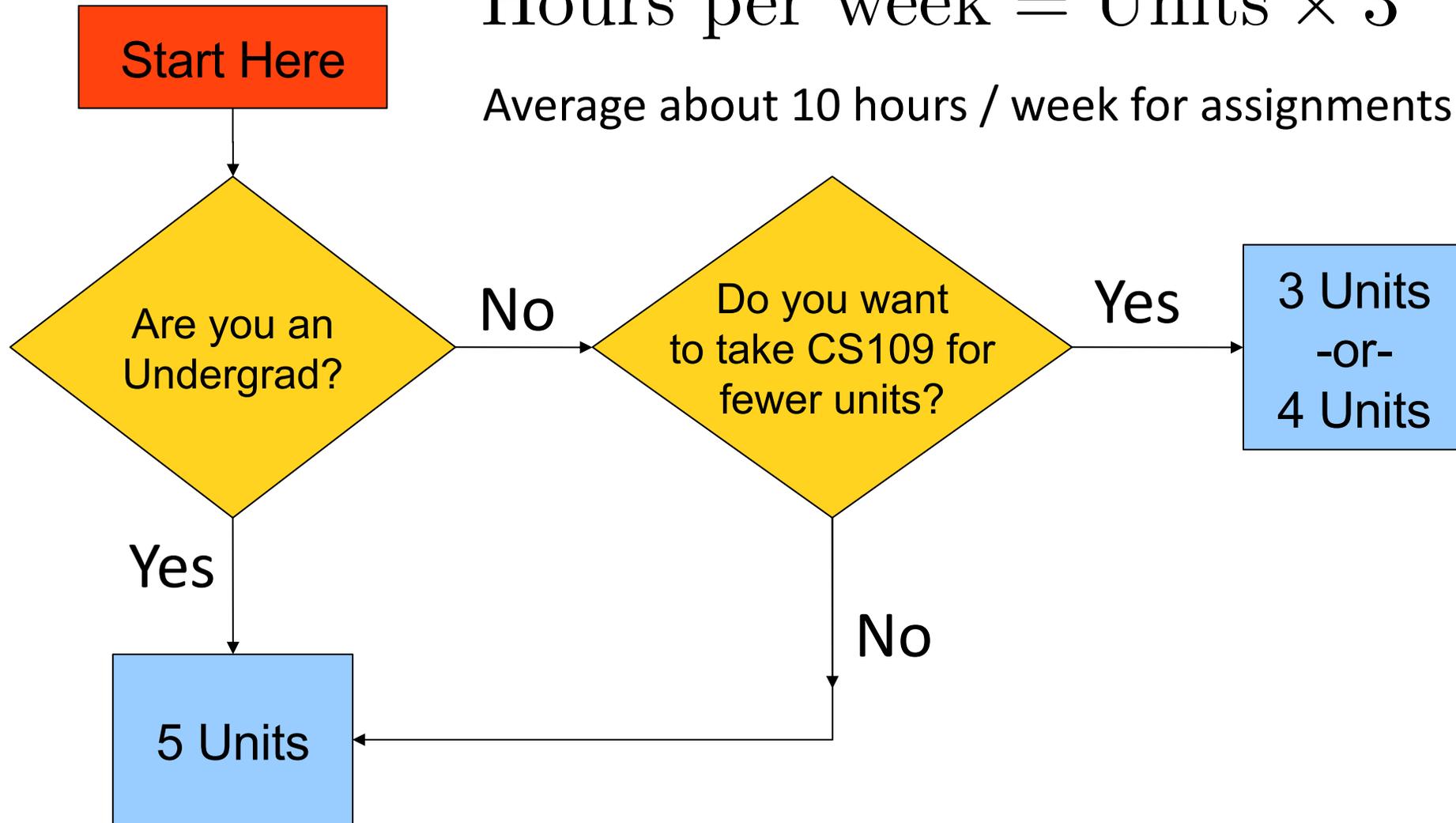
Review session on Friday



CS109 Units

$$\text{Hours per week} = \text{Units} \times 3$$

Average about 10 hours / week for assignments



Class Breakdown

40%

6 Assignments

20%

Midterm

2 hour exam, Nov 1st, 7pm

30%

Final

3 hour exam, Dec 13th, 8:30am

10%

Section Participation



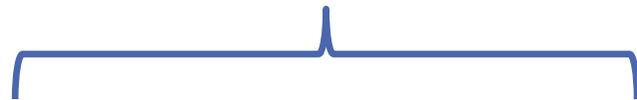
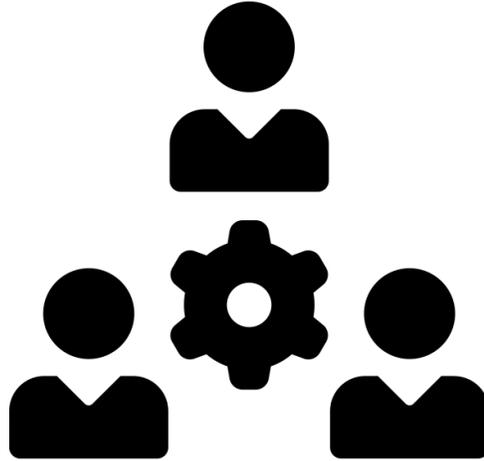
Is Class Online?



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



Ask questions



Q&A forum
All announcements



“Working” office hours
start on Thursday



Email cs109@cs.stanford.edu



Chris has 1:1 CS109 office hours

From Last Year: Course Reader!

Probability for Computer Scienc

chrispiech.github.io/probabilityForComputerScientists/en/

Course Reader for CS109

Search book...

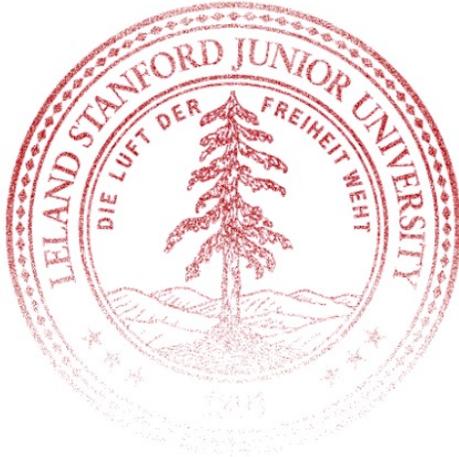
Part 1: Core Probability

- 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

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



ACE Companion Course



CS ACE Program Application - Winter 2021-22

Additional Calculus for Engineers (ACE) is designed to provide the skills and solid foundation in mathematics, computational math in engineering, and computer science to students interested in pursuing an engineering major.

Students participating in ACE are required to attend an additional weekly section and enroll in their selected course for 1 additional unit. ACE participants will also receive access to additional exam review sessions and other resources.

NOTE ON CONCURRENT ENROLLMENT: Concurrent class enrollment and completion is REQUIRED. ACE courses cannot be standalone courses.

Please complete the following application to be considered for an ACE section this quarter.

Note that space in each section is limited.

The PRIORITY deadline is 5PM on Friday, December 17, 2021.

The FINAL deadline for ACE applications is at 5pm on Friday, January 7, 2022

Applicants will be notified of their application result on a rolling basis.

<https://forms.gle/QPhjeoTnDJMaTAh3A>



Julia Kadie



CS109 | Syllabus

localhost:8001/handouts/syllabus.html

CS109 Course Problem Sets Lecture Section Resources Schedule

Syllabus

Syllabus

UPDATED 4 DAYS AGO

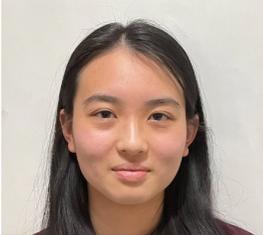
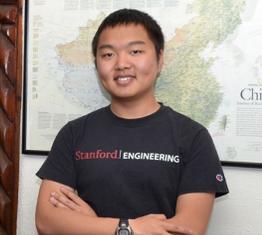
If you have any questions after reading this Syllabus, post on our [discussion forum](#), or email us at our mailing list: cs109@cs.stanford.edu.

Teaching Team



Professor: Chris Piech
✉ piech@cs
🏠 Durand 305
🕒 OH: Monday 4:00 to 5:30p

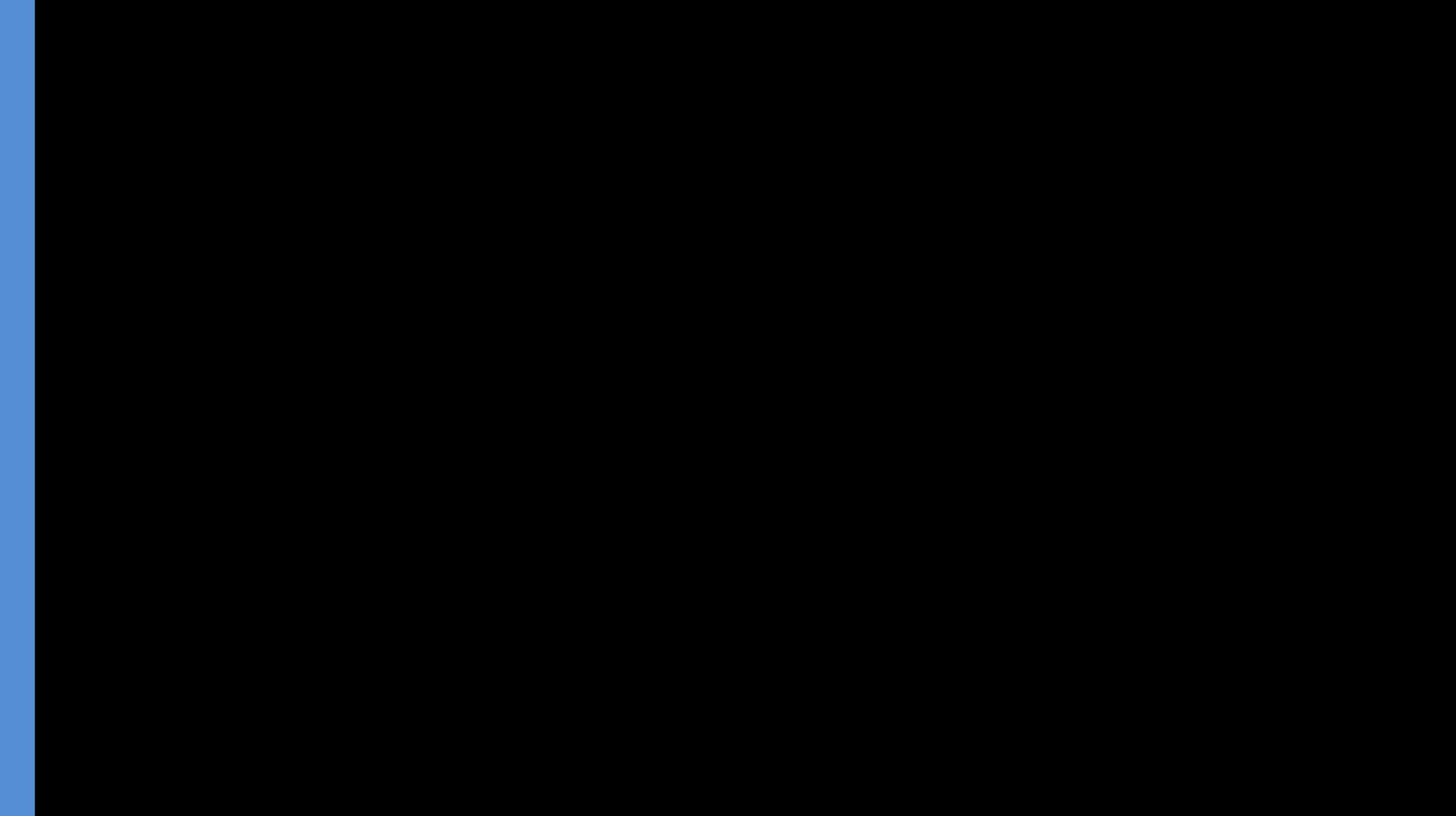
We are lucky to have a phenomenal group of Course Assistants:



Teaching Team

- I. Course Overview
- II. Course Structure
- III. Course Resources
- IV. Honor Code
- Looking Forward to a Great Quarter

localhost:8001/handouts/syllabus.html



Story of Modern AI

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

Brief History

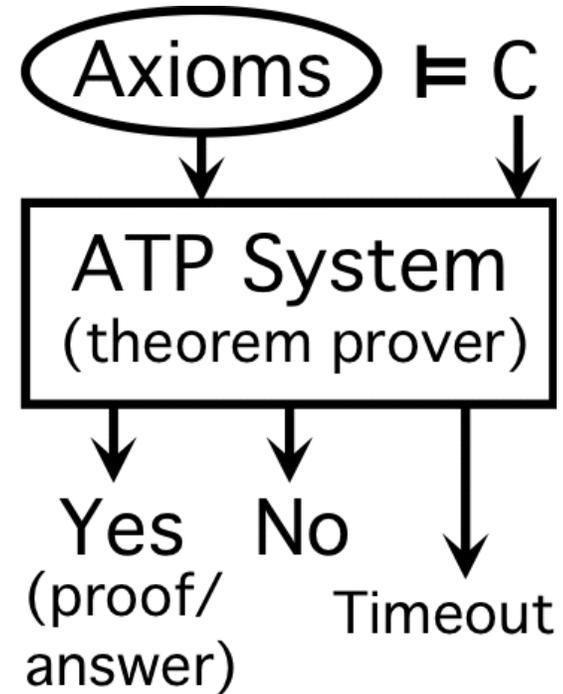


Early Optimism 1950s

1952



1955



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

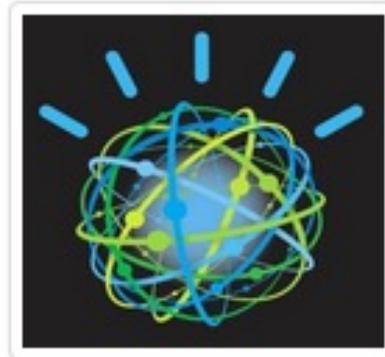
Big Milestones Part 1



1997 Deep Blue

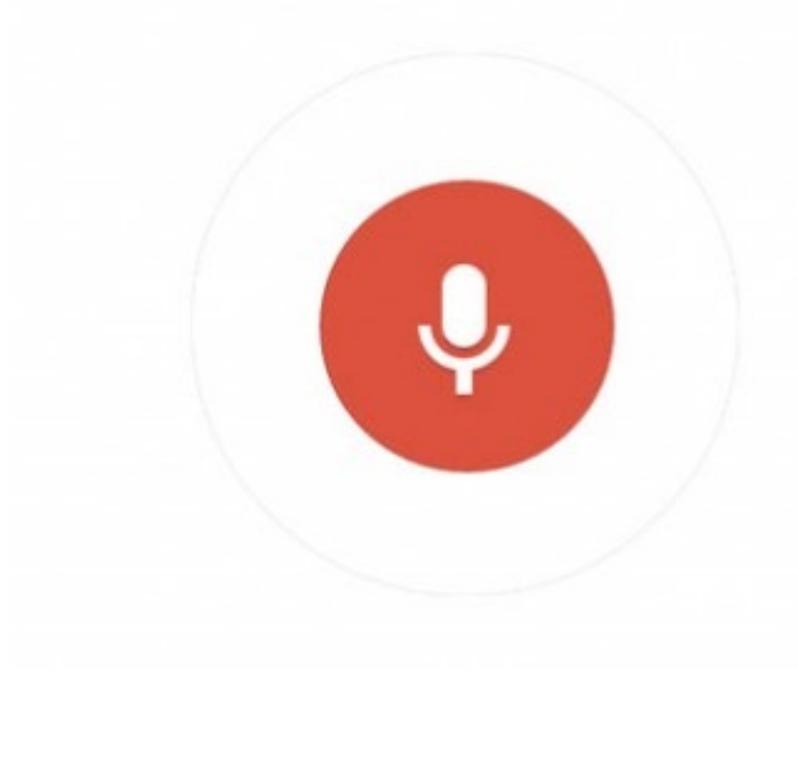


2005 Stanley



2011 Watson

I was told speech was 30 years out



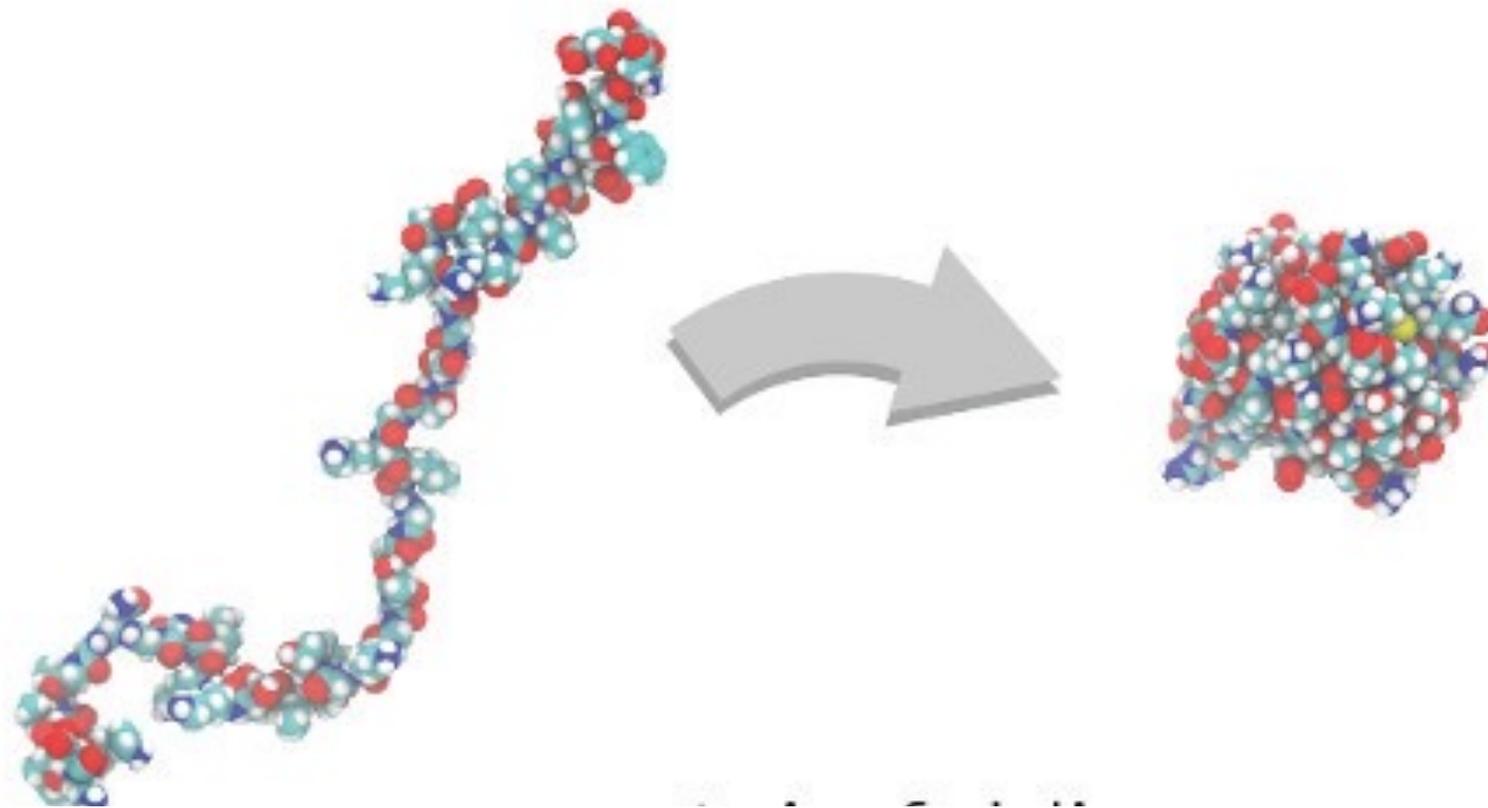
Almost perfect...



The last remaining board game



Protein Folding



Directions From A to B

The screenshot shows Google Maps with a route from Sacramento, CA to Mexico City, Mexico. The sidebar on the left provides the following information:

- Home (2600 Columbia St)** to **Mexico City, Mexico**
- Leave now
- Send directions to your phone
- via I-10 E**: 37 hr, 2,191 miles. Includes warnings: "This route has tolls.", "This route crosses a country border. Verify COVID-19 restrictions.", and "Your destination is in a different time zone." Details link is present.
- via México 15D**: 38 hr, 2,207 miles
- via I-5 S and I-10 E**: 38 hr, 2,405 miles

At the bottom of the sidebar, there are icons for Restaurants, Hotels, Gas stations, and Parking Lots, along with a "Layers" button.



3:31



Google Translate



English



Ukrainian



ENGLISH



Please translate this into Ukrainian.
Thank you



Camera



Conversation



Transcribe

UKRAINIAN



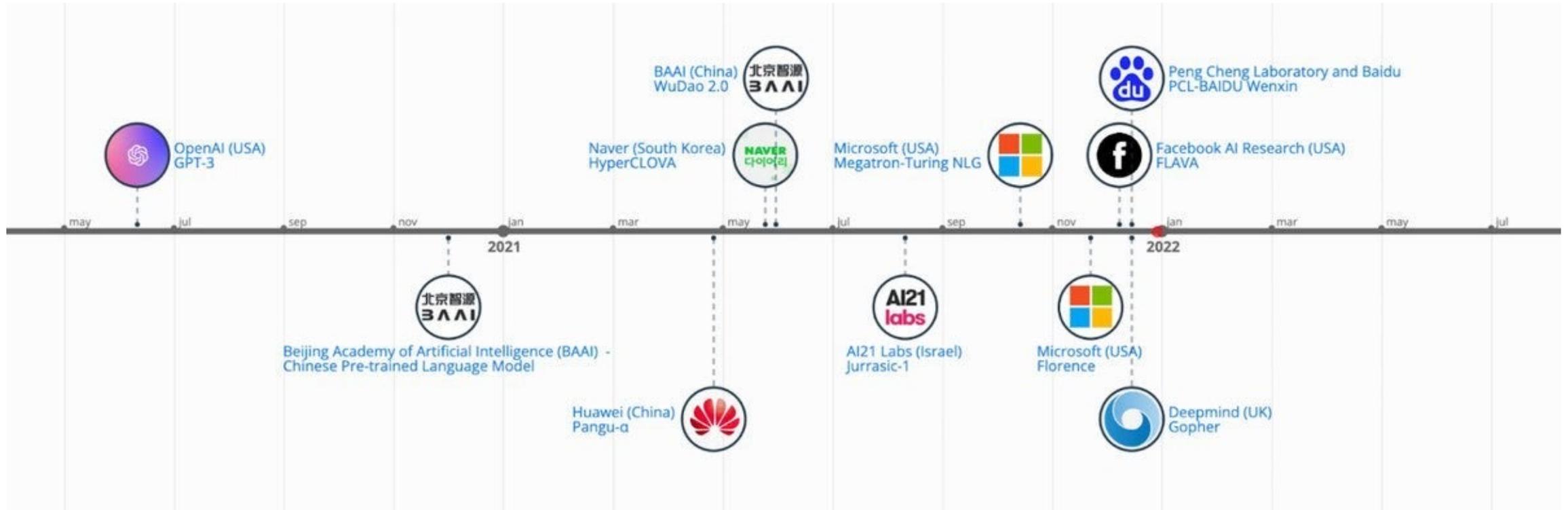
Будь ласка, переведіть це

Self Driving Cars



And then this year, everything changed, again

AI that (seems) to understand language



Midjourney Images made by Freya (20 months)

Three birdies happy



Robot Maasai



What is going on?

[suspense]

Focus on one problem

Computer Vision



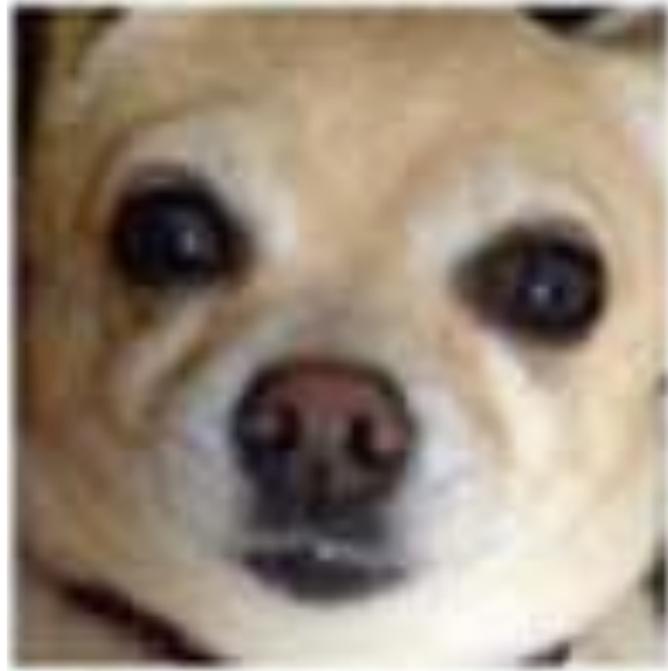
Chihuahua or muffin?

Piech, CS109, Stanford University



Can you do it?

Chihuahua or Muffin?



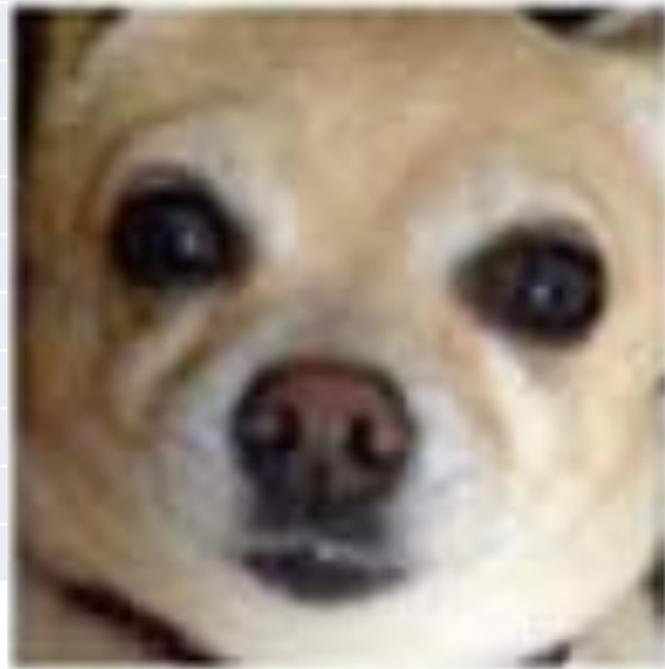
Chihuahua or Muffin?



How about now?

What a computer sees

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



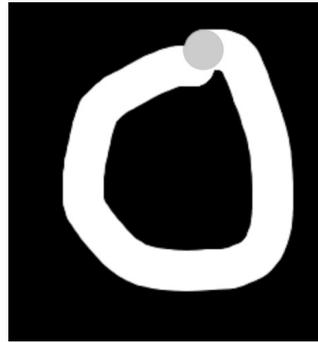
That is a picture
of a **one**



Classification



That is a picture
of a **zero**



Classification



That is a picture
of an **zero**



* 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	0	1
1	0	0	1	0	1	1	1	0	1	0	0	0	0	0
1	1	1	0	1	0	0	1	1	0	0	1	0	1	0
1	1	1	1	1	0	0	0	0	0	1	1	0	1	1
0	0	0	1	1	0	0	1	0	0					
1	0	0	1	1	0	0	0	1	0					
1	1	0	1	1	0	0	1	1	0					
1	0	1	0	0	1	0	0	1	0					
0	0	0	0	1	0	1	0	1	1					
0	1	1	0	0	0	0	0	1	1					
0	0	1	0	1	1	1	0	0	0					
0	1	1	1	0	1	0	0	1	0					
1	1	0	0	0	0	0	0	0	0					
0	0	0	0	0	0	0	0	1	1					
0	0	1	1	1	0	1	0	1	1					



What a human sees



Very hard to Program



```
public class DigitDetector extends ConsoleProgram {  
  
    public void run() {  
        println("Todo: Write program");  
    }  
  
}
```



Perhaps there is an insight?

Two Great Ideas

1. Artificial Neurons

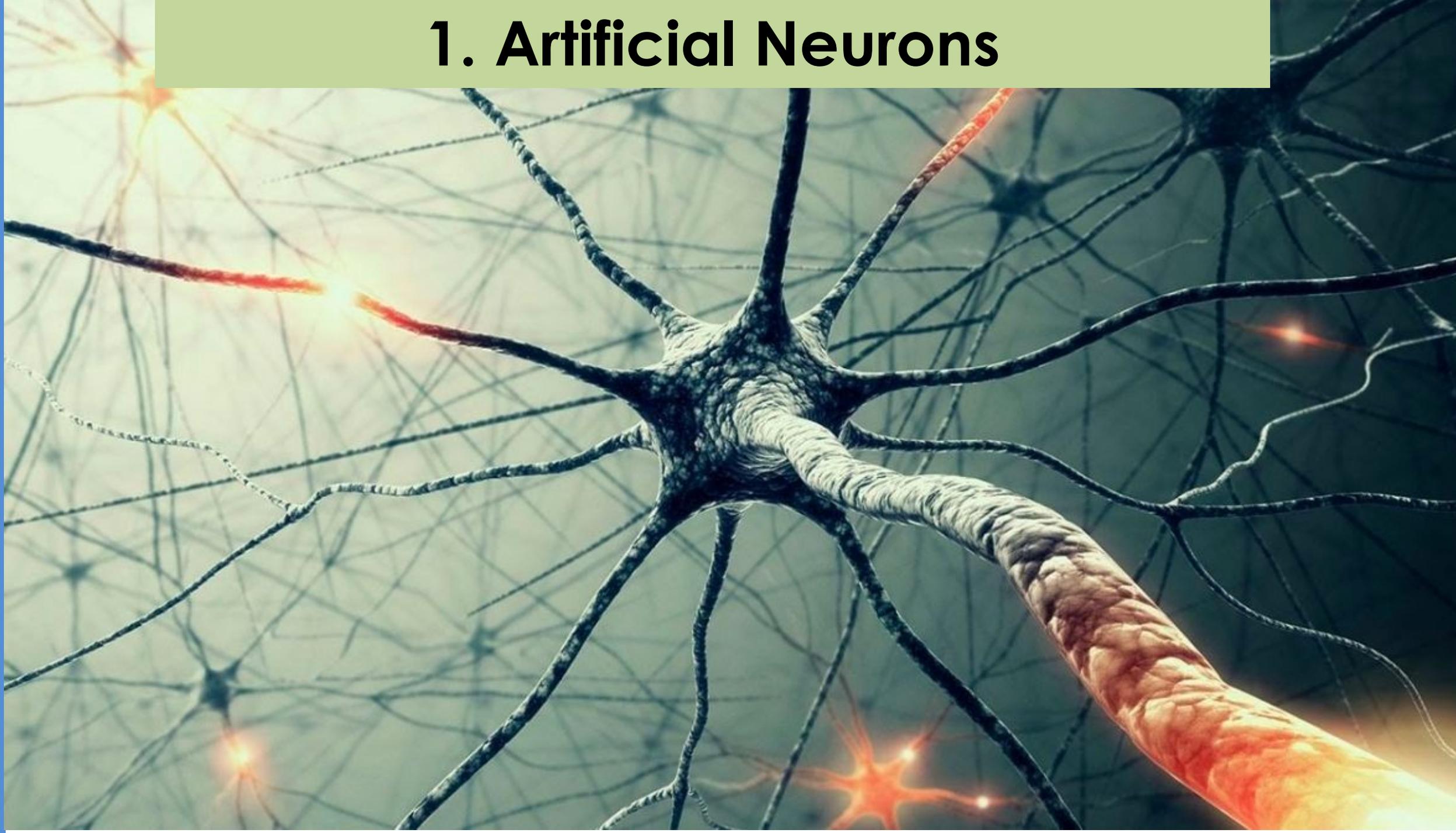
2. Learn by Example

Two Great Ideas

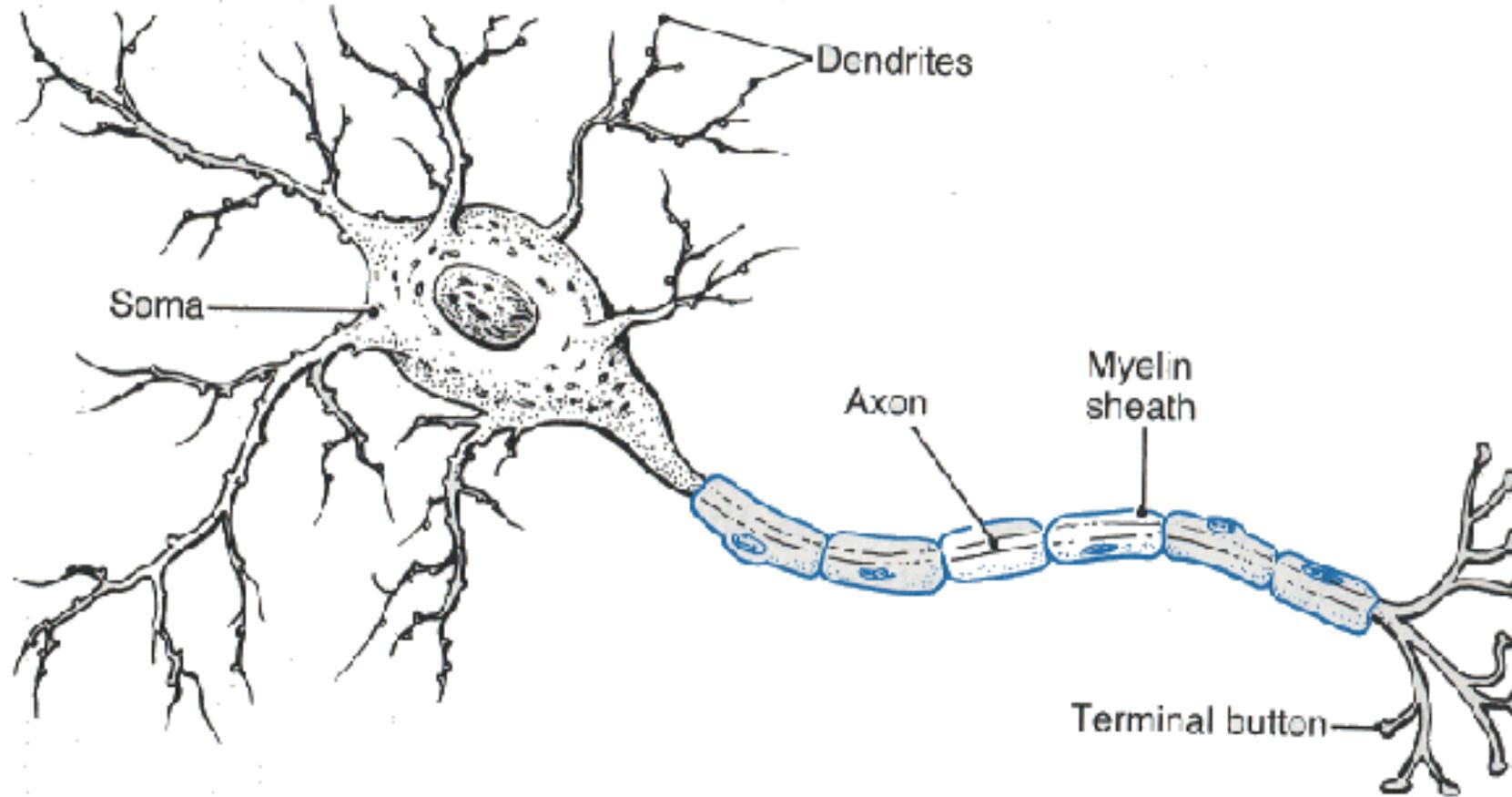
1. Artificial Neurons

2. Learn by Example

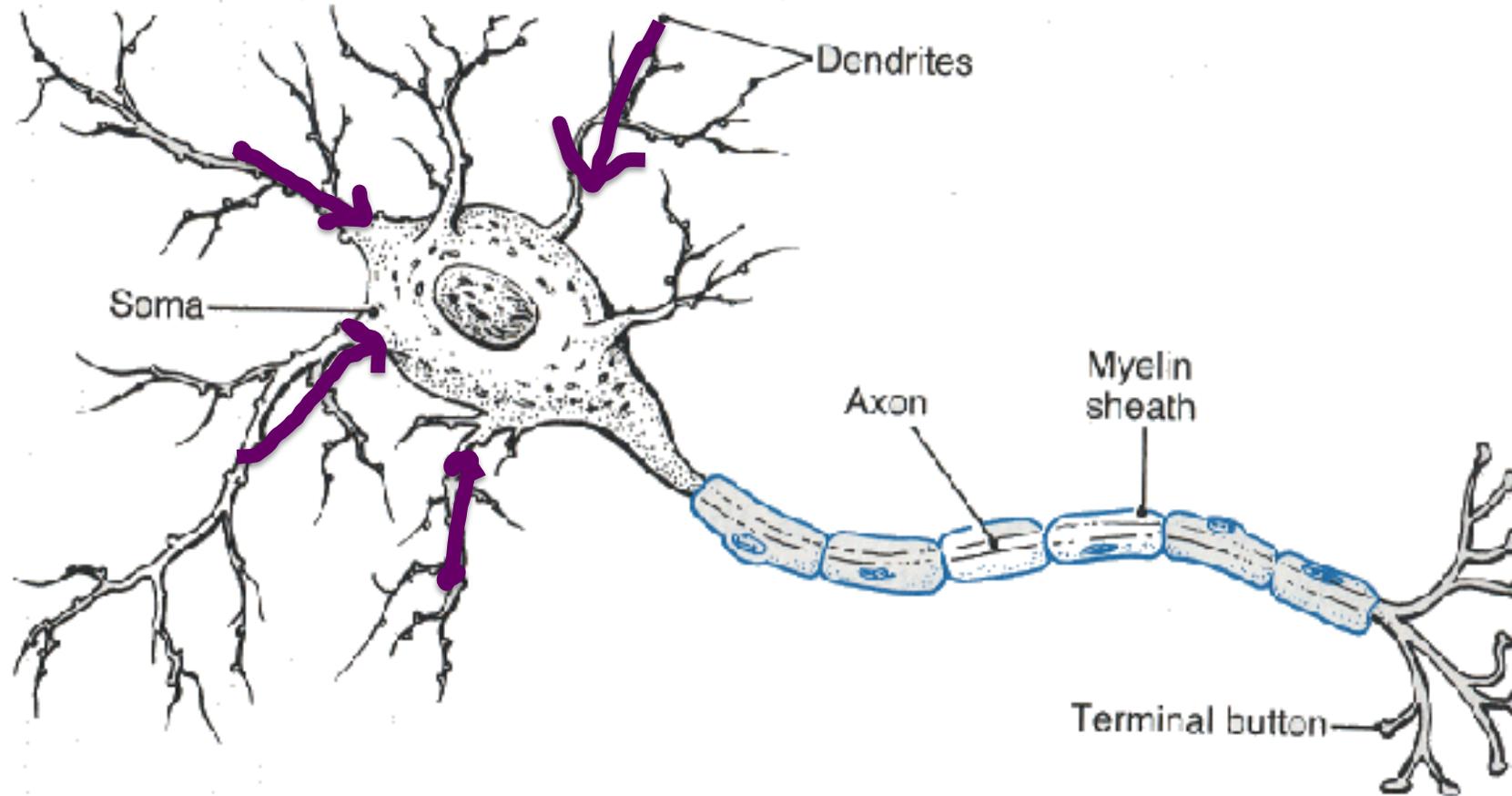
1. Artificial Neurons



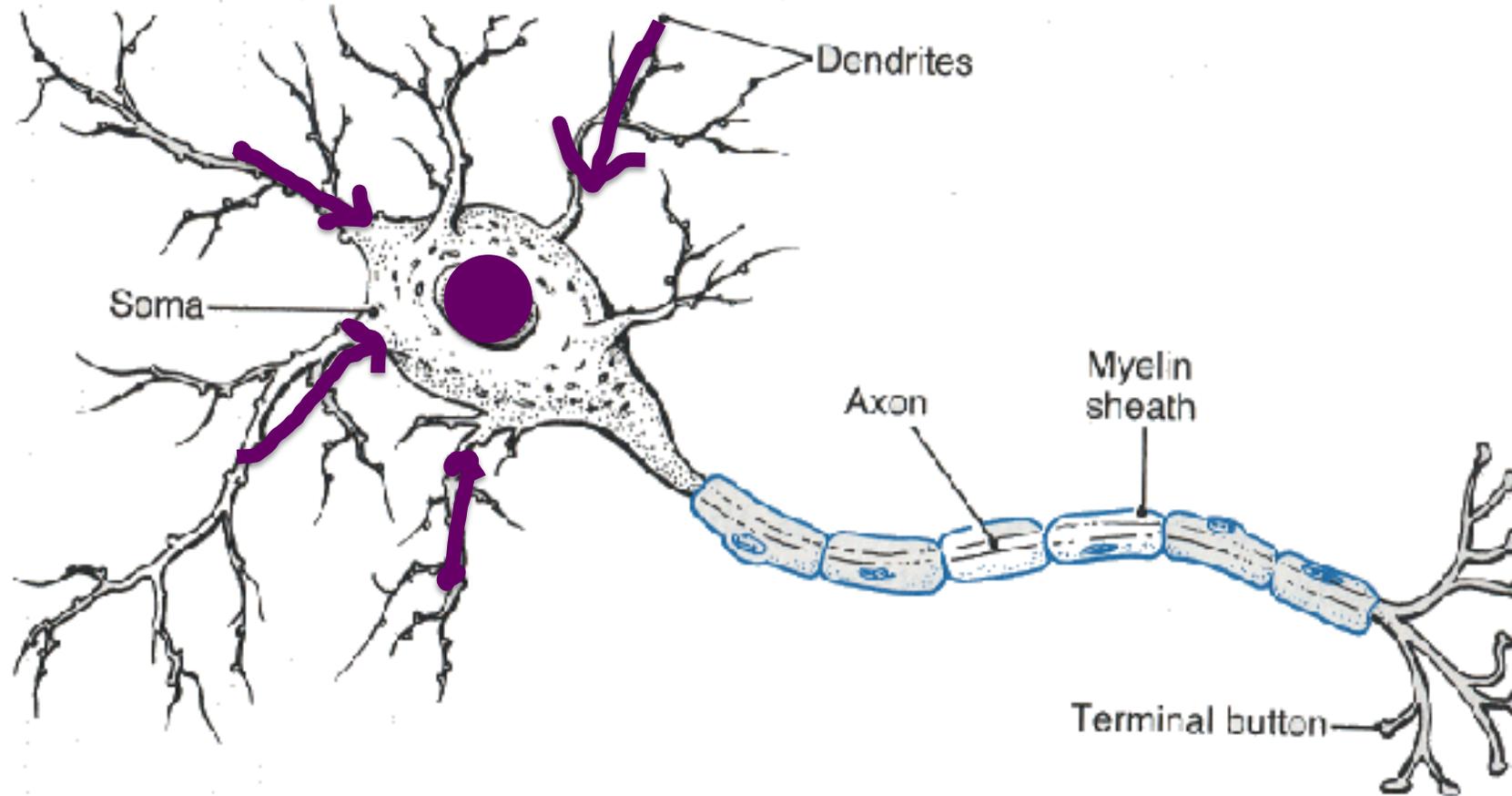
Neuron



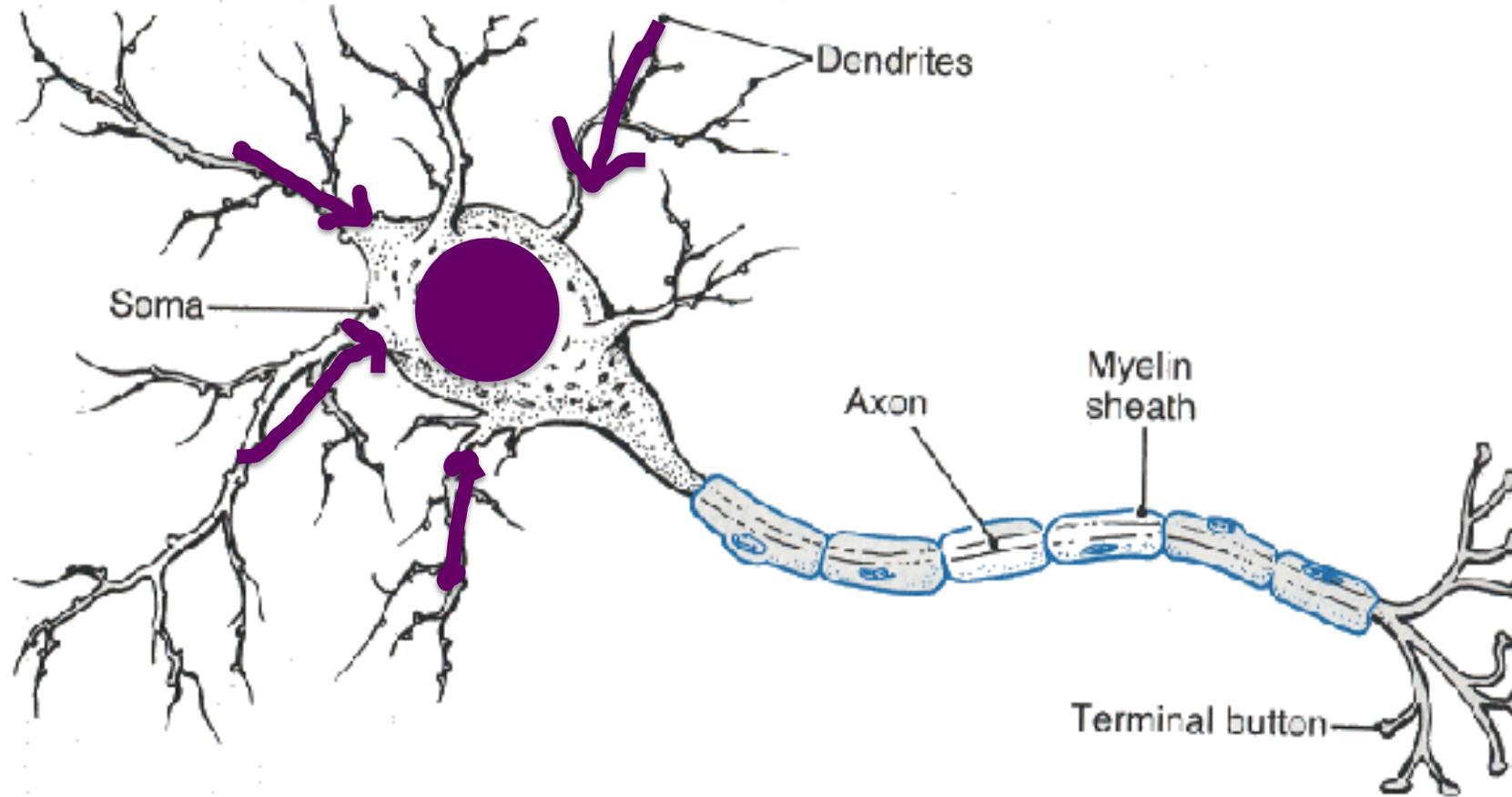
Neuron



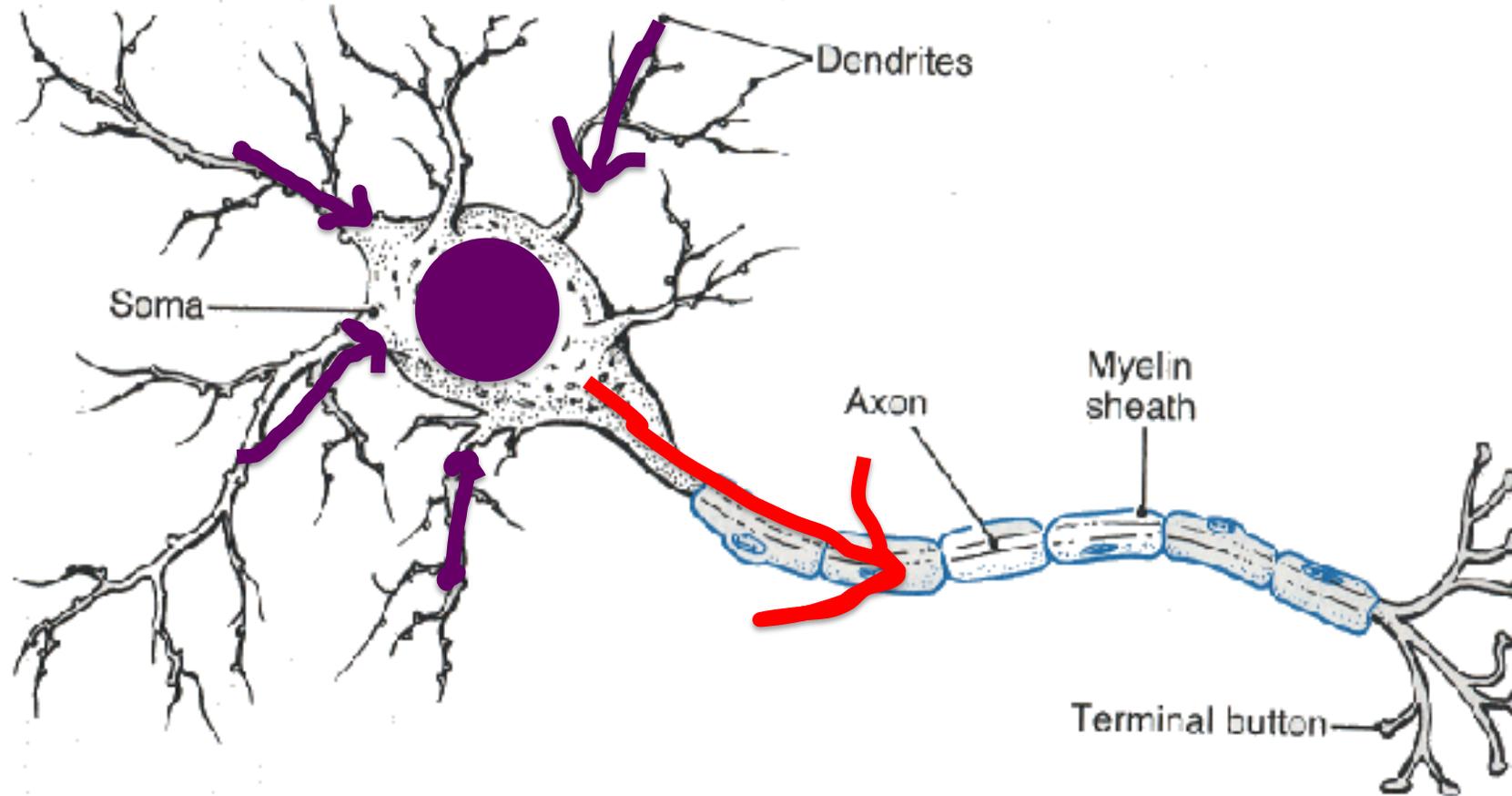
Neuron



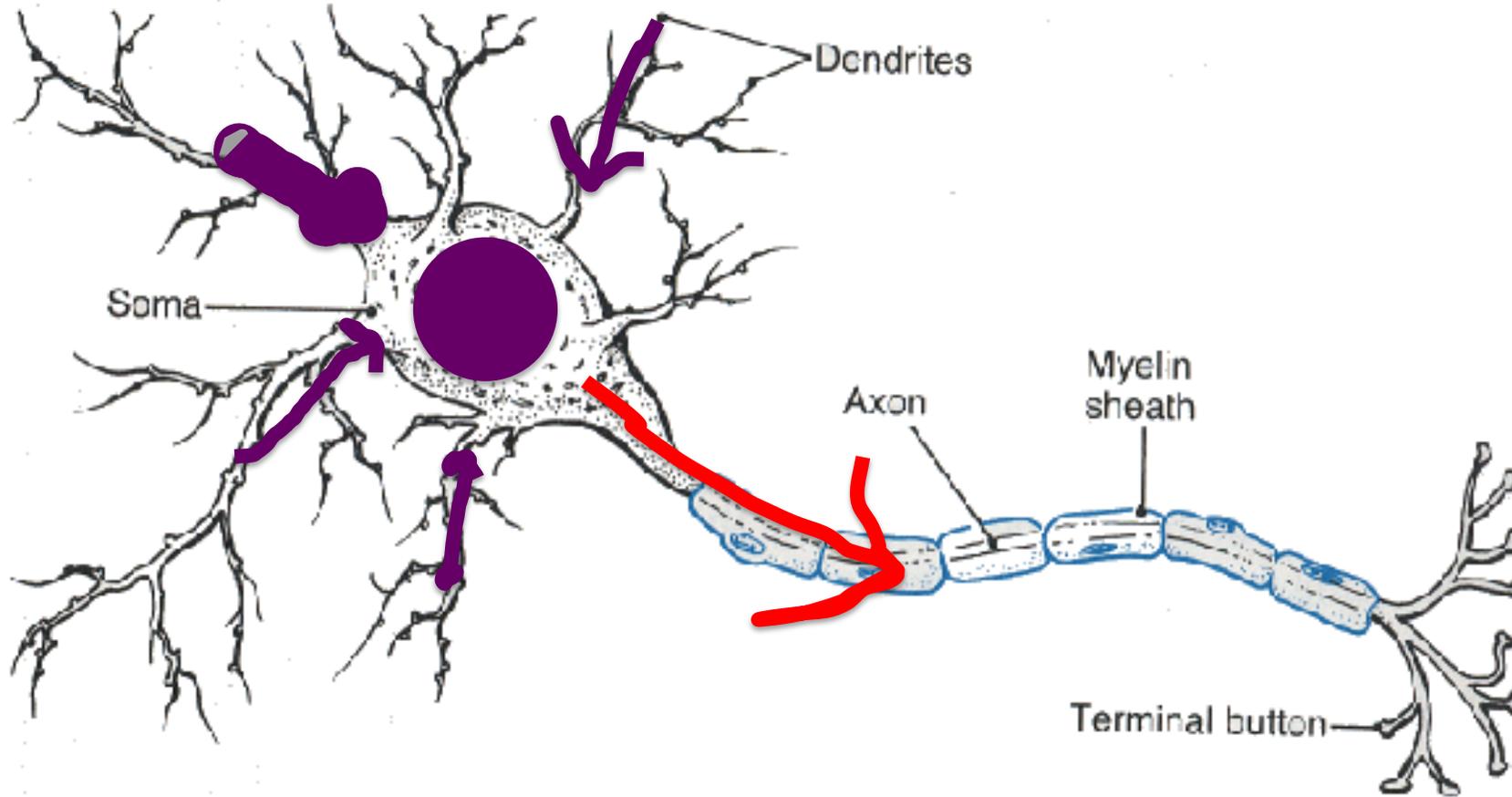
Neuron



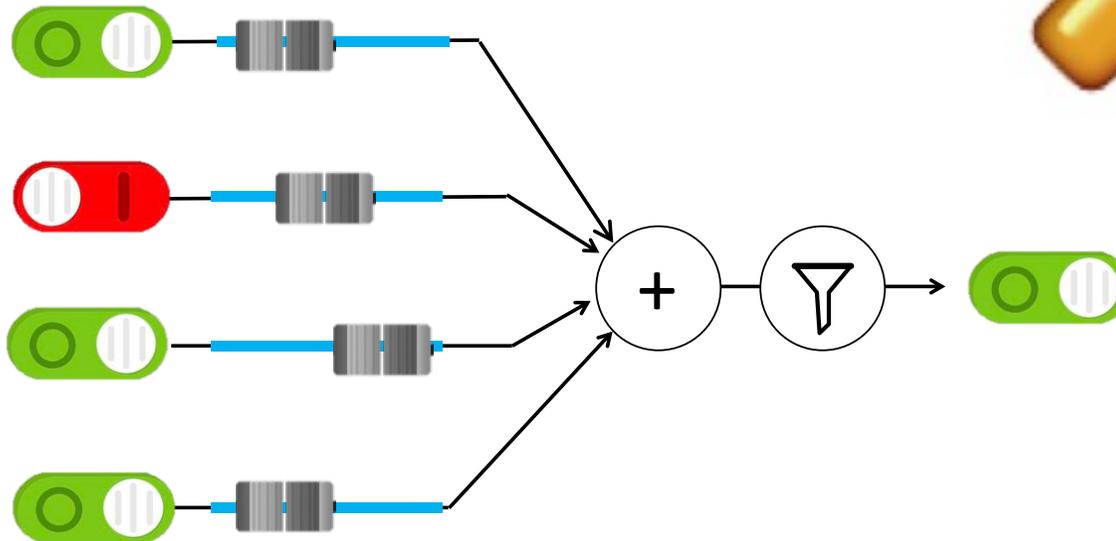
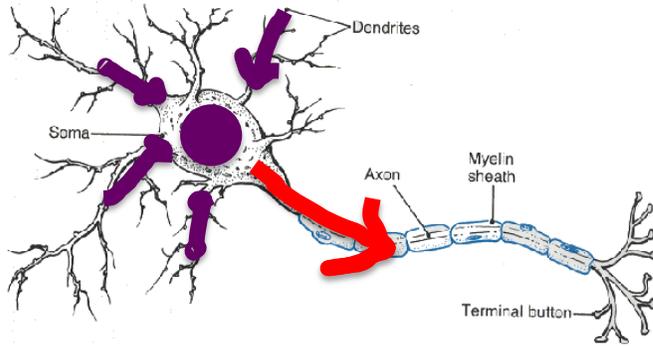
Neuron



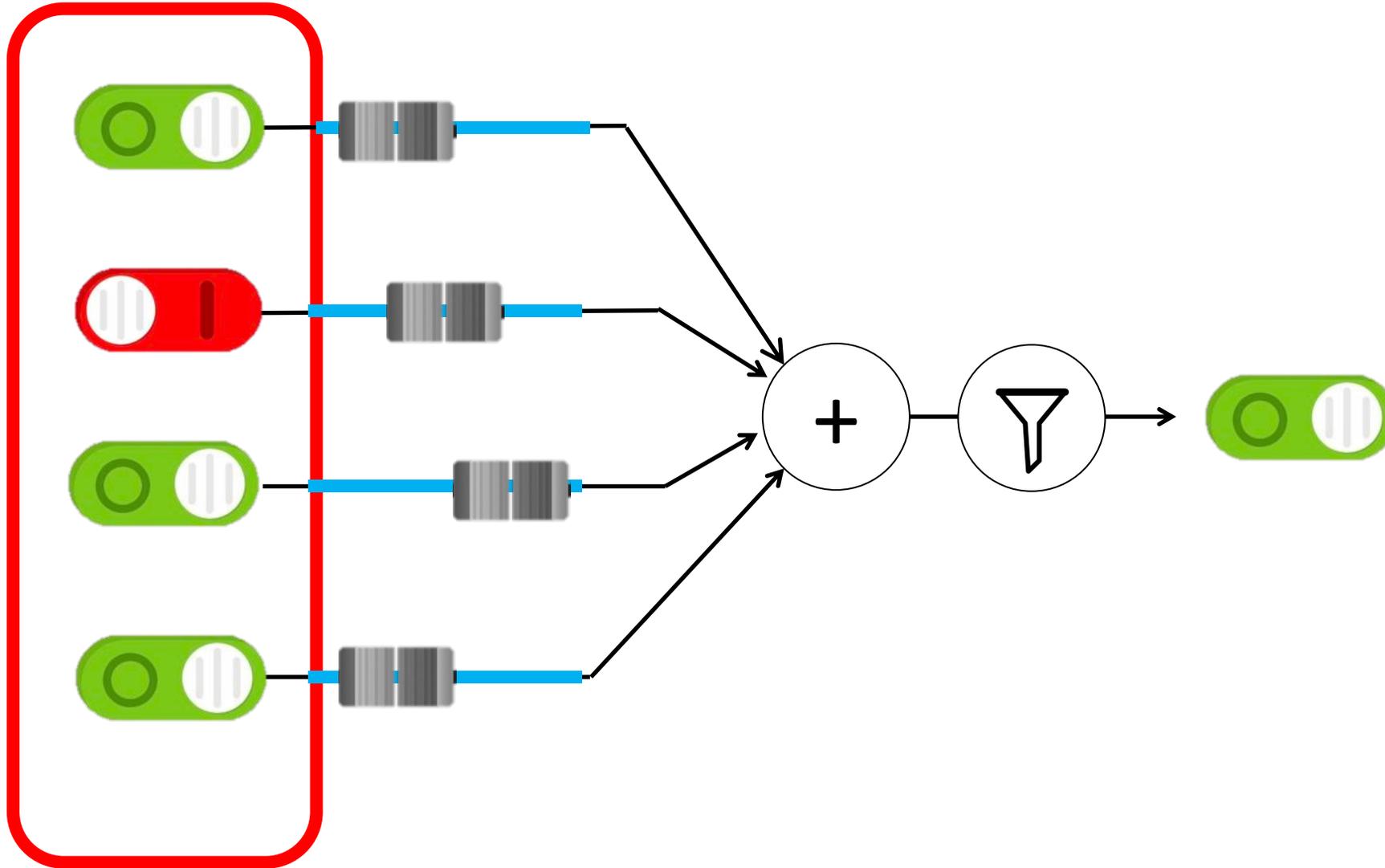
Some Inputs are More Important



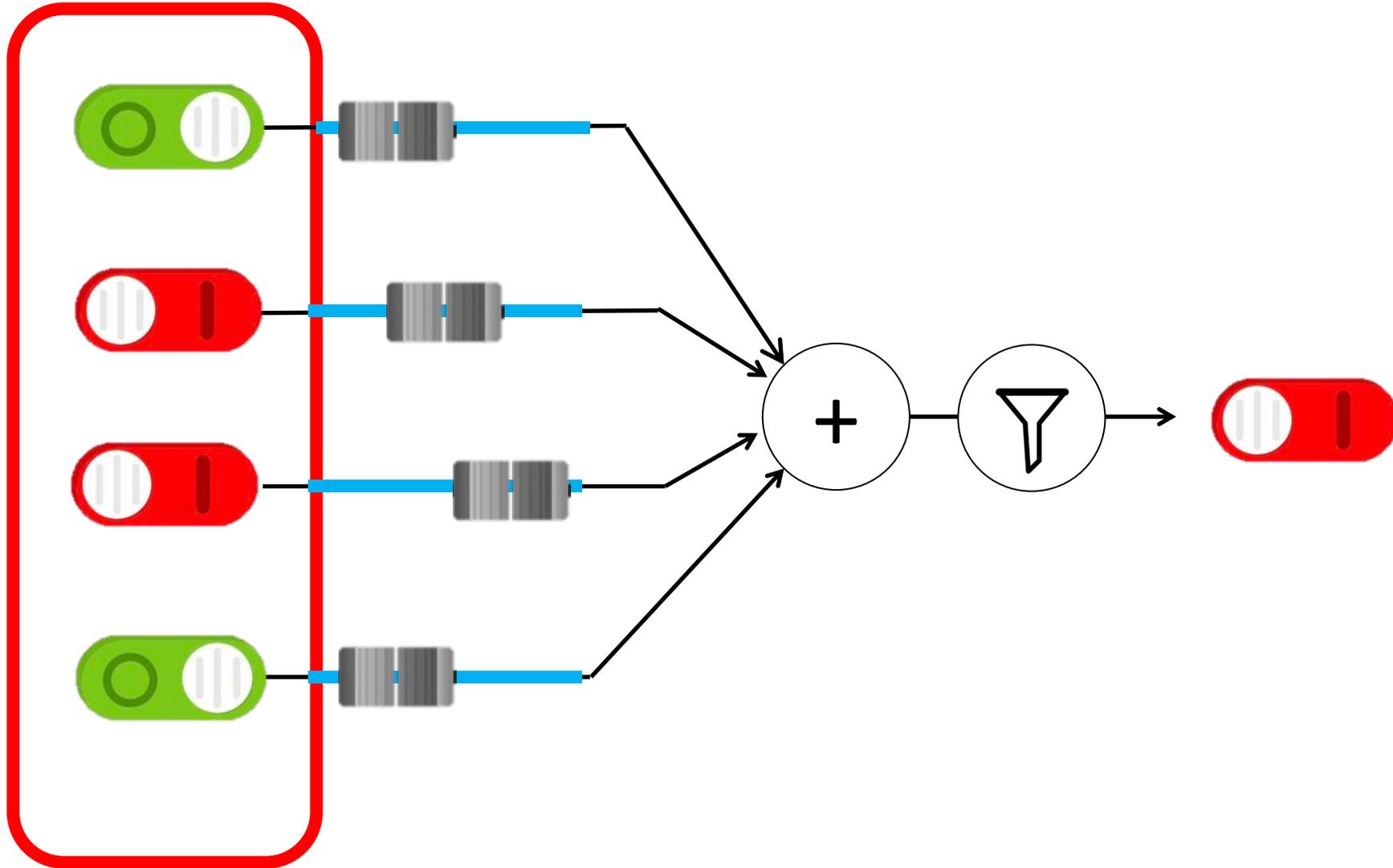
Artificial Neuron



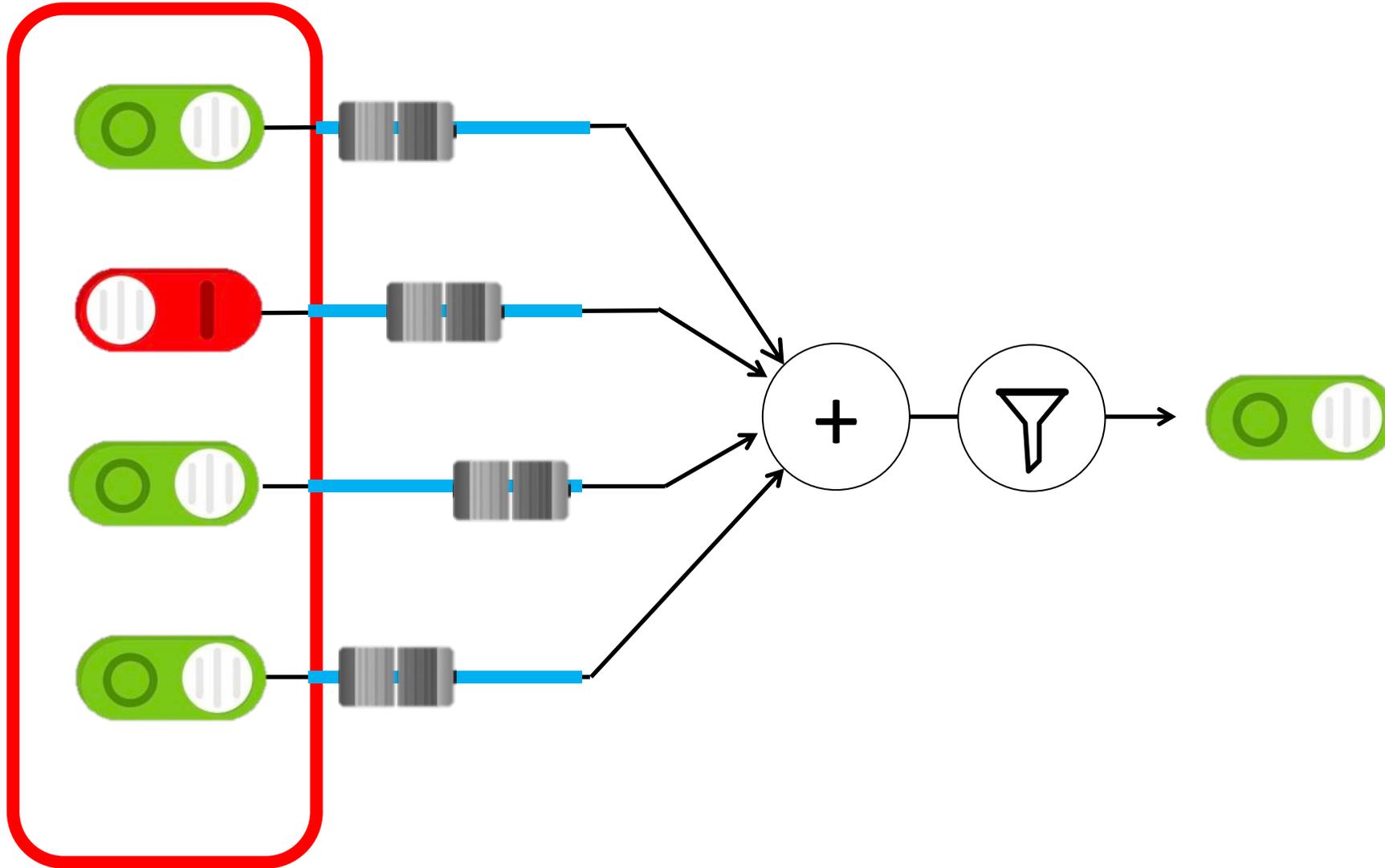
Inputs



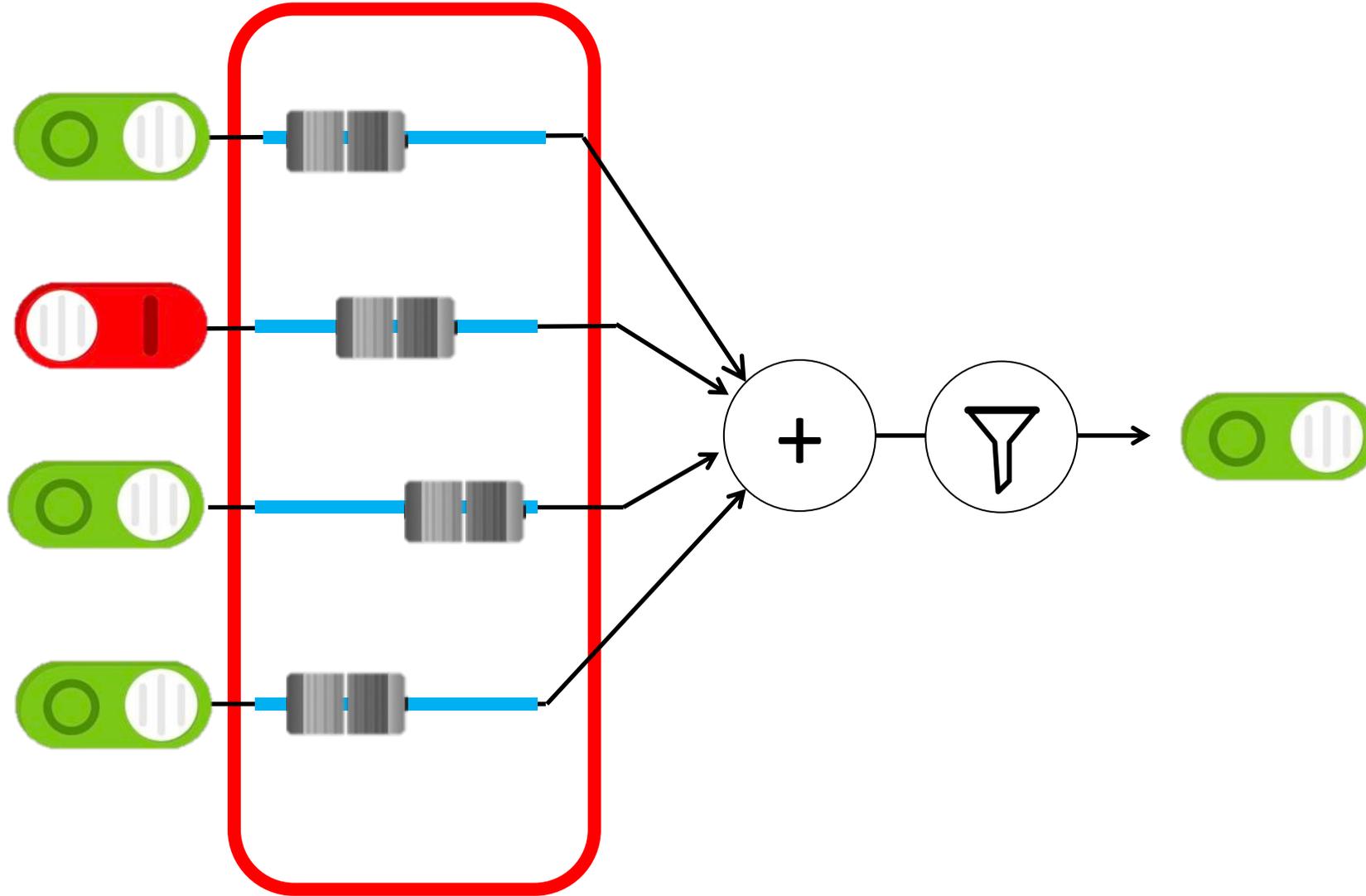
Inputs



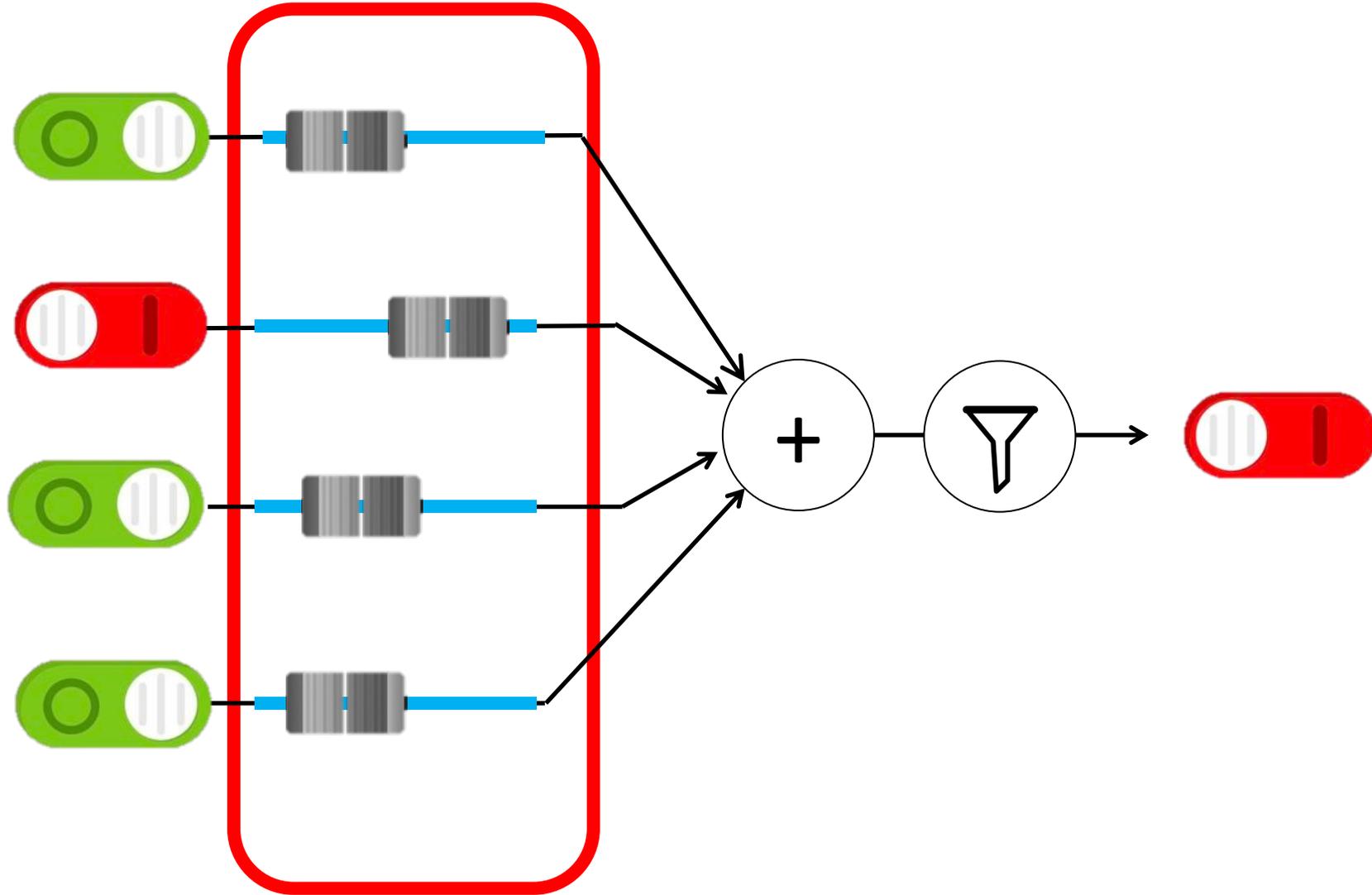
Inputs



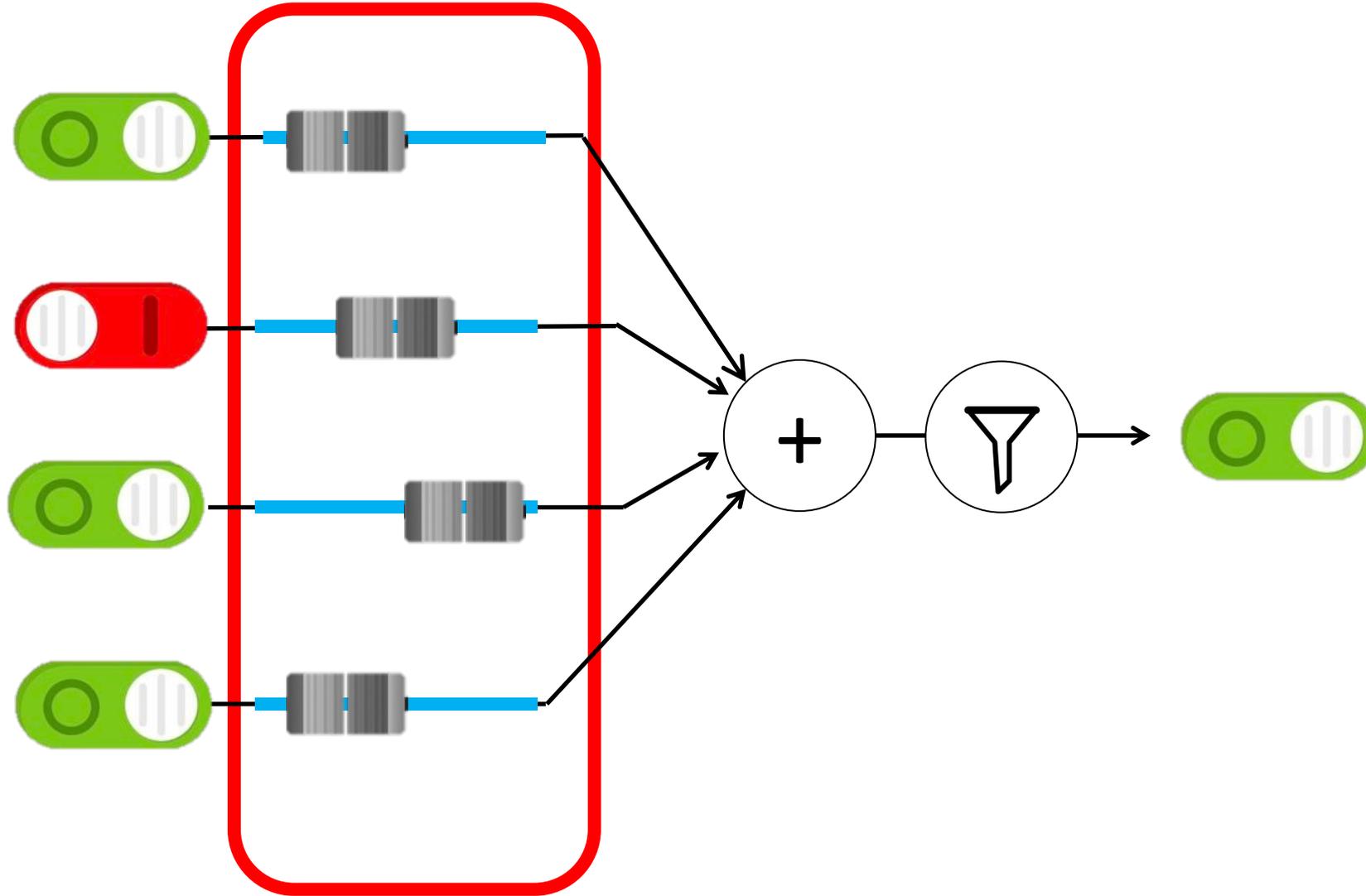
Weights



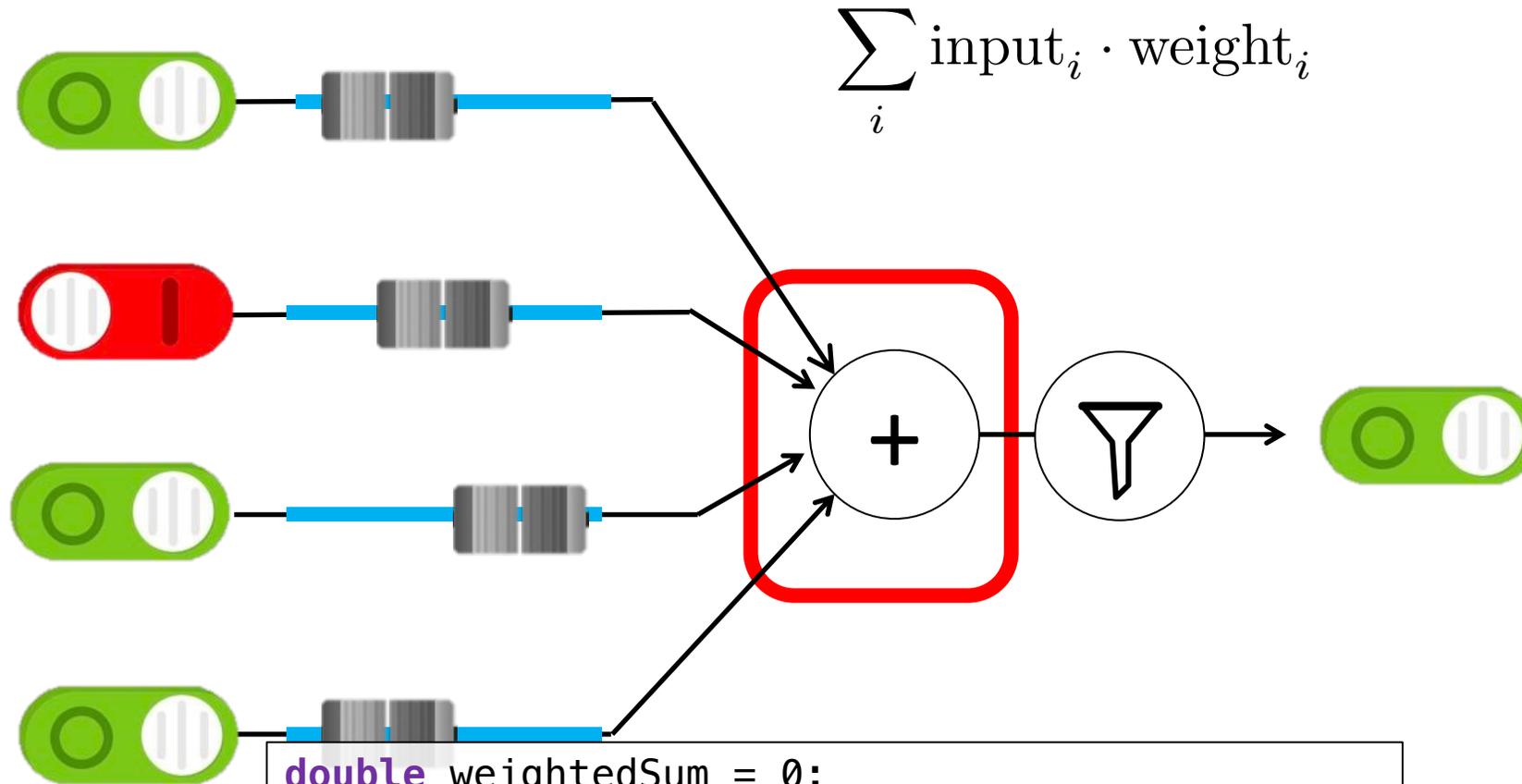
Weights



Weights



Weighted Sum

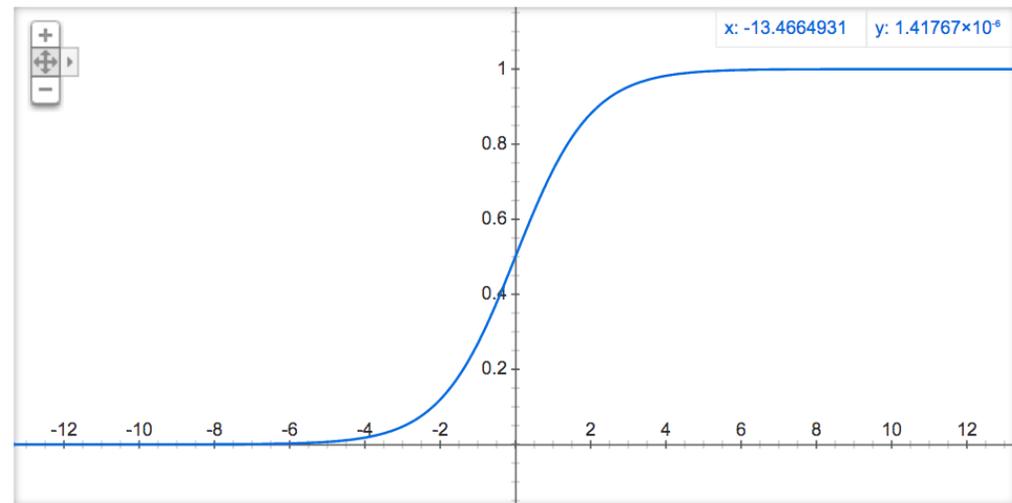
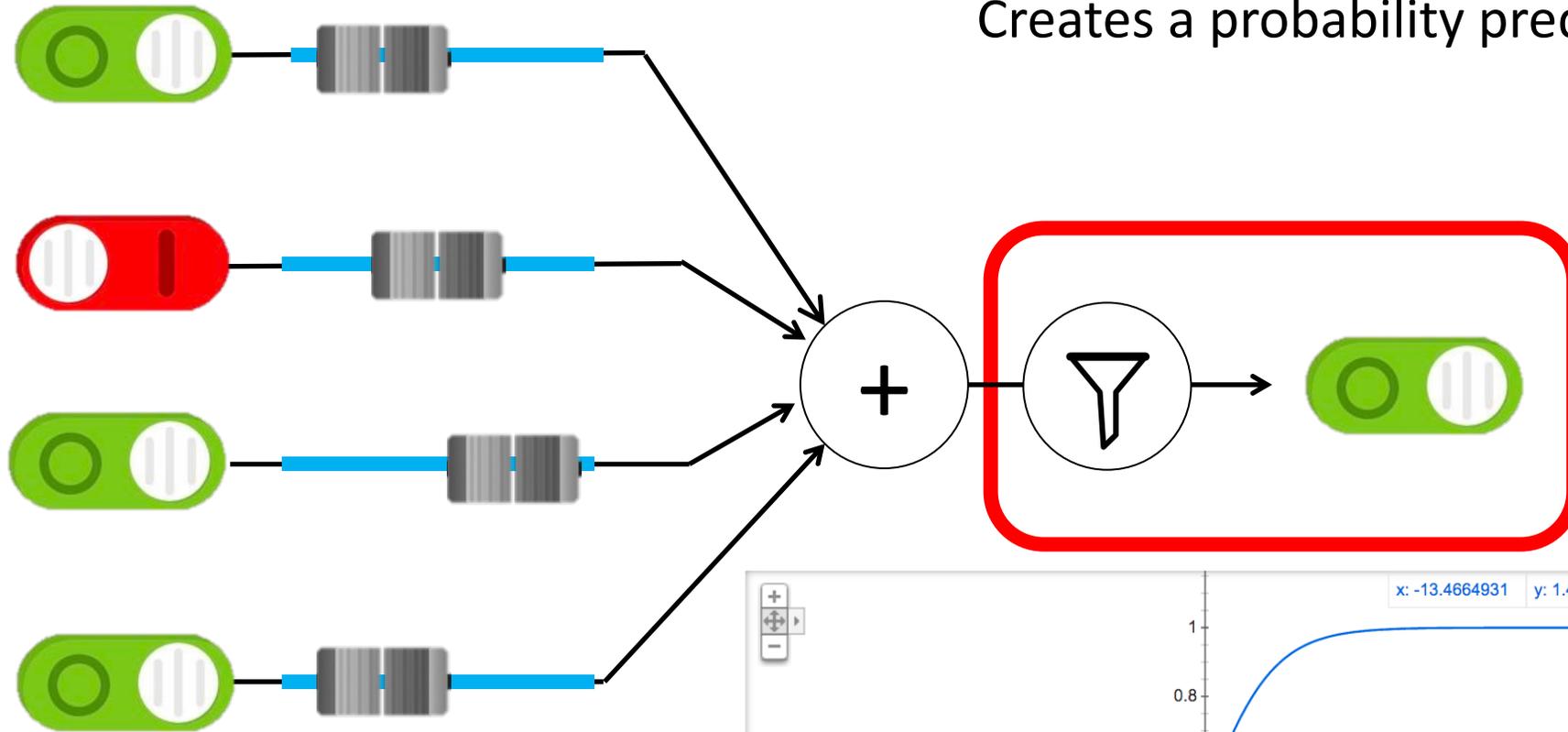


```
double weightedSum = 0;  
weightedSum += input0 * weight0;  
weightedSum += input1 * weight1;  
weightedSum += input2 * weight2;  
weightedSum += input3 * weight3;
```



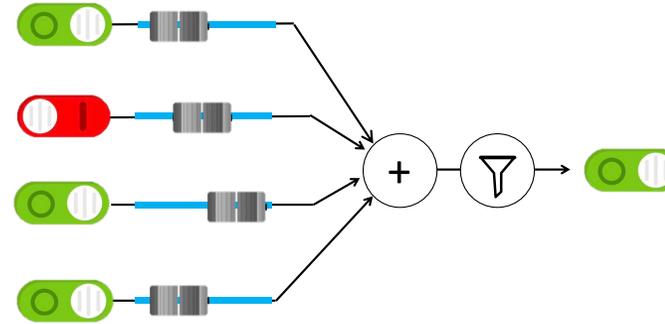
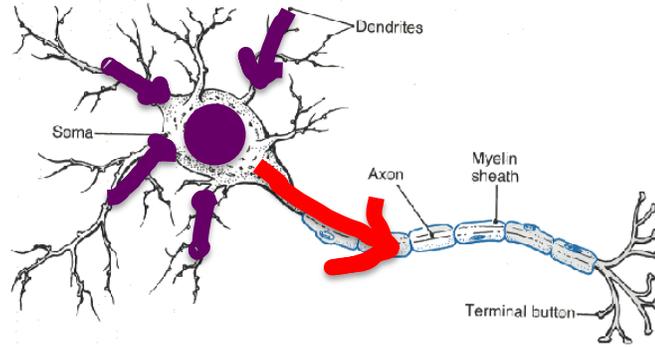
Filter and Output

Creates a probability prediction

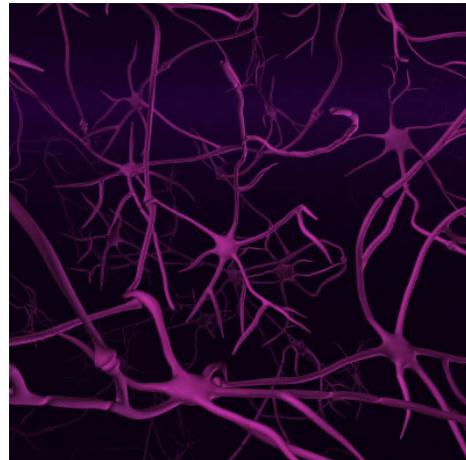


Biological Basis for Neural Networks

- A neuron



- Your brain

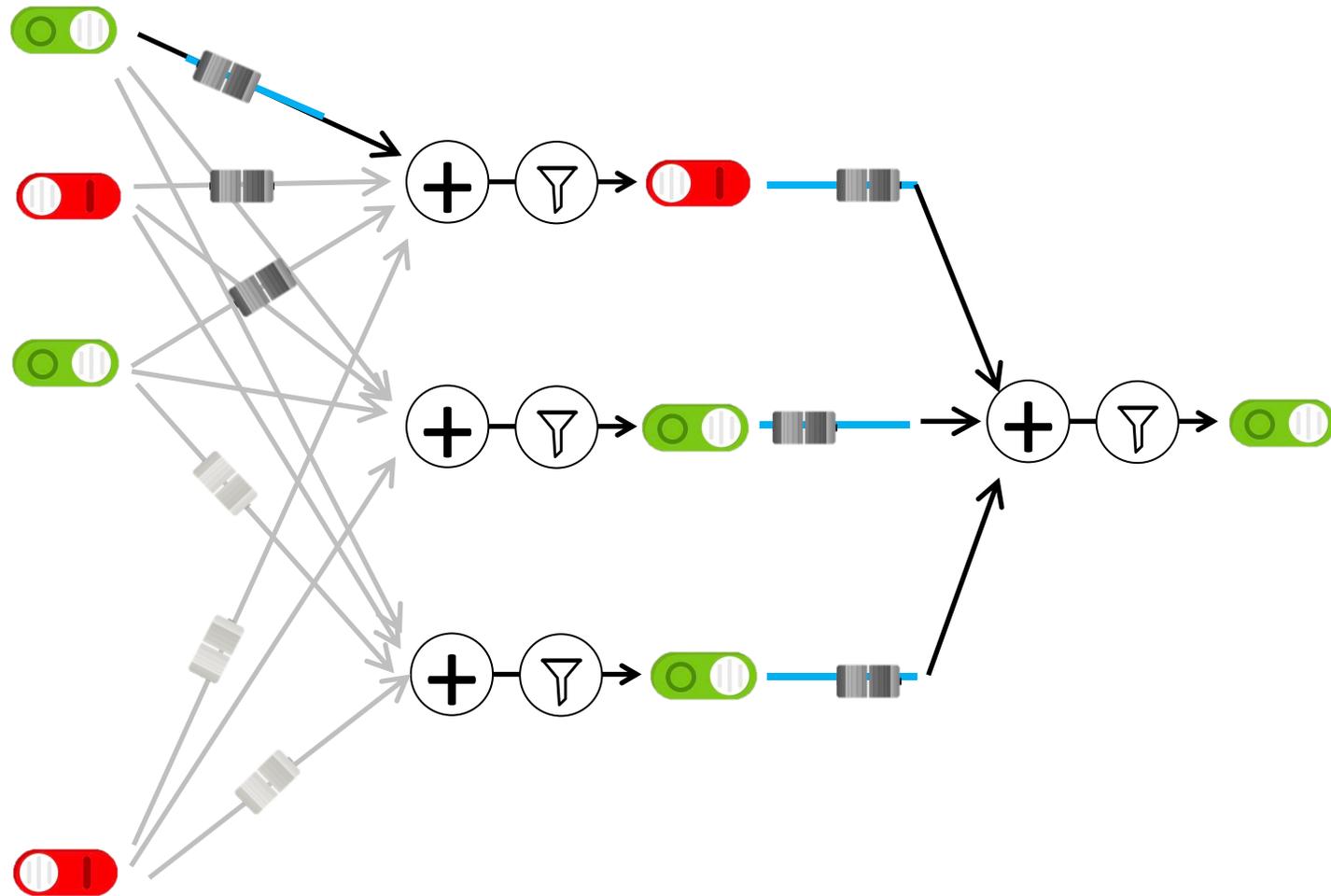


???

Actually, it's probably someone else's brain

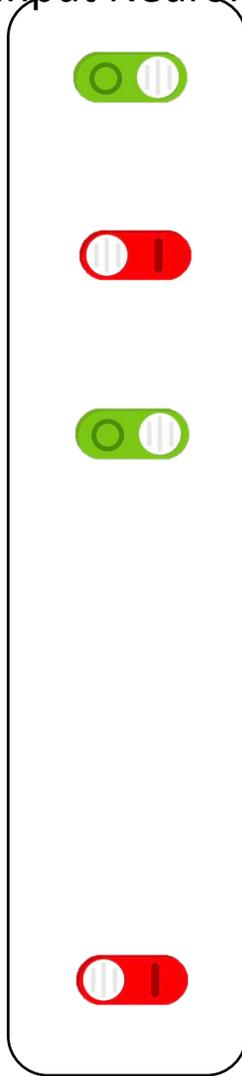


Put Many Together

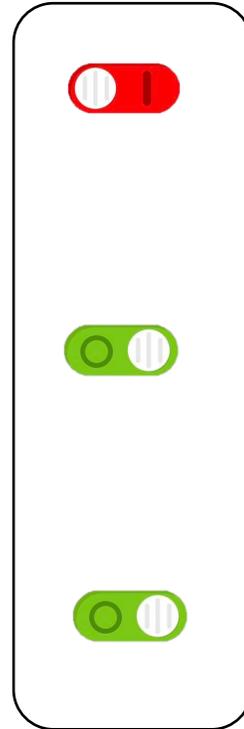


Put Many Together

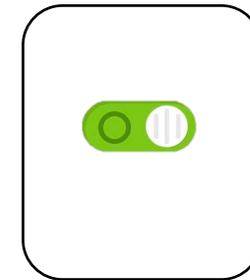
Input Neurons



Hidden Neurons



Output Neurons

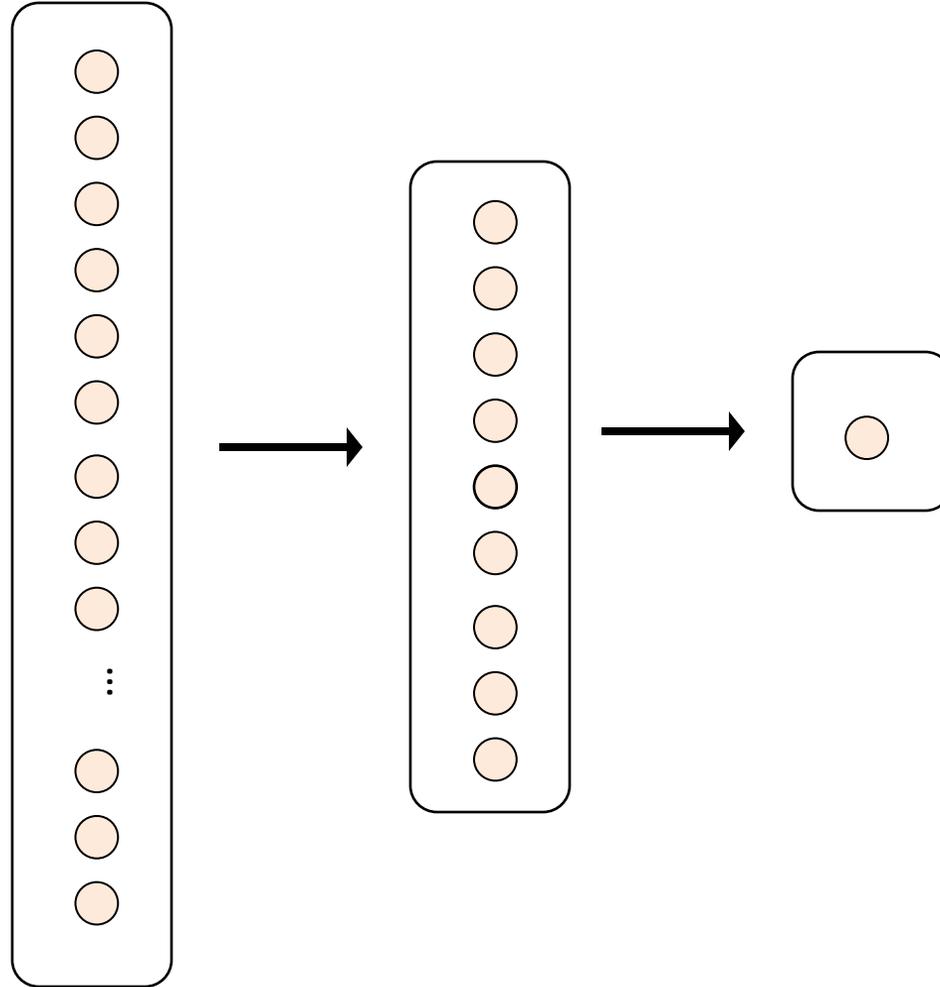


Making a Prediction

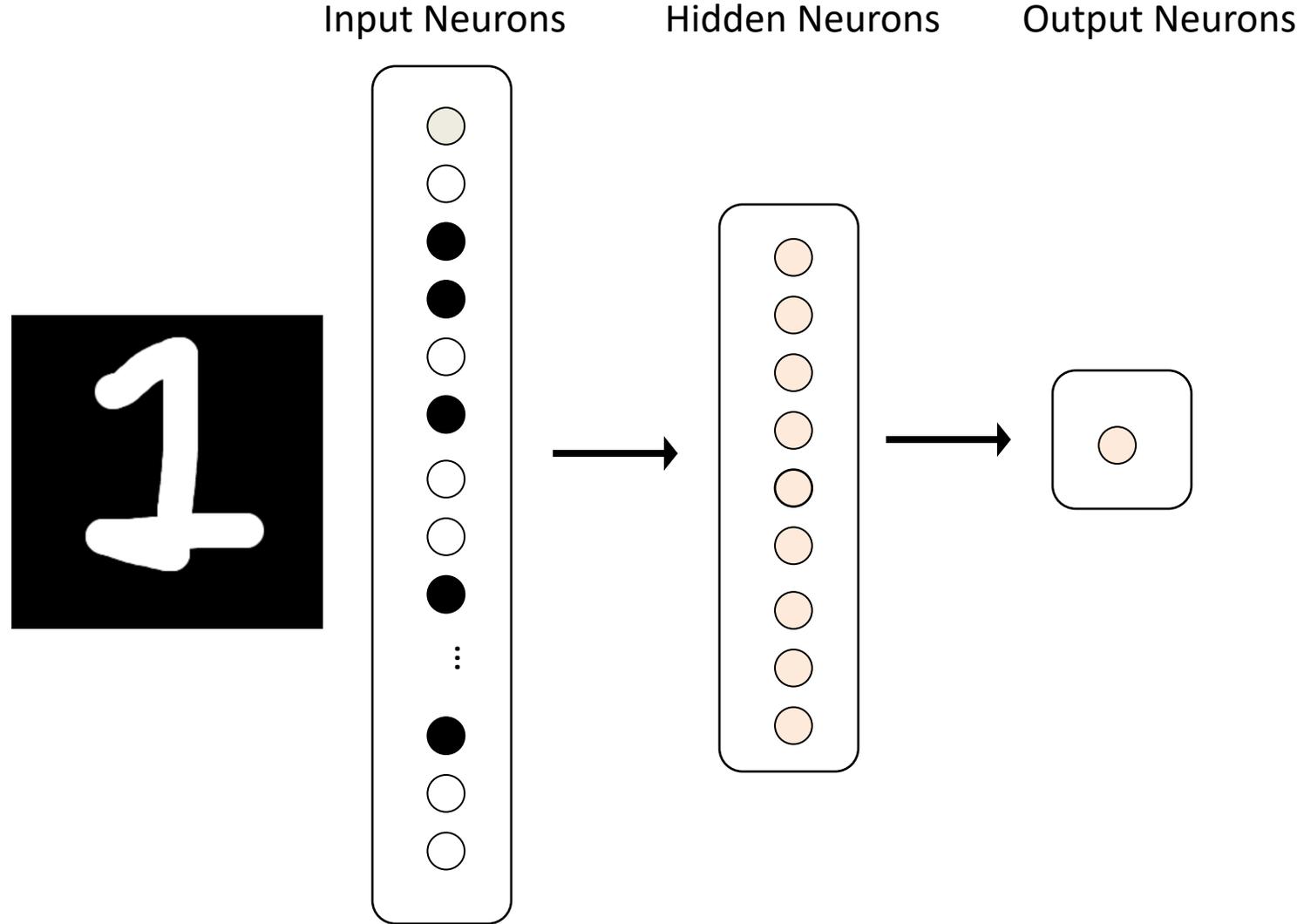
Input Neurons

Hidden Neurons

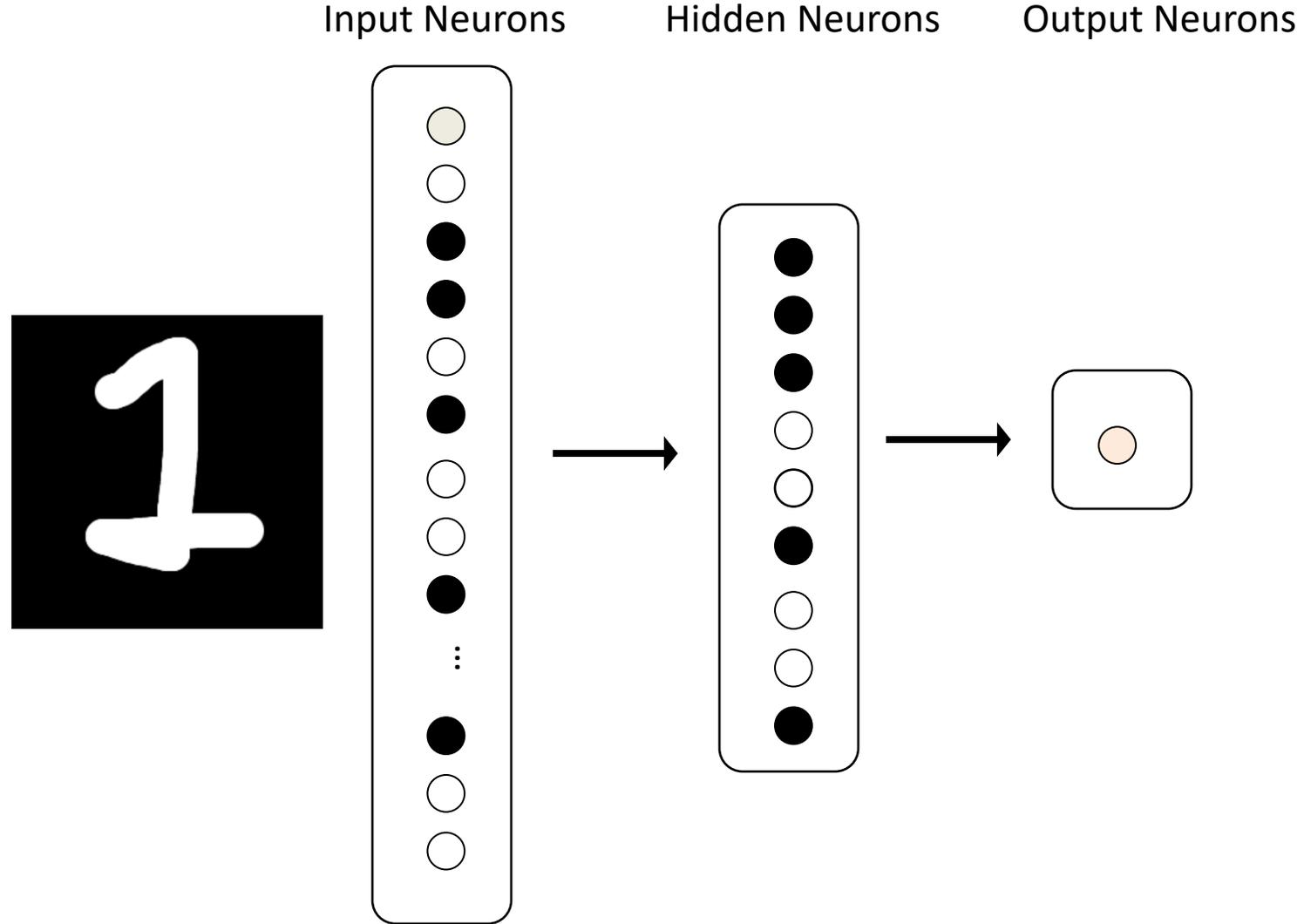
Output Neurons



Making a Prediction



Making a Prediction

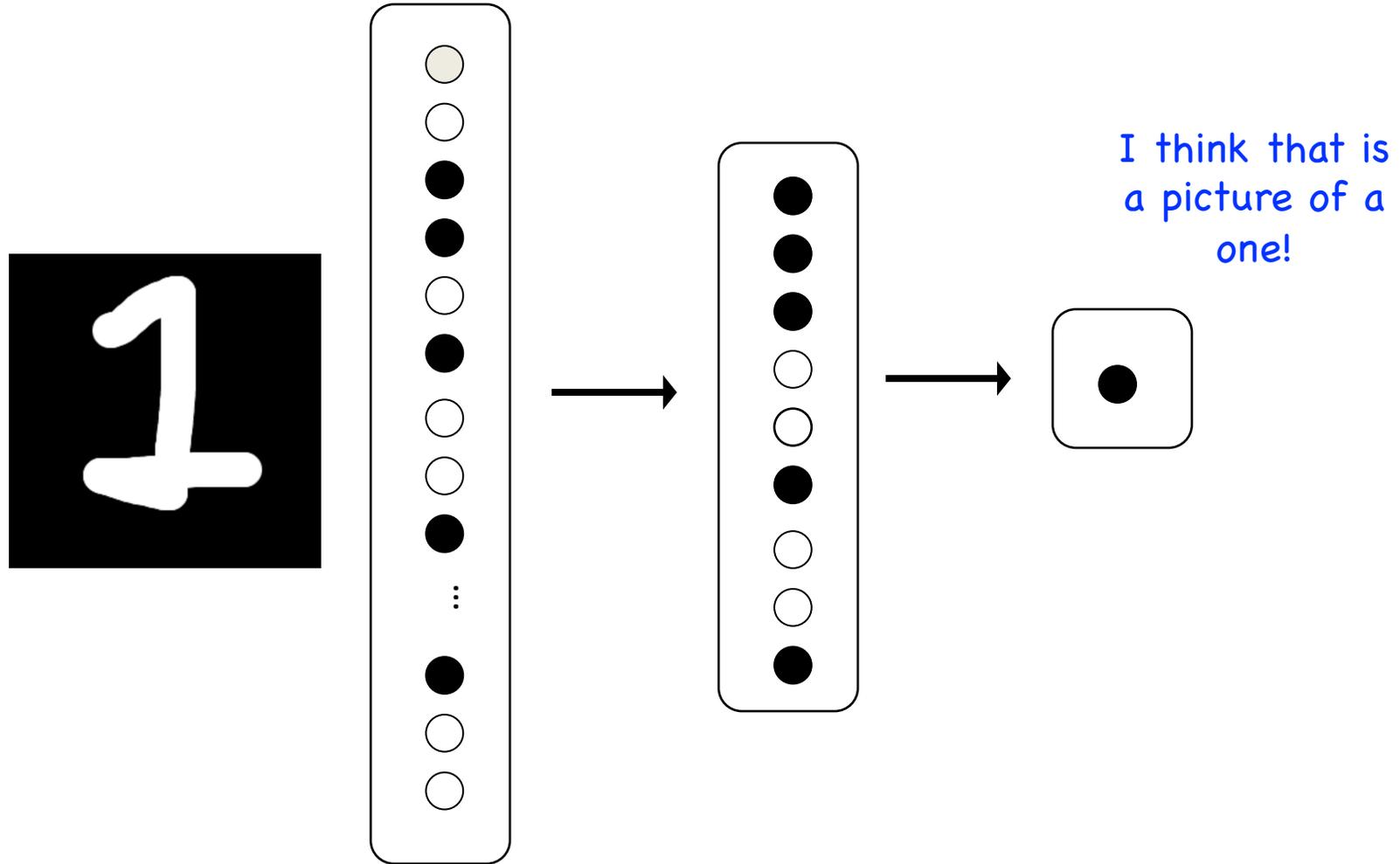


Making a Prediction

Input Neurons

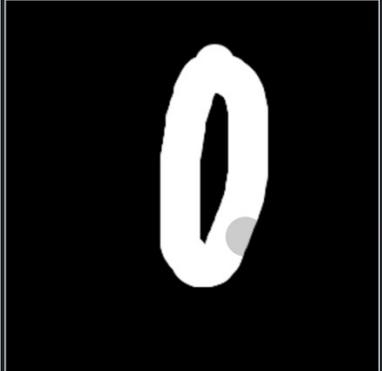
Hidden Neurons

Output Neurons



Demonstration

Draw your number here



X  

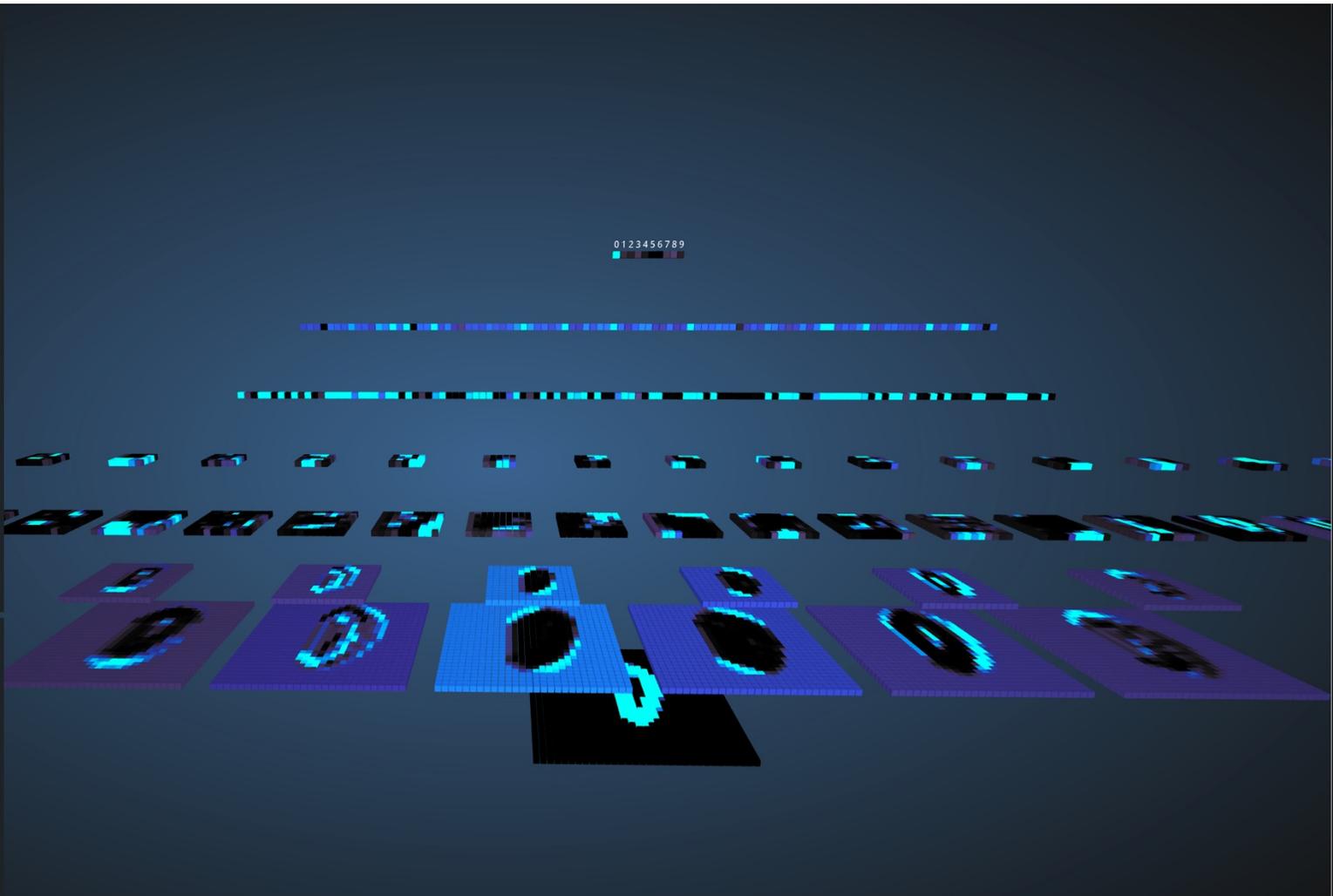
Downsampled drawing:

First guess:

Second guess:

Layer visibility

Input layer	Show
Convolution layer 1	Show
Downsampling layer 1	Show
Convolution layer 2	Show
Downsampling layer 2	Show



<http://scs.ryerson.ca/~aharley/vis/conv/>



Interpret the Output as Prediction

Draw your number here



0 1 2 3 4 5 6 7 8 9



Downsampled drawing:

First guess: 3

Second guess: 3

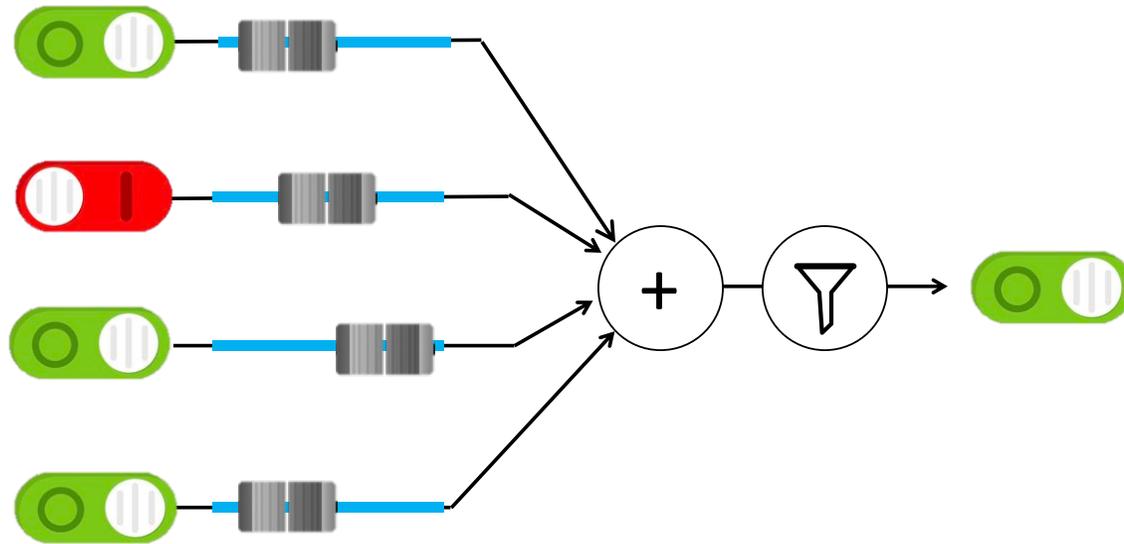
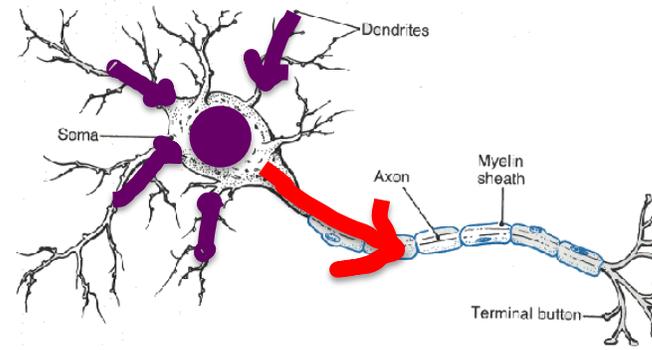
8

Layer visibility

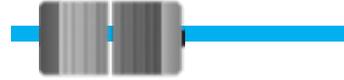
Input layer	Show
Convolution layer 1	Show
Downsampling layer 1	Show
Convolution layer 2	Show



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

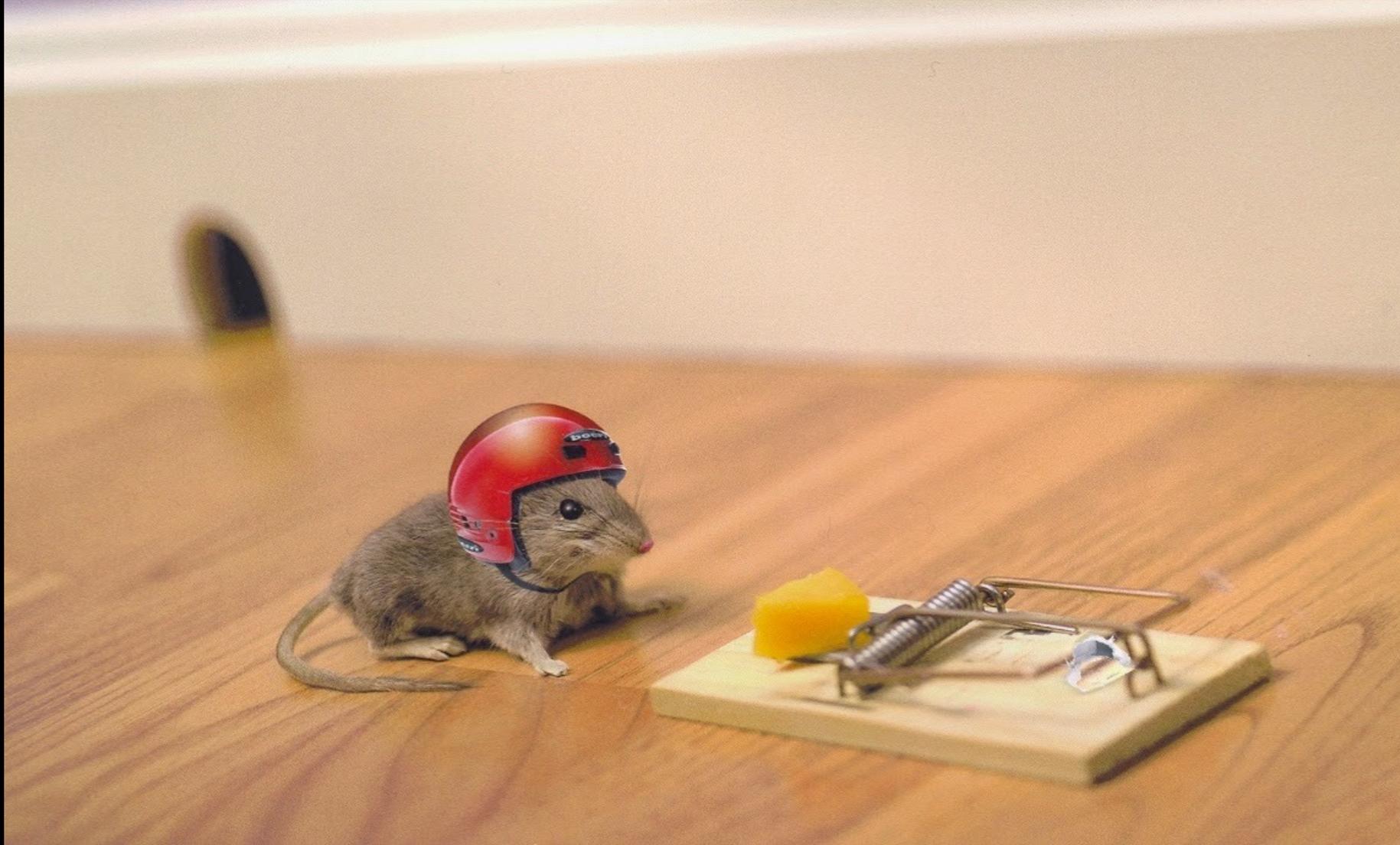
2. Learn by Example

Two Great Ideas

1. Artificial Neurons

2. Learn by Example

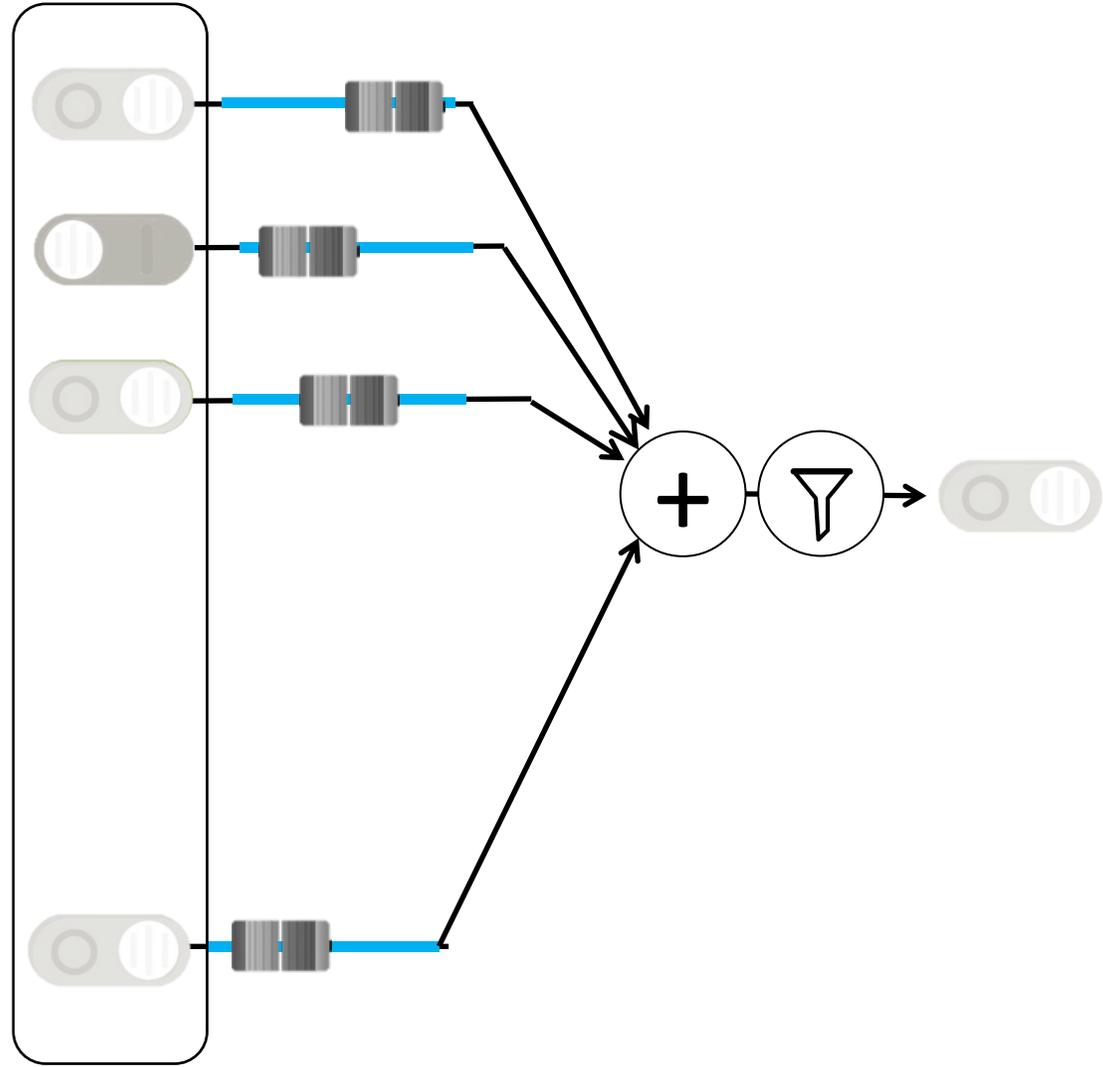
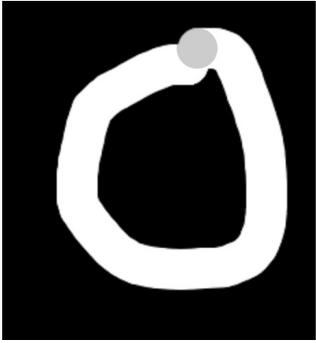
2. Learn From Experience

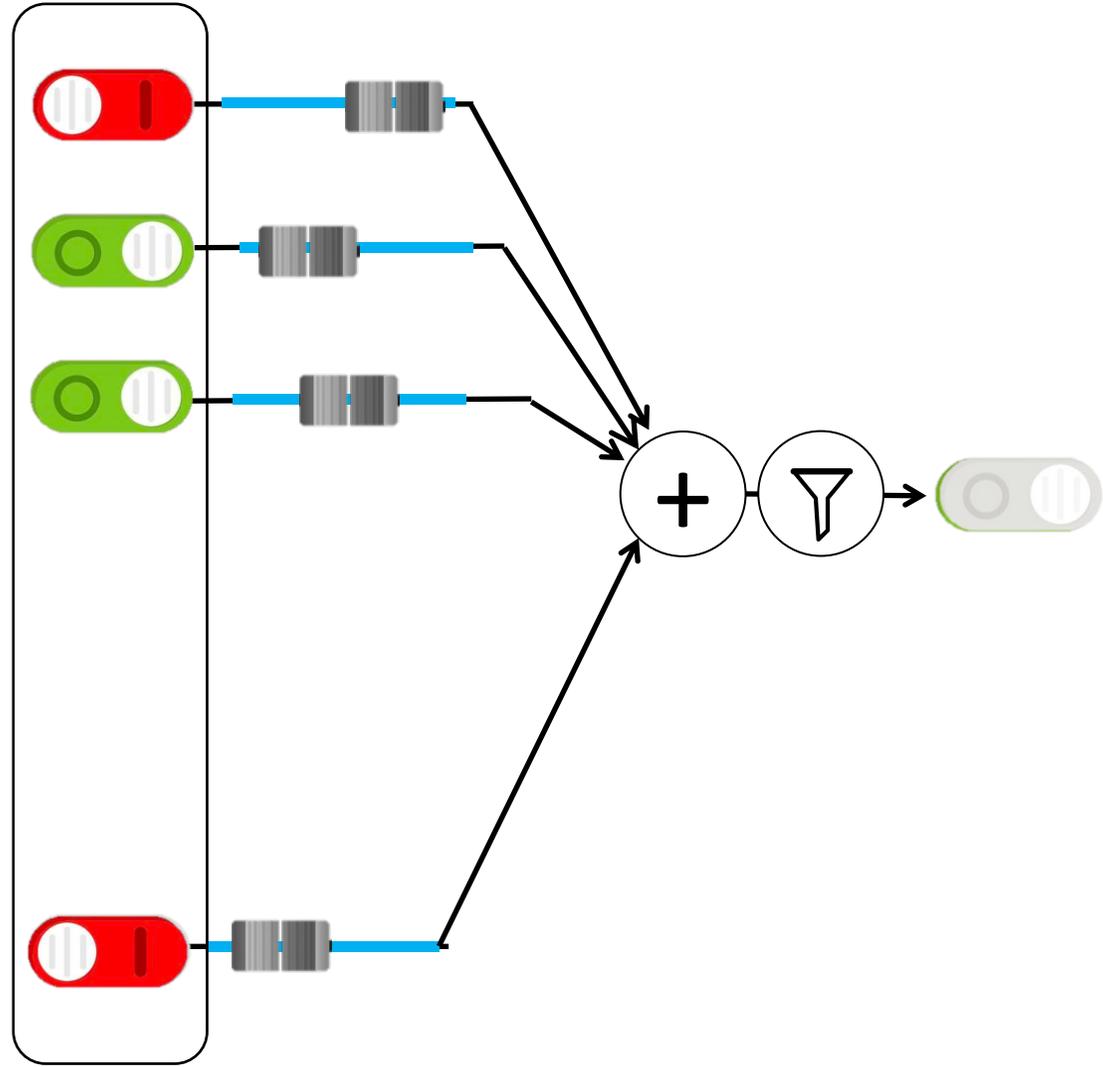
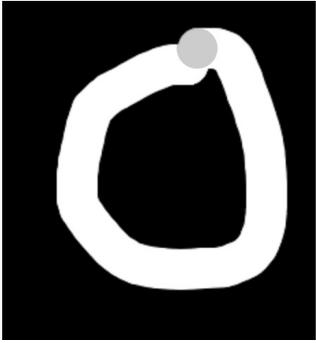


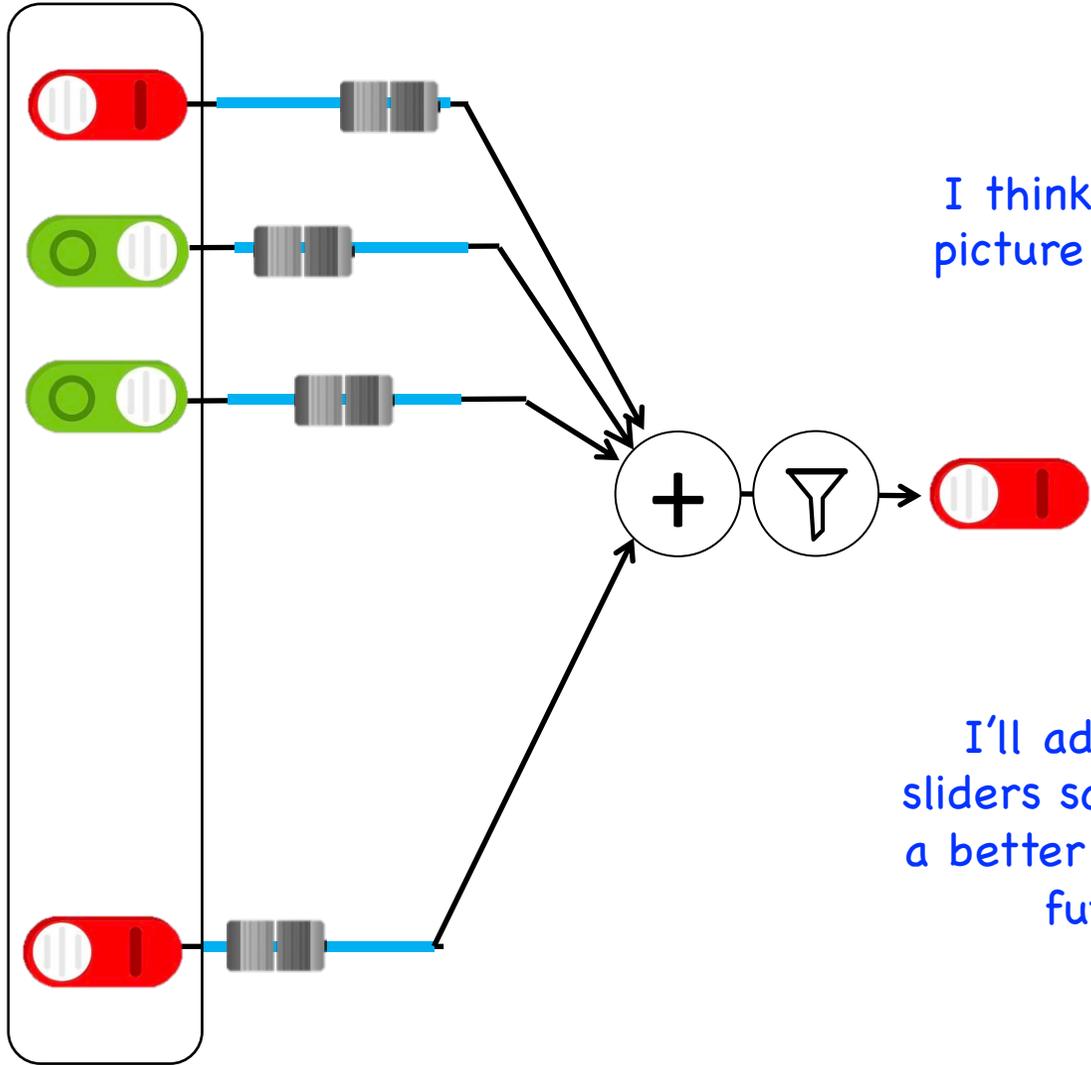
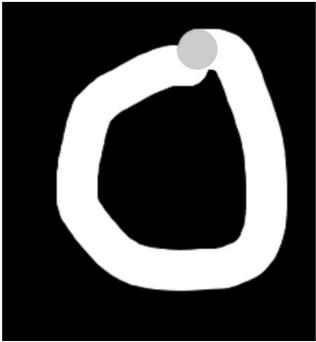
Learn by Example

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
3 3 3 3 3 3 3 3 3 3 3 3 3 3 3
4 4 4 4 4 4 4 4 4 4 4 4 4 4 4
5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
6 6 6 6 6 6 6 6 6 6 6 6 6 6 6
7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
8 8 8 8 8 8 8 8 8 8 8 8 8 8 8
9 9 9 9 9 9 9 9 9 9 9 9 9 9 9







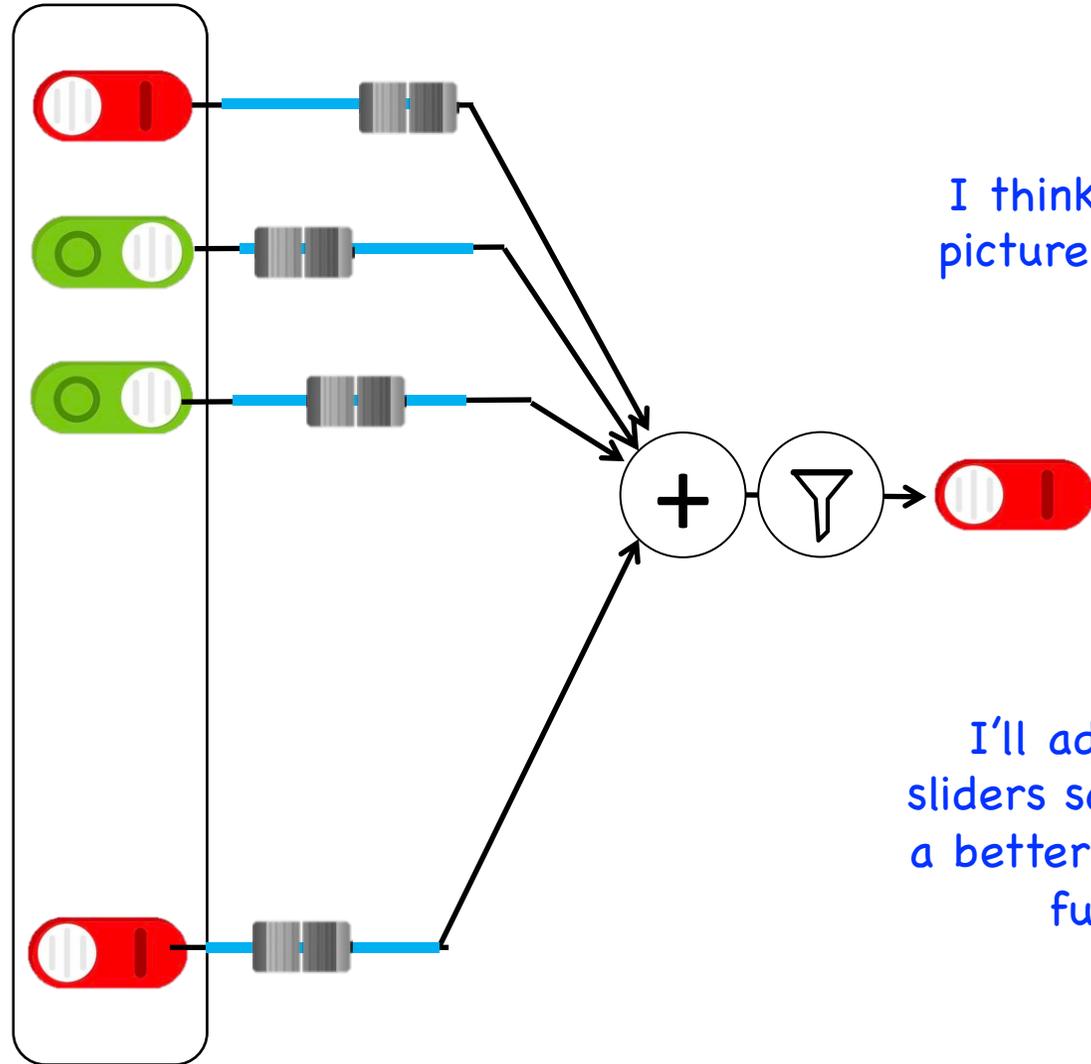
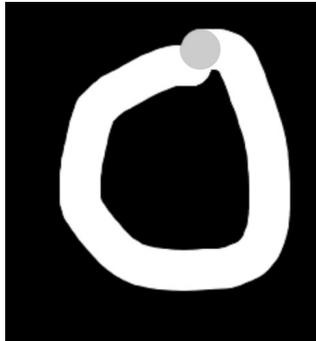


I think that is a picture of a **One!**

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

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



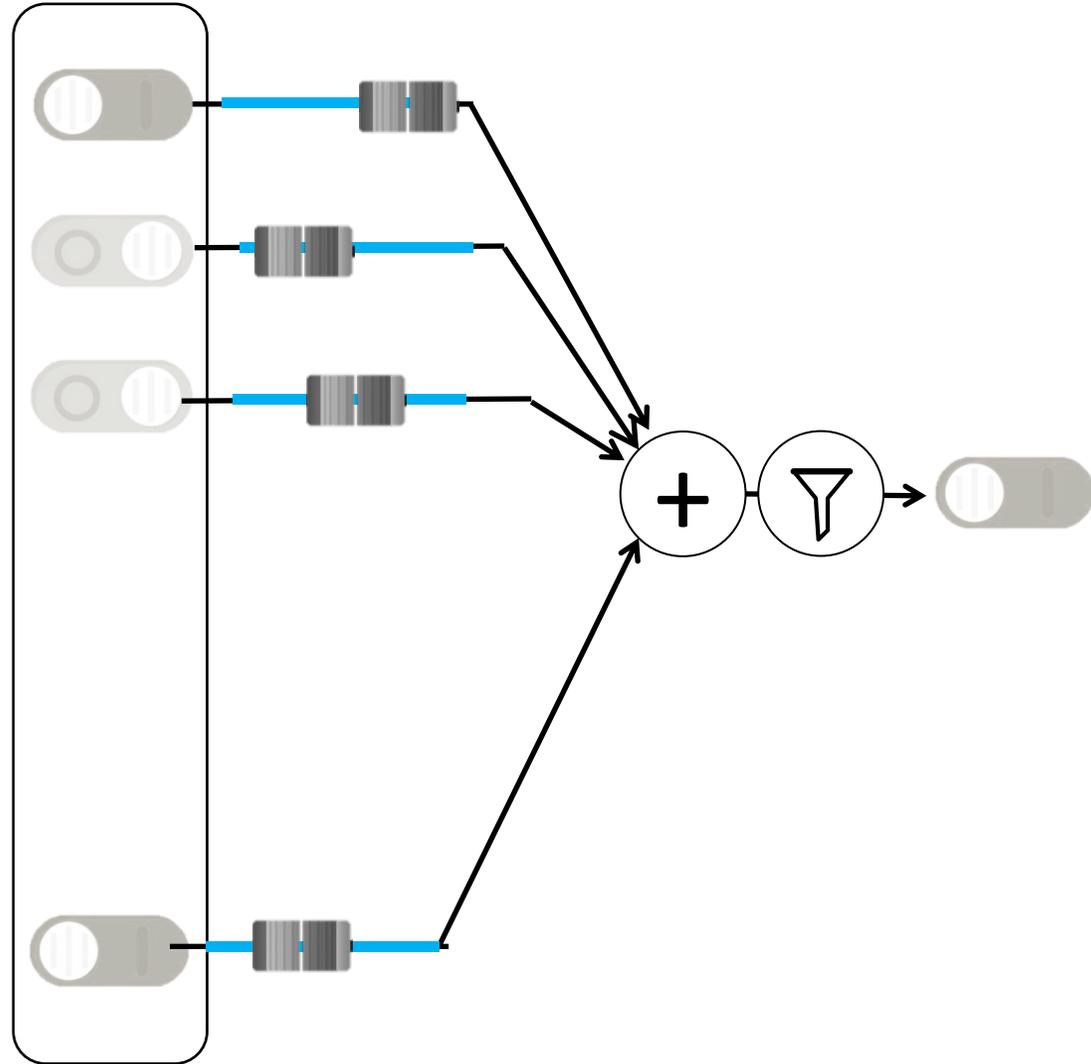


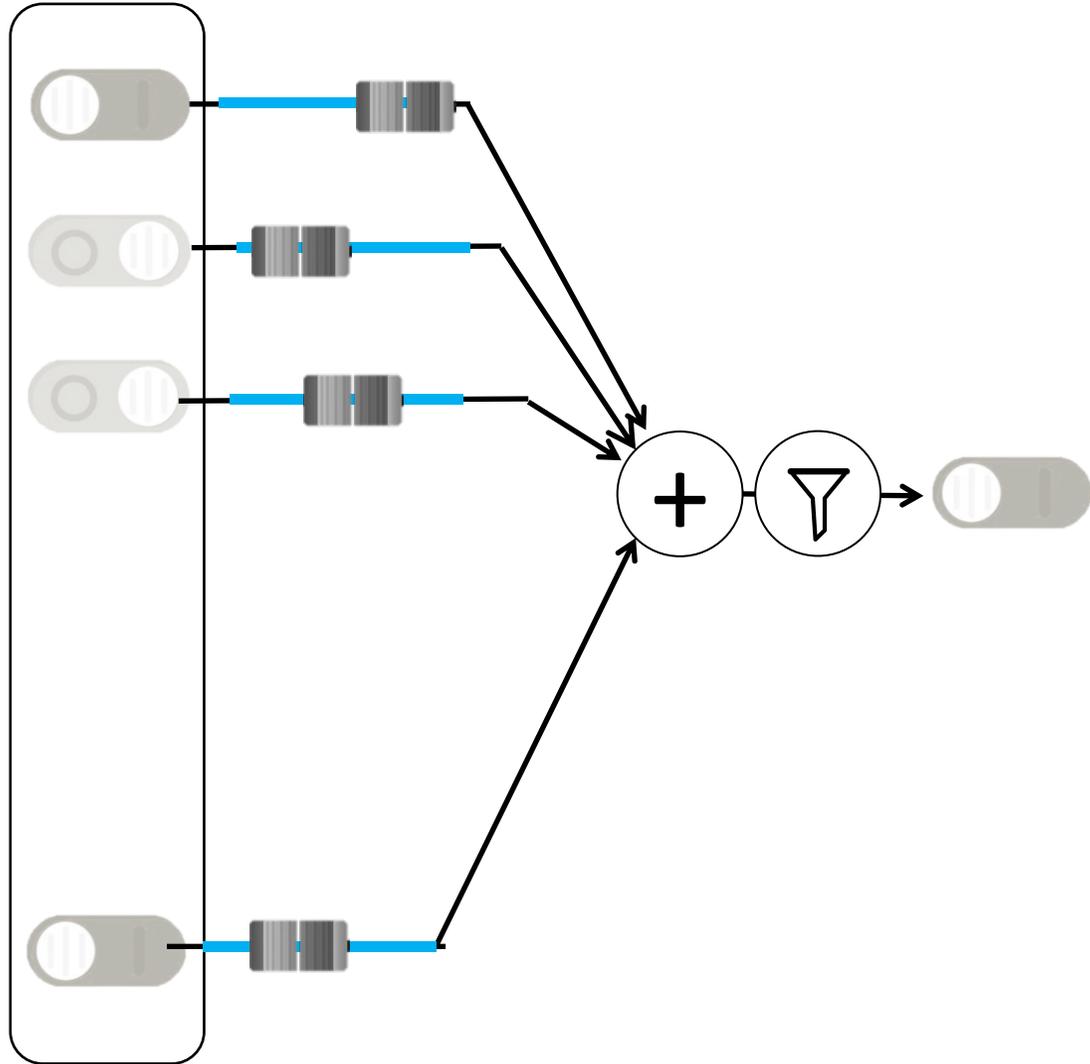
I think that is a picture of a **One!**

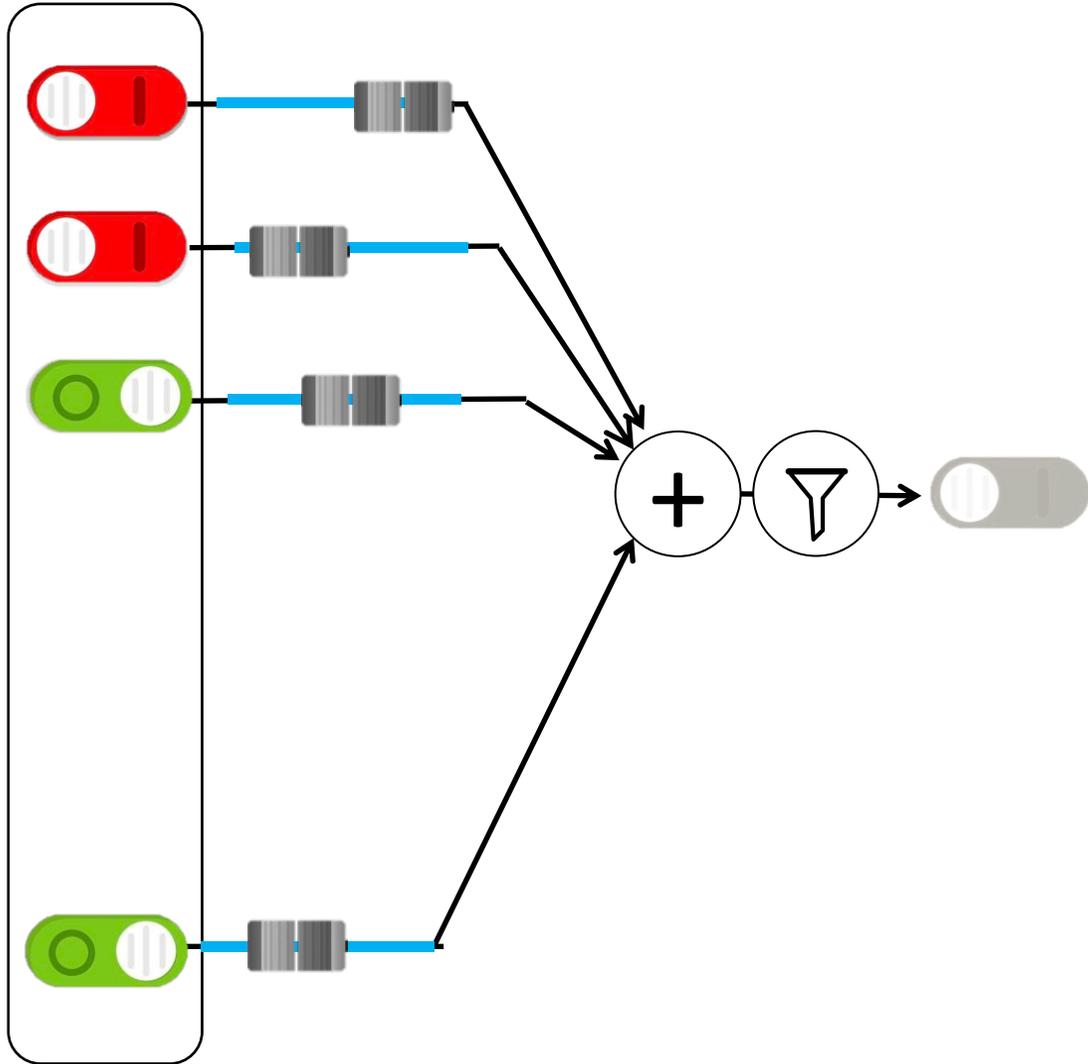
What do you mean it's actually a **Zero**?

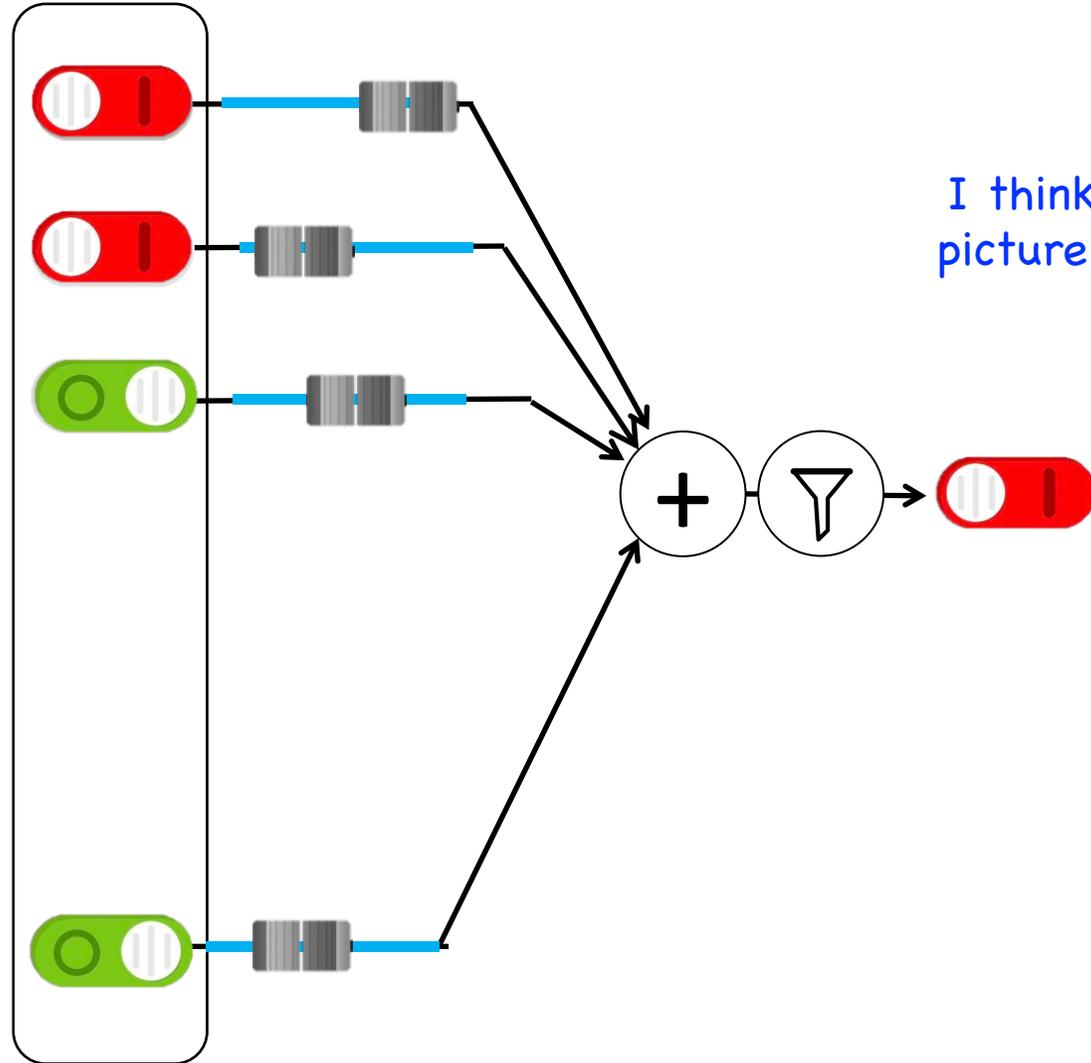
I'll adjust my sliders so that I do a better job in the future







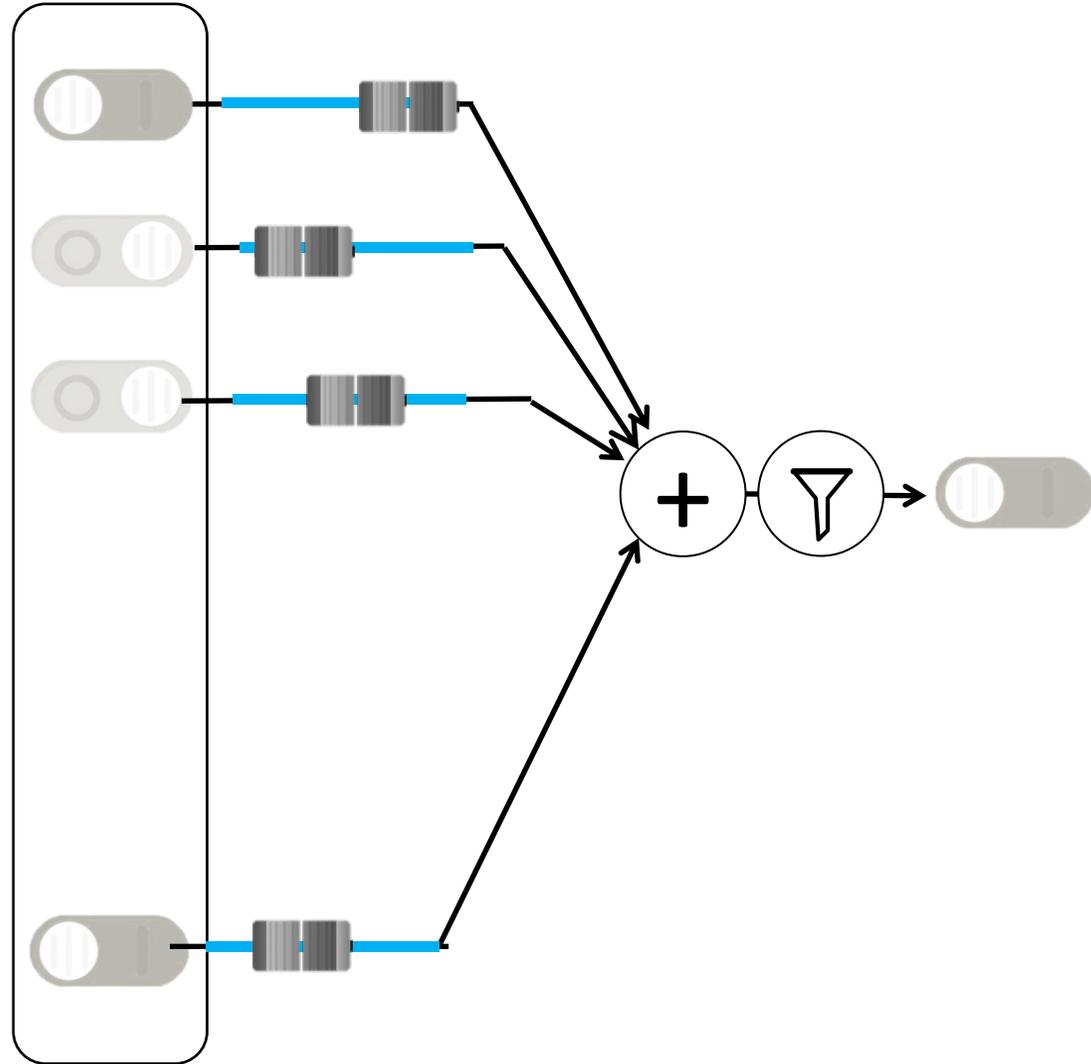


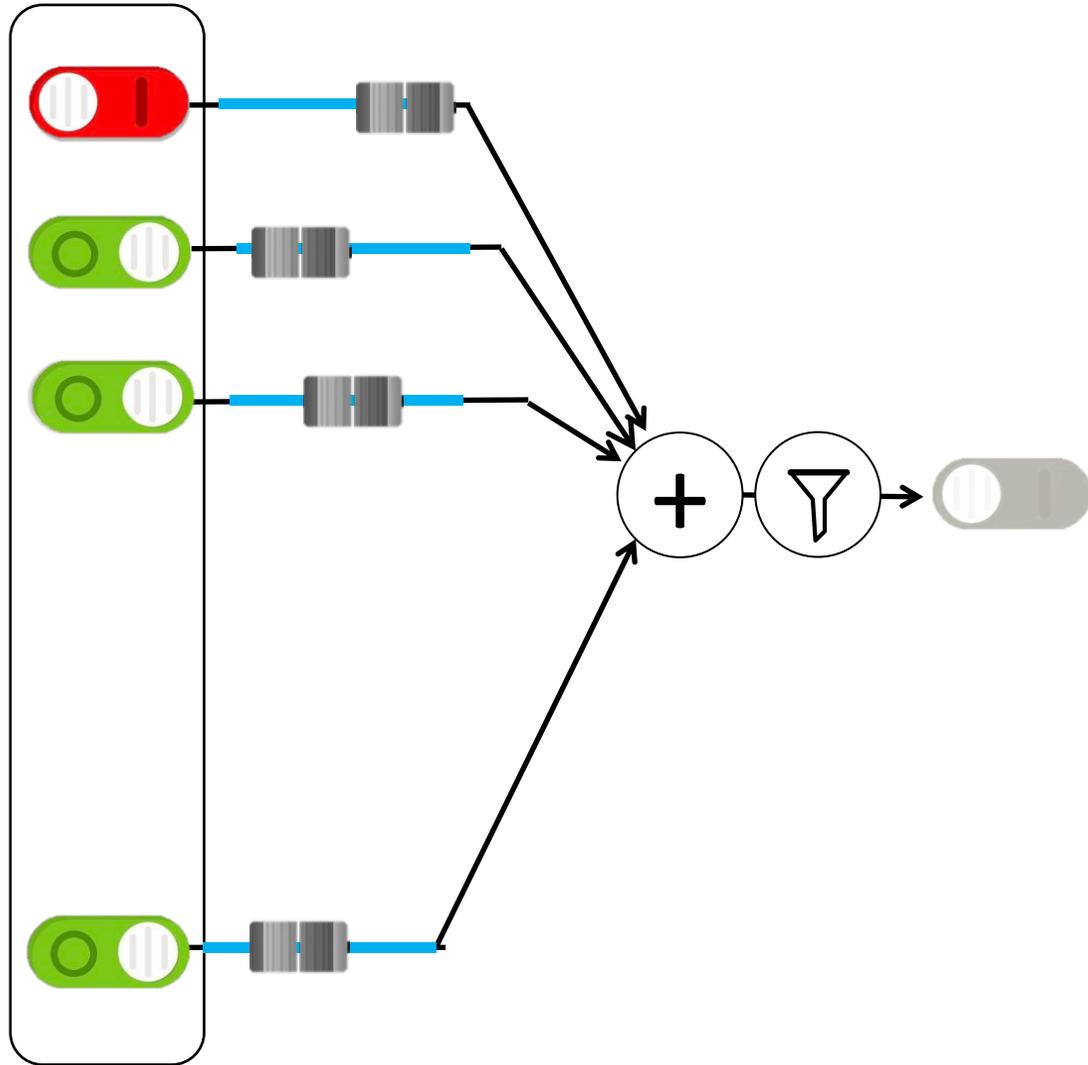


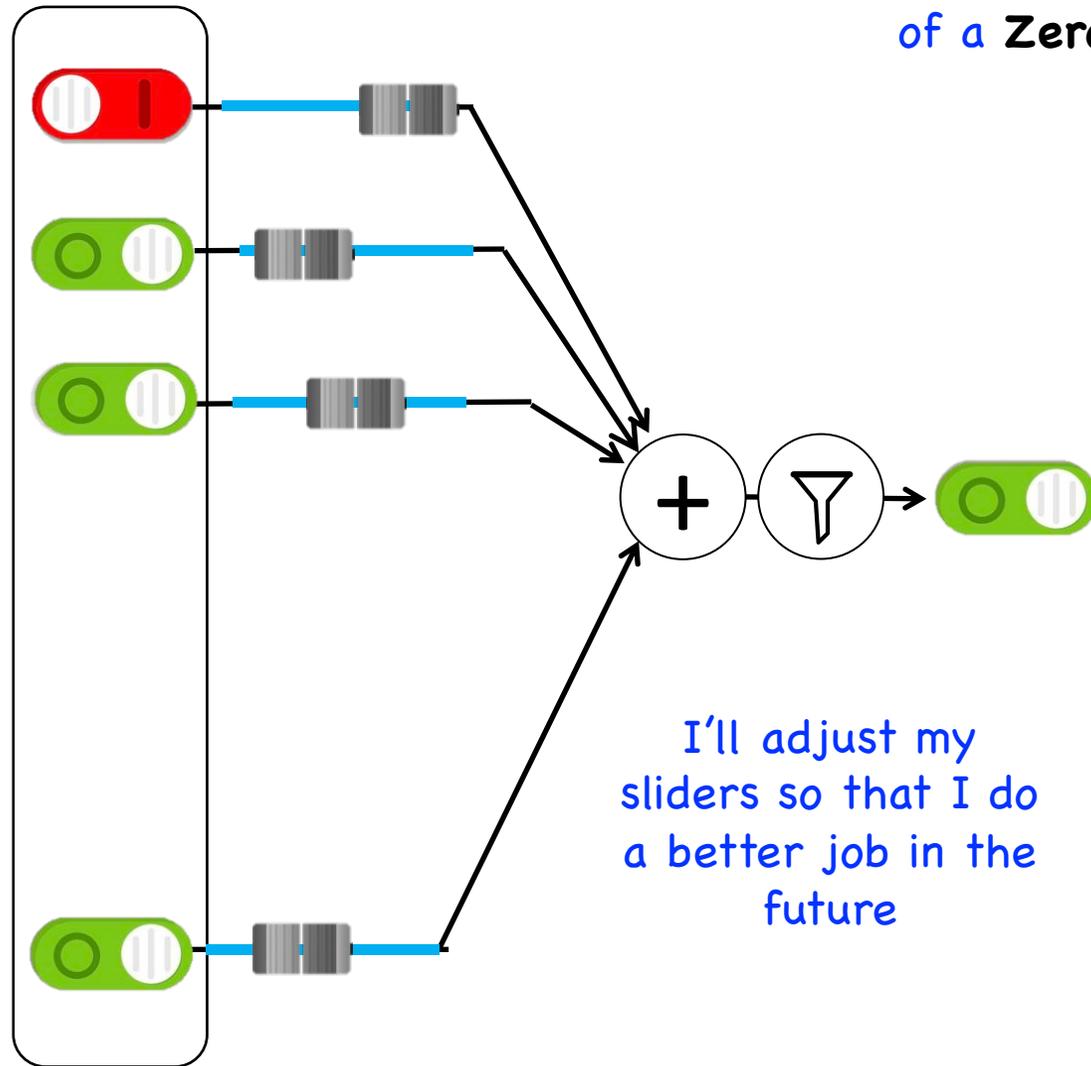
I think that is a picture of a **One!**

Wahoo I got it right!







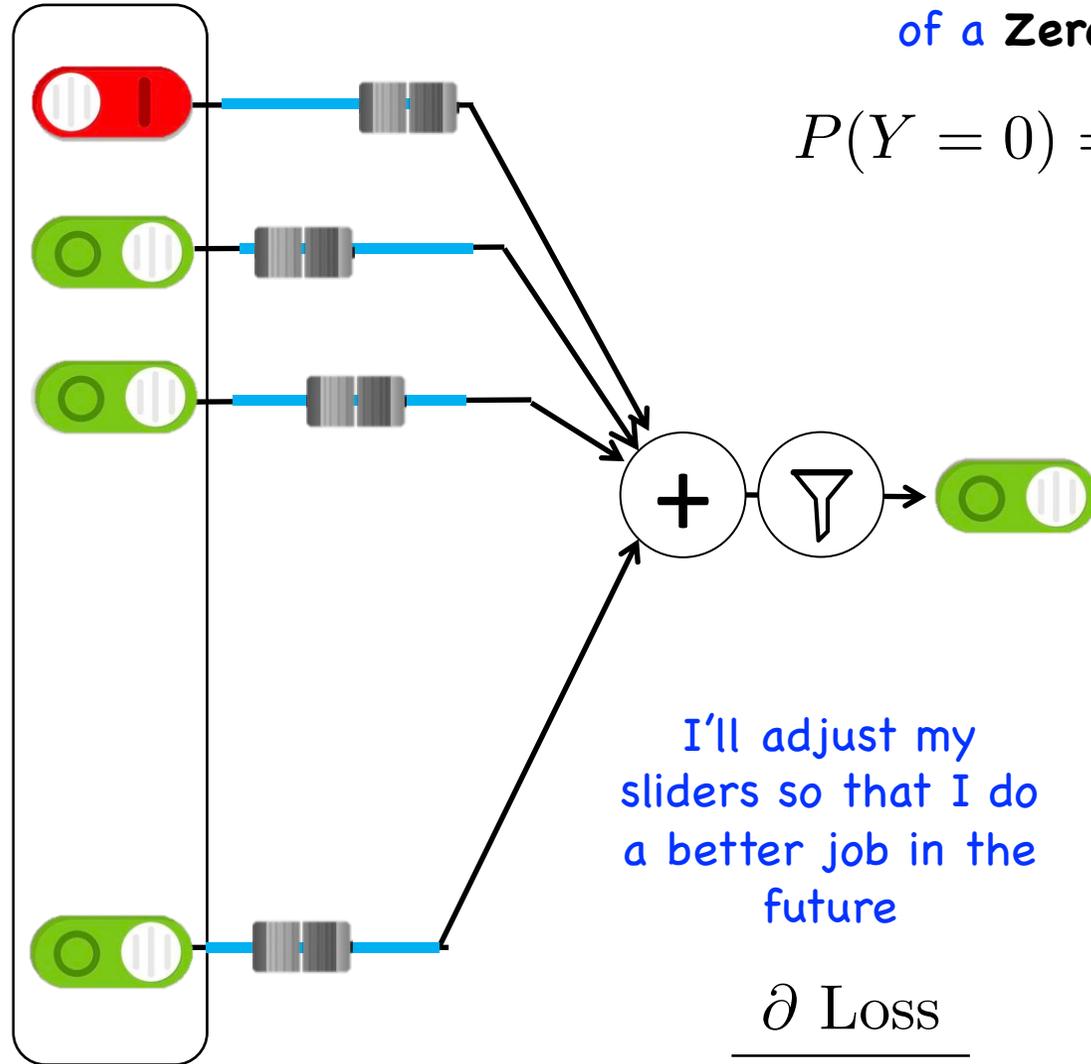


I think that is a picture of a **Zero**!

But it is actually a **Zero**

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





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

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

But it is
actually a **Zero**

$$\text{Loss} = 1$$

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

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



Update Neural Network

$$P(Y = 1|X = \mathbf{x}) = \hat{y} \qquad \hat{y} = \sigma \left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})} \right)$$

For one datum

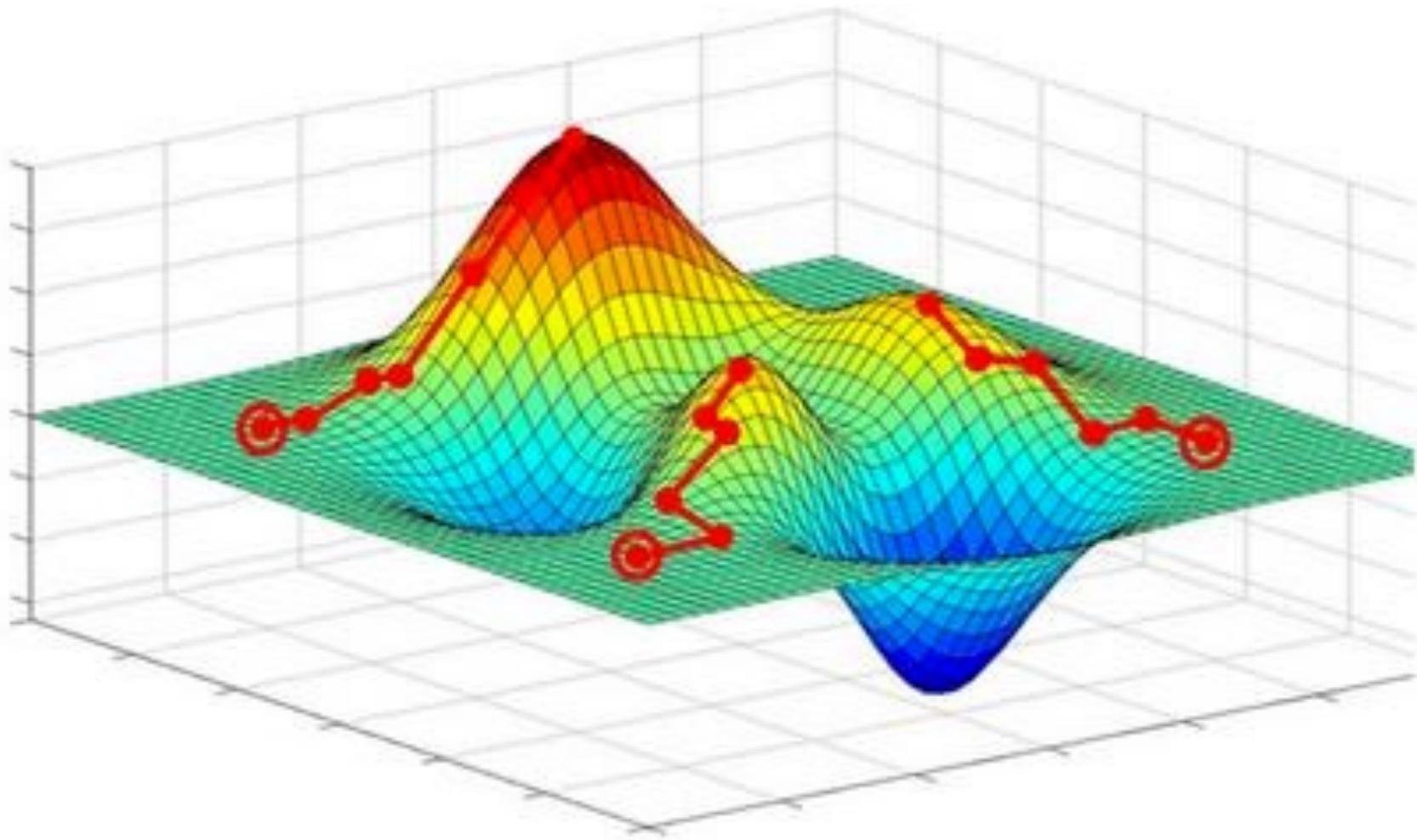
$$P(Y = y|X = \mathbf{X}) = (\hat{y})^y (1 - \hat{y})^{1-y}$$

For IID data

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n P(Y = y^{(i)} | X = \mathbf{x}^{(i)}) \\ &= \prod_{i=1}^n (\hat{y}^{(i)})^{y^{(i)}} \cdot \left[1 - (\hat{y}^{(i)}) \right]^{(1-y^{(i)})} \end{aligned}$$



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

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

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

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

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

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

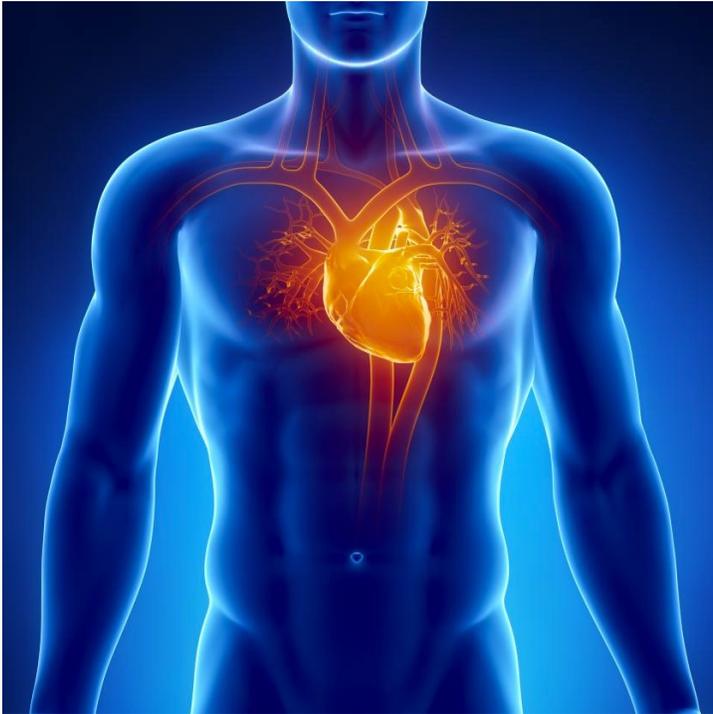
You will be able to do this.



Where you will be by the end of class

CS109: Theory Class focused on Applications

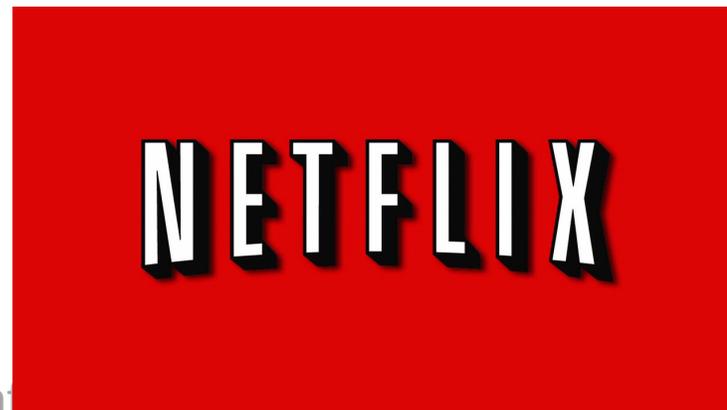
Heart



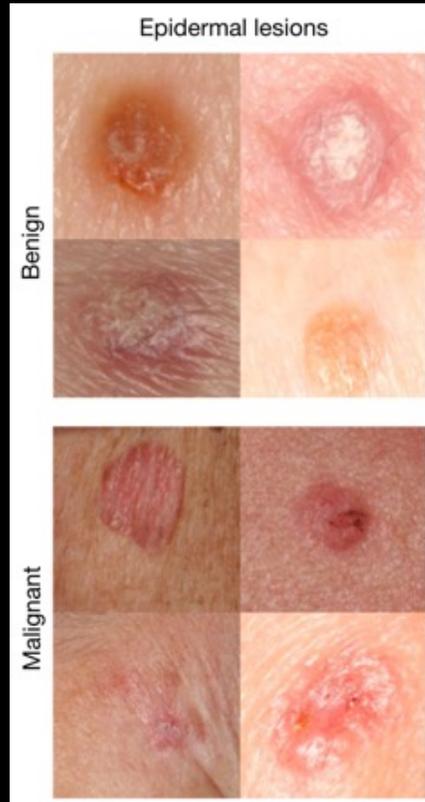
Ancestry



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 convolutional neural networks." *Nature* 542.7639 (2017): 115-118.

Not once, but twice, AI was 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



Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.

Human-level concept learning through probabilistic program induction.



४	५	६	७	८
९	०	१	२	३

Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory



Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.

Human-level concept learning through probabilistic program induction.

ಬ

ಹ	ಬ	ಗ	ಜ	ಝ
ಞ	ಠ	ಱ	ಠ	ಢ

Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory



Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.

Human-level concept learning through probabilistic program induction.

ಬ

ಹ	ಬ	ಗ	ಜ	ಝ
ಞ	ಠ	ಱ	ಠ	ಢ

Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory



Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.

Human-level concept learning through probabilistic program induction.

பு

ஊ	உ	பு	பு	பு
கூ	கூ	கூ	கூ	கூ

Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory



Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.

Human-level concept learning through probabilistic program induction.

ಬ

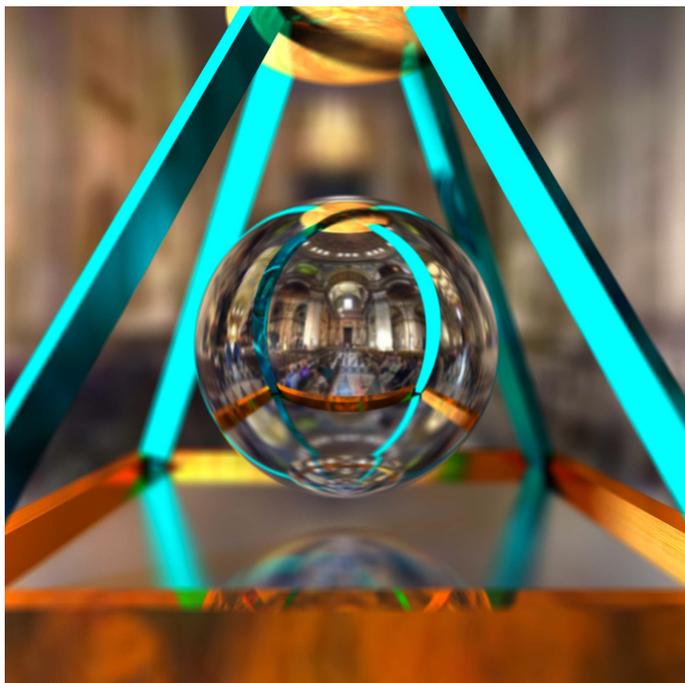
ಹ	ಖ	ಗ	ಬ	ಝ
಼	ಠ	ಞ	ತ	ದ

Current deep learning methods are not enough to move the needle as far as we want, **especially on socially relevant problems** that often do not have the benefit of massive public datasets. The best new ideas are coming from probability theory

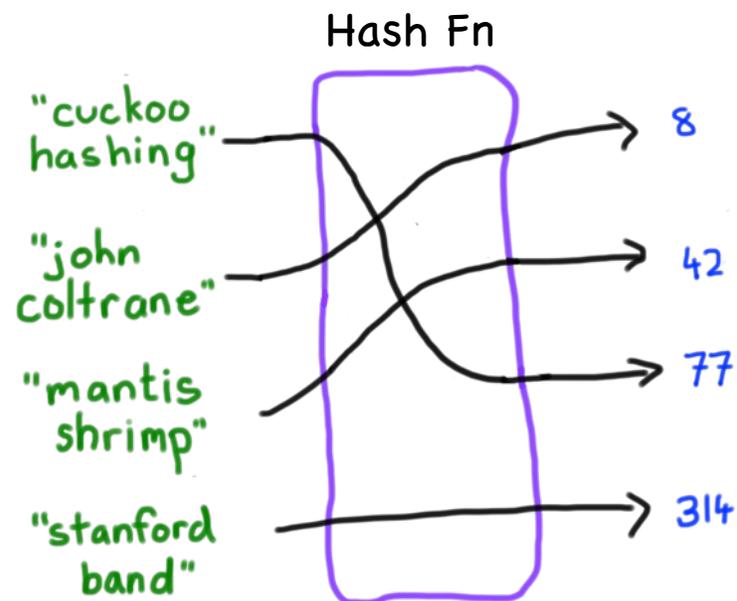


Algorithms and Probability

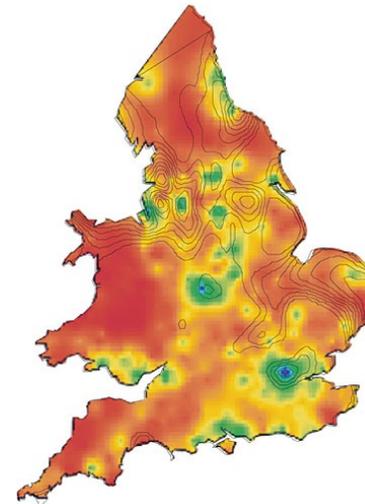
Eg Raytracing



Eg HashMaps



Understanding the world and building tools



Autocomplete



dinosaurs we

[Advanced Search](#)
[Language Tools](#)

- dinosaurs **websites for kids**
- dinosaurs **we're back**
- dinosaurs **webcomic**
- dinosaurs **webquest**
- dinosaurs **were made up by the cia to discourage time travel**
- dinosaurs **website**
- dinosaurs **went extinct**
- dinosaurs **weight**
- dinosaurs **we are scientists**
- dinosaurs **weed episode**



Recommender Systems

amazon.com Hello. Sign in to get [personalized recommendations](#). New customer? [Start here](#). **FREE 2-Day Shipping, No Minimum Purchase**

Your Amazon.com Today's Deals Gifts & Wish Lists Gift Cards Your Account | Help

Shop All Departments Search Books GO 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
[Add to Shopping Cart](#)
or
[Sign in](#) to turn on 1-Click ordering.
or
[Add to Cart with FREE Two-Day Shipping](#)
Amazon Prime Free Trial required. Sign up when you check out. [Learn More](#)

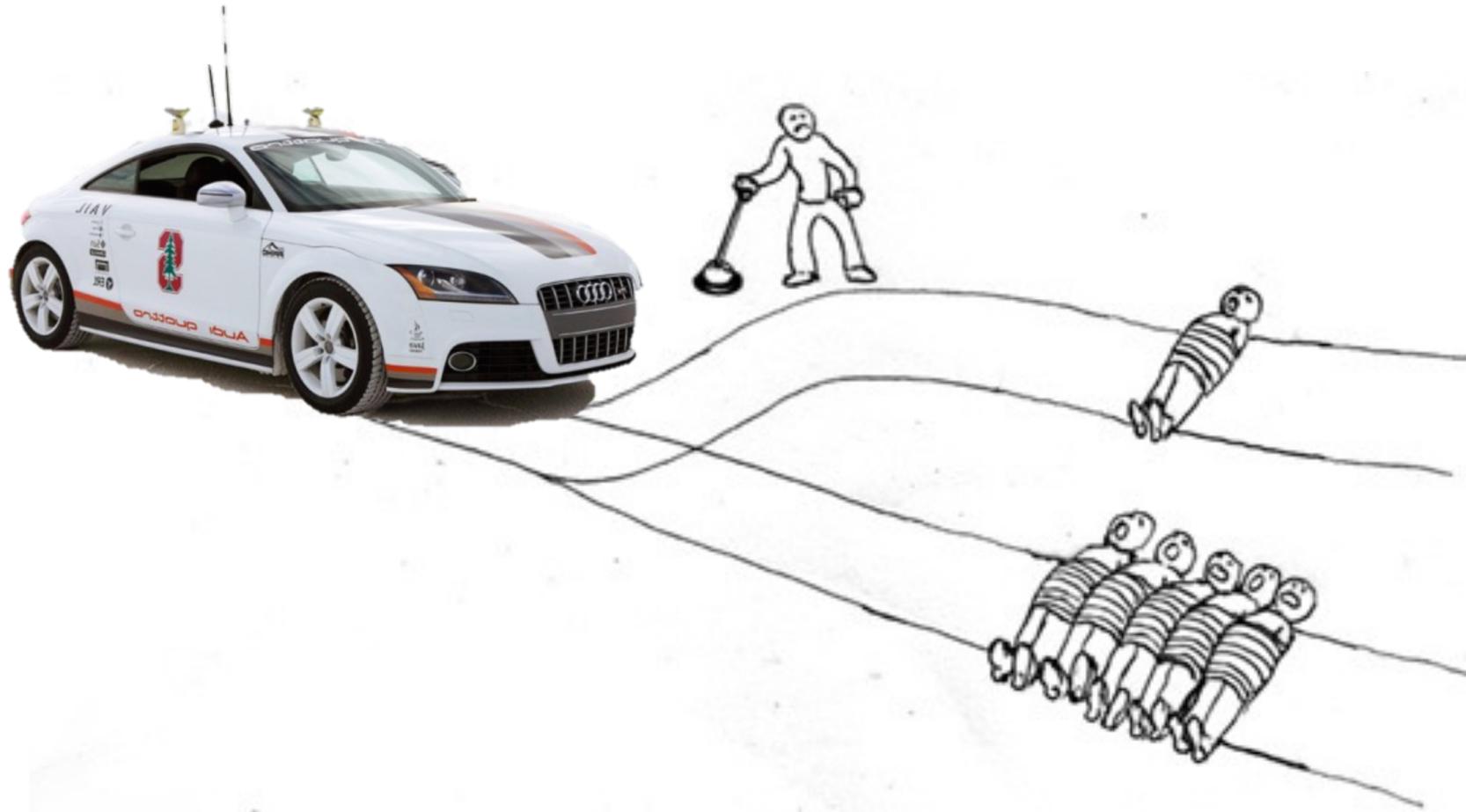
Customers Who Bought This Item Also Bought

Page 1 of 20

- Harry Potter and the Prisoner of Azkaban (Book 3)** by J.K. Rowling
★★★★☆ (2,599) \$16.49
- Harry Potter and the Goblet of Fire (Book 4)** by J.K. Rowling
★★★★☆ (5,186) \$19.79
- Harry Potter and the Order of the Phoenix (Book 5)** by J. K. Rowling
★★★★☆ (5,876) \$10.18
- Harry Potter and the Half-Blood Prince (Book 6)** by J.K. Rowling
★★★★☆ (3,597) \$10.18
- The Tales of Beedle the Bard, Collector's Ed...** by J. K. Rowling
★★★★☆ (176)

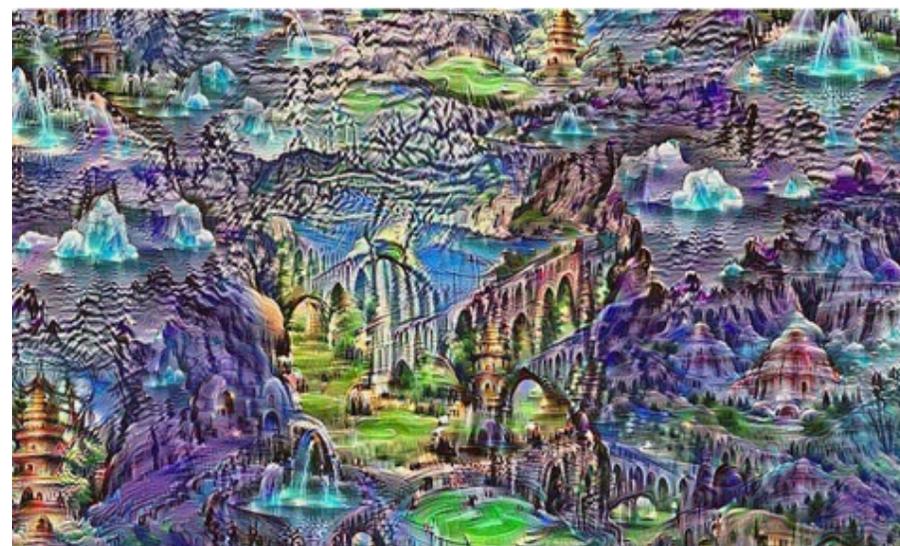
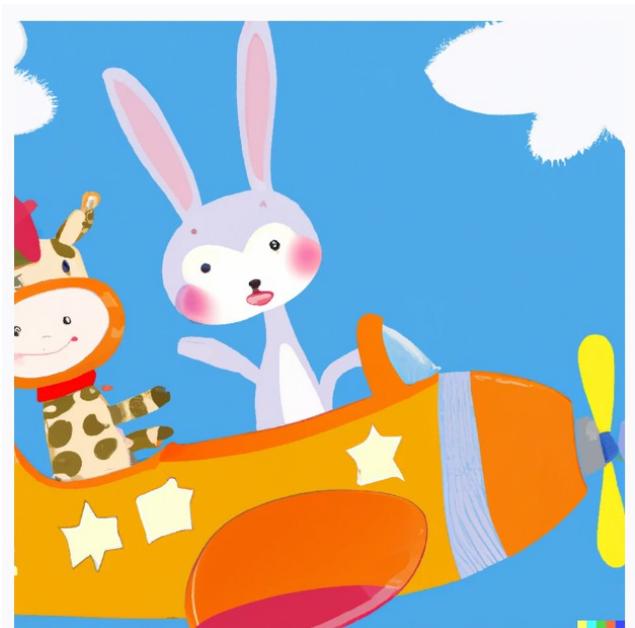


Philosophy and Ethics



Art and Probability

Create a probabilistic sample from the distribution of images, given a prompt



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
t
in



- /Administration
- /Human Resources
- /Legal
- /Accounting
- /Finance
- /Marketing
- /Publicity
- /Promotion
- /Research
- /Business
- /Development
- /Engineering
- /Manufacturing
- /Planning

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:

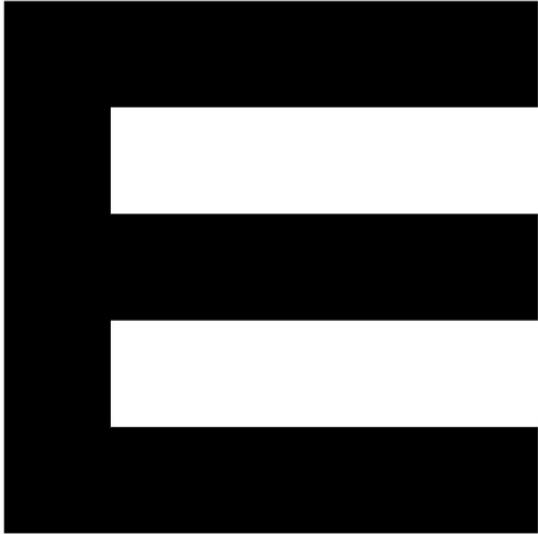
“Better understanding of probability”



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:

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

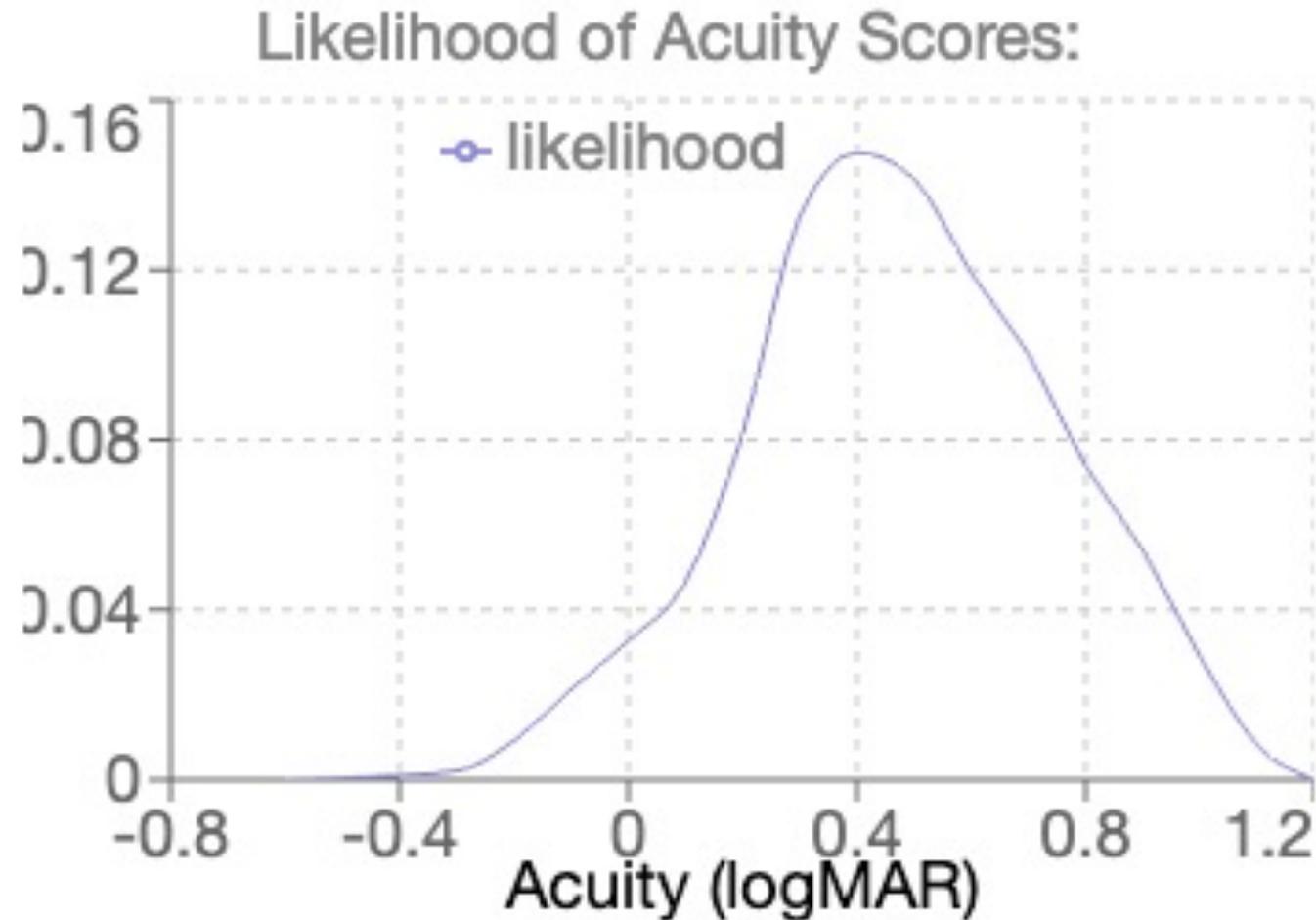
Where θ is the user's vision score and f is the font size of the letter.

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.

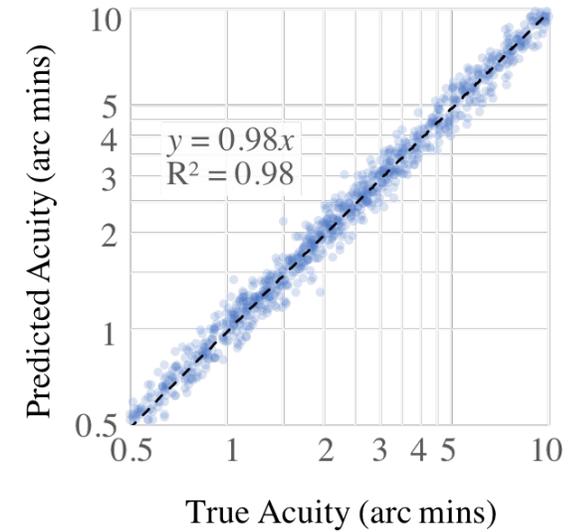
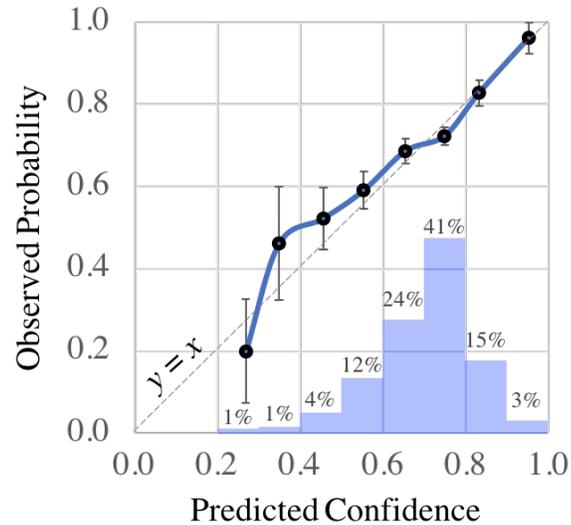
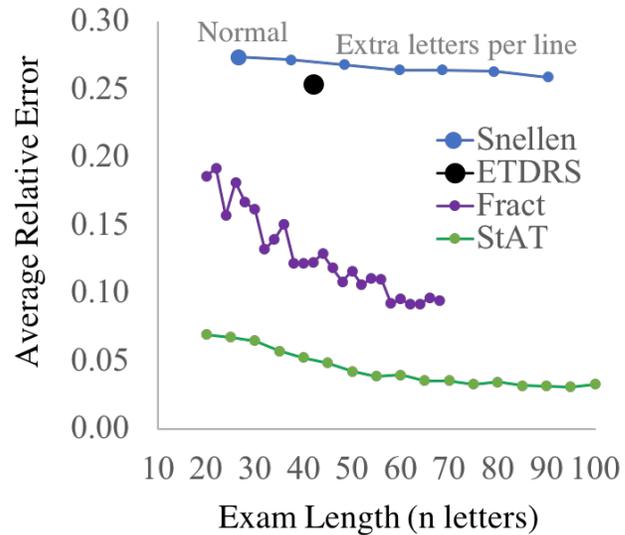


Learn Real Skills in CS109

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,² Robert T Chang,² Charles Lin²

¹Department of Computer Science, Stanford University

²Department 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

^{*}Equal contribution
Copyright © 2021
Intelligence (www

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 discretise letter sizes on a chart. On the other hand, digital exams can show letters 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.

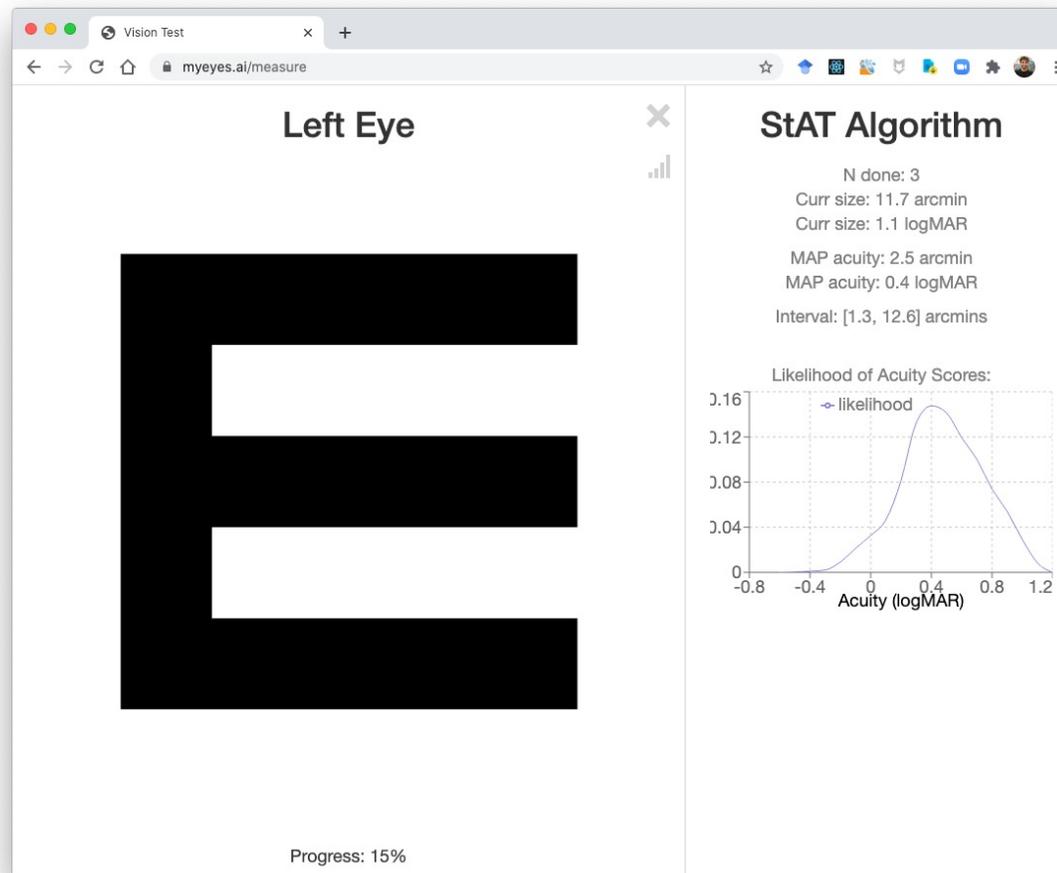
While promising, contemporary digital exams have yet to dramatically improve accuracy over traditional chart-based approaches. The current best digital exam uses a psychometric Item Response Theory (IRT) algorithm for both selecting the next letter size to query and for making a final prediction of acuity. 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, with a confidence in the final

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

STACT was named after Ed. We continue in this

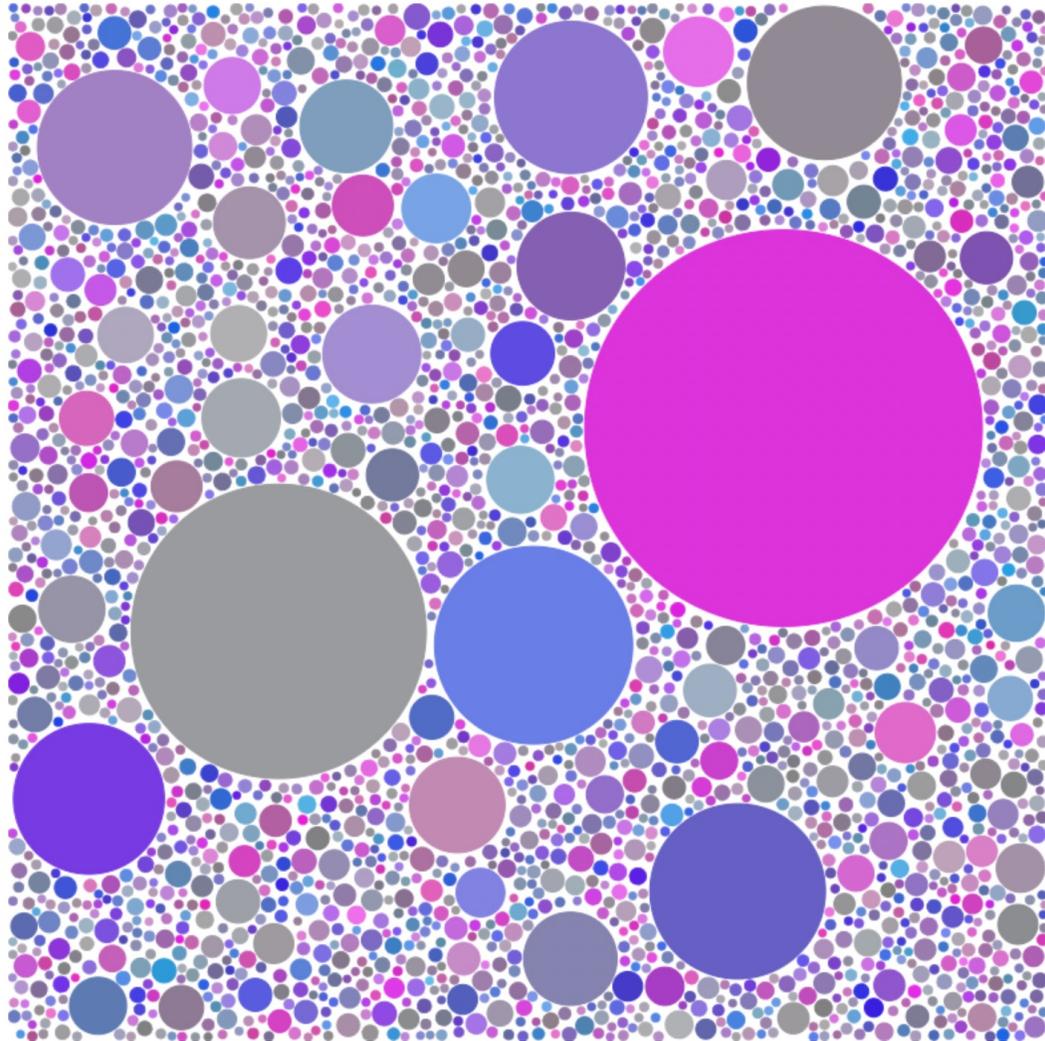


Science

THE LANCET



What about last years final?



Regenerate

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

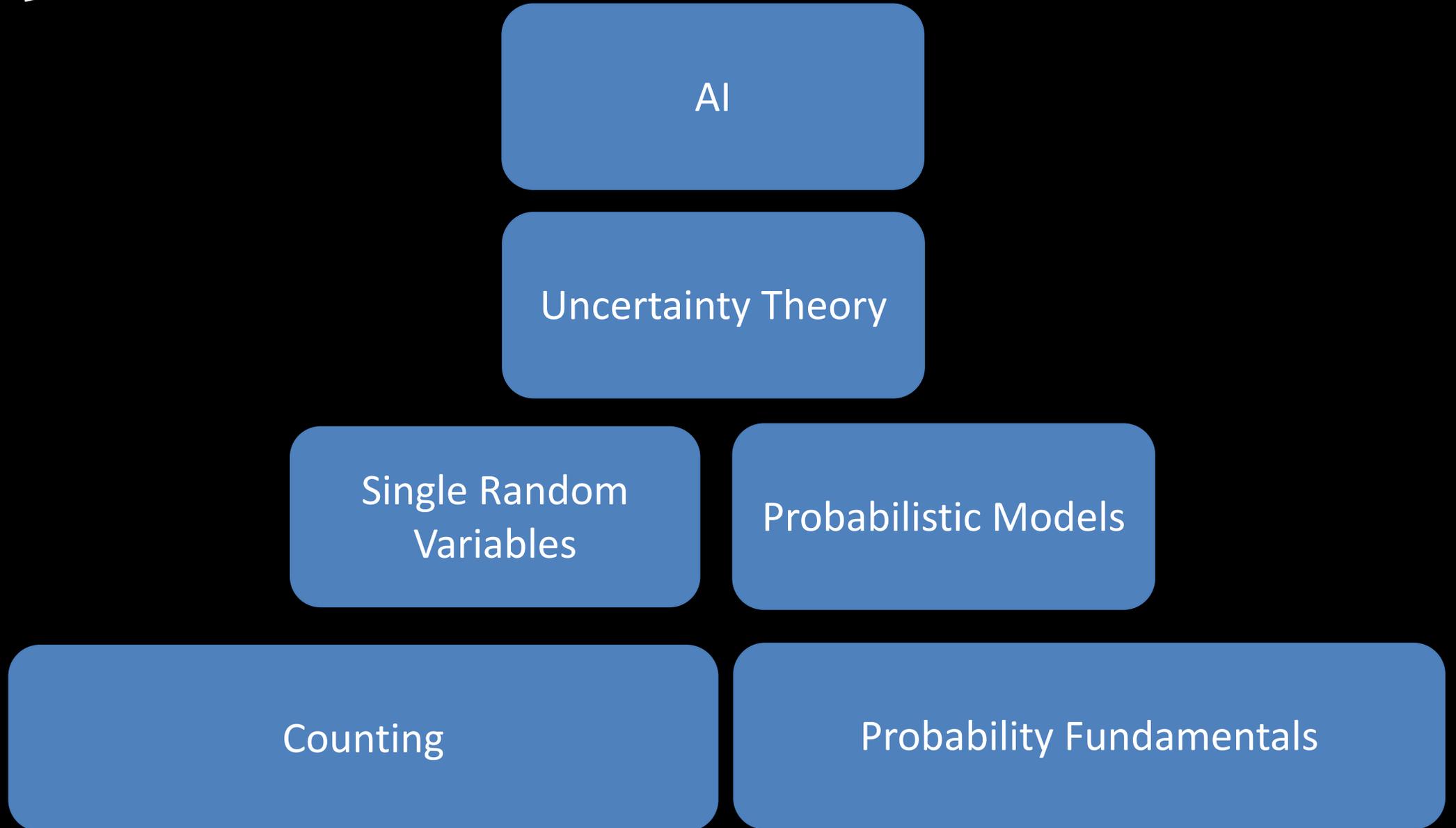
Delayed gratification

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.



Lets dive in...

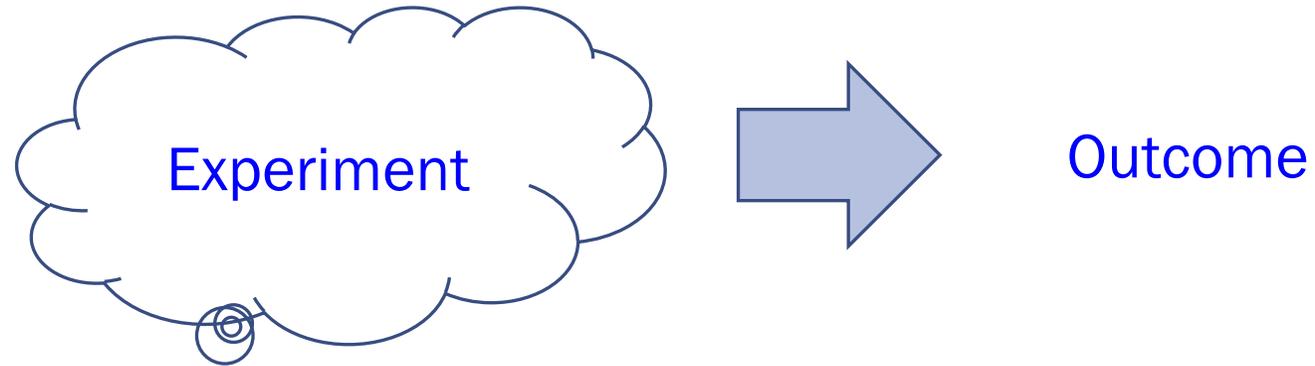
2 min pedagogic pause.



Counting I

What is Counting?

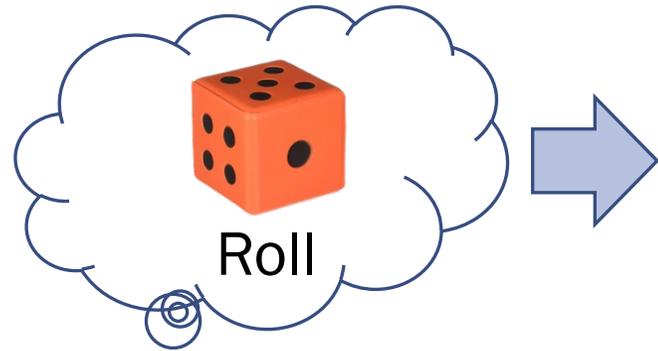
An experiment
in probability:



Counting:

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

What is Counting?



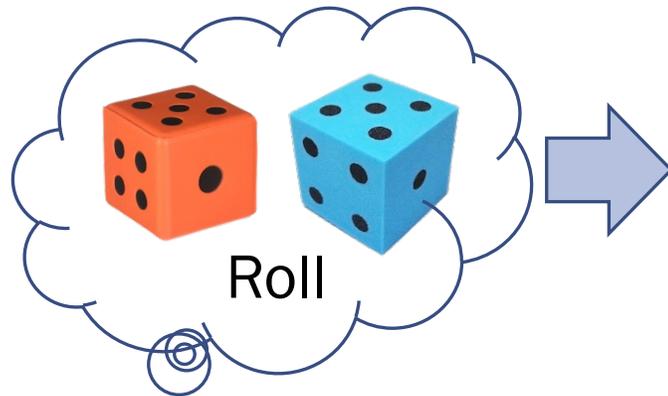
6

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



3

{2, 4, 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)}

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?

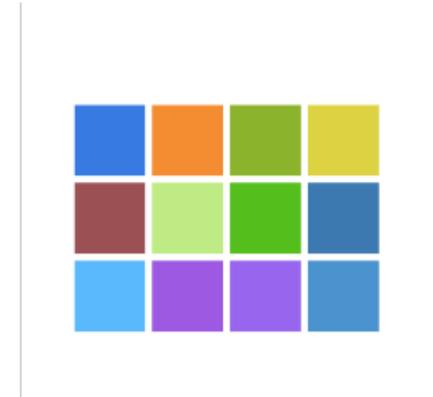
Each pixel can be one of 17 million distinct colors



(a) 12 million pixels



(b) 300 pixels



(c) 12 pixels

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



How Many Unique Images?

Each pixel can be one of 17 million distinct colors



(a) 12 million pixels

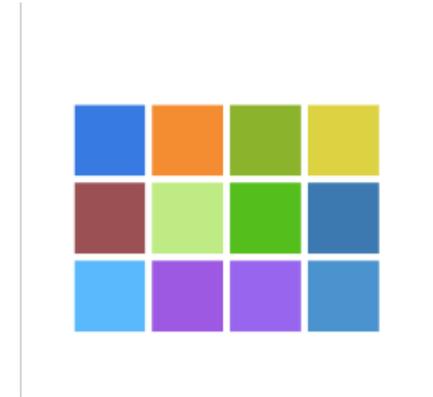
$$\approx 10^{86696638}$$



(b) 300 pixels

$$\approx 10^{2167}$$

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



(c) 12 pixels

$$\approx 10^{86}$$

Sum Rule of Counting

If the outcome of an experiment can be either from

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

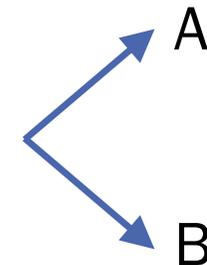
or Set B , where $|B| = n$,

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: 20 + 10



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
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
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 $|A| = m$,

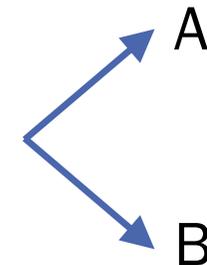
or Set B , where $|B| = n$,

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 and school at
which to study probability!
Exciting.