CS109: Probability for Computer Scientists

Gather around and let me tell you a story.

Something BIG is happening

ChatGPT Claude Gemini

But first…
who am I?

Juliette Woodrow – Instructor for CS109

Juliette Woodrow – what really describes me

My Life Goal is to Be a Teacher ☺

Juliette Woodrow

Ph.D Candidate at Stanford University

jwoodrow@stanford.edu

My Life Goal is to Be a Teacher ☺

In Computer Science

My Life Goal is to Be a Teacher ☺

In Computer Science

I have been teaching in some capacity for over 15 quarters at Stanford.

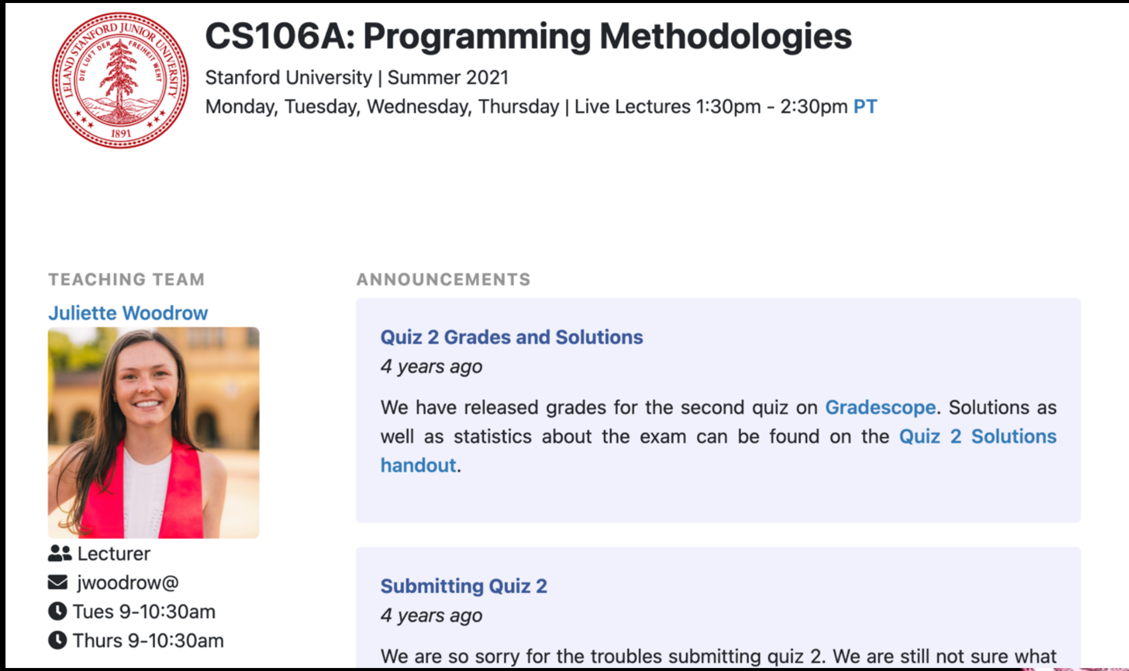
My Life Goal is to Be a Teacher ☺

In Computer Science

I have been teaching in some capacity for over 15 quarters at Stanford.

I was the instructor for CS106A in 2021.

I have been a TA and a head TA for CS109.



CS106A: Programming Methodologies

Stanford University | Summer 2021

Monday, Tuesday, Wednesday, Thursday | Live Lectures 1:30pm - 2:30pm PT

TEACHING TEAM

Juliette Woodrow

Lecturer

jwoodrow@

Tues 9-10:30am

Thurs 9-10:30am

ANNOUNCEMENTS

Quiz 2 Grades and Solutions

4 years ago

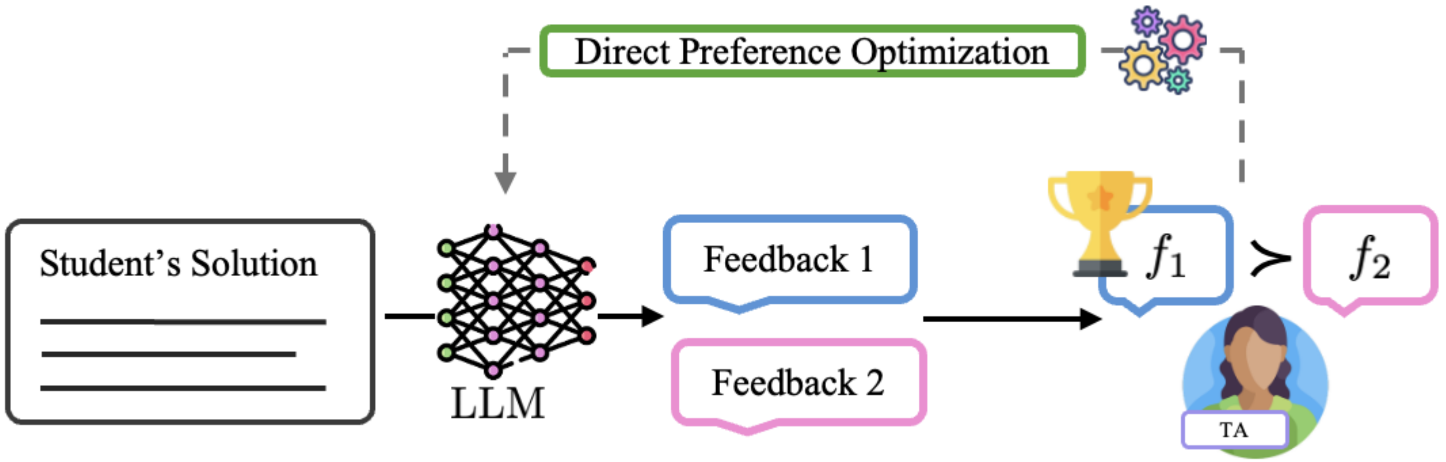
We have released grades for the second quiz on Gradescope. Solutions as well as statistics about the exam can be found on the Quiz 2 Solutions handout.

Submitting Quiz 2

4 years ago

We are so sorry for the troubles submitting quiz 2. We are still not sure what

Research Journey – Aligning Large Language Models



$$\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\log\sigma\left(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)}-\beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\right)\right]$$

Using **Direct Preference Optimization**, we can fine-tune an LLM to quickly learn our course.

How can we efficiently **align Large Language Models** to match our course preferences?

In a blind controlled study of experts, we significantly **outperform** GPT-4o

Research Journey – From Probability to LLM Prompts

$$
\begin{aligned}
\log p^\star &= \log\left(\frac{P(y^\star \mid x)}{P(y^\star)}\right) - \log\left(\sum_{k=1}^{N} \frac{P(y_k \mid x)}{P(y_k)}\right) \\
&= \text{PointwiseMutualInfo}(x, y^\star) - \log\left(\sum_{k=1}^{N} e^{\text{PMI}(x, y_k)}\right) \\
&\geq \text{PointwiseMutualInfo}(x, y^\star) - \log N.
\end{aligned}
$$

A carefully constructed prompt that allows us to estimate PMI

ChatGPT Claude Gemini Your favorite LLM

$$PM\hat{I}(x, y^*)$$

Can probability and math turn into natural language text prompts to LLMs?

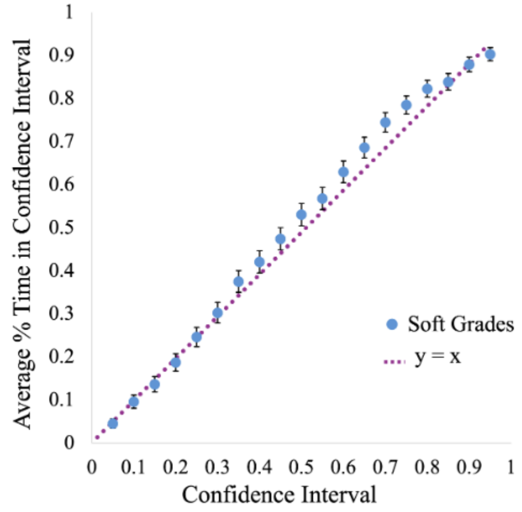
YES! And it turns out this framing increases the bounds of what types of problems we can use LLMs to solve!

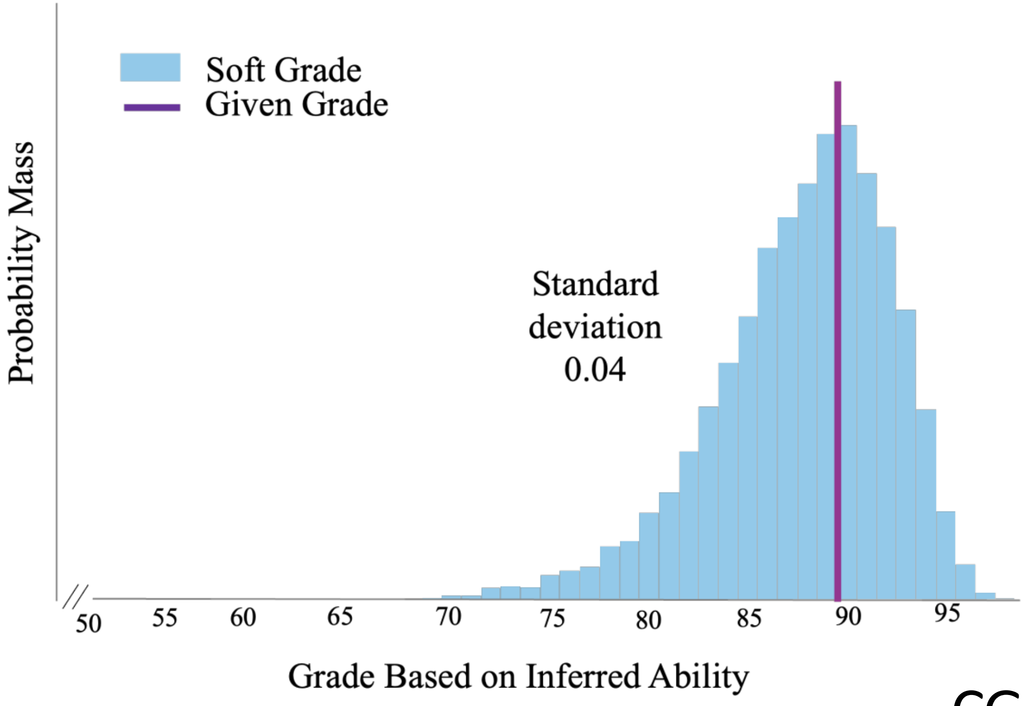
Research Journey – Representing Uncertainty

Can we **represent the uncertainty** around course grades? Like ones given at Stanford.

$$\sigma^2_{\text{posterior}} = \left(\frac{1}{\sigma^2_{\text{prior}}} + \sum_{j=1}^{k} \frac{1}{\epsilon_j^2} \right)^{-1},$$

$$\mu_{\text{posterior}} = \sigma^2_{\text{posterior}} \left(\frac{\mu_{\text{prior}}}{\sigma^2_{\text{prior}}} + \sum_{j=1}^{k} \frac{a_{ij}}{\epsilon_j^2} \right).$$





	Likelihood			Expected Calibration Error
	Exact	±1	±3	(ECE)
Oracle	0.119 ± 0.007	0.324 ± 0.013	0.612 ± 0.016	0.018
Soft Grades	0.118 ± 0.007	0.323 ± 0.014	0.605 ± 0.016	0.020
FMN Baseline	0.066 ± 0.007	0.192 ± 0.019	0.407 ± 0.027	0.131

Representing Grades as **Probability Distributions** leads to better predictions.

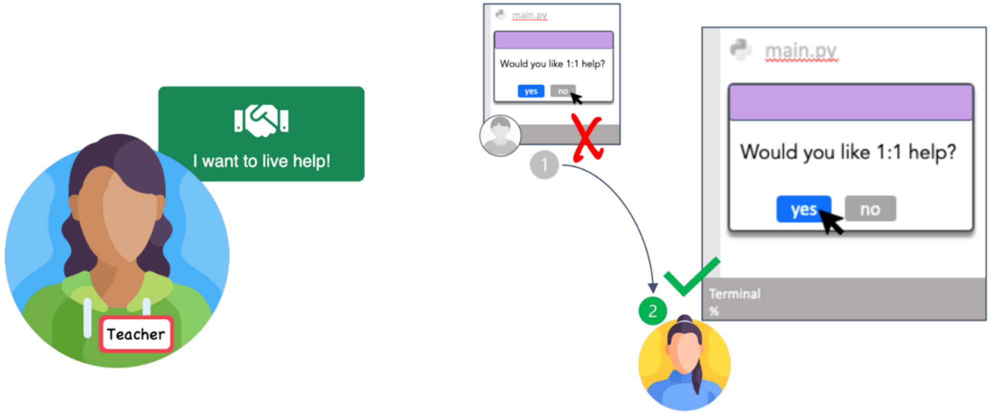
CGRT is state of the art for imputing missing grades.

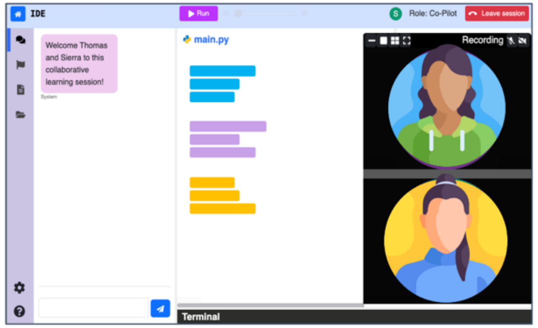
Research Journey – Code in Place

Helped create and run **Code in Place** – The course in the world with the most teachers

5,000+ Teachers
60,000+ Students
Since 2020

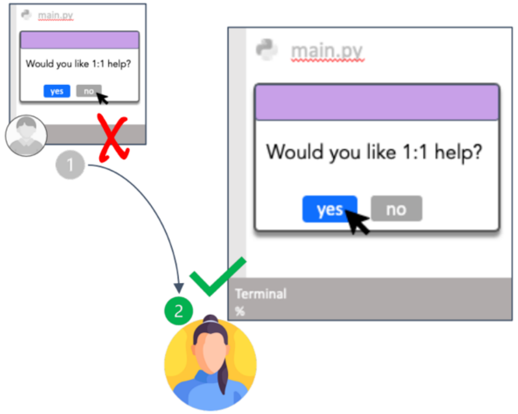
Research Journey – Code in Place

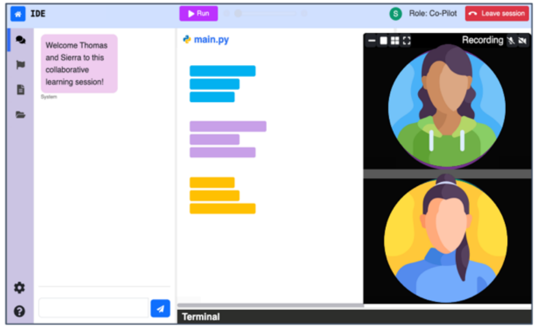




Co-created **TeachNow** – a system that **spontaneously matches teachers with students**. A single TeachNow session increased course retention by 15%.

Research Journey – Code in Place





Co-created **TeachNow** – a system that **spontaneously matches teachers with students**. A single TeachNow session increased course retention by 15%.

Who can be a teacher and when? We expanded TeachNow so that students can help other students.

Learning through teaching is a magical thing.

Fantastic Teaching Team









Fantastic Teaching Team









Supported by Professor Chris Piech

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 21 or equivalent (important, coreq ok)

- Differentiation
- Integration
- Basic facility with linear algebra (vectors)

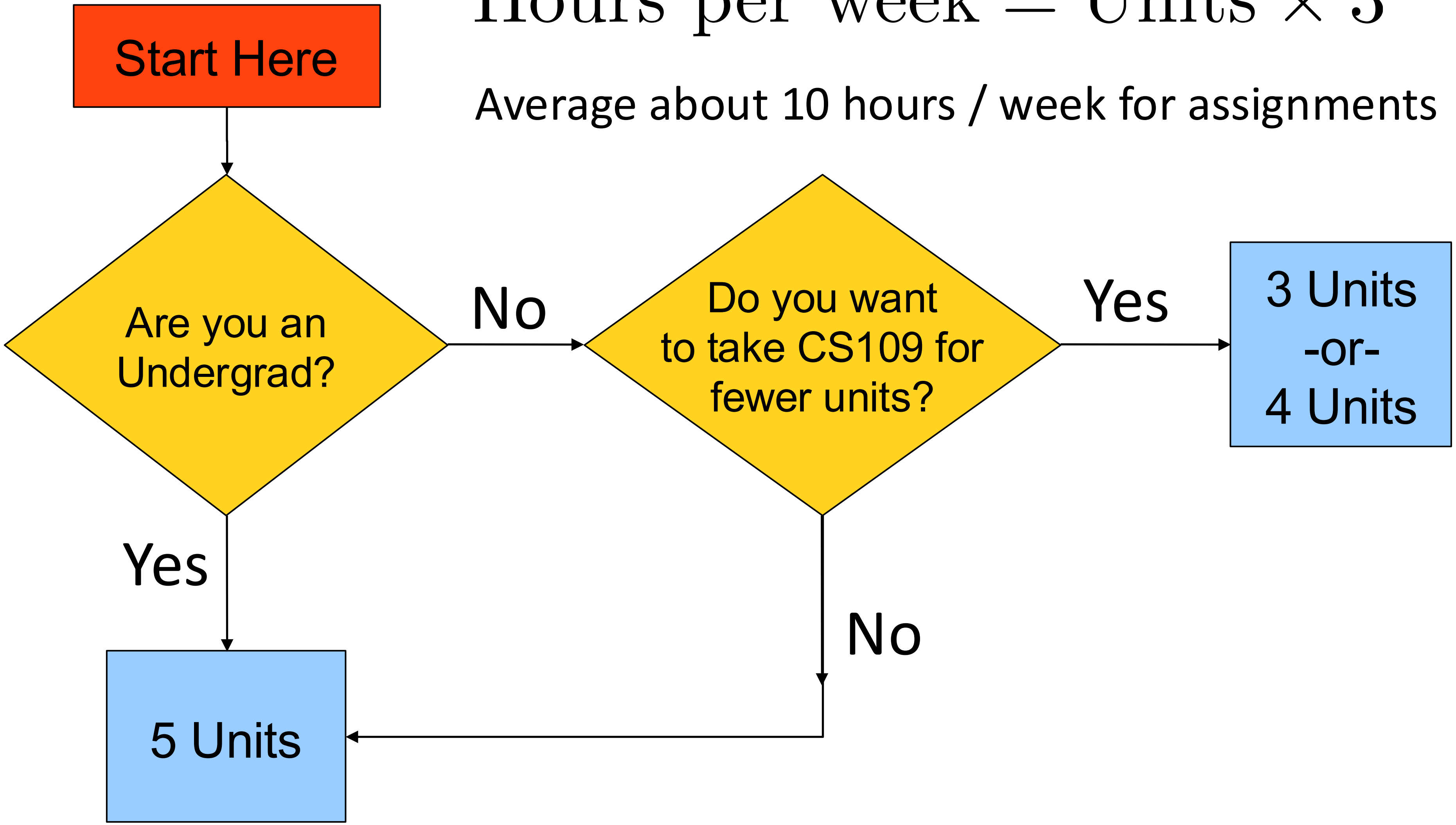
Coding in CS109

Review session on Friday

CS109 Units

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

Average about 10 hours / week for assignments



Class Breakdown

30%	**7 Assignments**
20%	**Midterm** 2 hour exam, Feb 10th, 7pm
30%	**Final** 3 hour exam, March 17th, 8:30am
20%	**Participation** Section, PEP

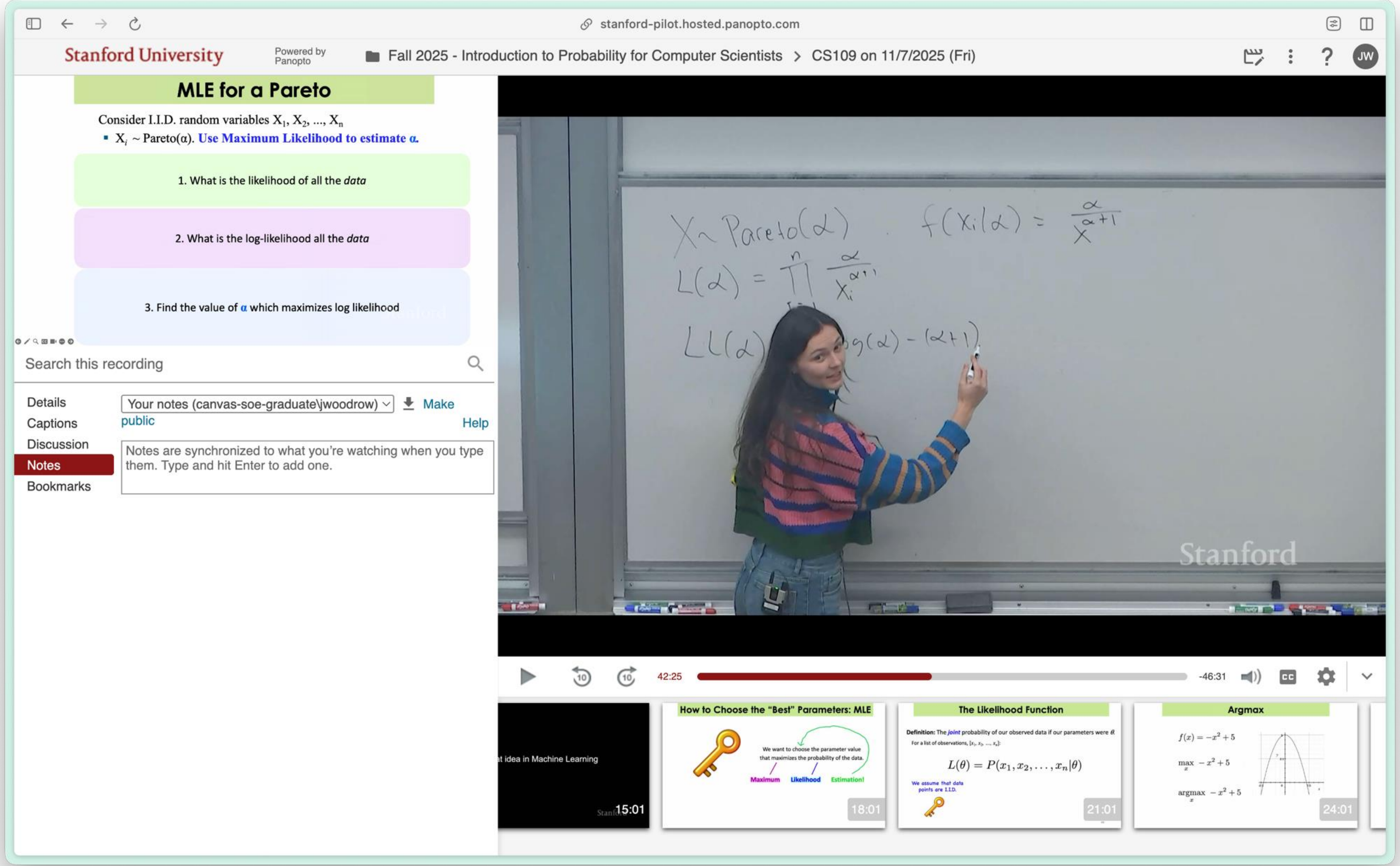
Personal Exam Prep (PEP)

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

Playlist of the Quarter

[coming on Wednesday]

Is Class Online?



TLDR: Yes. Come to live class. It's a good time (and good for you) and you'll get extra credit!

Come to Class

Section participation grade 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.

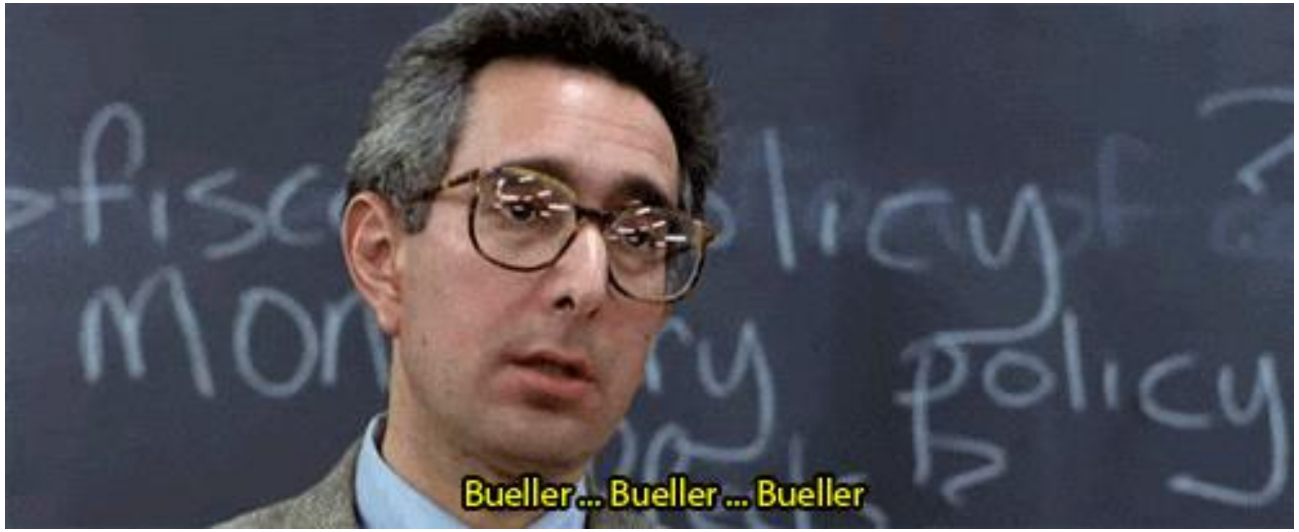
Come to Class – Extra Credit

To incentivize staying on track with the material……

Award extra credit to those who attend lecture in person!!

You need a phone (or a device with a camera) to record your attendance.

Attending 20 out of the 26 lectures gives you the full amount of credit.



* CGOE students, check the syllabus.
We didn't forget about you :)

Let's Try Out Attendance Tracking System

Step 1: On your phone to go - https://tinyurl.com/stanford-qr

Step 2: Scan the QR Codes

Step 3: Wait until you see "Attendance recorded"

Step 4: You can check our own attendance !!

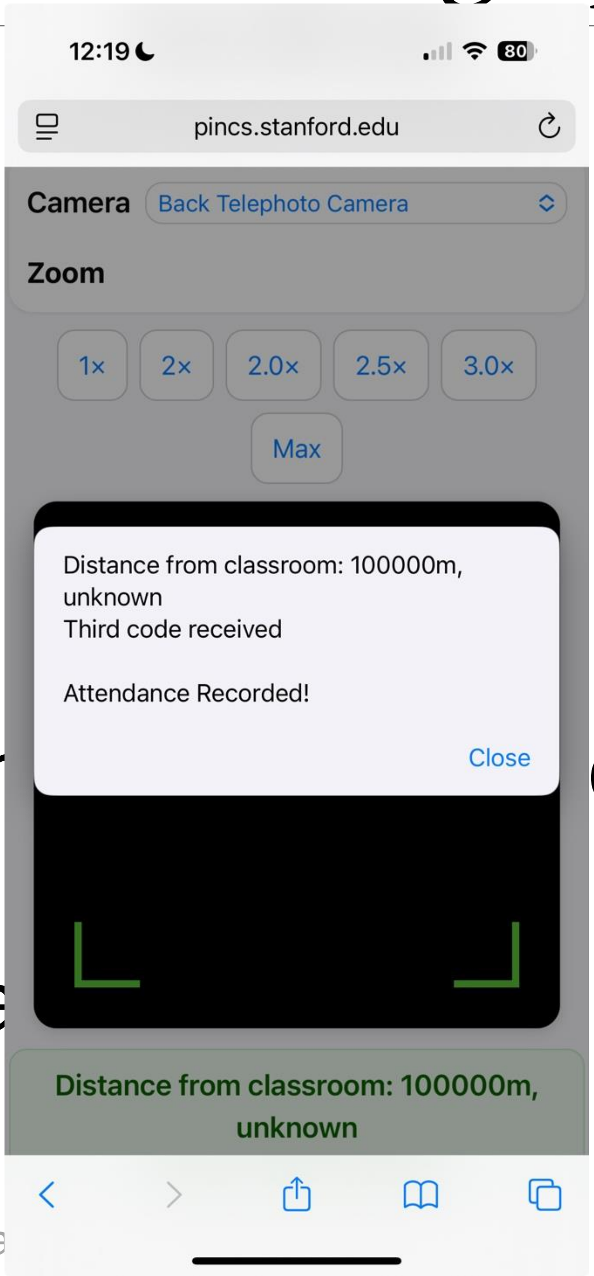
Let's Try Out Attendance Tracking System

Step 1: On your phone to go - https://tinyurl.com/stanford-qr

Step 2: Scan the QR Codes

Step 3: Wait until you see "Atter d"

Step 4: You can check our own a



Let’s Try Out Attendance Tracking System

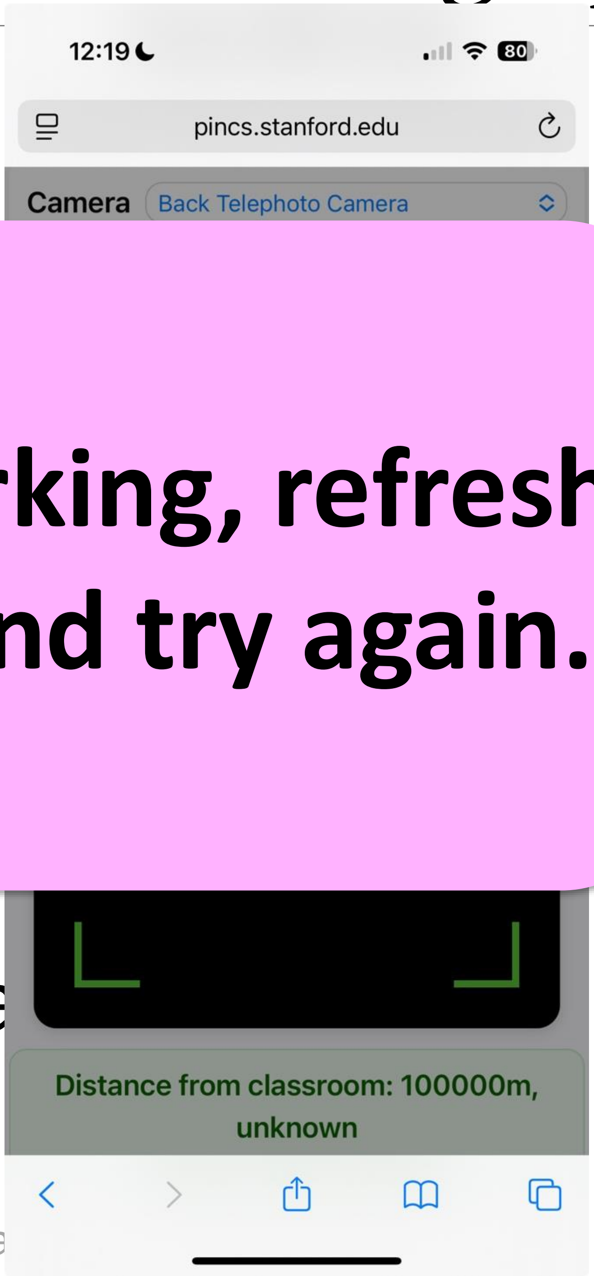
Step 1: On your phone to go -

https://tinyurl.

Step 2: Scan th

Step 3: Wait ur

Step 4: You can check our own a



If it isn’t working, refresh your page and try again.

Lecture Attendance

Ask questions

Q&A forum
All announcements

“Working” office hours
start on Wednesday

Email cs109@cs.stanford.edu

Answer Questions – Extra Credit

Extensive research has shown that one of the best ways to learn is to practice teaching.

You test your own knowledge, often realizing gaps and deepening our own understanding.

AND you help out a friend. Which is a beautiful thing.

Answer Questions – Extra Credit

Extensive research has shown that one of the best ways to learn is to practice teaching.

WIN

You test your own knowledge, often realizing gaps and deepening our own understanding.

WIN

AND you help out a friend. Which is a beautiful thing.

WIN

Answer Questions – Extra Credit

Extensive research has shown that one of the best ways to learn is to practice teaching.

You test your own knowledge, often realizing gaps and deepening our own understanding.

AND you help out a friend. Which is a beautiful thing.

ed

We will award extra credit for answering questions on Ed.

Answer Questions – Extra Credit

Extensive research has shown that one of the best ways to learn is to practice teaching.

You test your own knowledge, often realizing gaps and deepening our own understanding.

AND you help out a friend. Which is a beautiful thing.

This could be you, helping your peers!

We will award extra credit for answering questions on Ed.

Problem Sets – Extra Credit

More info to come on Wednesday, when the problem set is released.

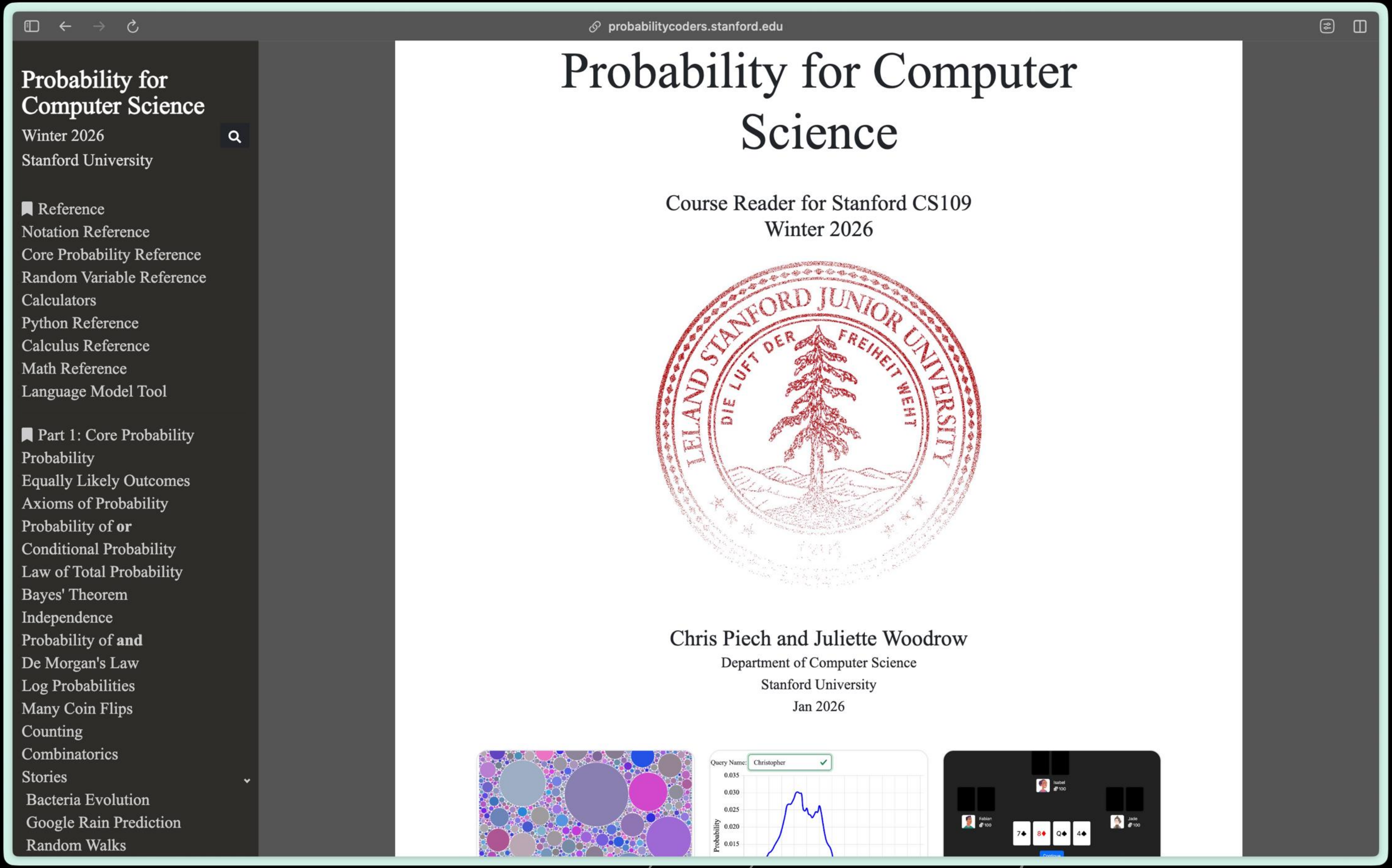
TLDR – can also get extra credit on problem sets.

Extra Credit – A philosophy

Extra Credit – A philosophy

Extra credit in CS109 is truly optional !!!

We Have A Course Reader!



ACE Companion Course


docs.google.com

CS ACE Program Application - Winter 2026

Additional Calculus for Engineers (ACE) is designed to provide a solid foundation in mathematics, computational math in engineering, and computer science to undergraduate students interested in pursuing an engineering degree. The mission of ACE is to provide our students, particularly those from under-resourced and/or minoritized backgrounds, with the opportunities and support to succeed in their course and in their major. ACE sections, instruction, and the ACE Course Assistants (CA's) are intentional about fostering environments based on equity, purposeful engagement, and inclusive teaching practices. Students selected to participate in ACE are those whose application demonstrates the most significant need or have not had access to institutional resources, exposure, or instruction related to Calculus and/or Computer Science.

Undergraduate students participating in ACE are required to attend an additional weekly section and enroll in their selected course for 1 additional unit. ACE participants also receive access to additional exam review sessions, individual tutoring, and other study resources. We accept applications from all students who believe they may benefit from participating in small active-learning sessions led by a highly trained graduate student.

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

Please complete the following application to be considered for a CS ACE this quarter.

Note that space in each section is limited and that section times may be subject to change.

The priority deadline is 11:59 PM Friday January 2, 2026.

The FINAL deadline for ACE applications is Friday, January 9, 2026, by 11:59 PM. Applicants will be notified of their application result on a rolling basis.

[LINK TO SIGN UP FORM]

web.stanford.edu

bus

DAYS AGO

ny questions after reading this Syllabus, post on our discussion forum.

g Team

Instructor: Juliette Woodrow
jwoodrow @ cs
CoDa Basement

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

Isabel Michel

Tae Kyu Kim

Sudharsan Sundar

Story of Modern AI

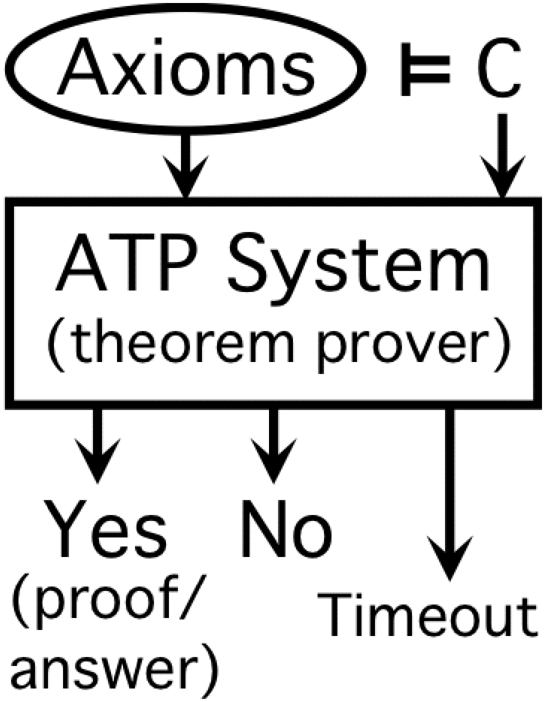
Modern AI
or, How we learned to combine probability and programming

Brief History

Early Optimism 1950s

1952

1955



Early Optimism 1950s

"Machines will be capable, within twenty years, of doing any work a $\boxed{\text{man}}$ 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

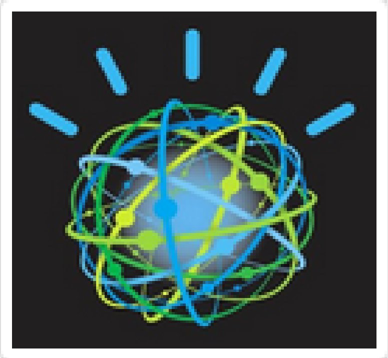

BRACE YOURSELVES
WINTER IS COMING

Something is going on in the world of AI

Big Milestones Part 1

1997 Deep Blue

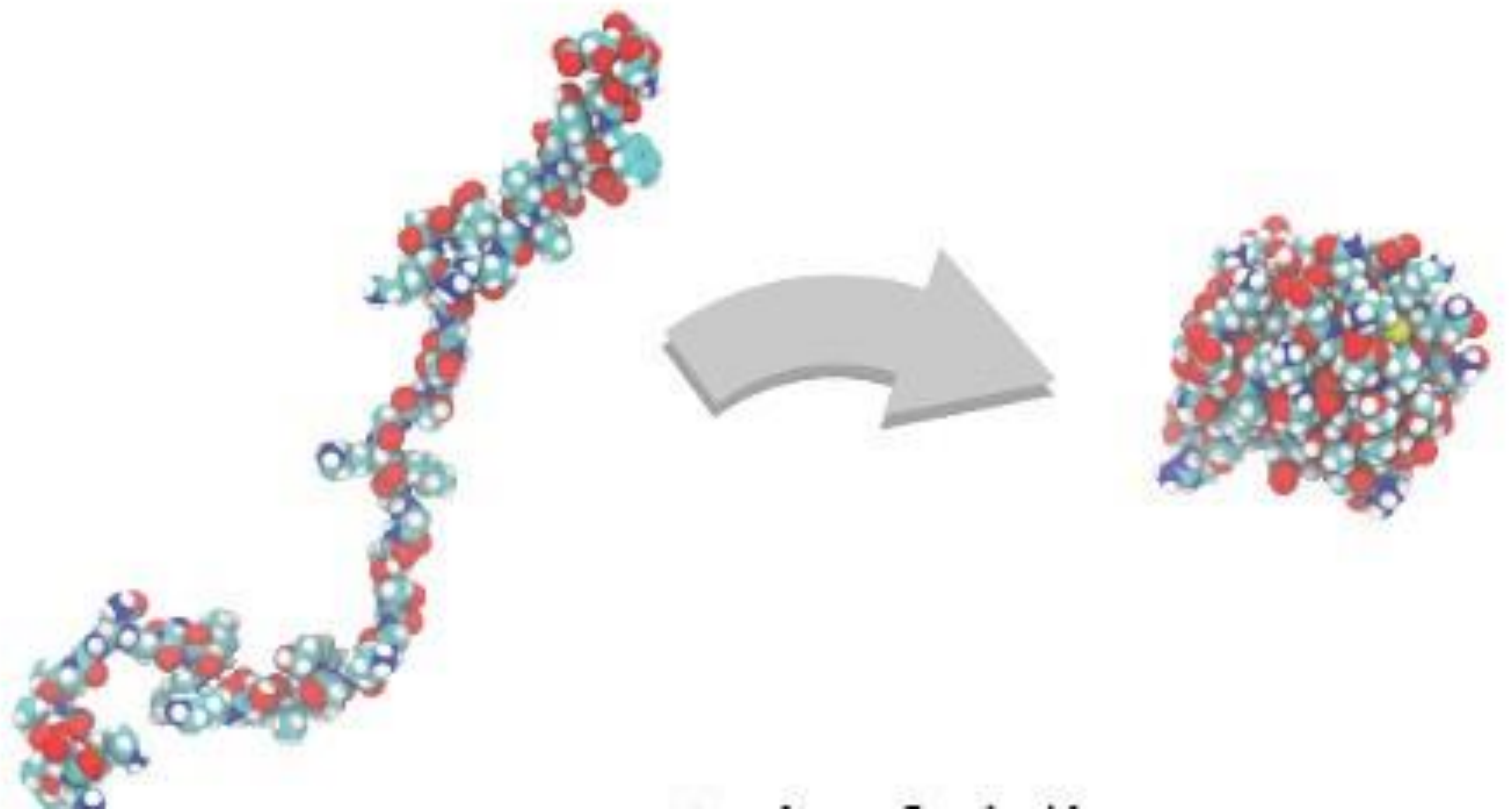
2005 Stanley



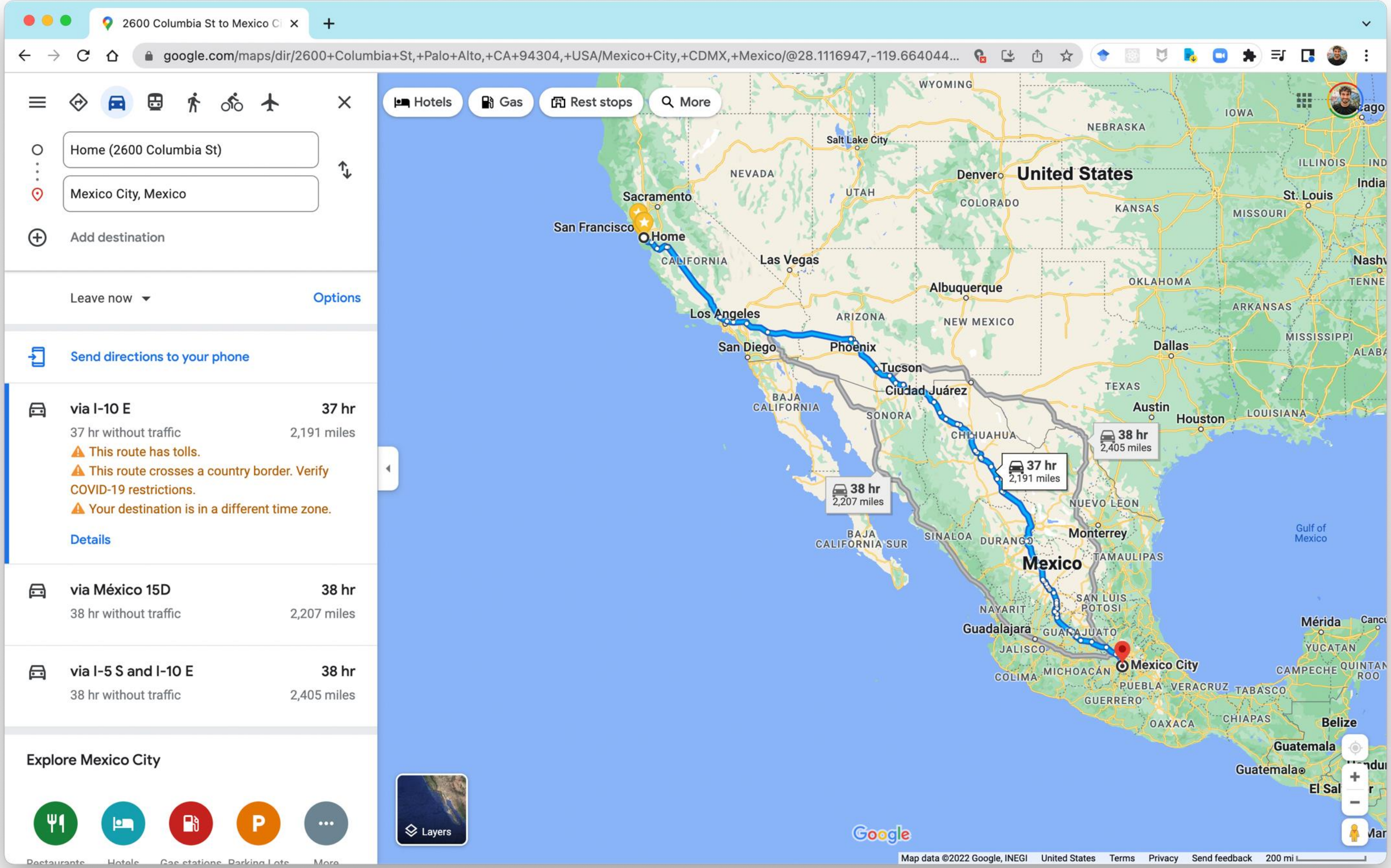
2011 Watson

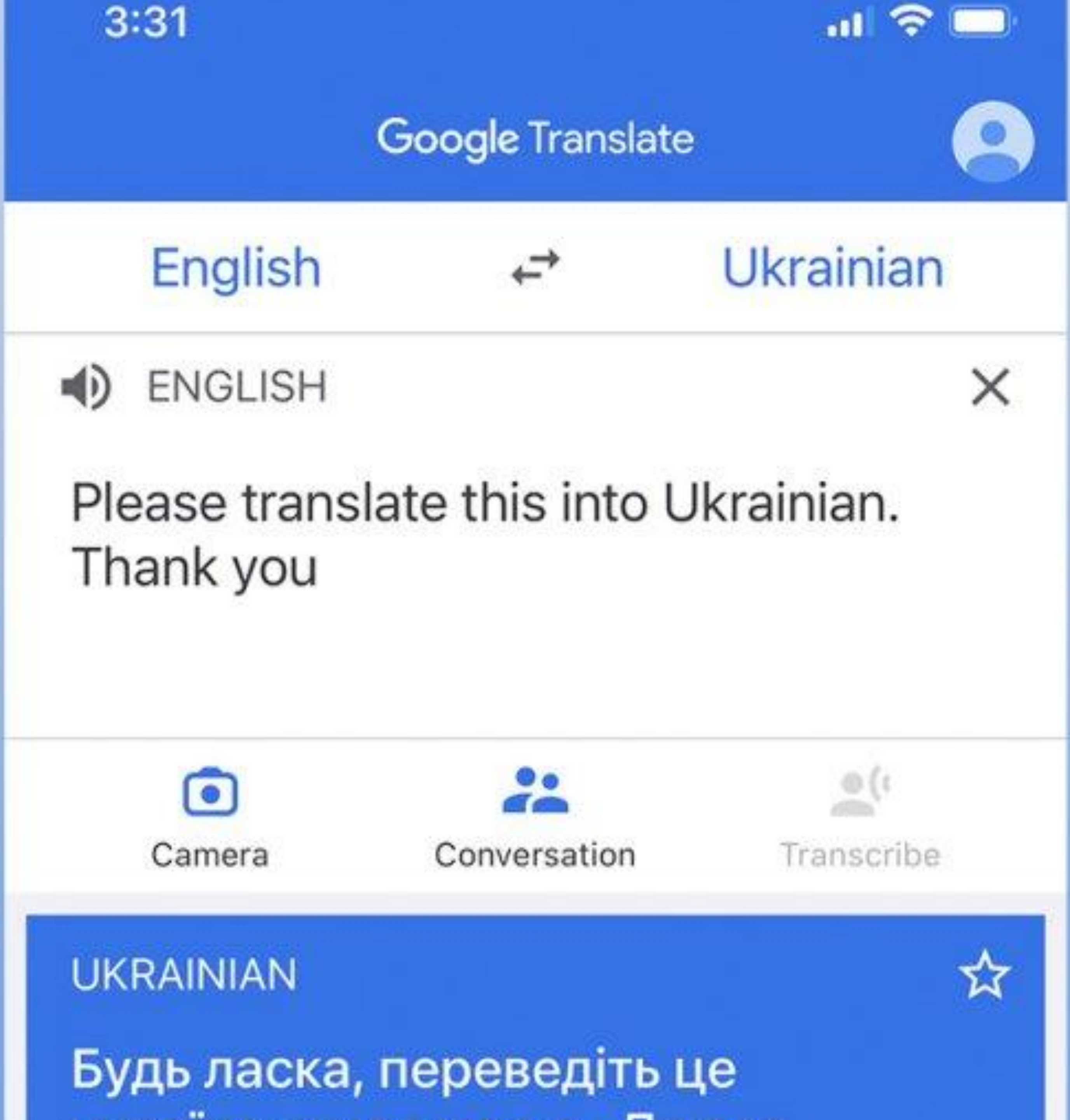
The last remaining board game

Protein Folding



Directions From A to B

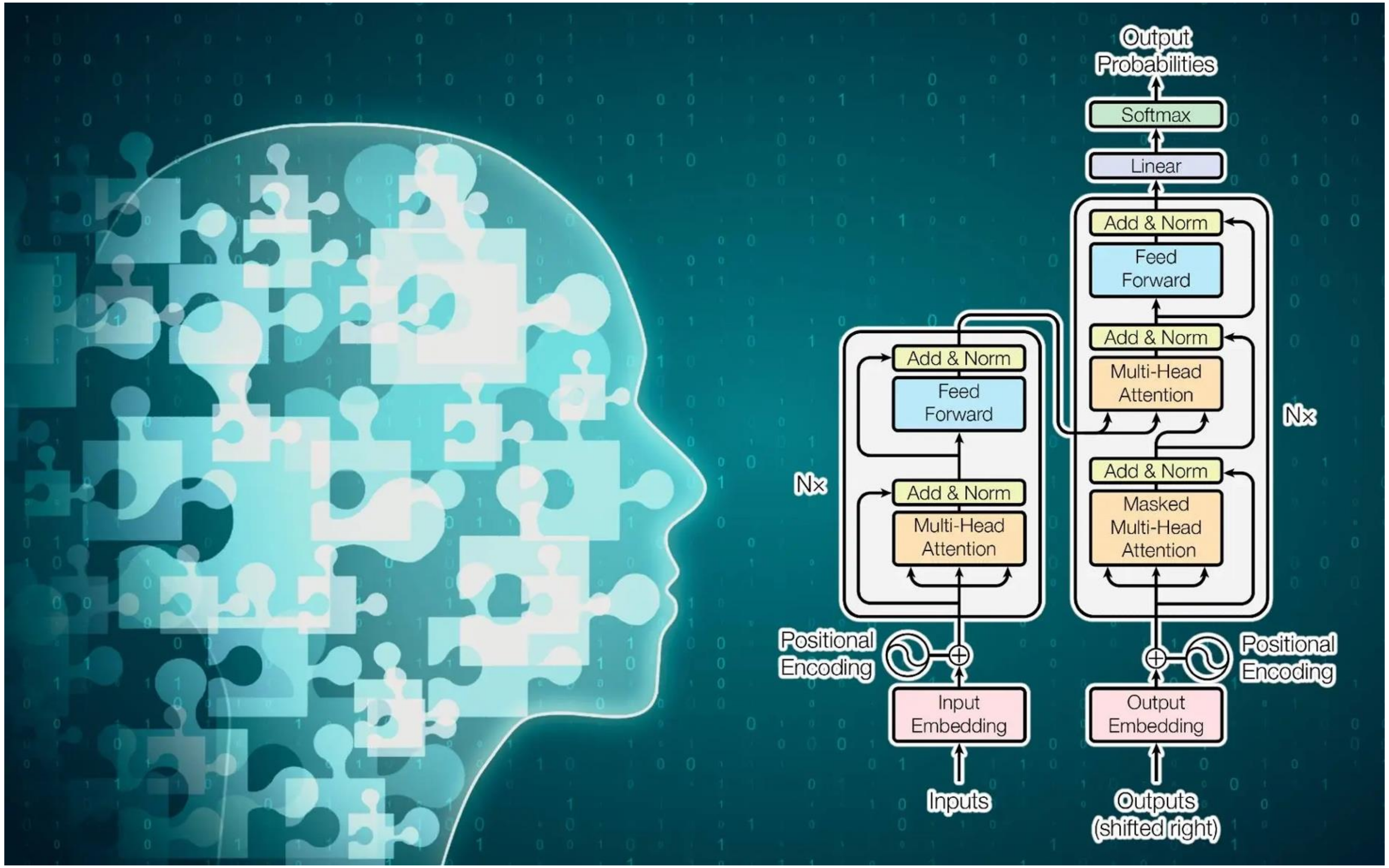



3:31
Google Translate
English
Ukrainian
ENGLISH
Please translate this into Ukrainian.
Thank you
Camera
Conversation
Transcribe
UKRAINIAN
Будь ласка, переведіть це

Self Driving Cars

And then in 2022, everything changed, again

AI that (seems) to understand language



What is going on?

[suspense]

Focus on one problem

Computer Vision

Chihuahua or muffin?

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

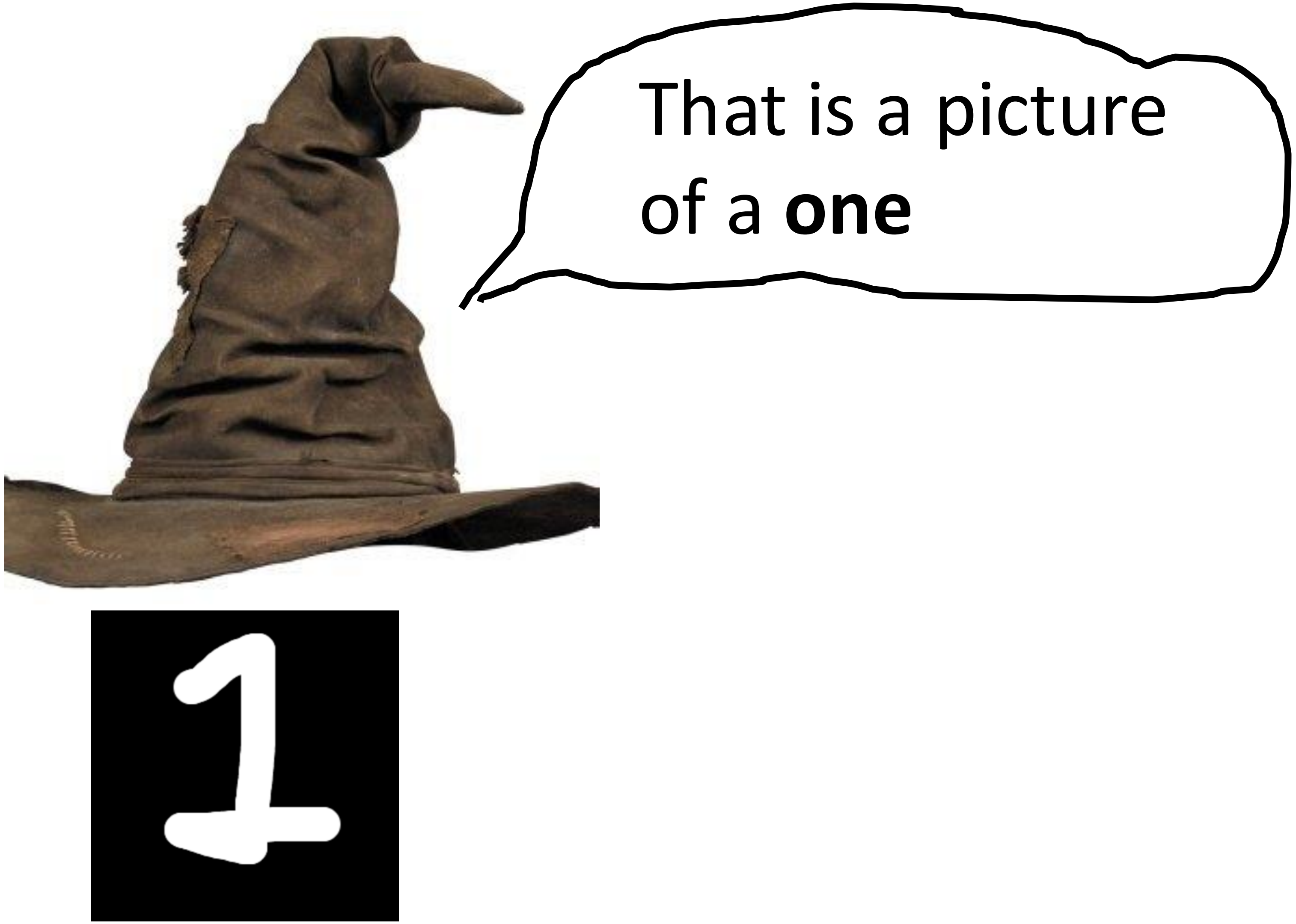
Why is it easy for Humans?

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

1

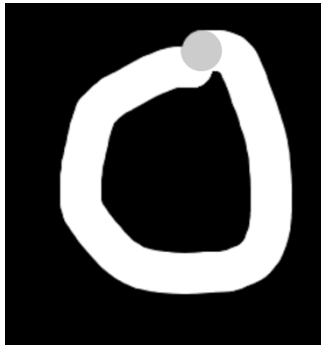
Make a Harry Potter Sorting Hat

Classification



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


```
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

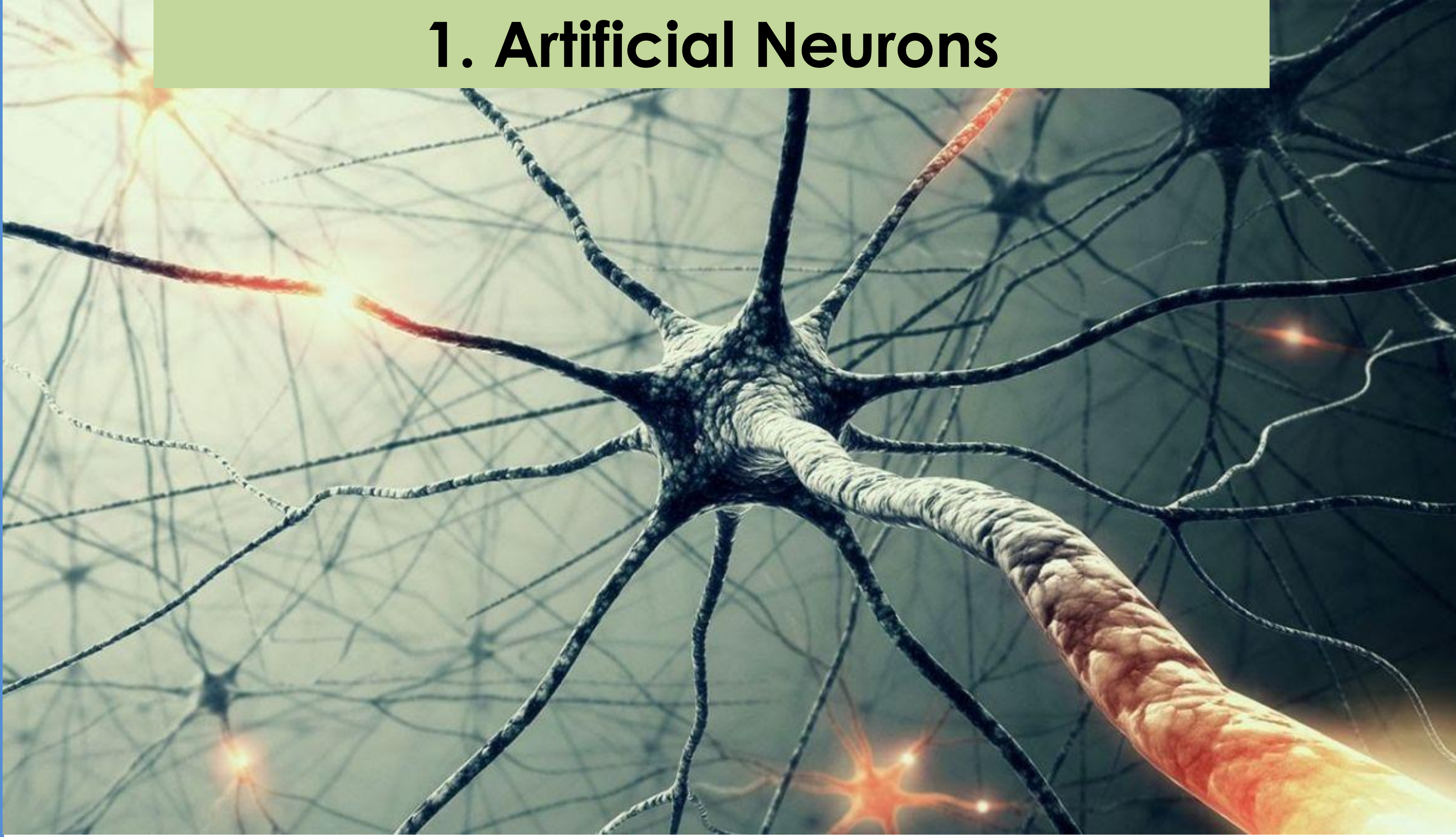
2. Learn by Example

Two Great Ideas

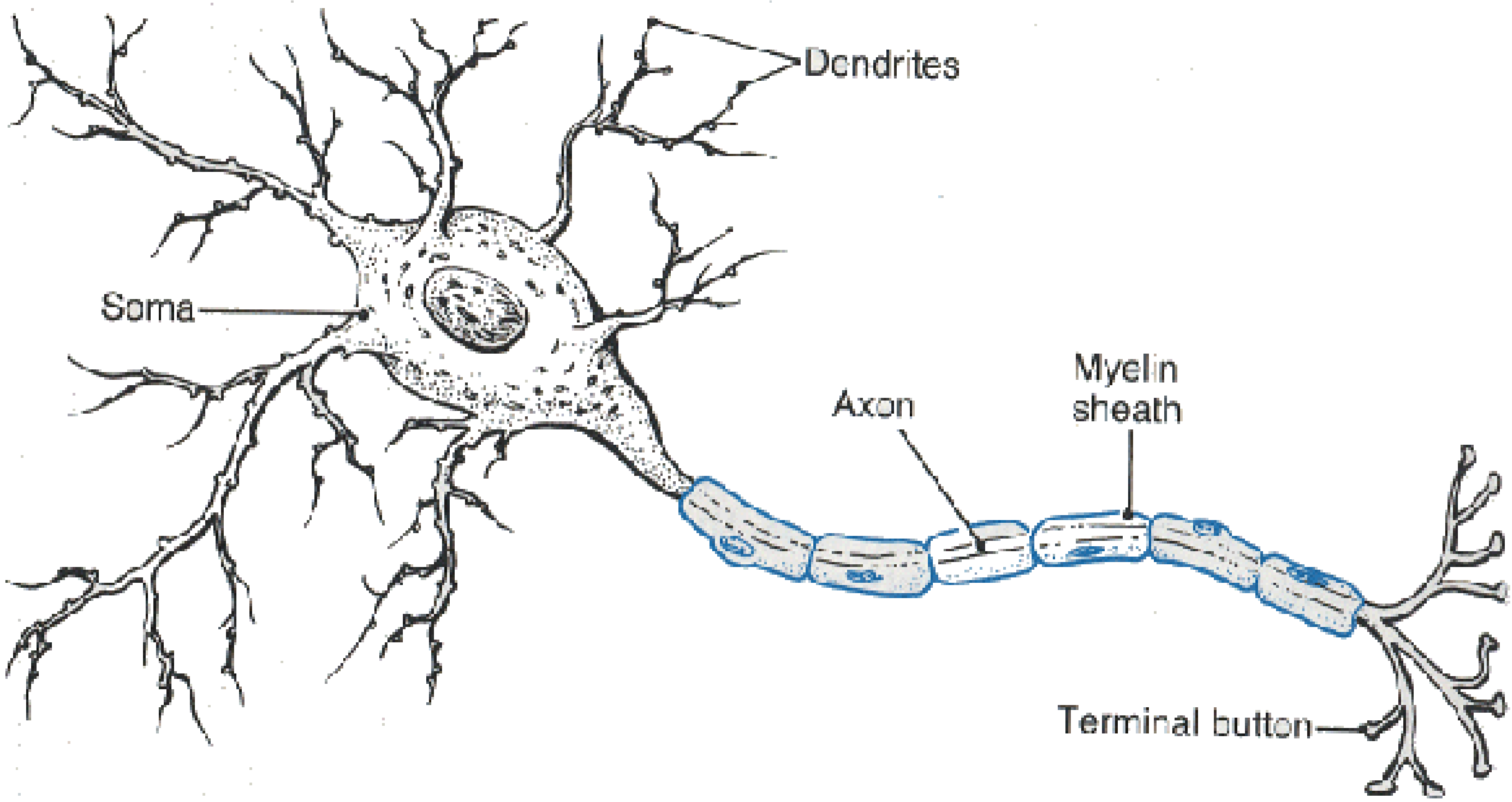
1. Artificial Neurons

2. Learn by Example

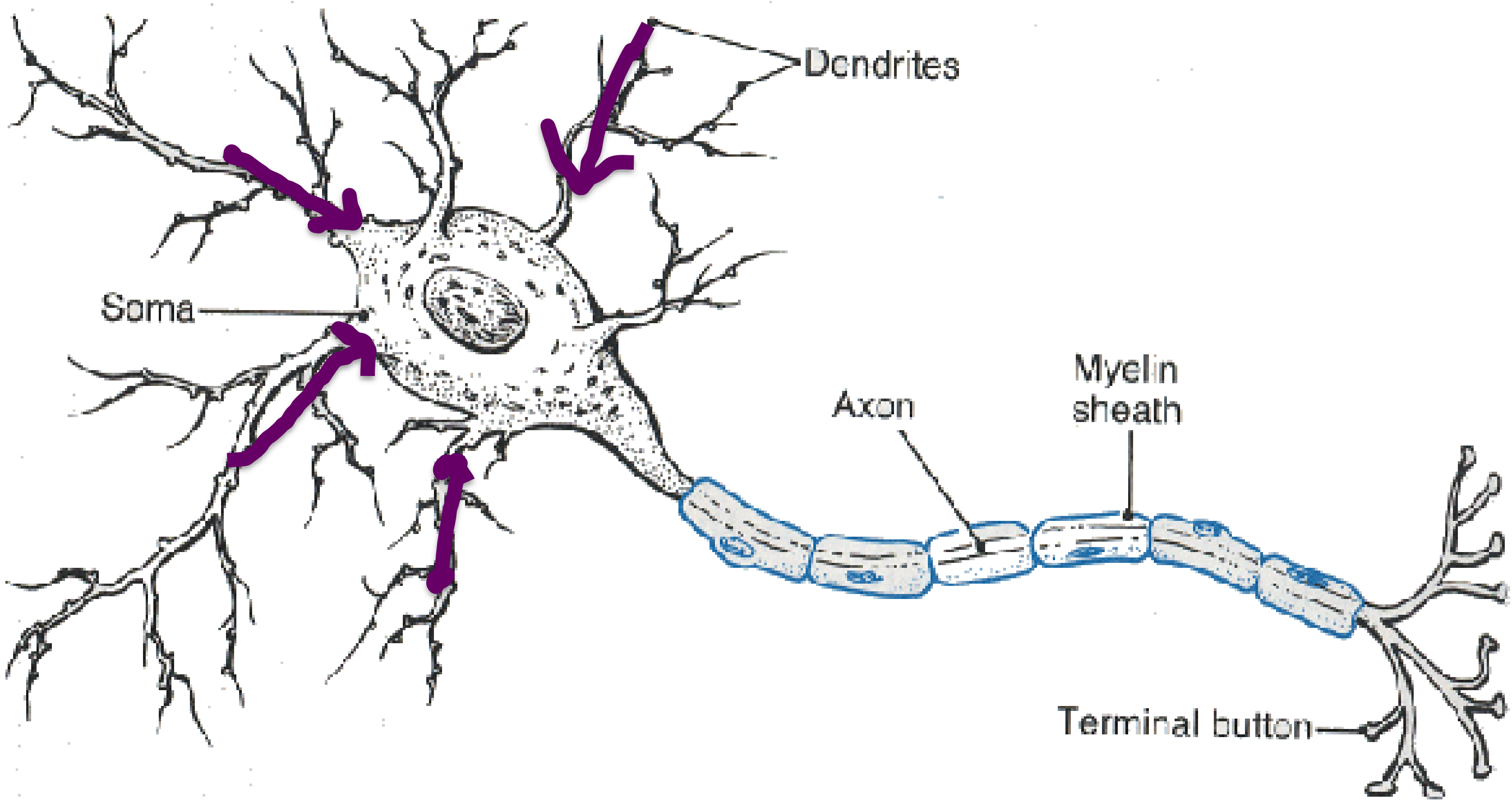
1. Artificial Neurons



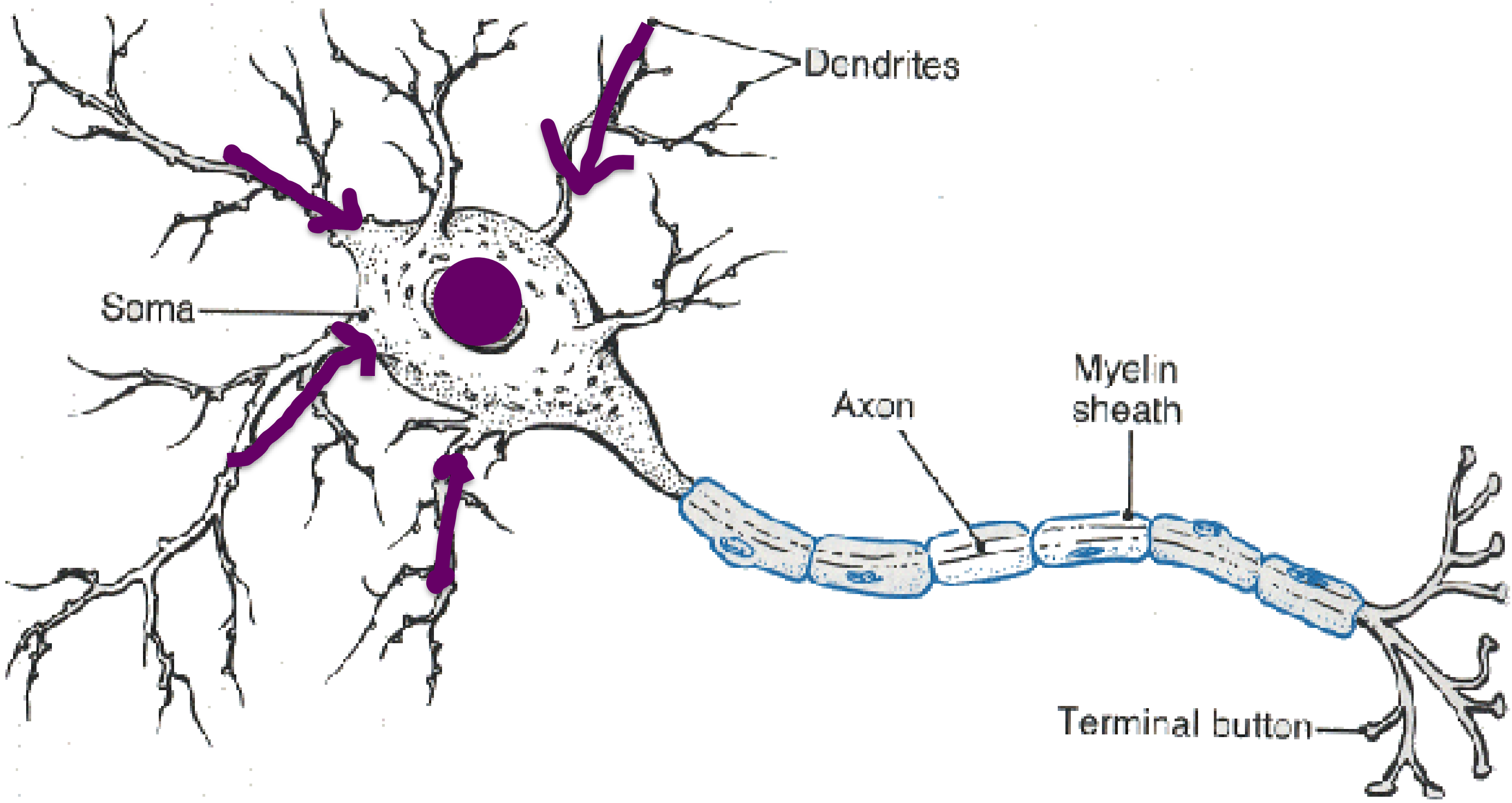
Neuron



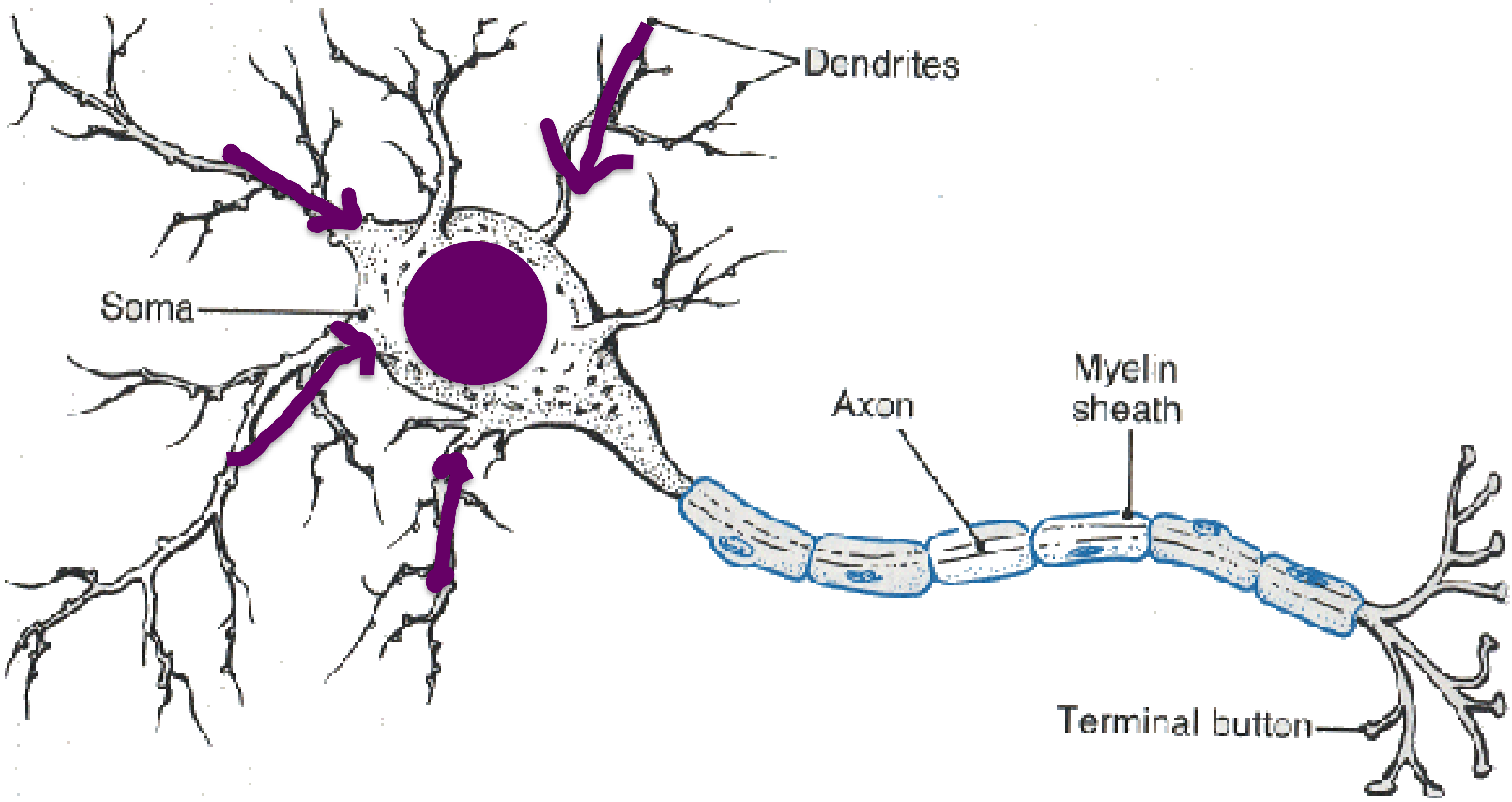
Neuron



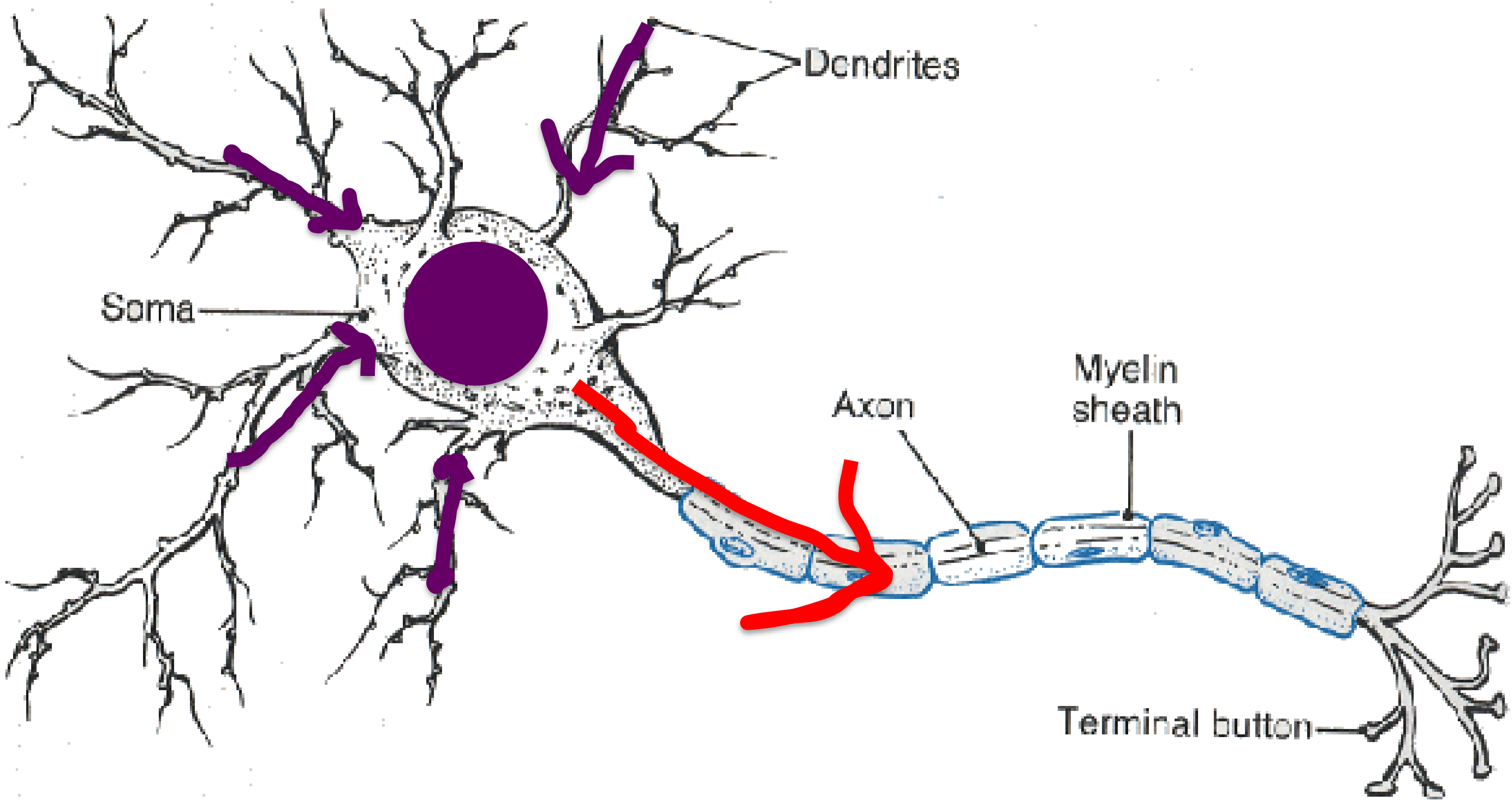
Neuron



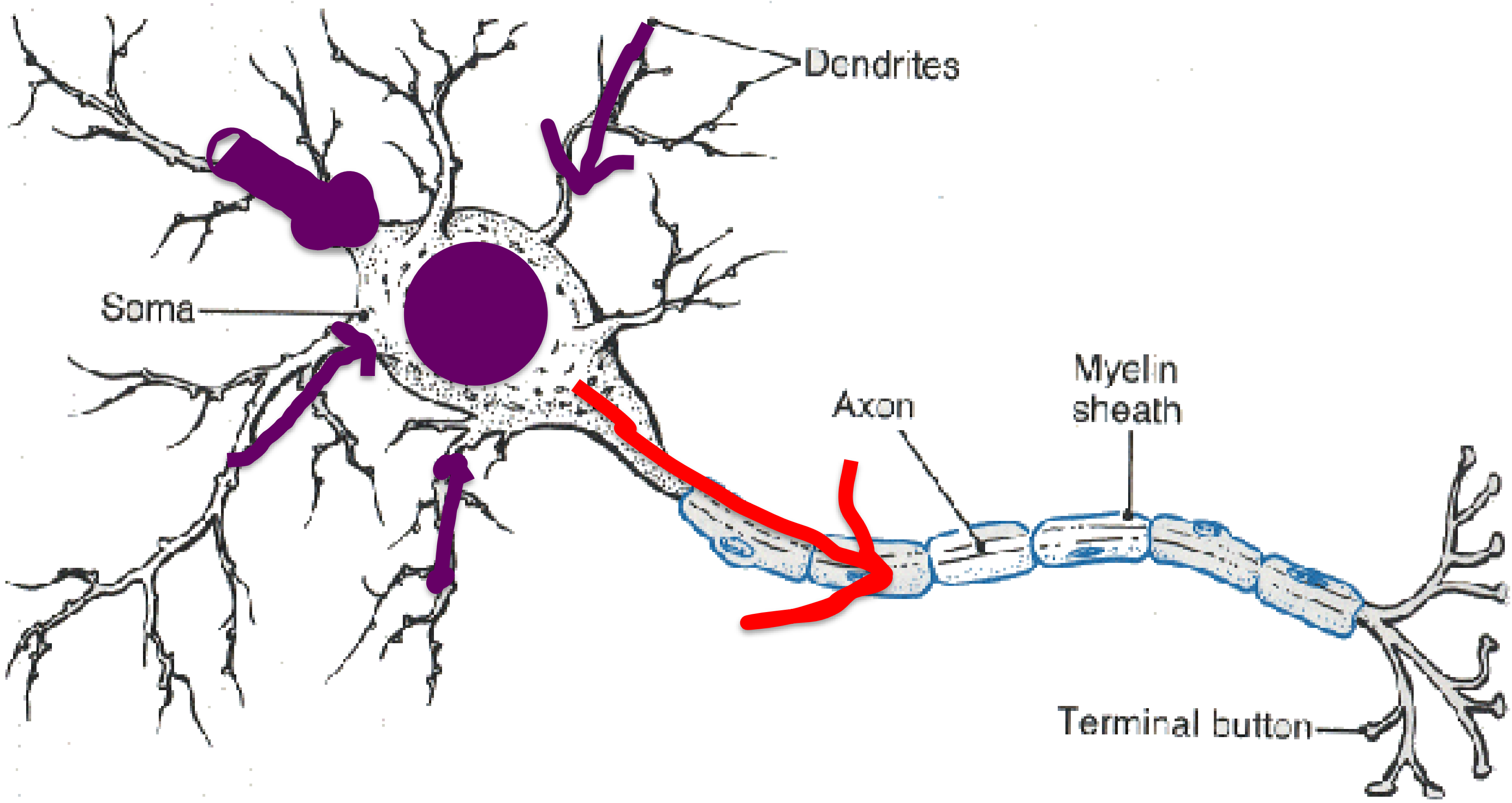
Neuron



Neuron

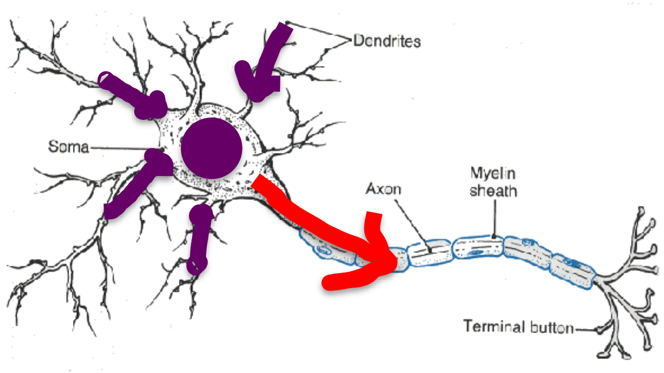


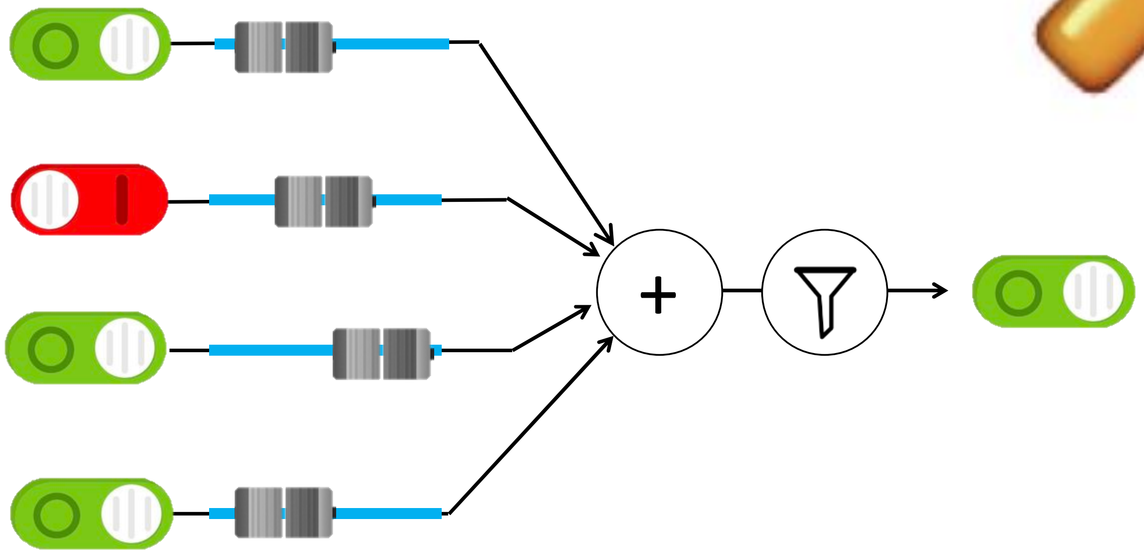
Some Inputs are More Important



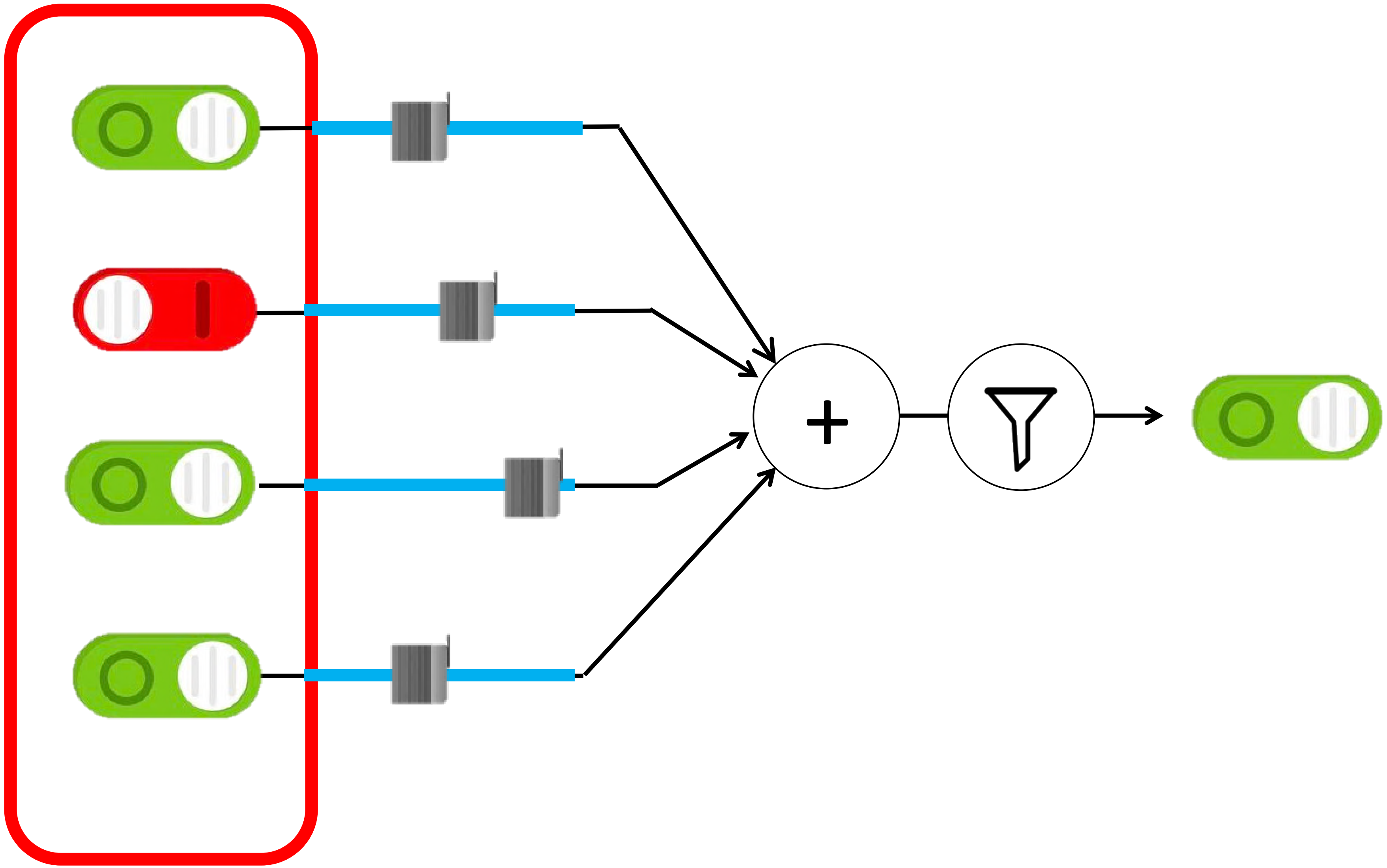
Piech + Woodrow

Artificial Neuron

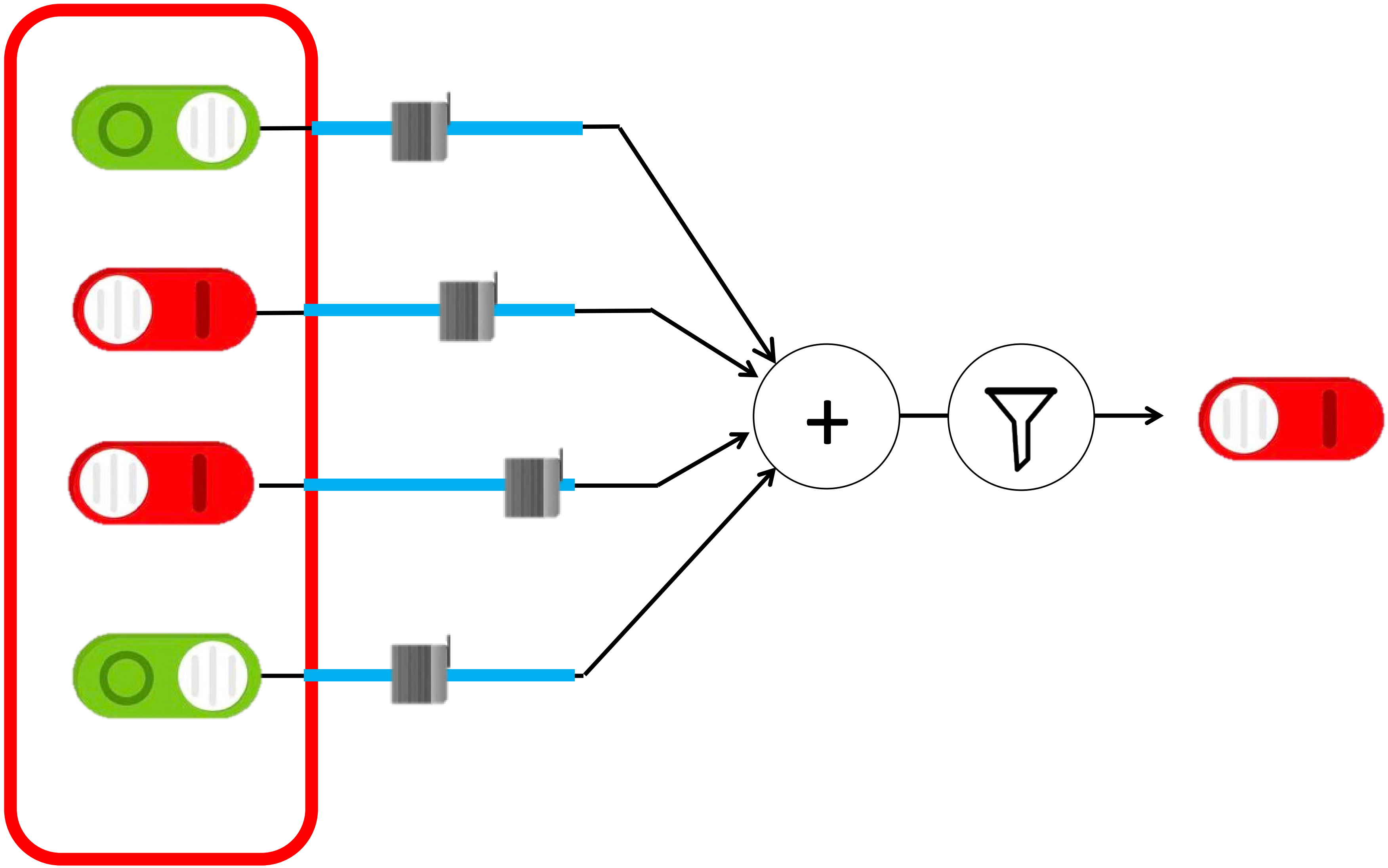




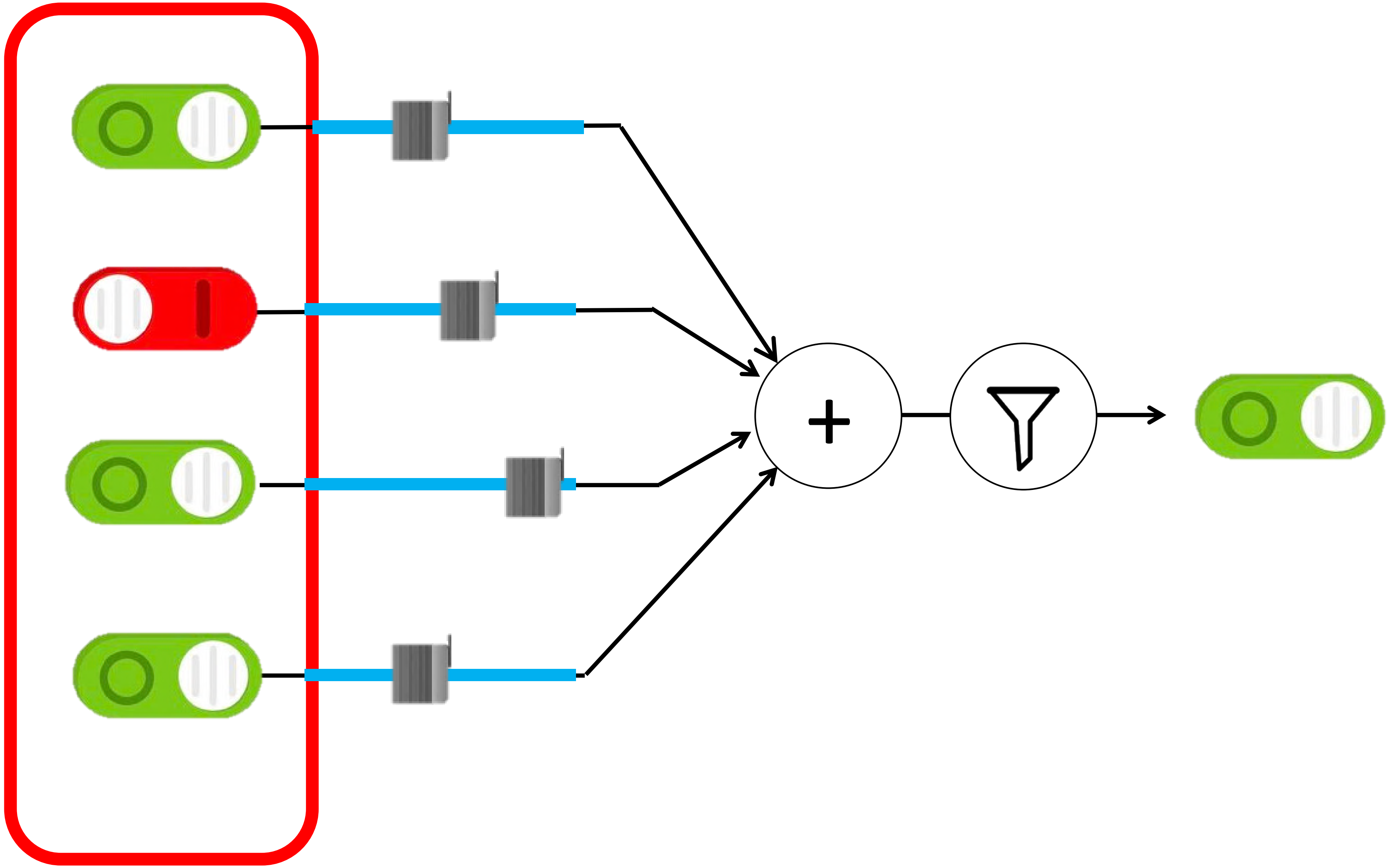
Inputs



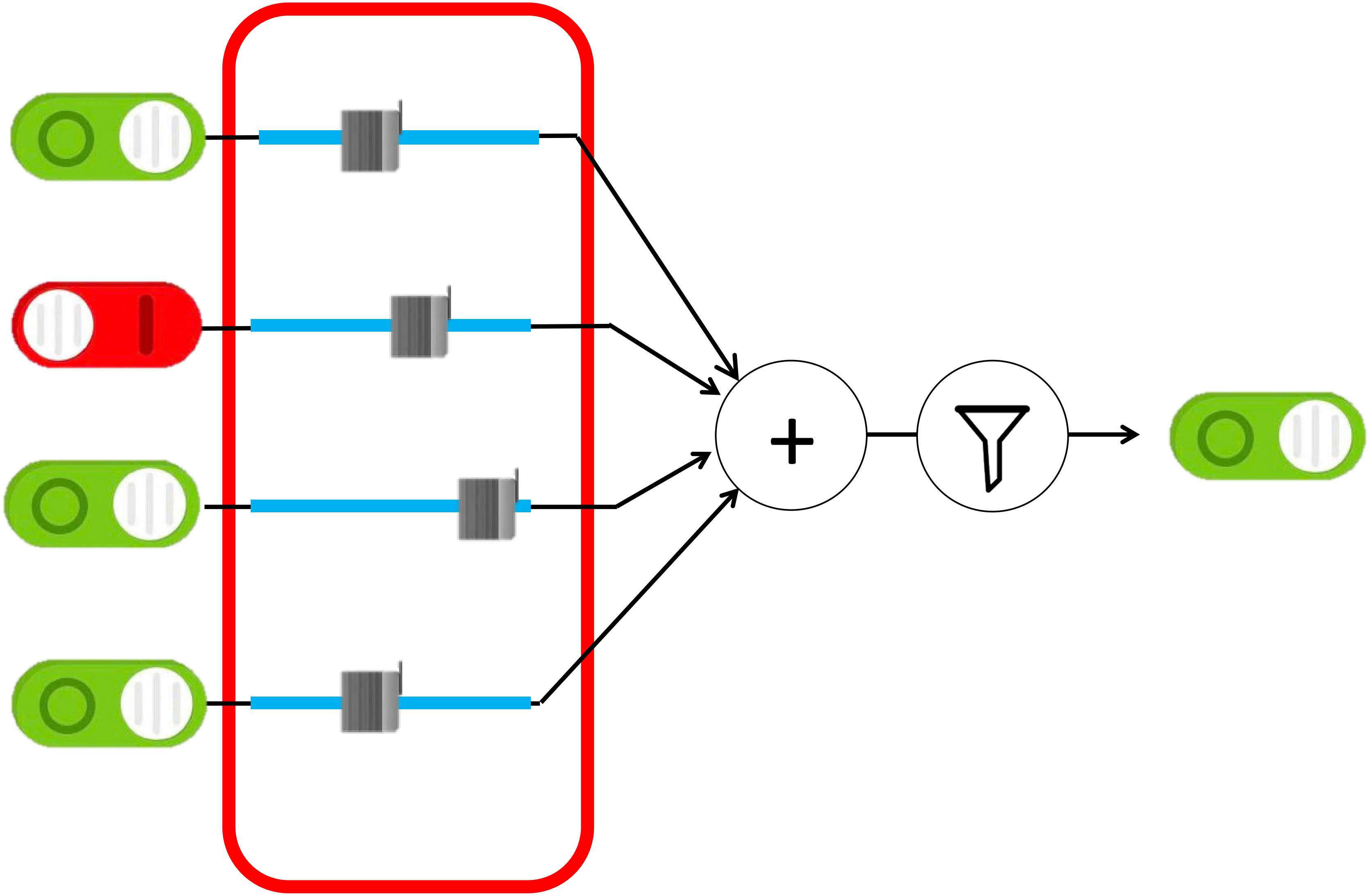
Inputs



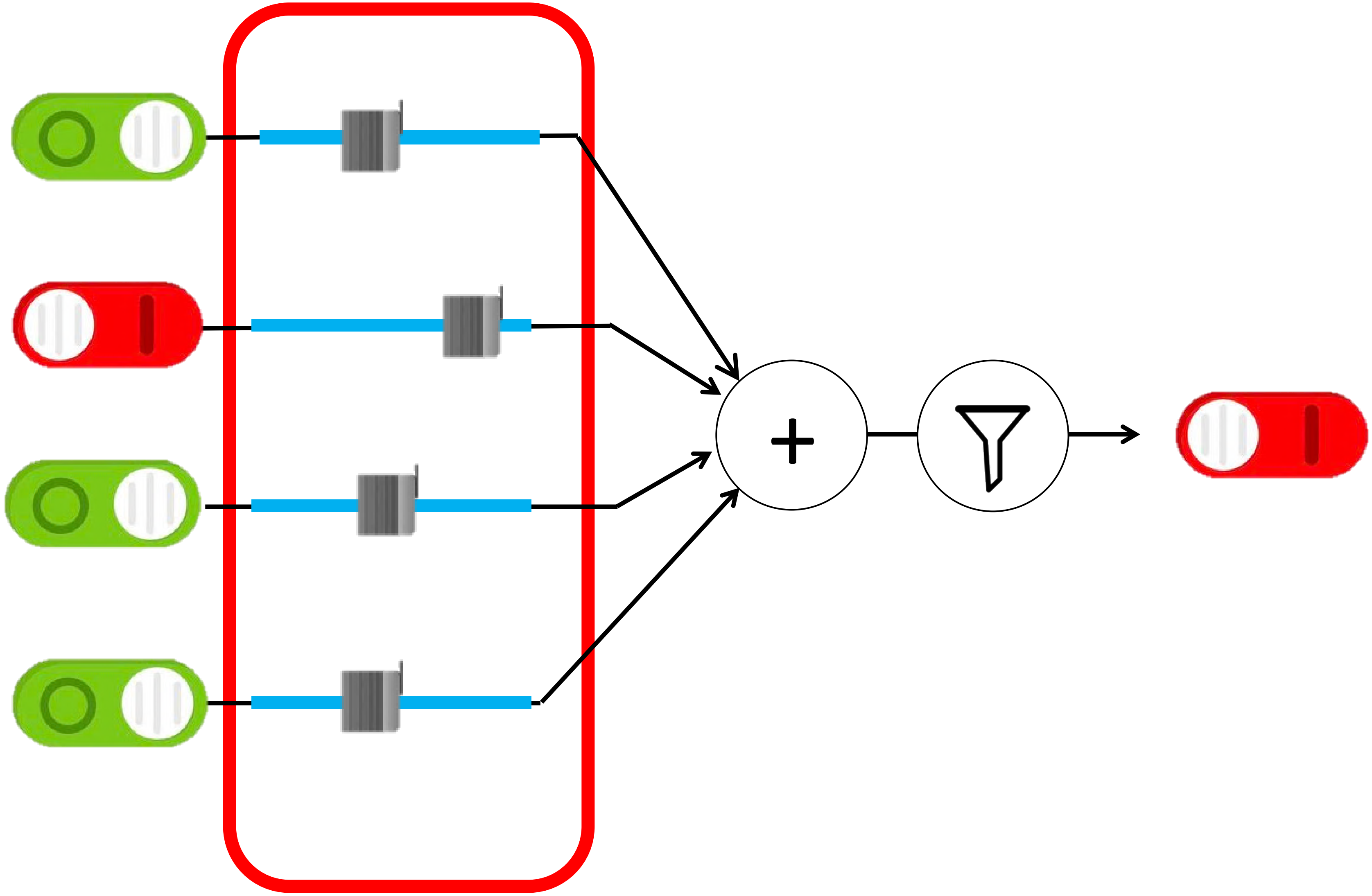
Inputs



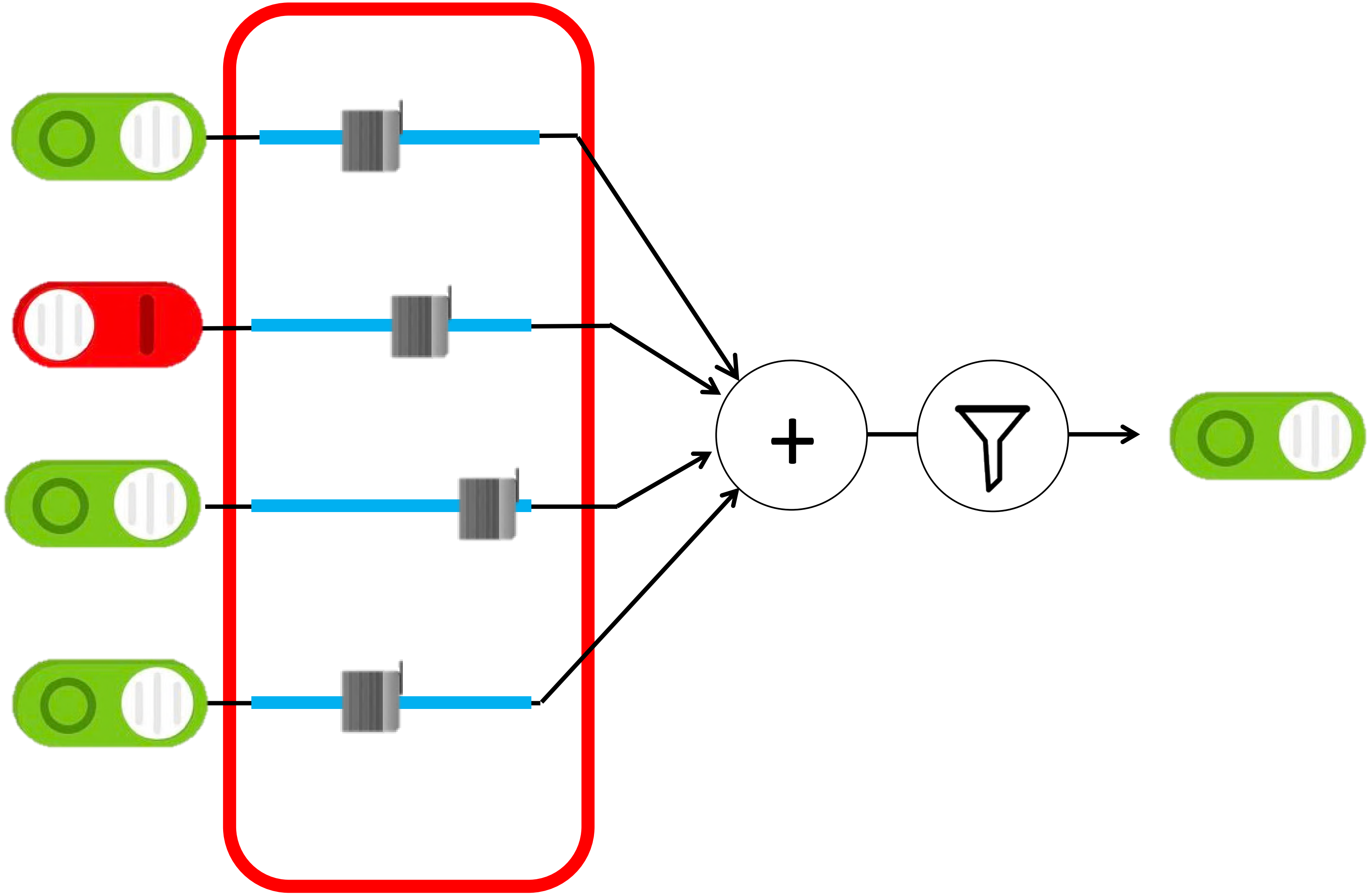
Weights



Weights

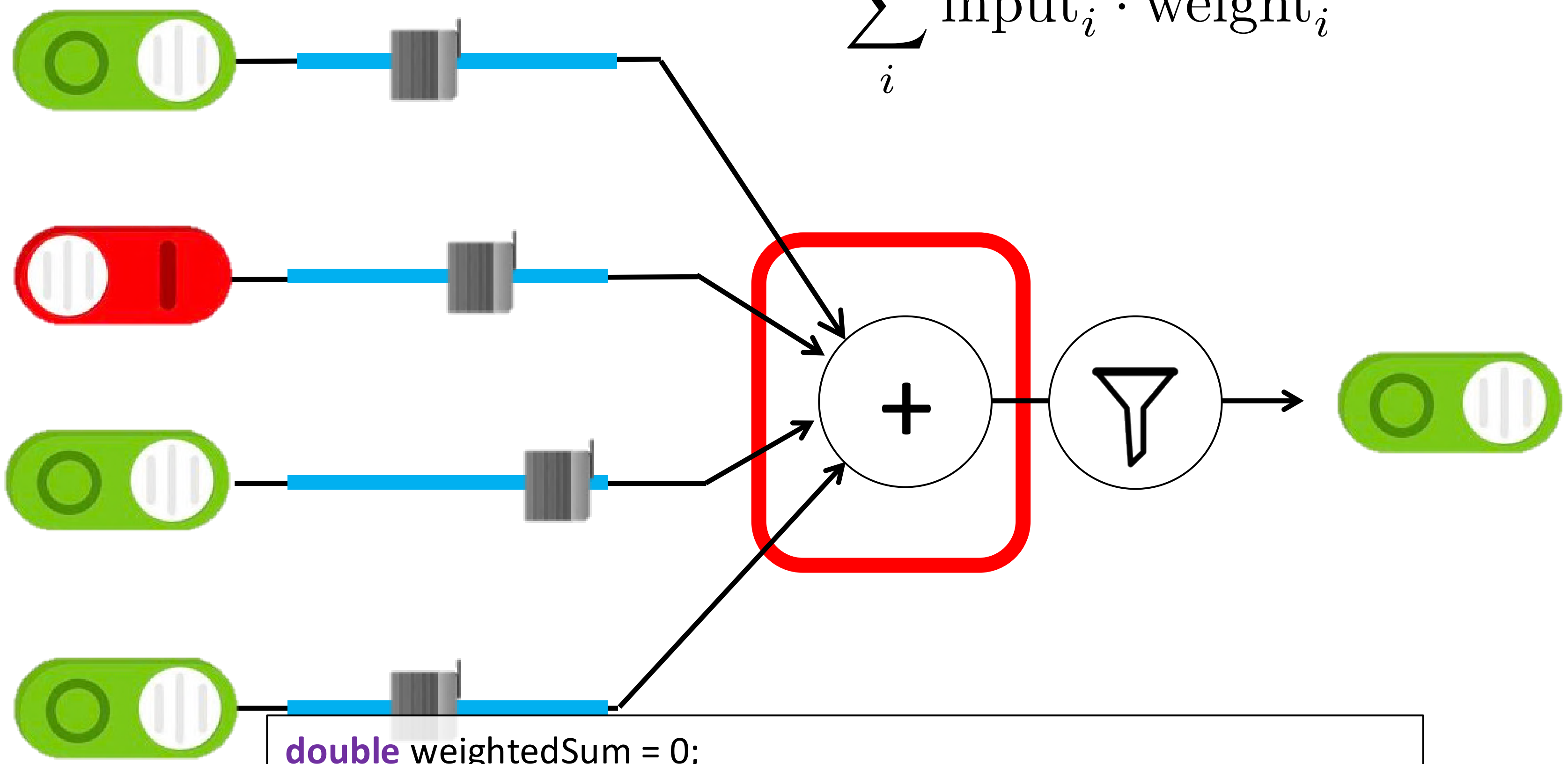


Weights



Weighted Sum

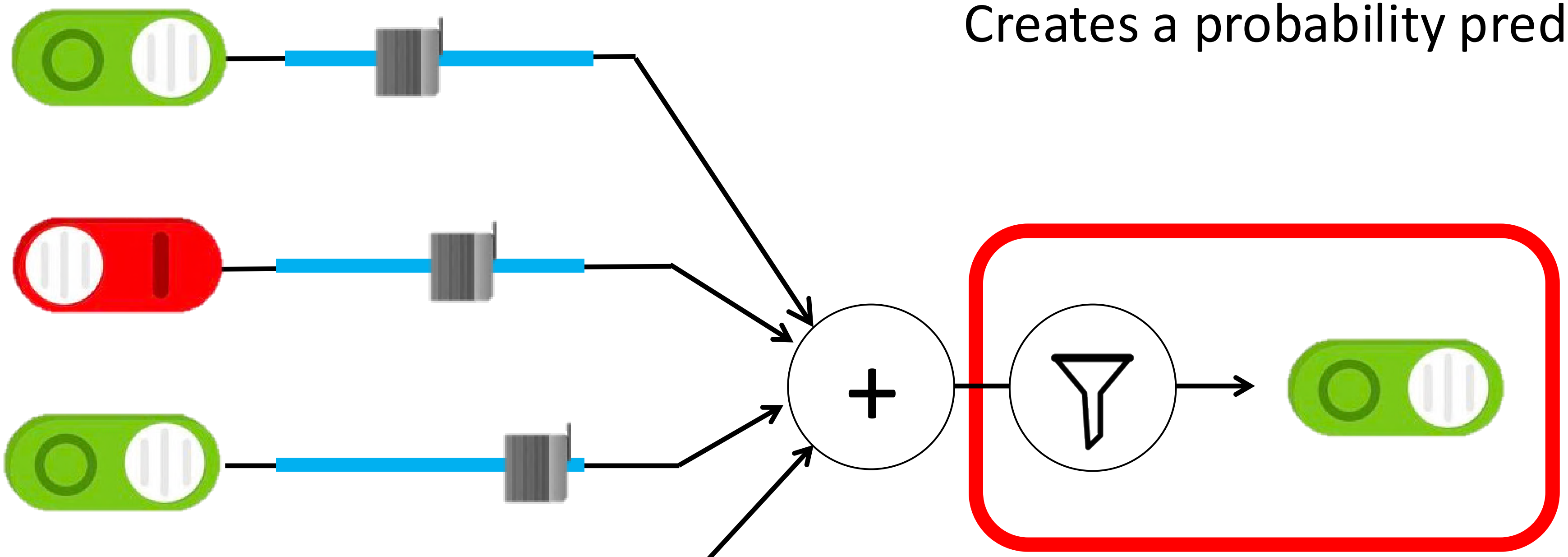
$$\sum_i \text{input}_i \cdot \text{weight}_i$$

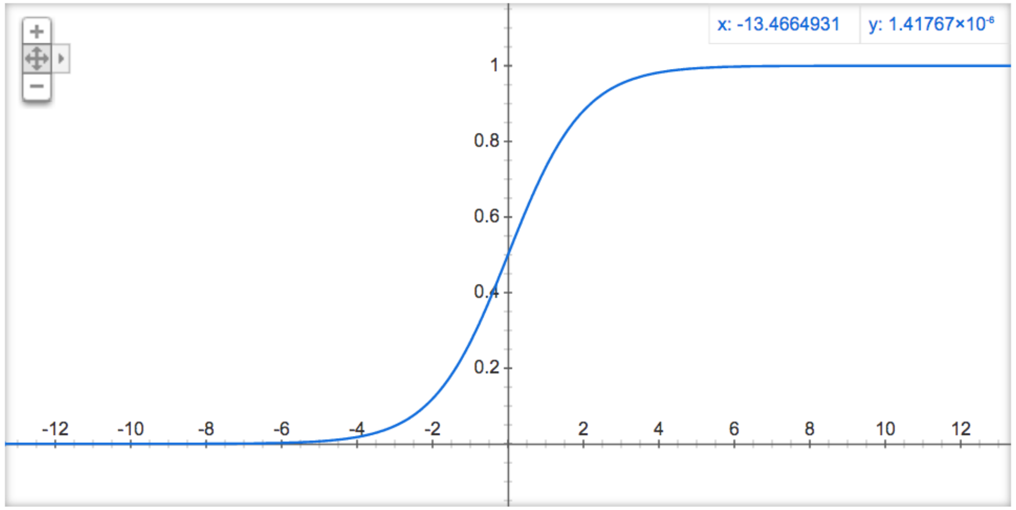


```
double weightedSum = 0;
weightedSum += input0 * weight0;
weightedSum += input1 * weight1;
weightedSum += input2 * weight2;
weightedSum += input3 * weight3;
```

Filter and Output

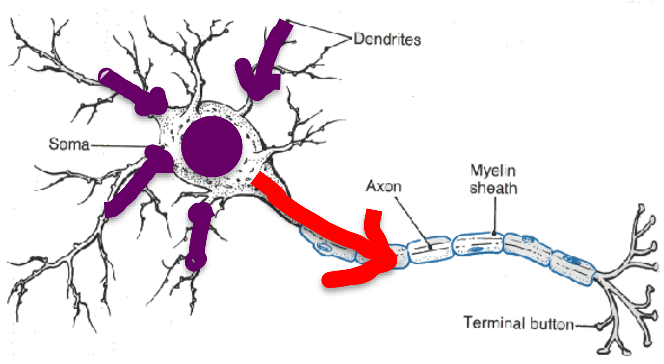
Creates a probability prediction

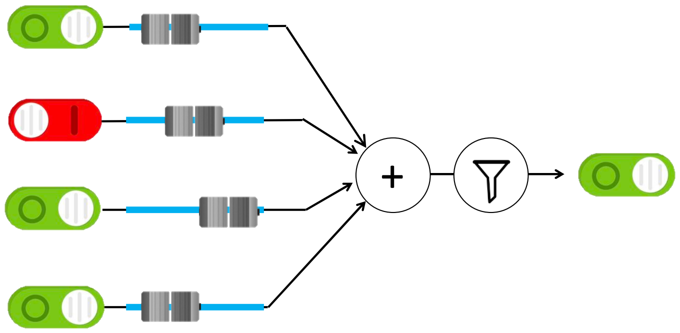




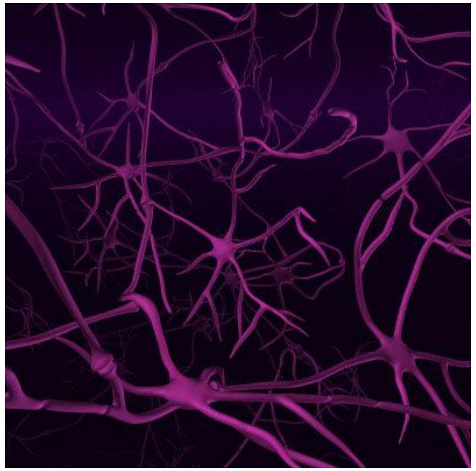
Biological Basis for Neural Networks

- A neuron





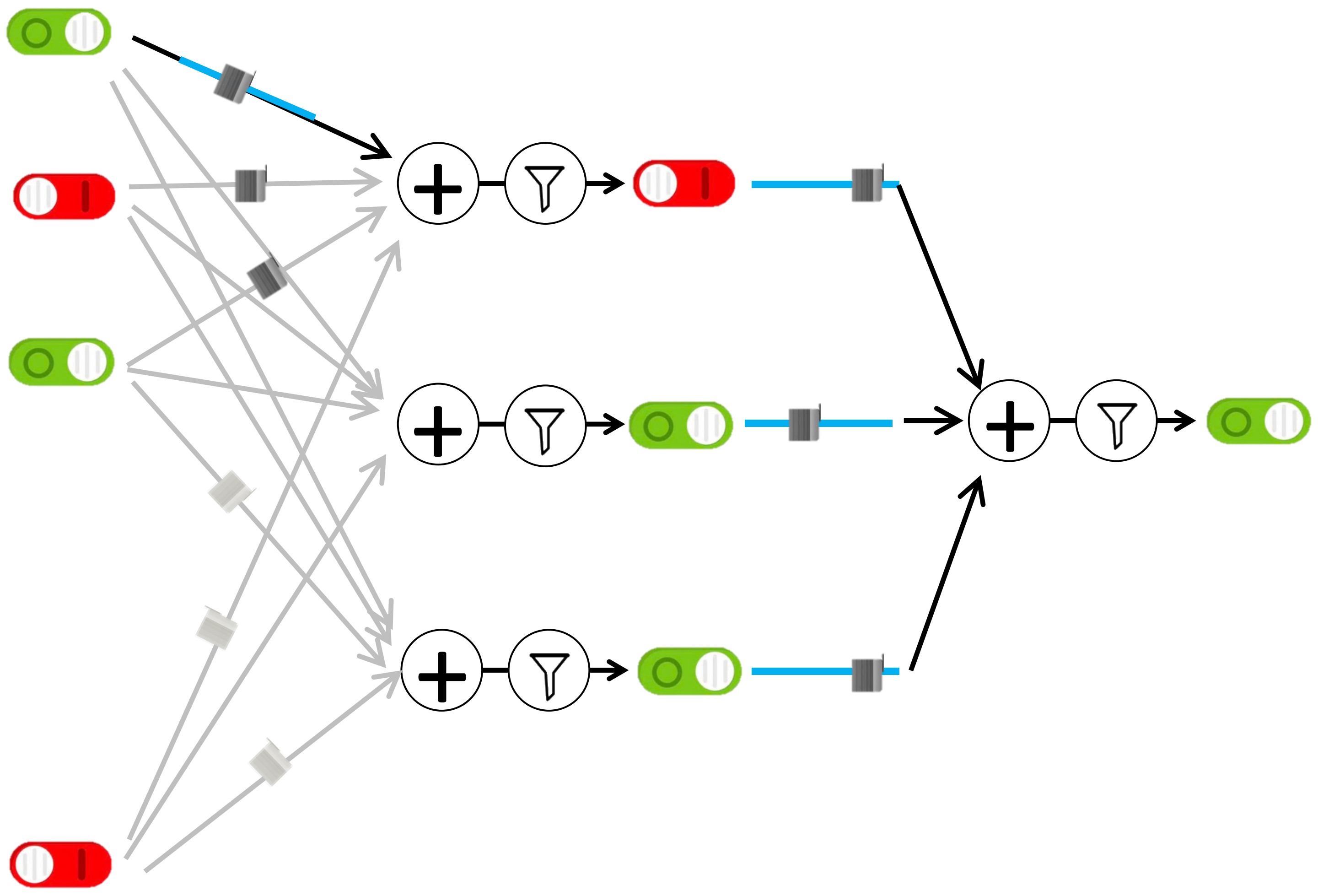
- Your brain



???

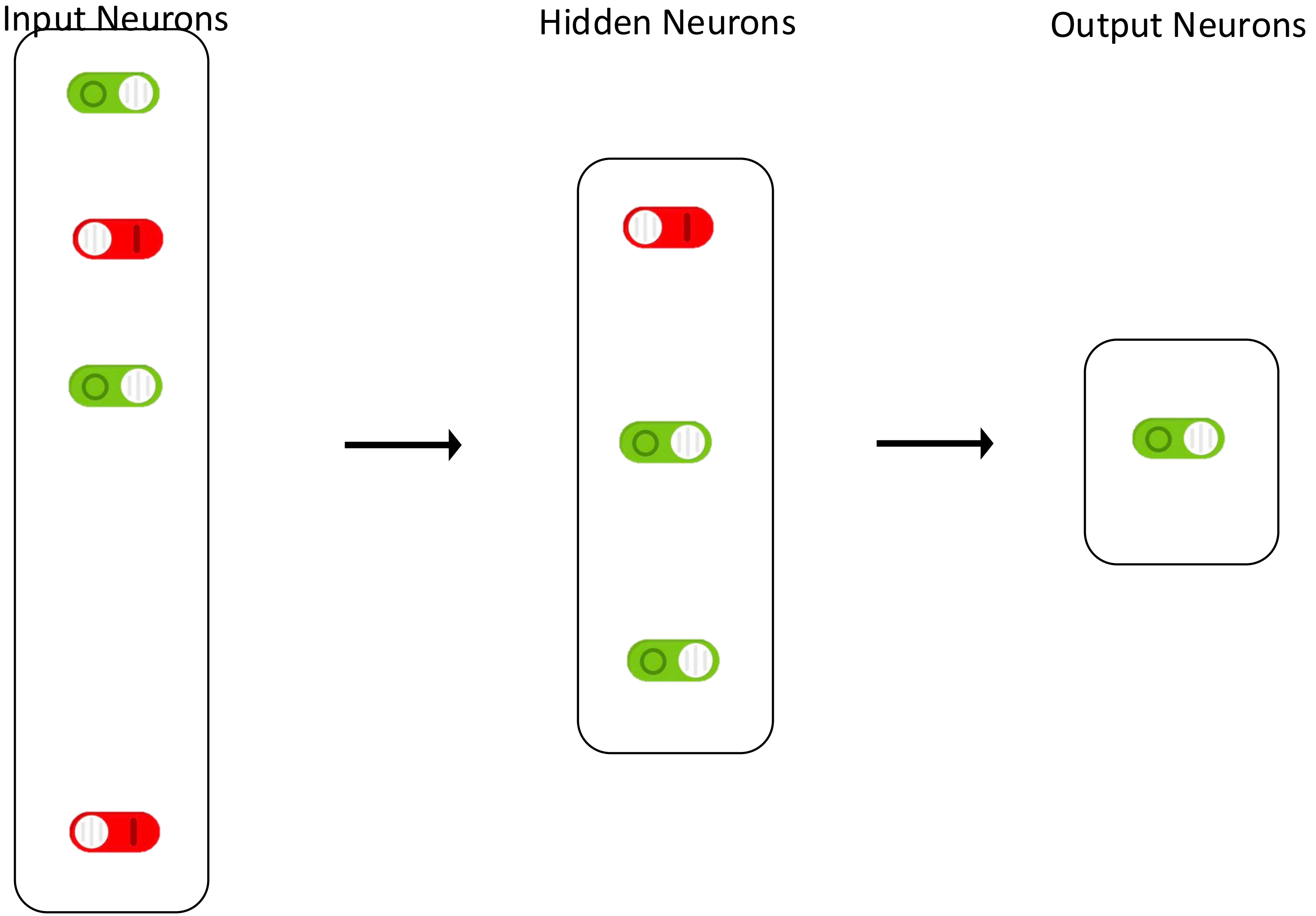
Actually, it's probably someone else's brain

Put Many Together

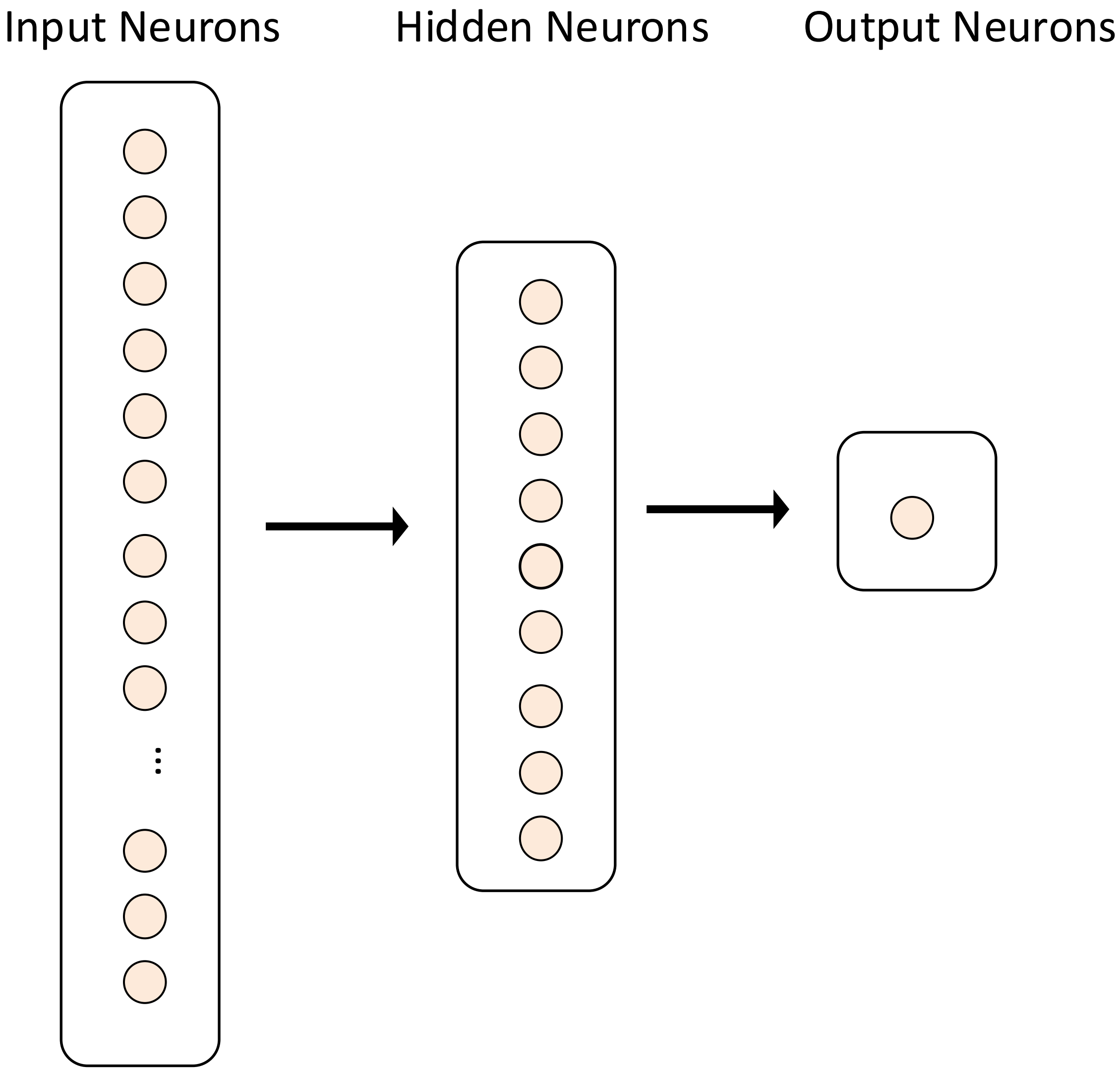


Piech + Woodrow

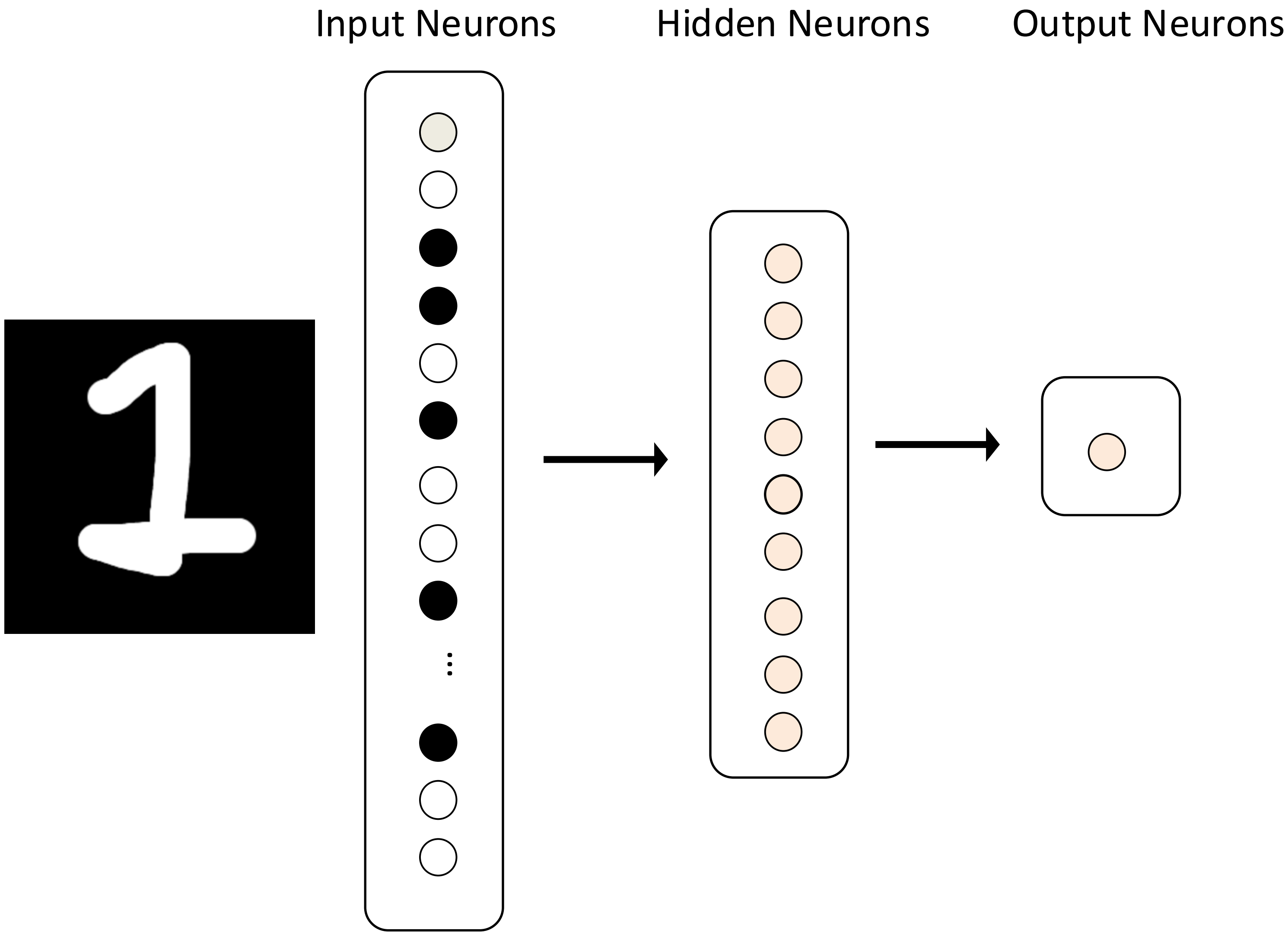
Put Many Together



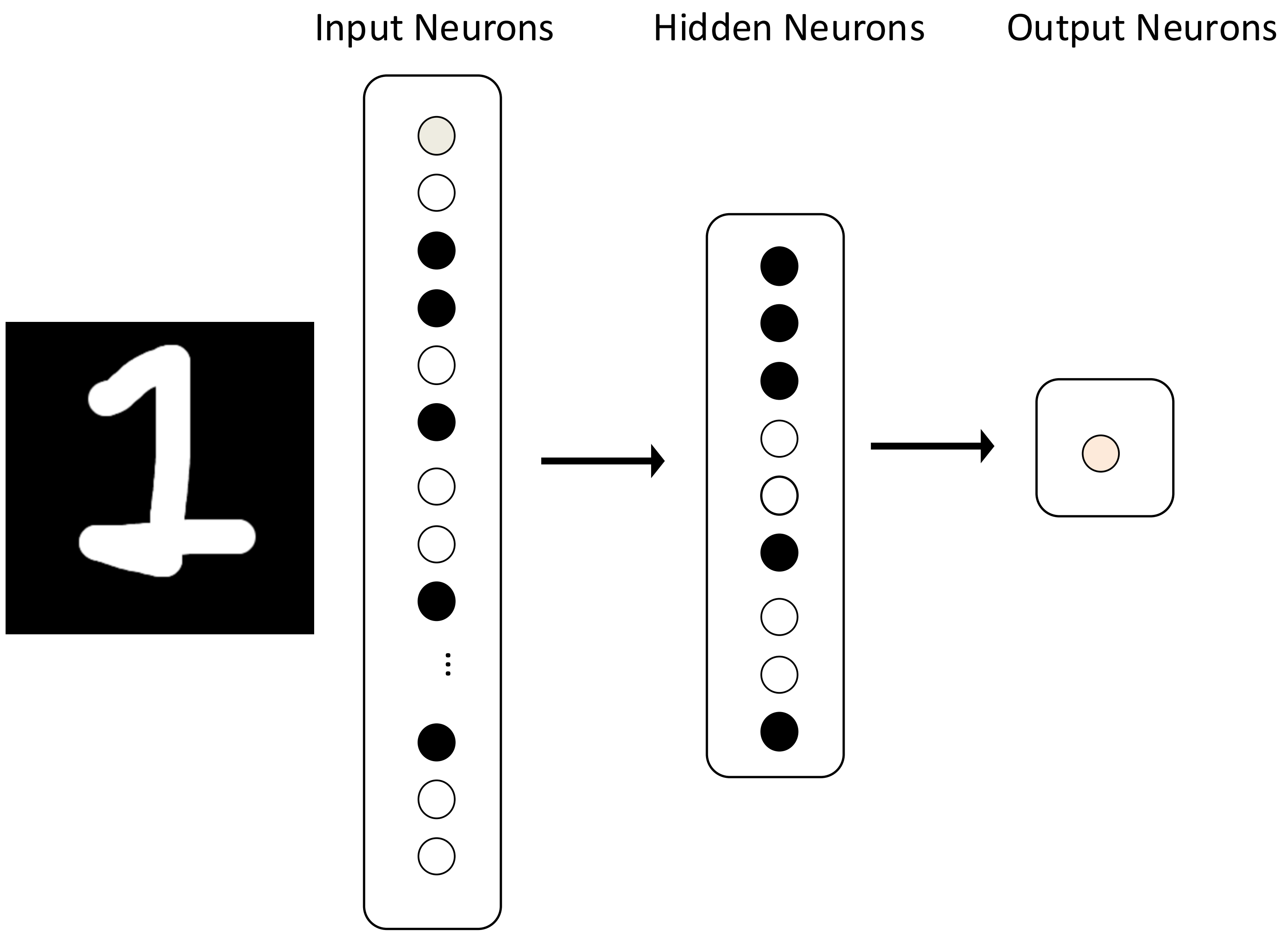
Making a Prediction



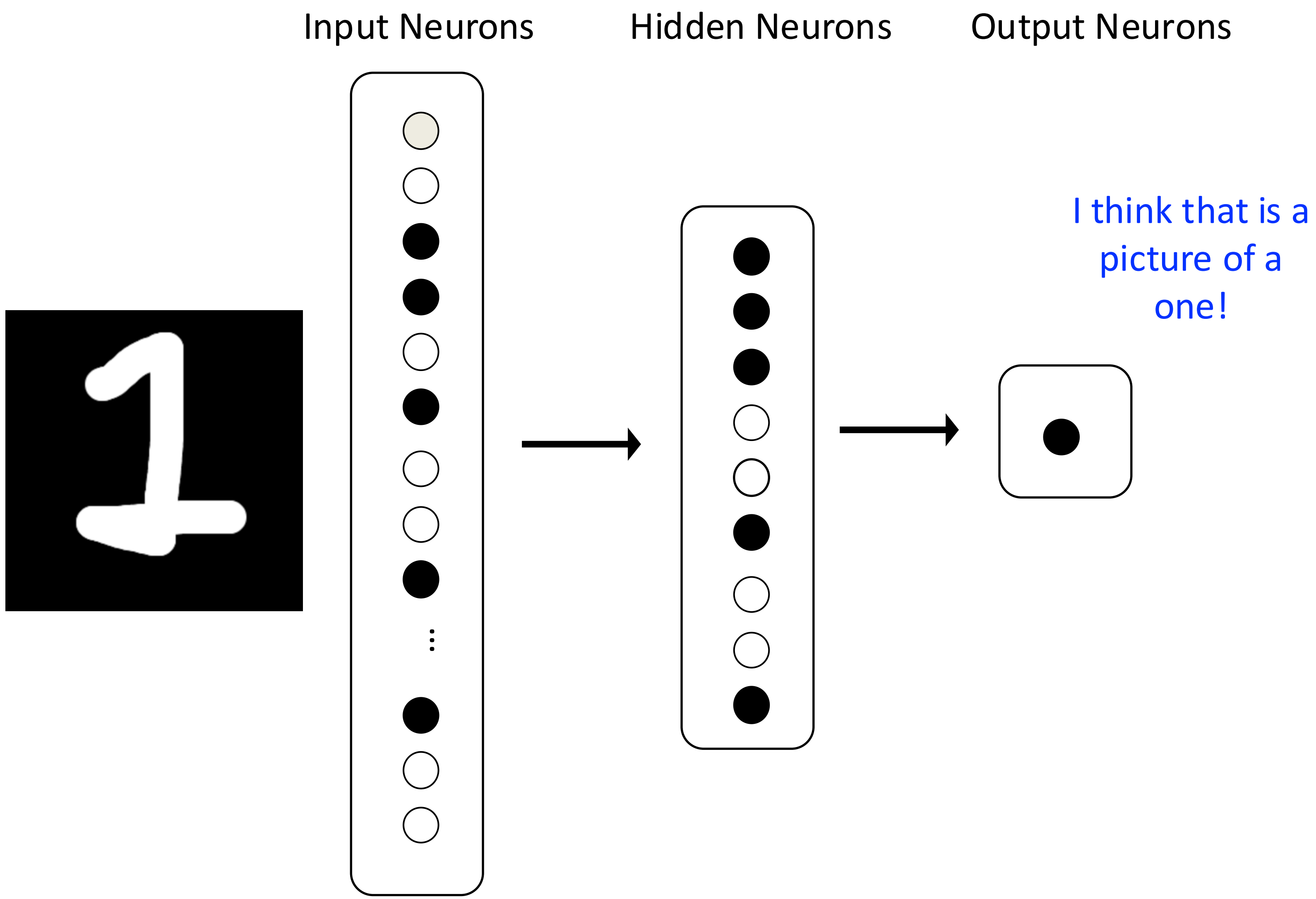
Making a Prediction



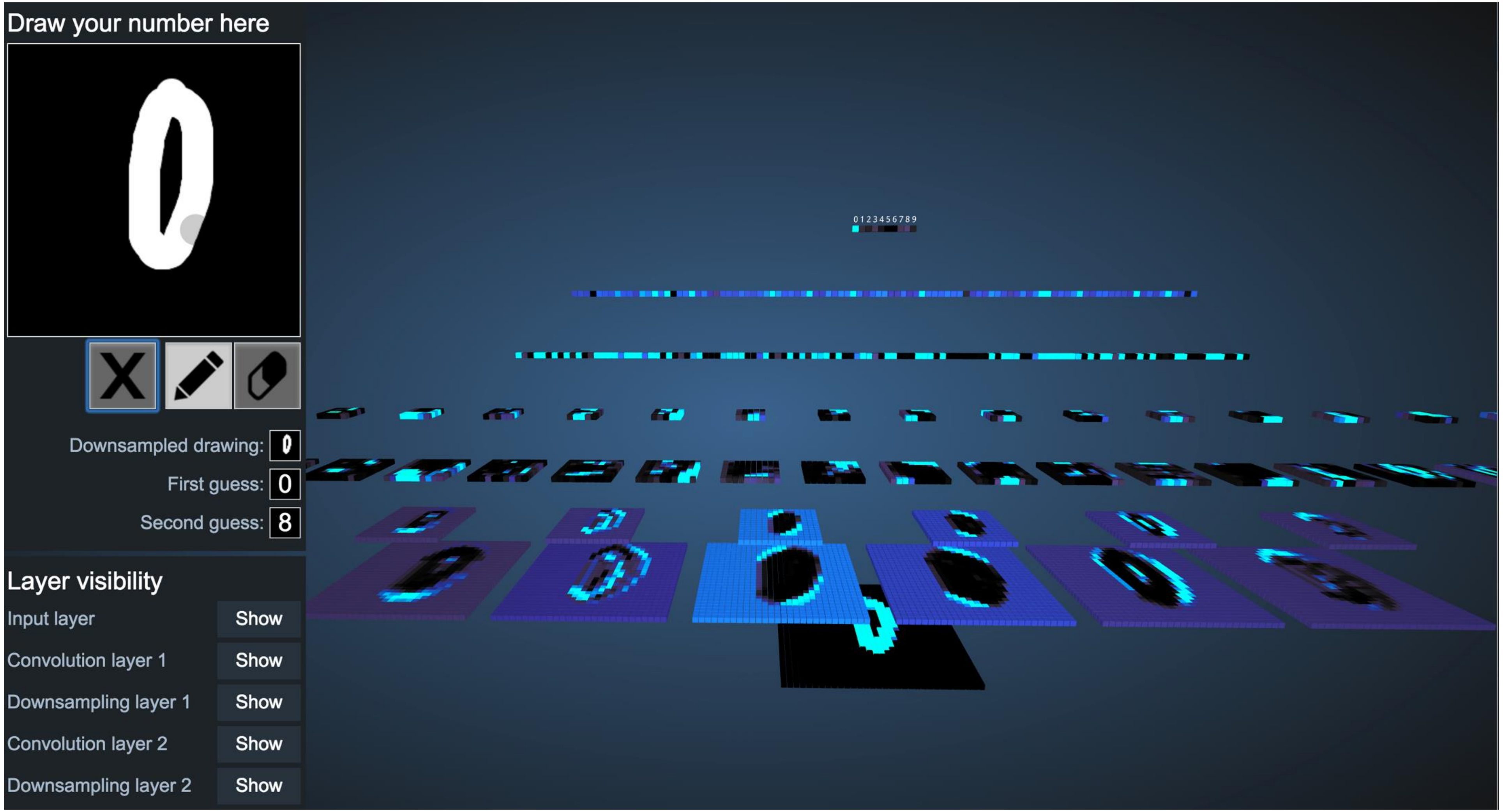
Making a Prediction



Making a Prediction

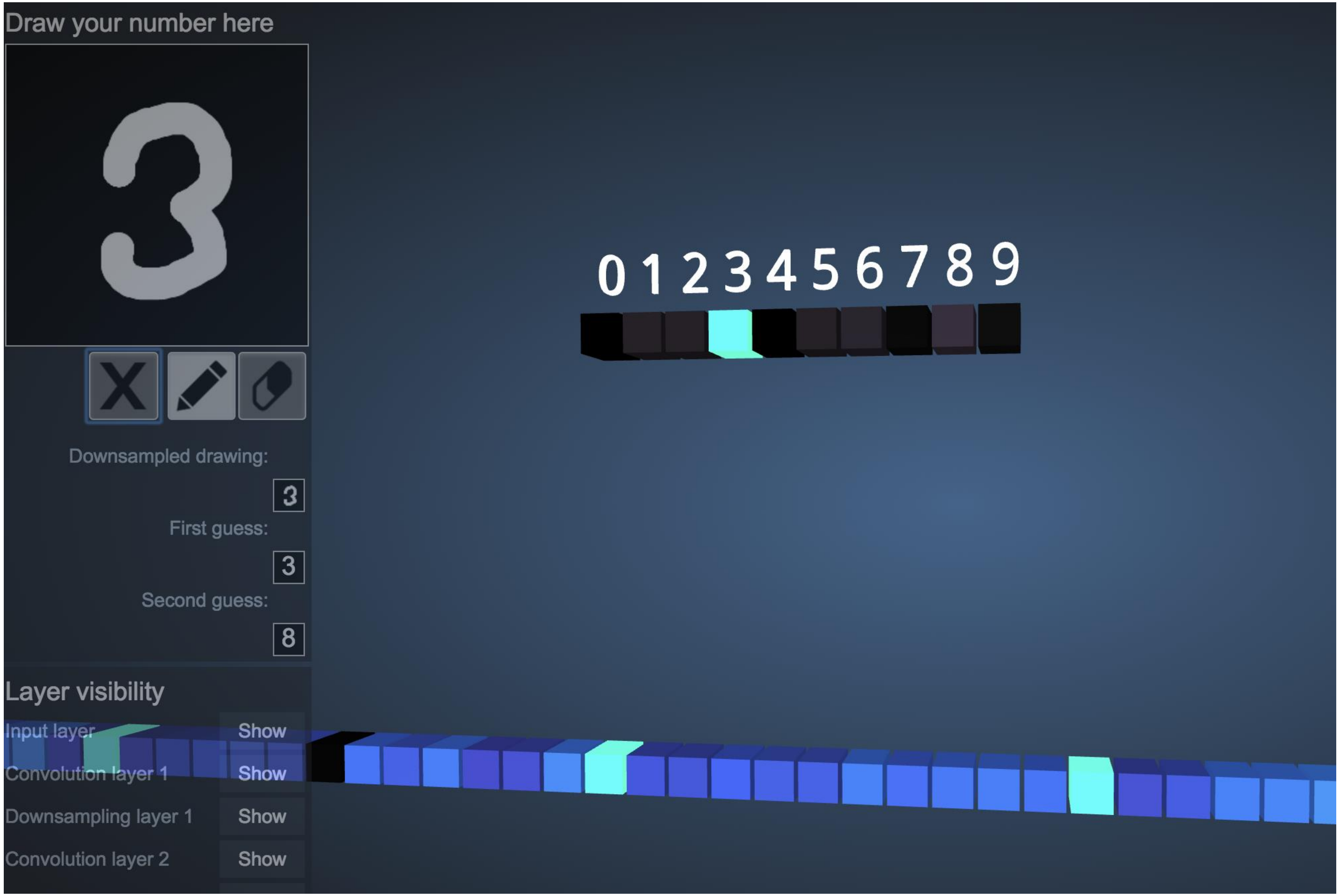


Demonstration

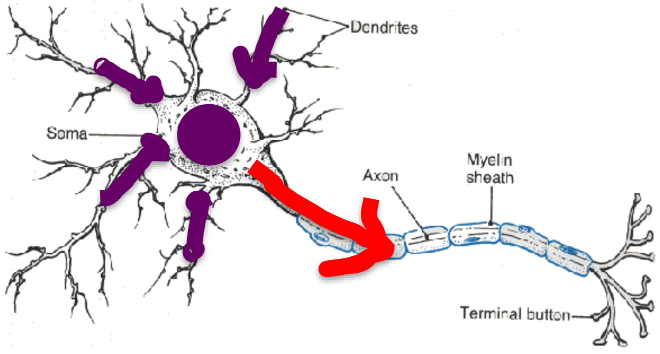


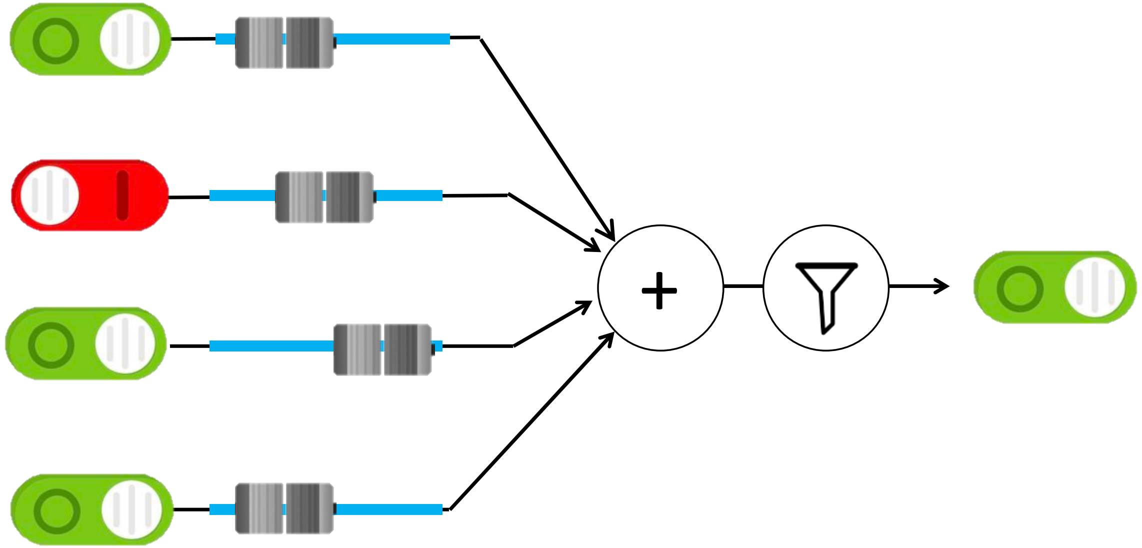
https://adamharley.com/nn_vis/cnn/2d.html

Interpret the Output as Prediction

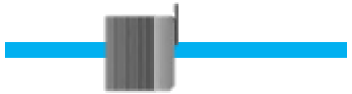


Great Idea: Artificial Neurons





Where do Artificial Neural Networks get their intelligence from?



Neural Networks get their intelligence from their sliders (parameters)

Two Great Ideas

1. Artificial Neurons

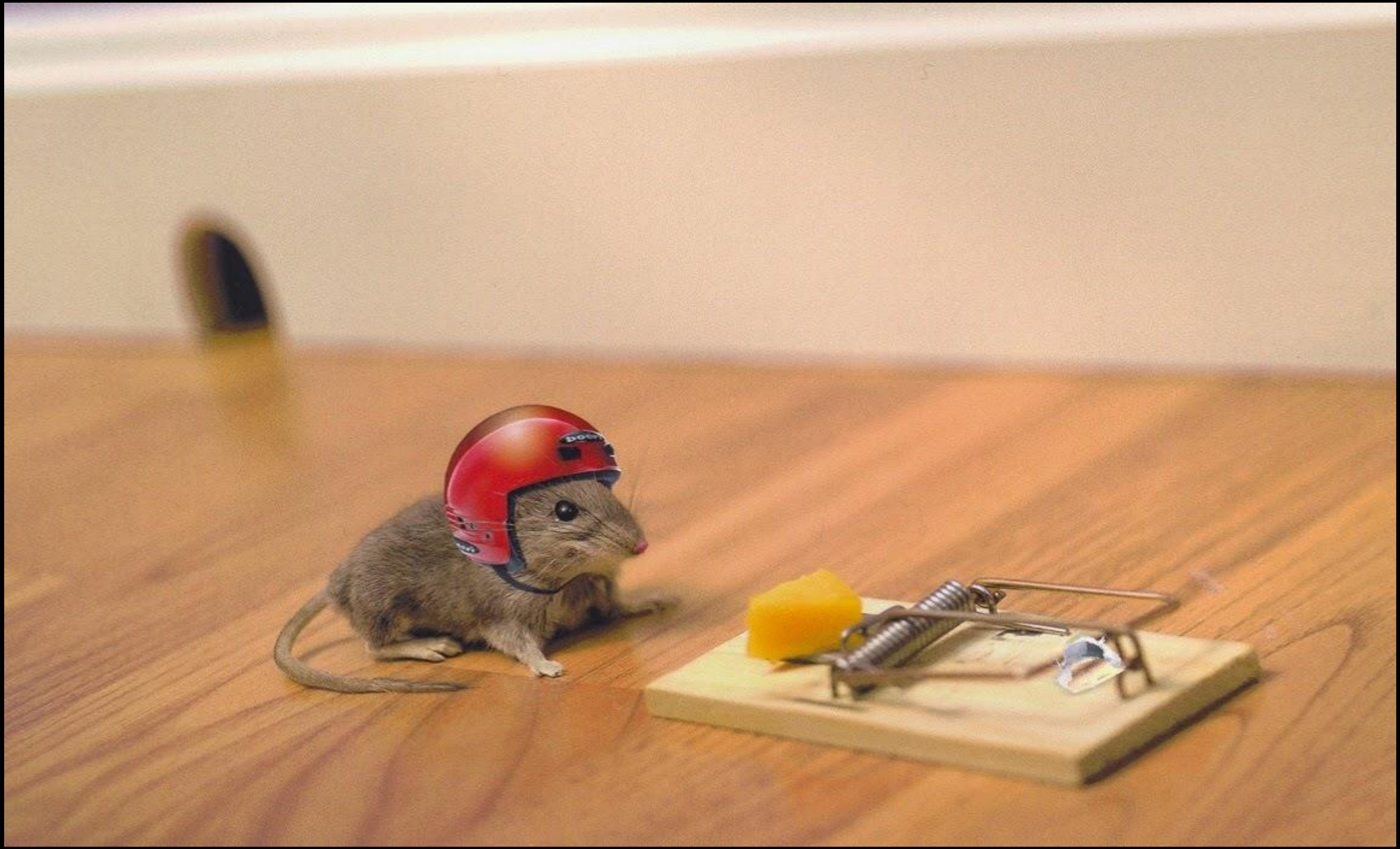
2. Learn by Example

Two Great Ideas

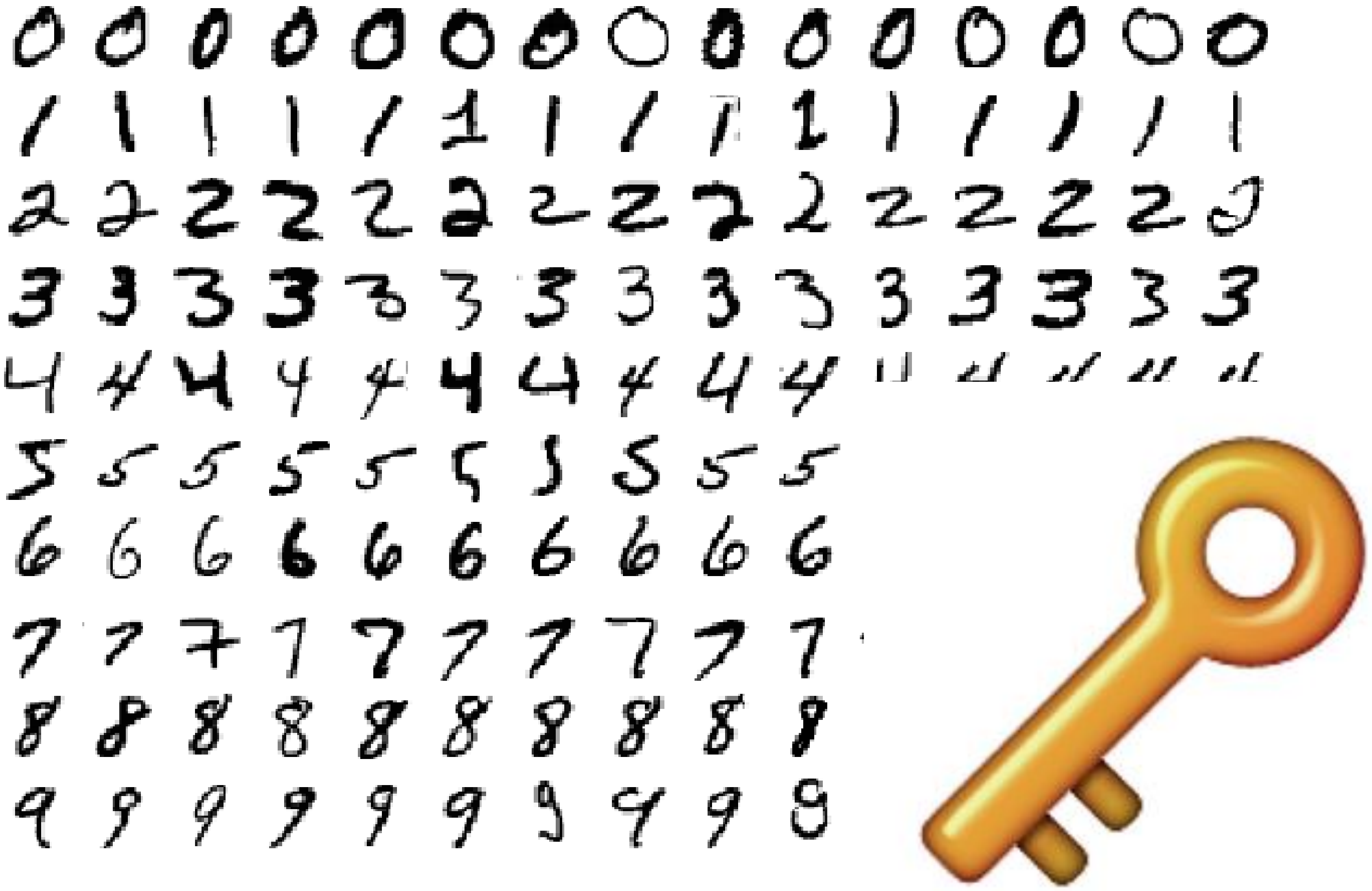
1. Artificial Neurons

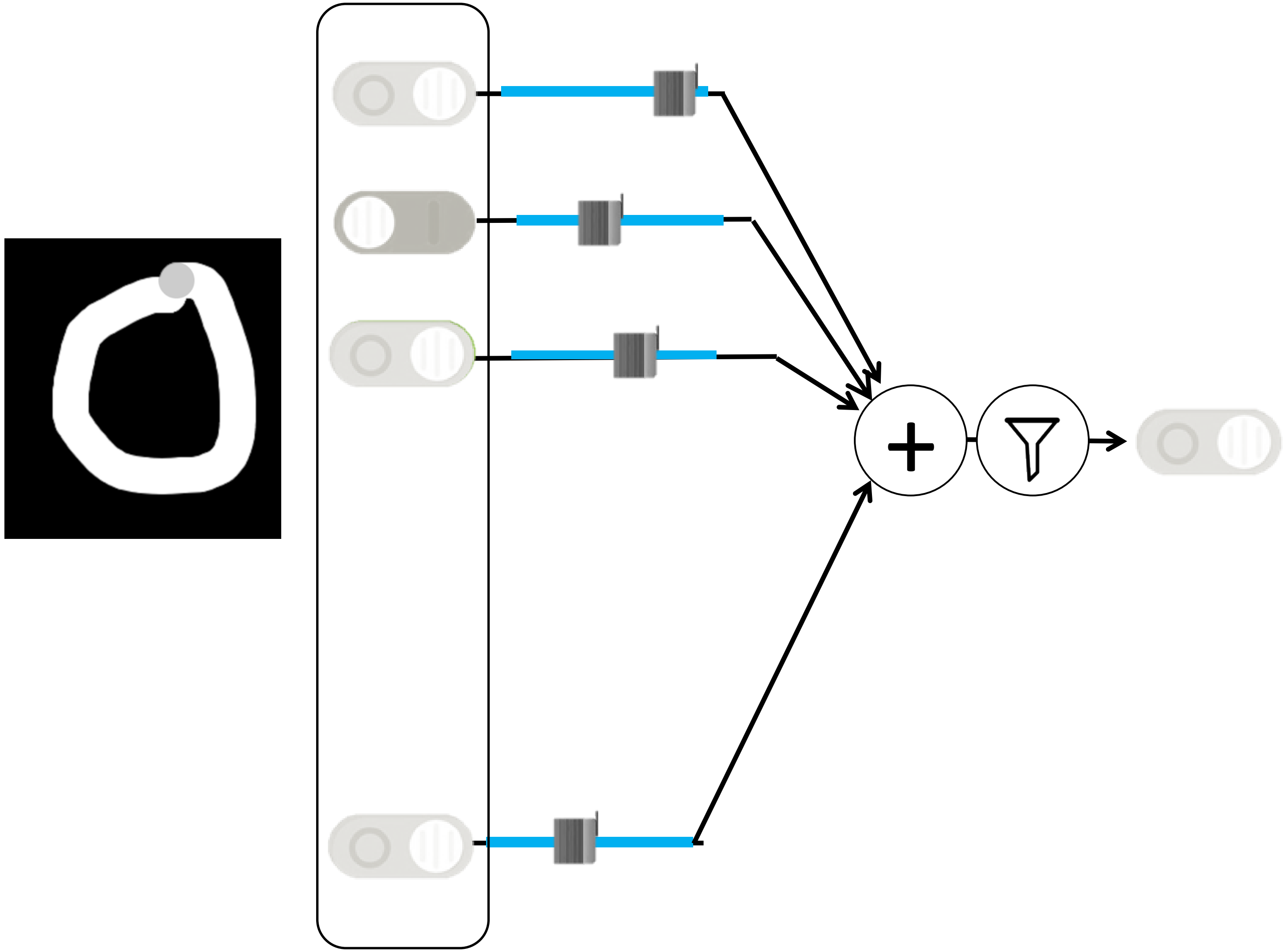
2. Learn by Example

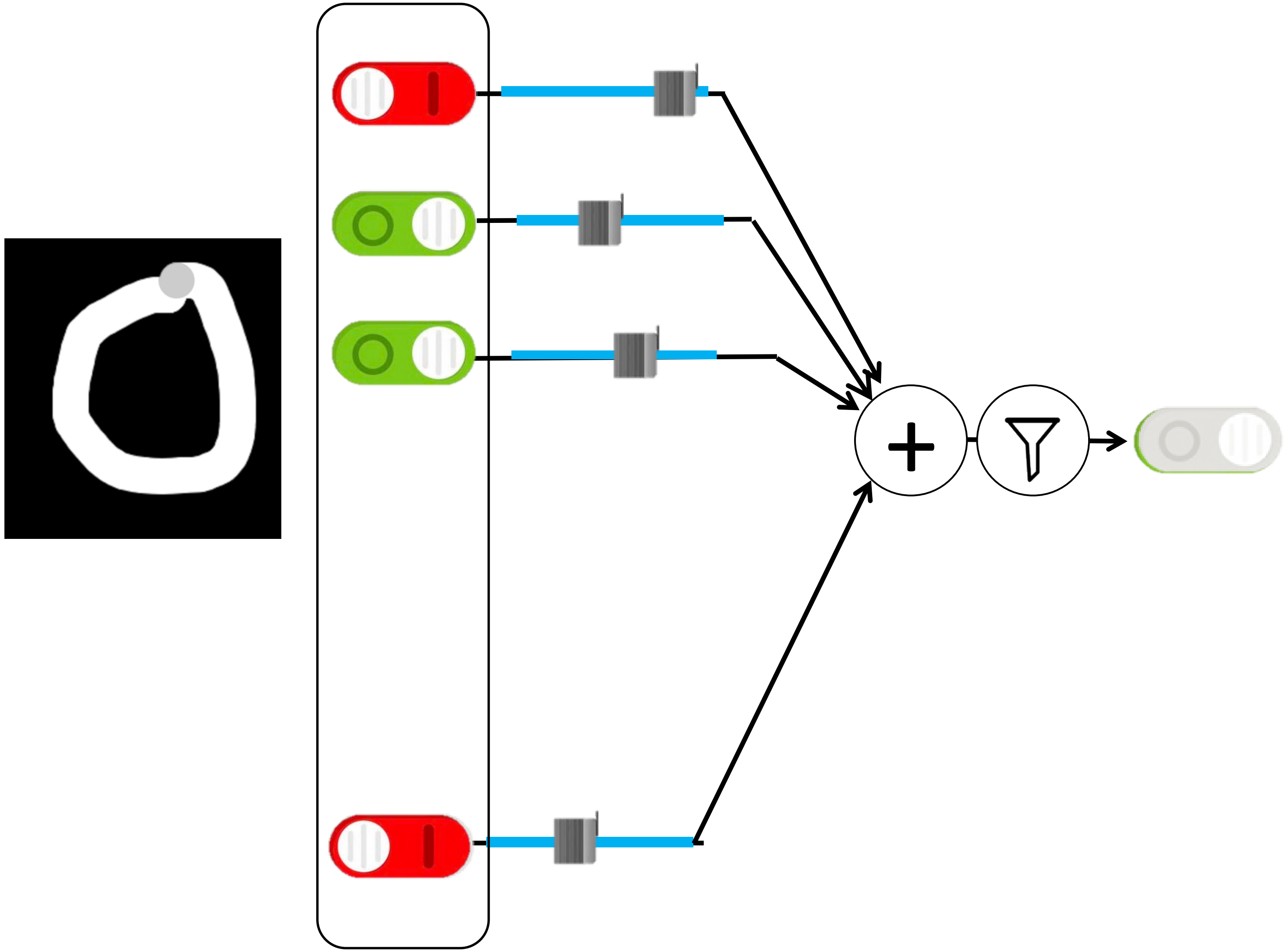
2. Learn From Experience

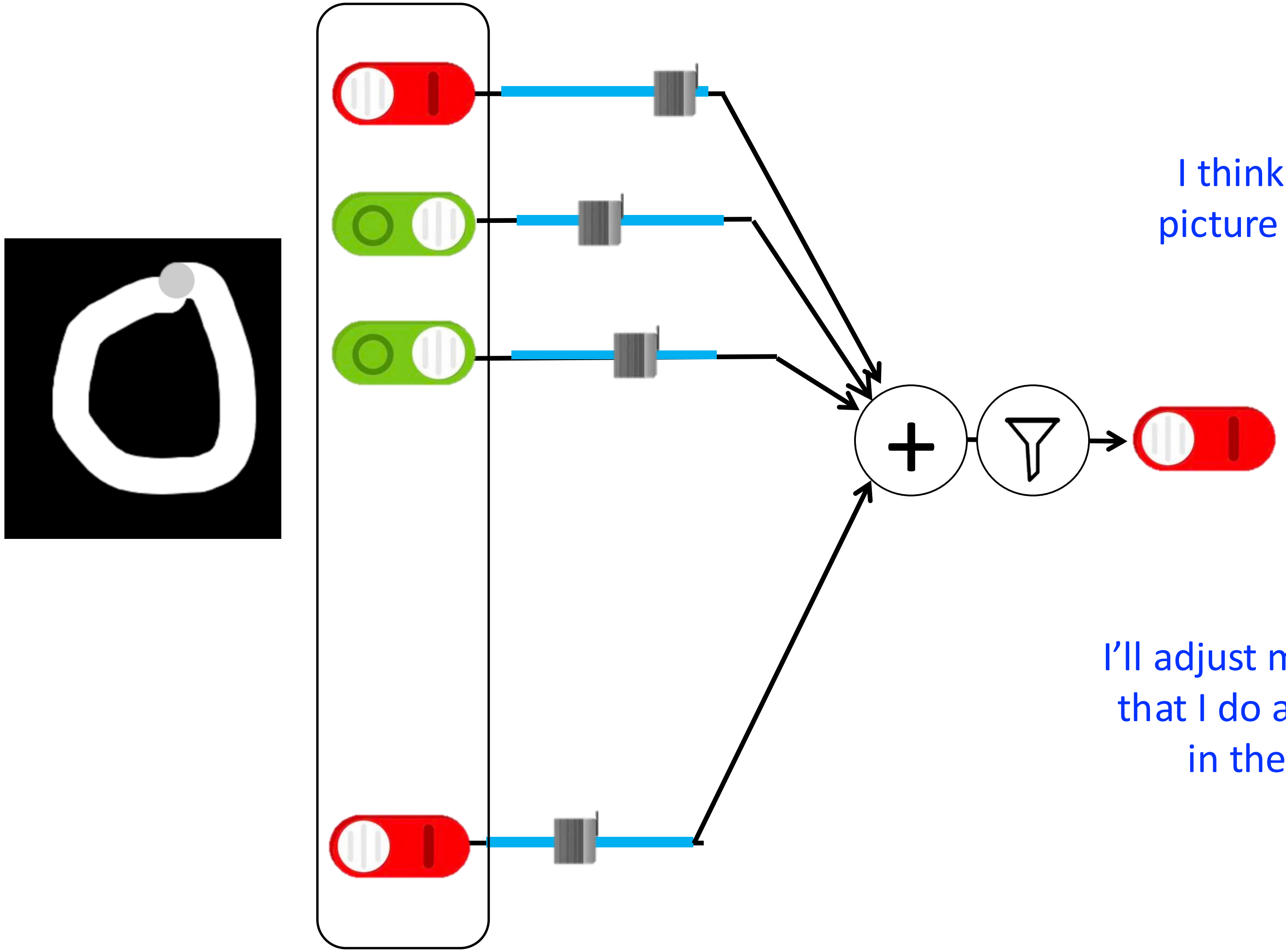


Learn by Example





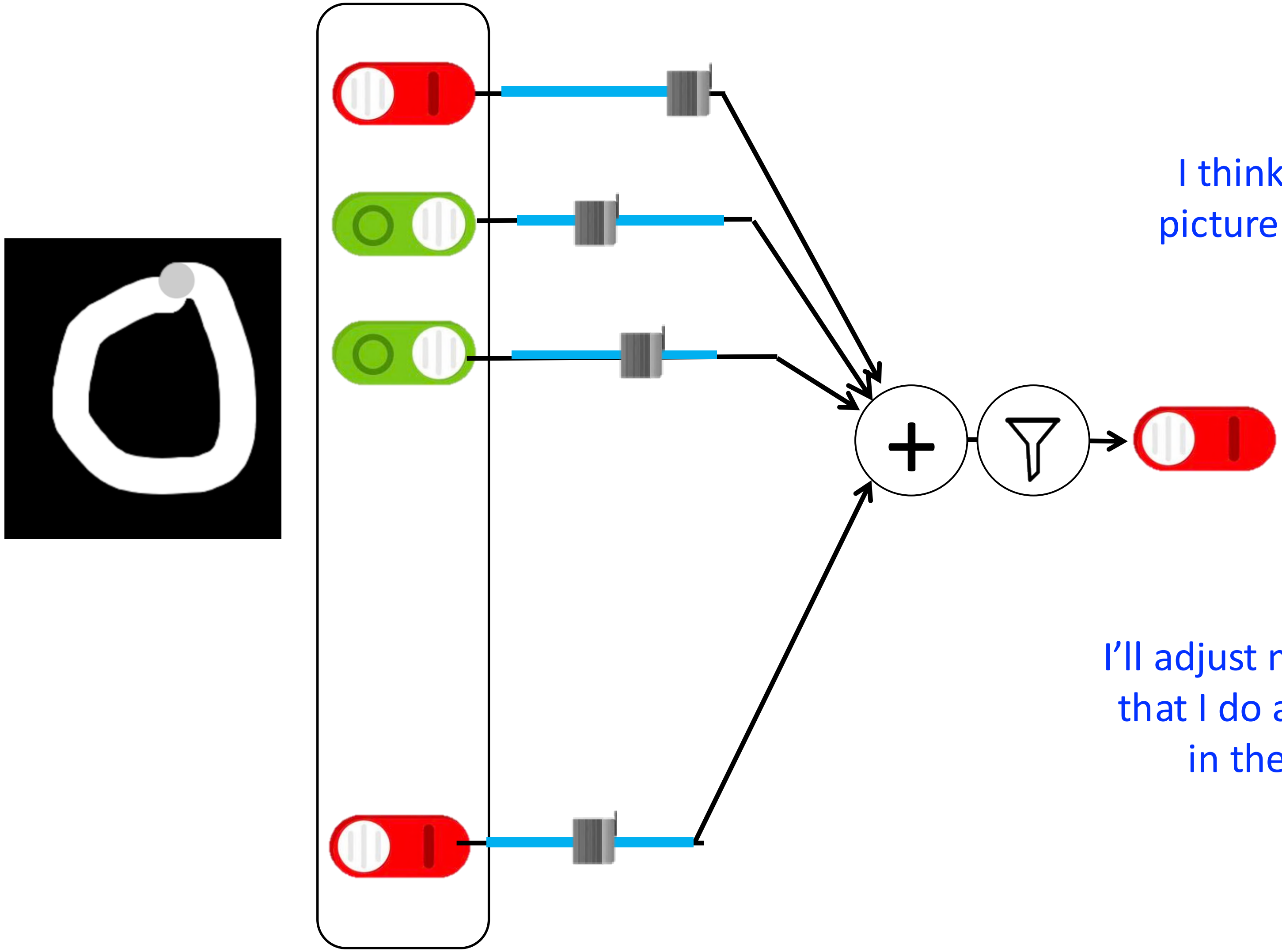




I think that is a picture of a **One**!

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

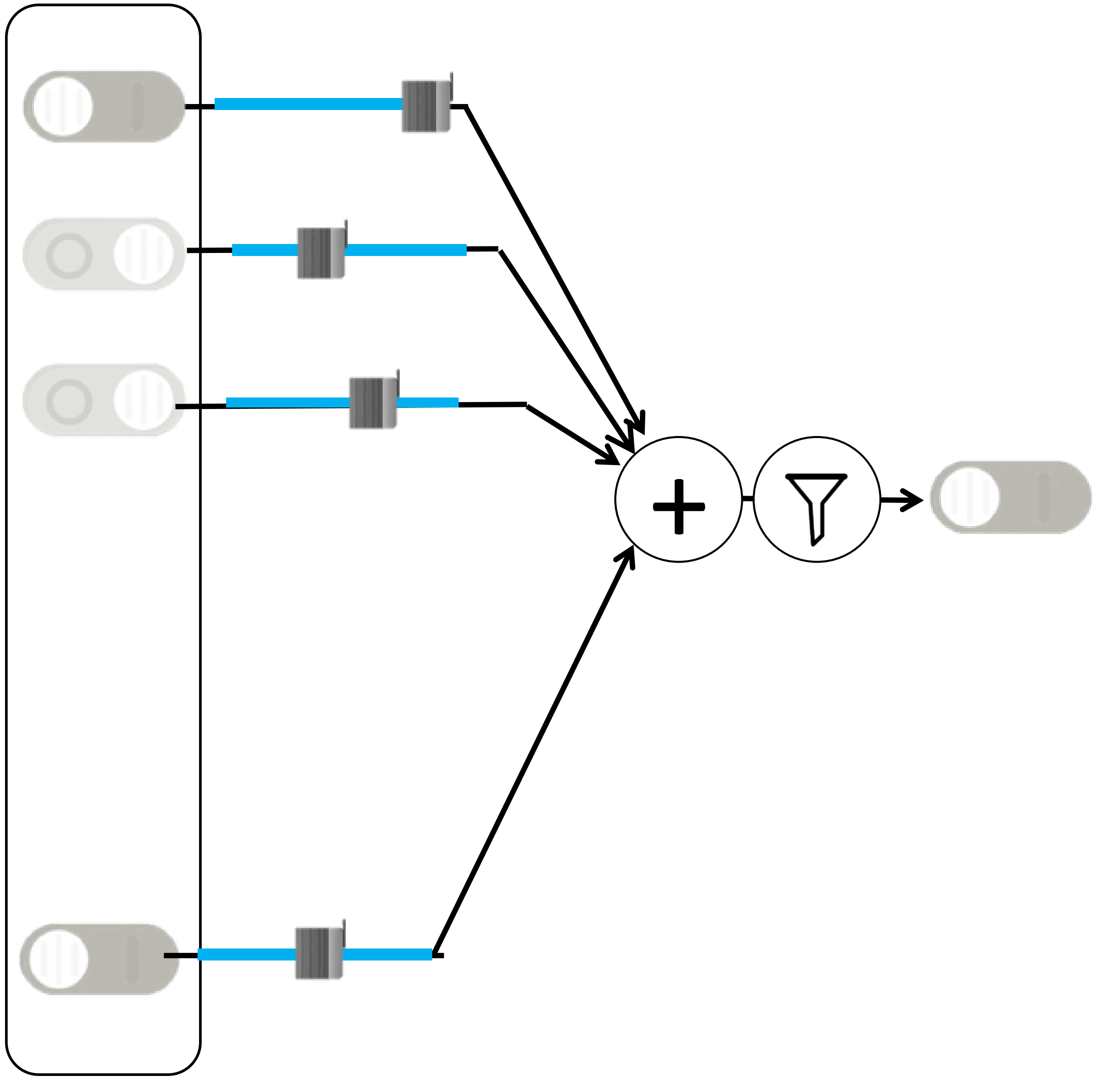
I'll adjust my sliders so that I do a better job in the future

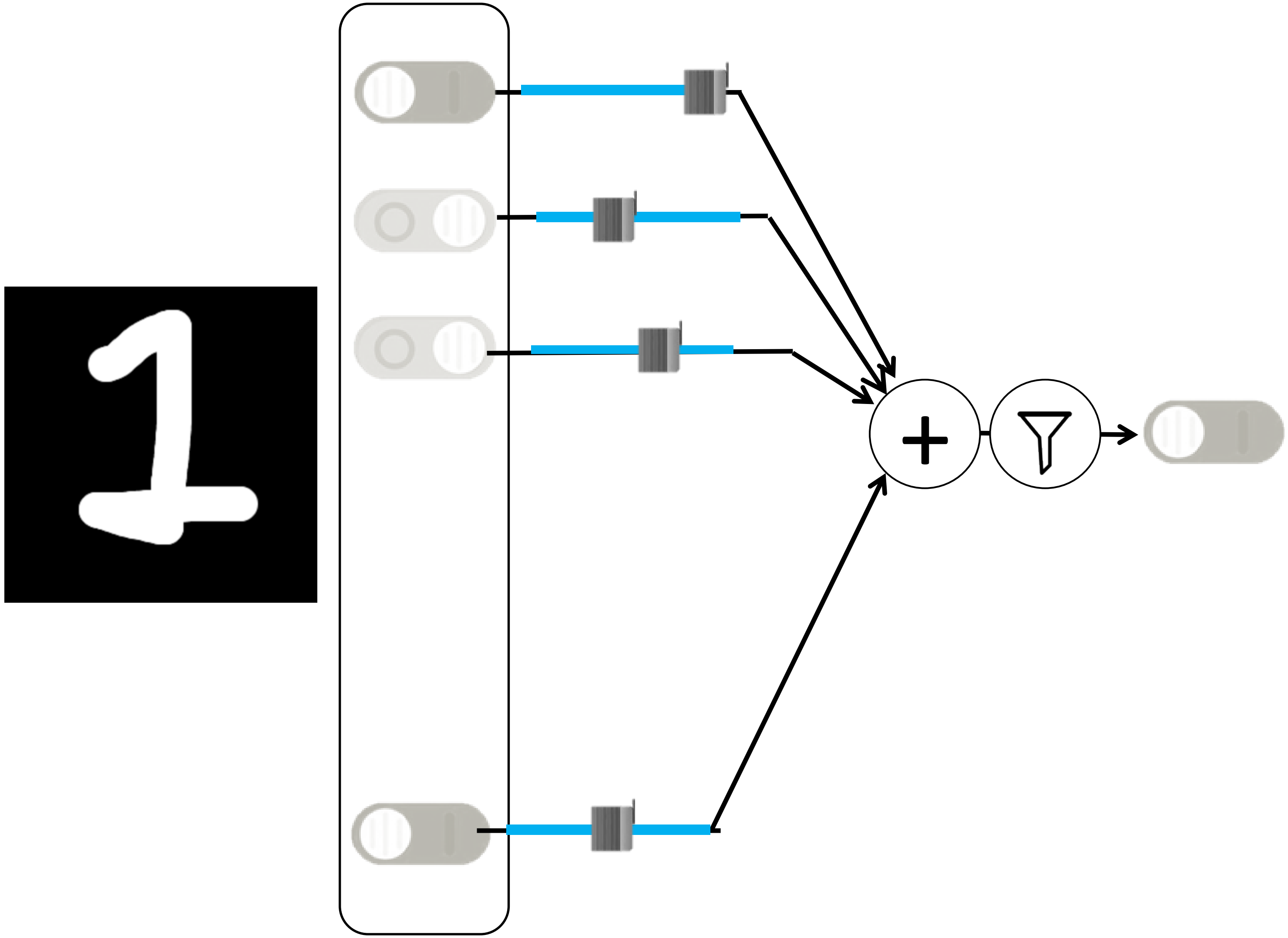


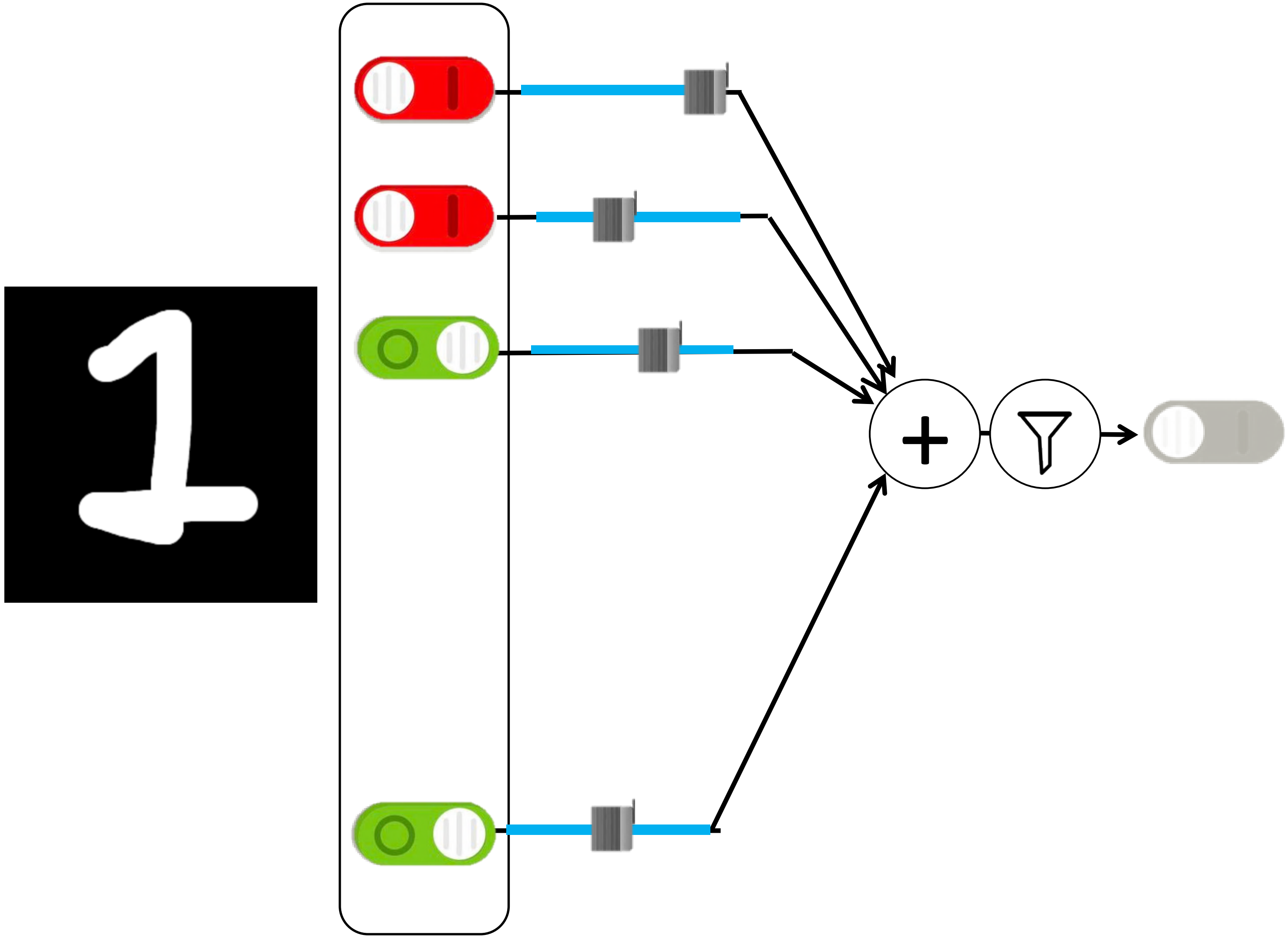
I think that is a picture of a **One**!

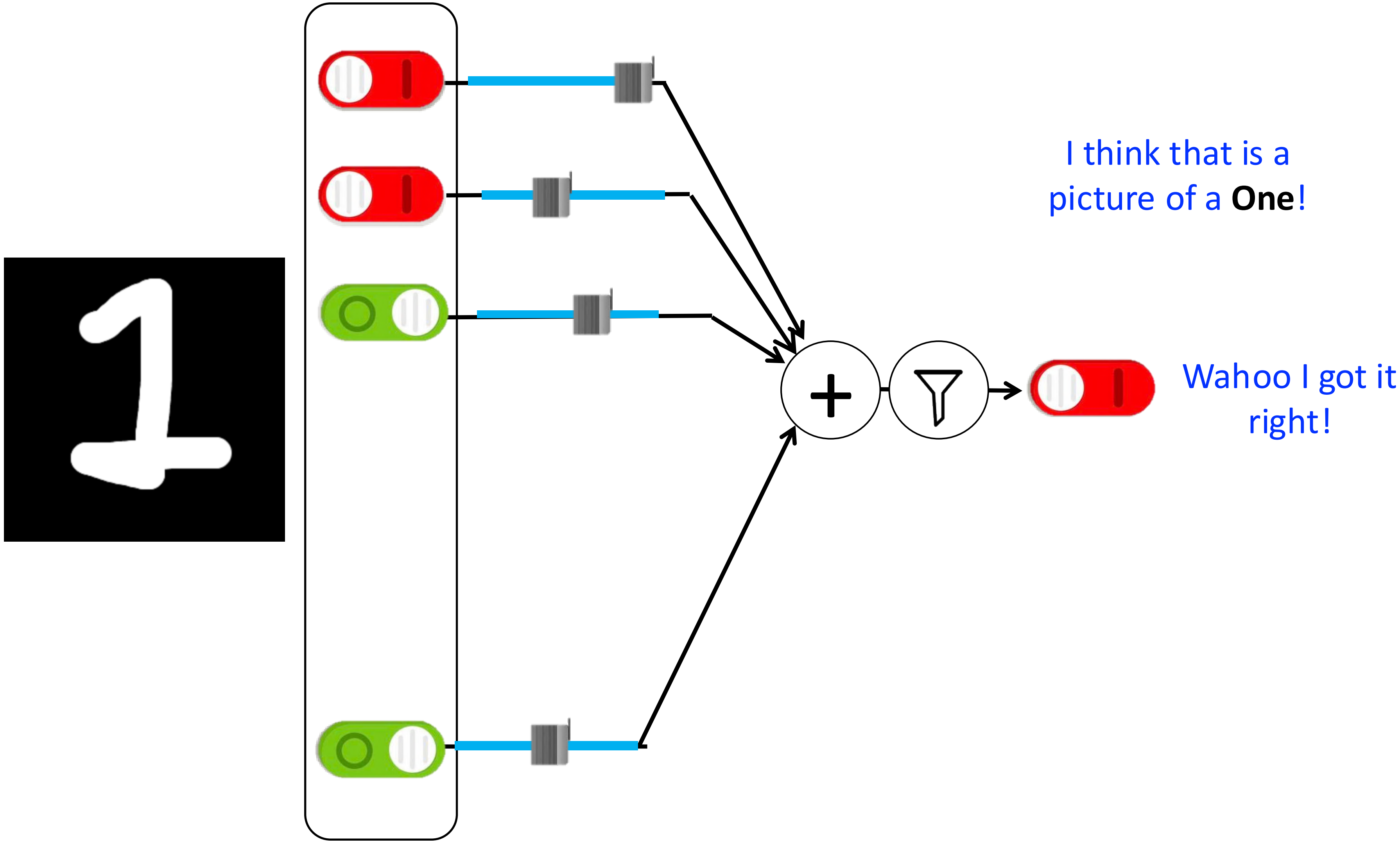
What do you mean it's actually a **Zero**?

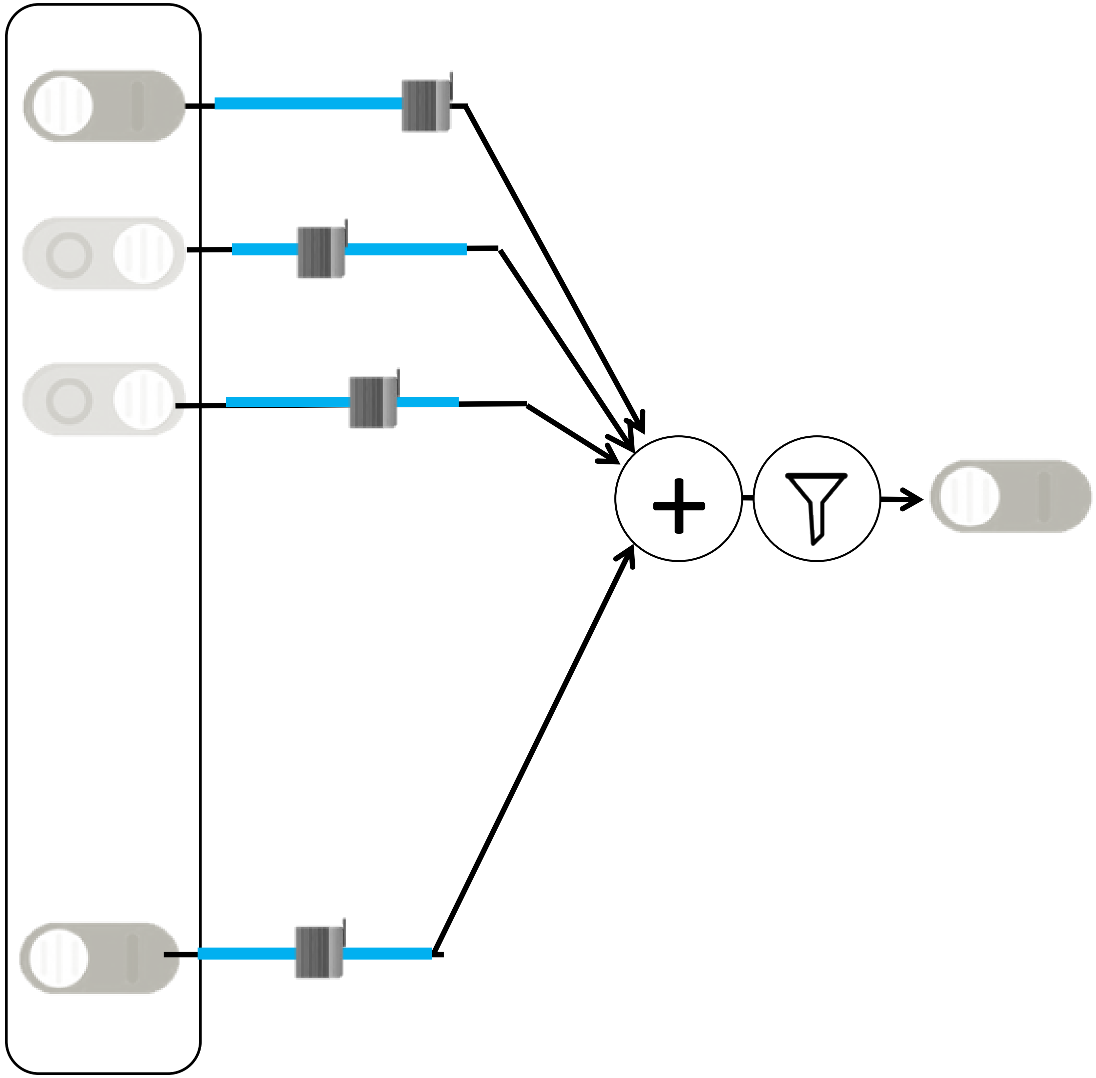
I'll adjust my sliders so that I do a better job in the future

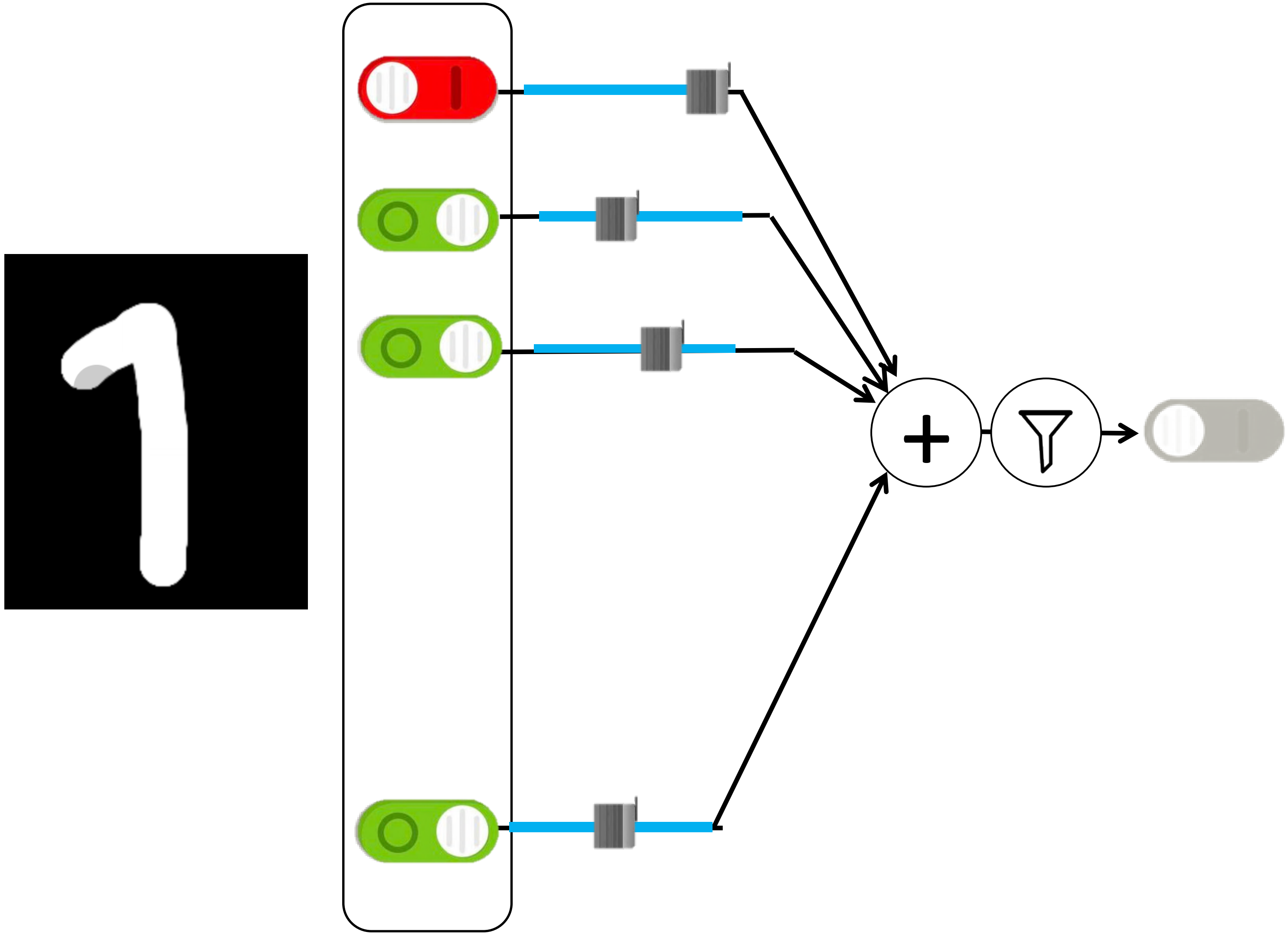


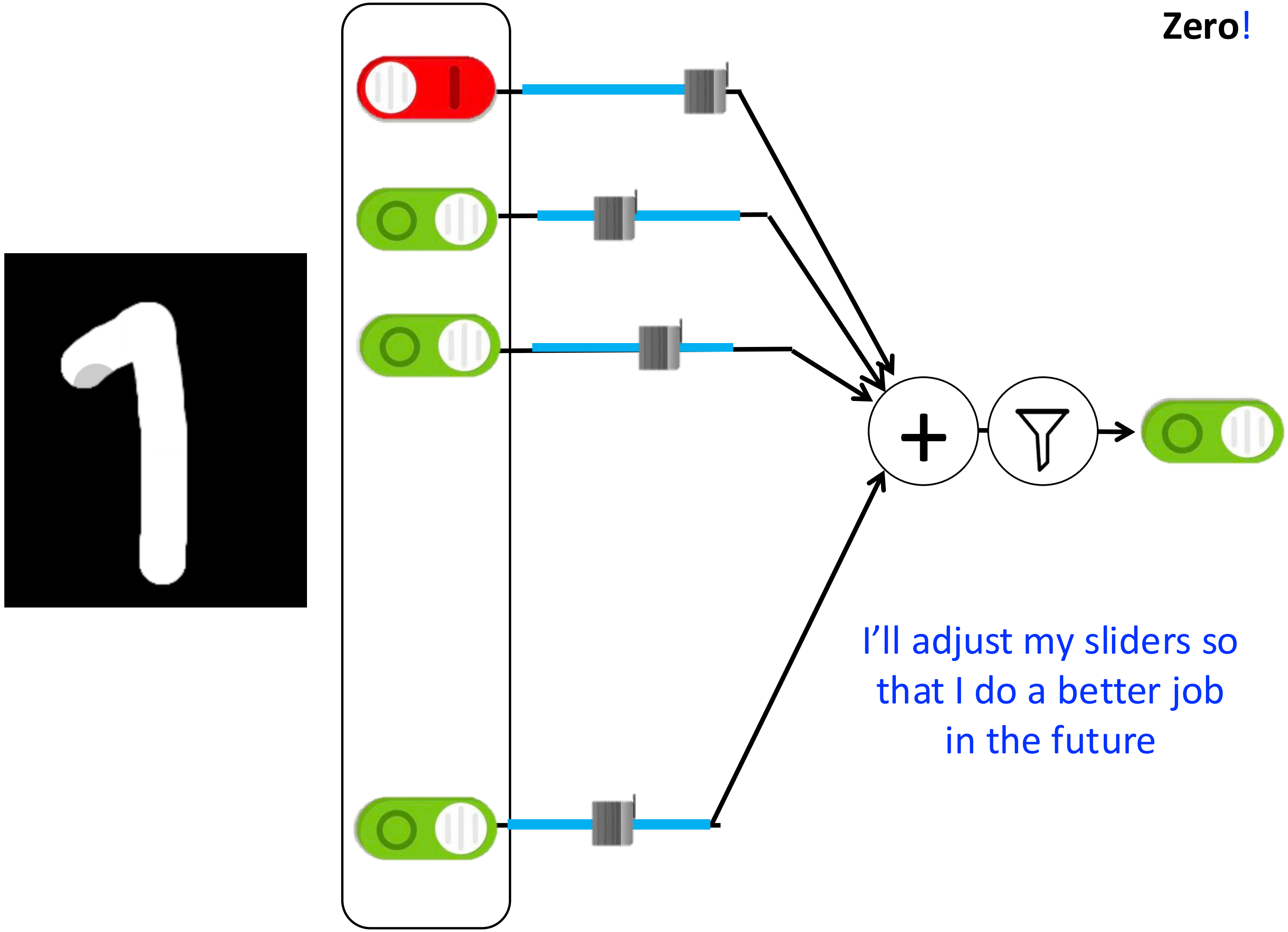





I think that is a
picture of a **One**!
Wahoo I got it
right!







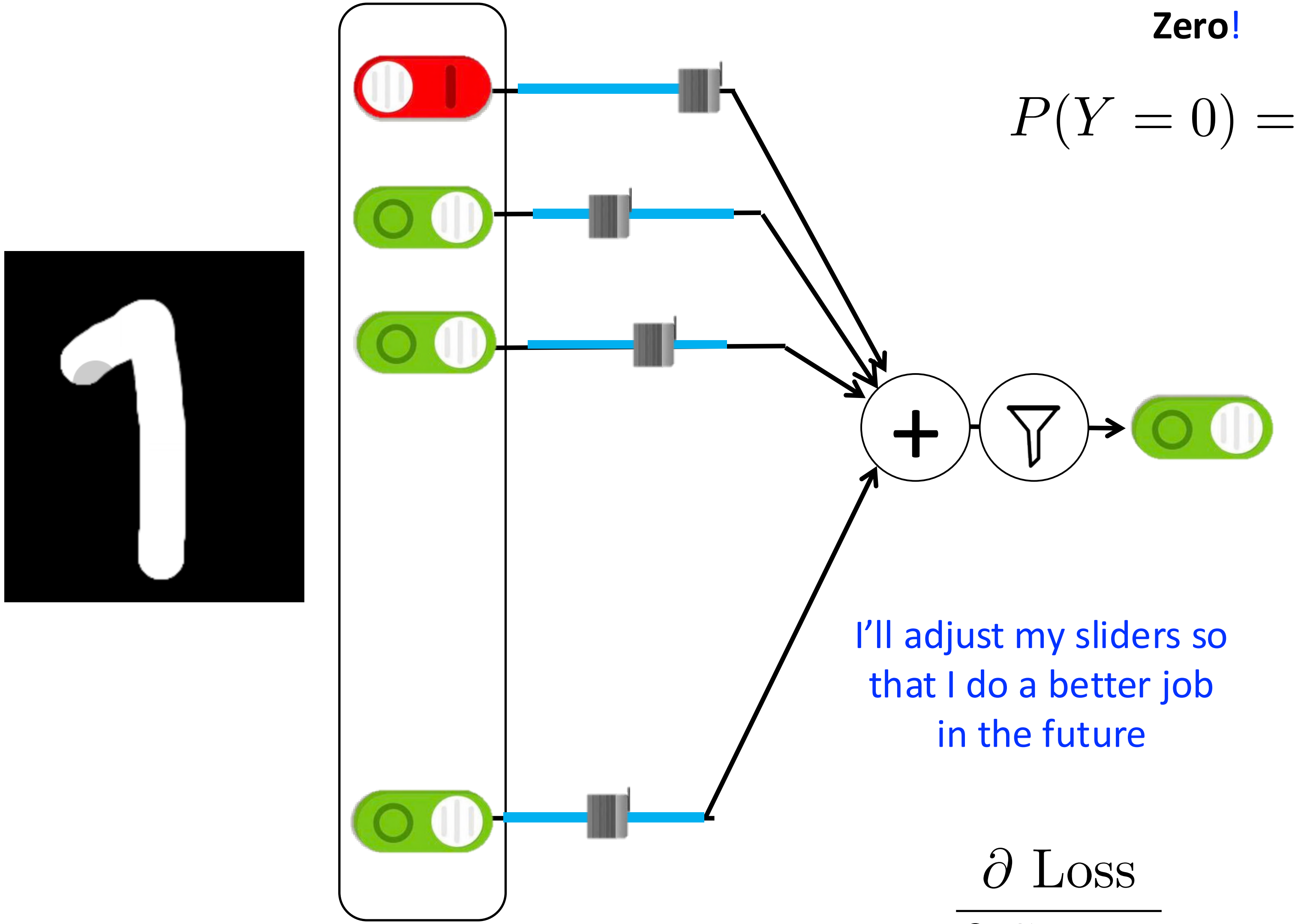
I think that is a picture of a **Zero**!

But it is actually a **One**

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

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

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



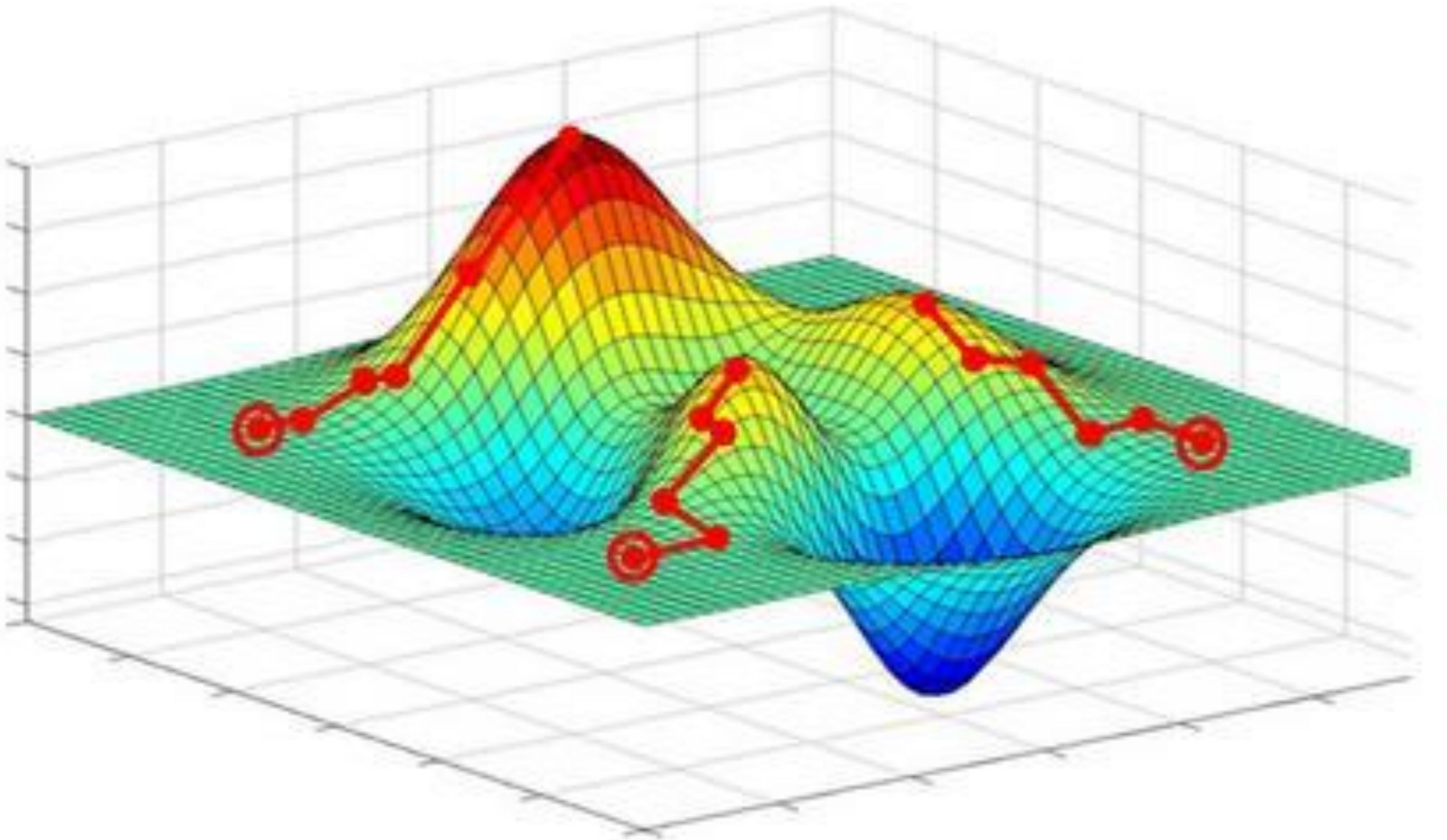
But it is actually a **Zero**

$$\text{Loss} = 1$$

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

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

Gradient Ascent



Walk uphill and you will find a local maxima
(if your step size is small enough)

Gradient of Probability

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

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

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

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

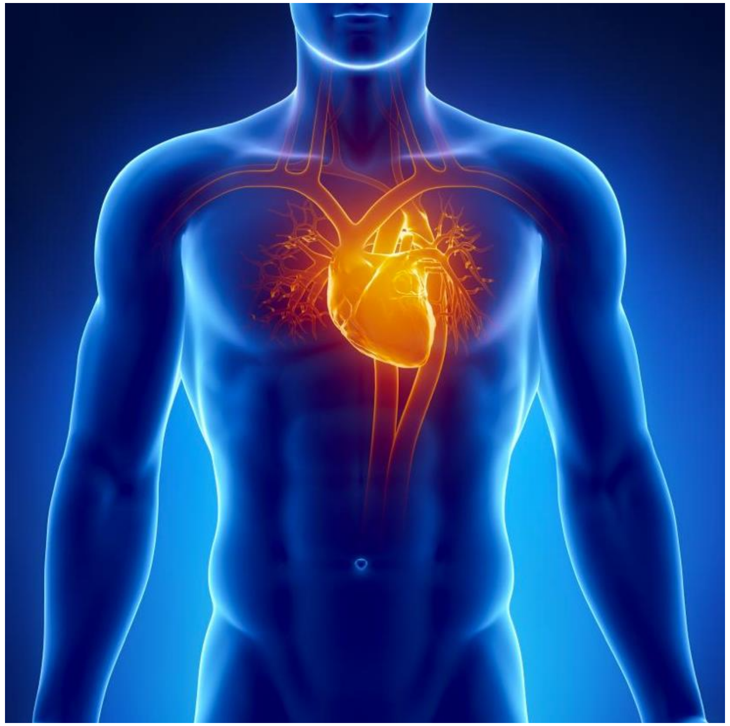
$$= \hat{y}[1 - \hat{y}] \cdot h_i$$

You will be able to do this.

Where you will be by the end of class

CS109: Theory Class focused on Applications

Heart



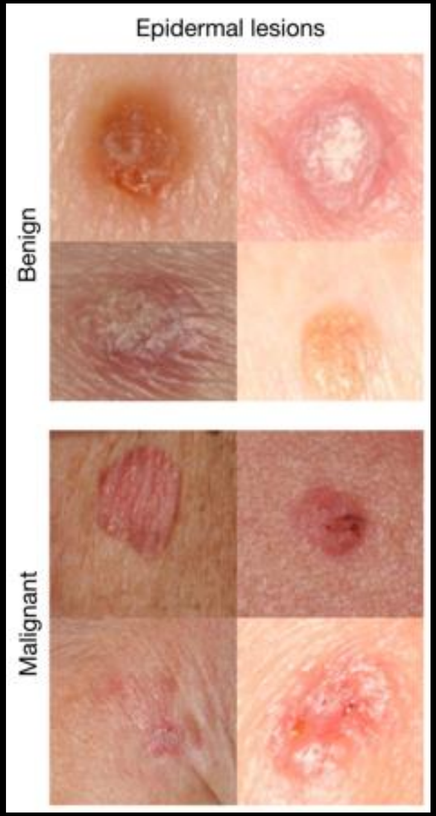
Ancestry

23andMe

Netflix

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 JASCHA@STANFORD.EDU
Stanford University

Eric A. Weiss EAWEISS@BERKELEY.EDU
University of California, Berkeley

Niru Maheswaranathan NIRUM@STANFORD.EDU
Stanford University

Surya Ganguli SGANGULI@STANFORD.EDU
Stanford University

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 effective[1].

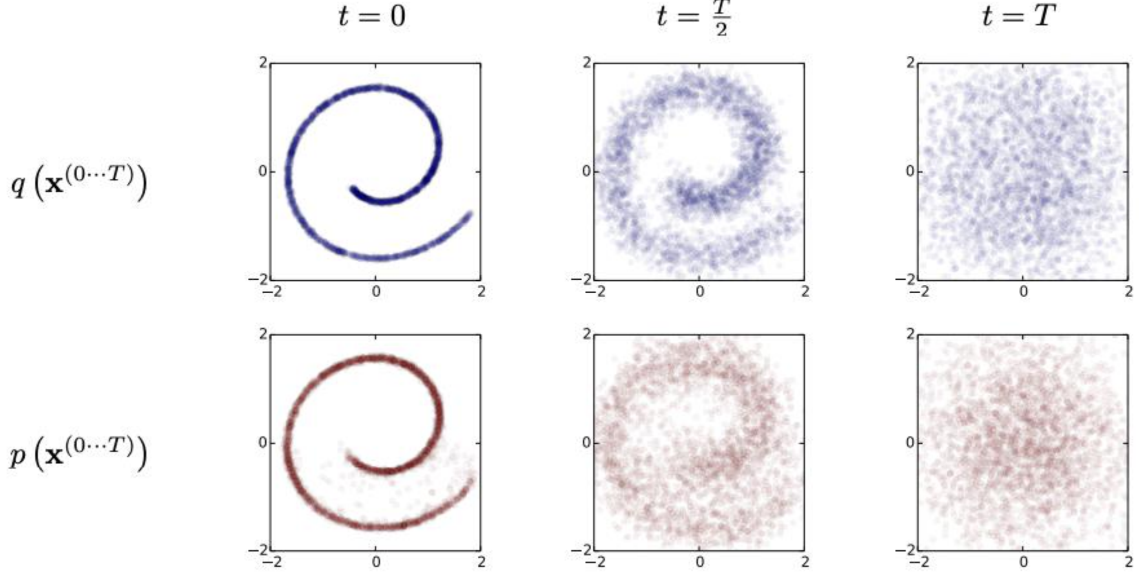
1.1. Diffusion probabilistic models

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

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

Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 2015. JMLR: W&CP volume 37. Copy-

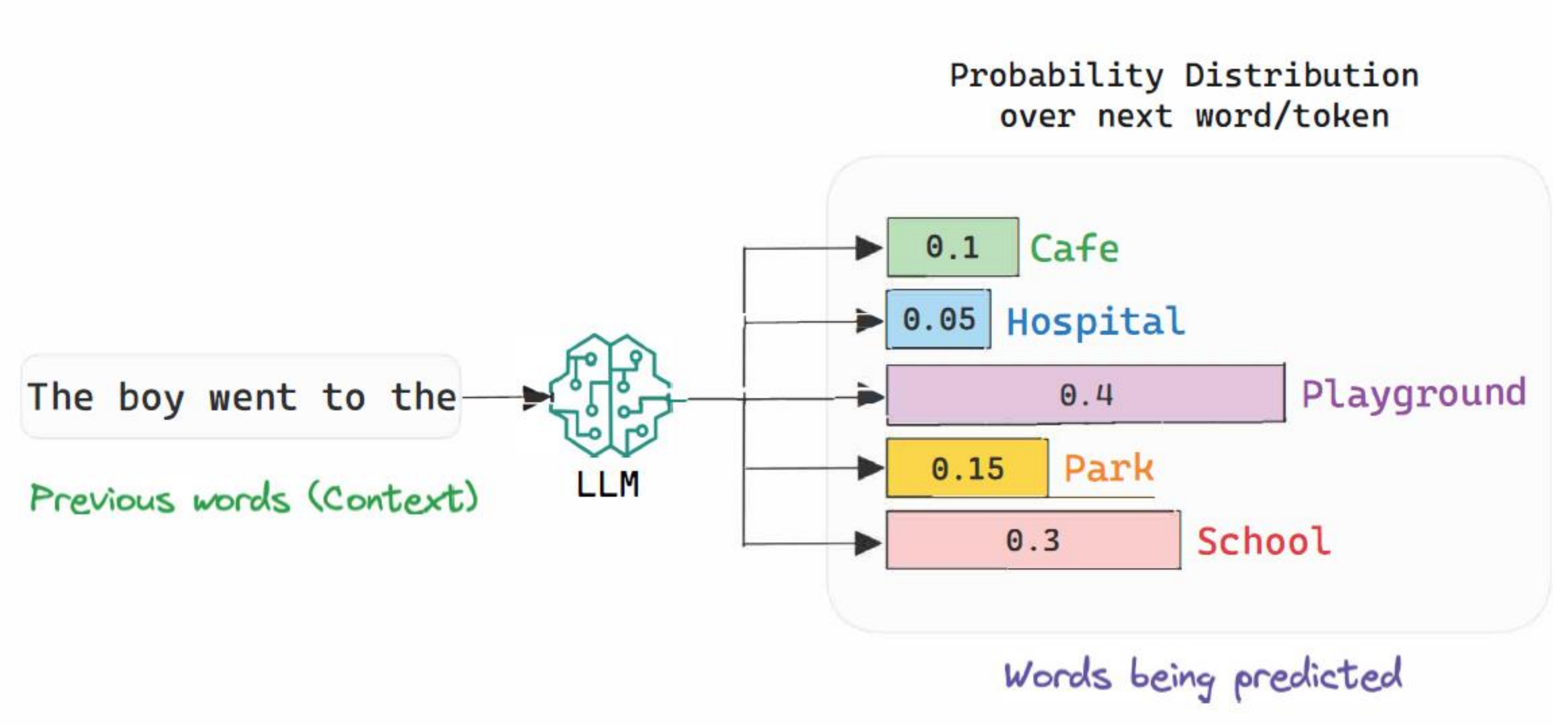
[1]Non-parametric methods can be seen as transitioning smoothly between tractable and flexible models. For instance, a non-parametric Gaussian mixture model will represent a small



Will teach this at the end of the quarter

Generative AI For Text – GPT5

How do Large Language Models (GPT5, Gemini, Claude) work?



Ethics for LLMs

One of the most important areas of active research.

Being fluent in the language of probability isn't just necessary to build these things, but also to regulate and reason about them as well.

Stanford University
Human-Centered
Artificial Intelligence

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

End of Story

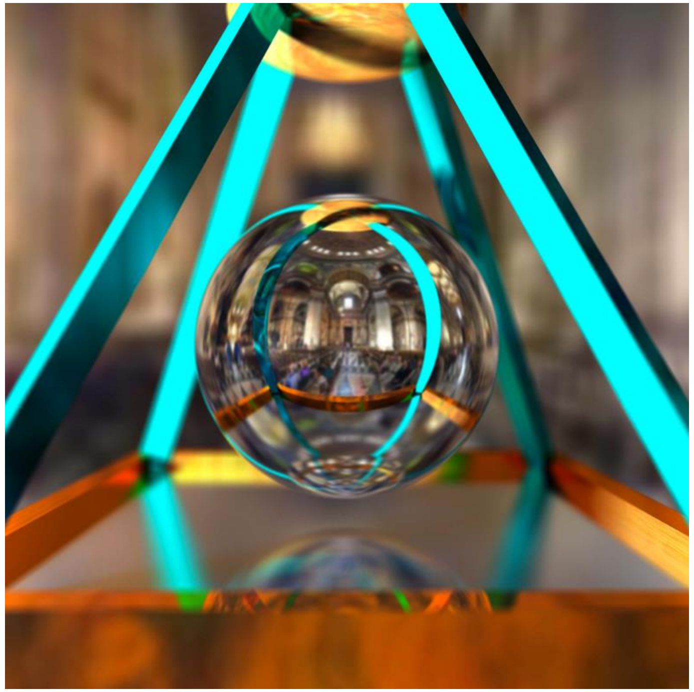
Except it isn't the end of the story…

Probability is more than just machine learning

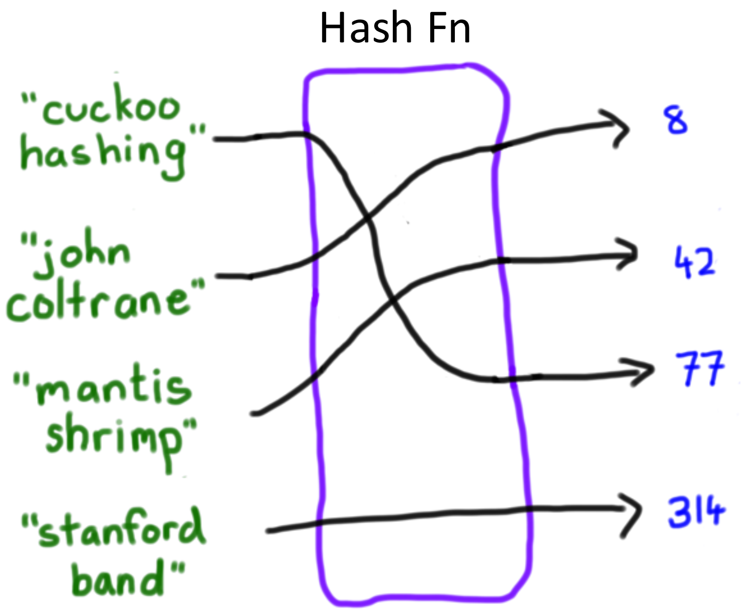
Abundance of important problems

Algorithms and Probability

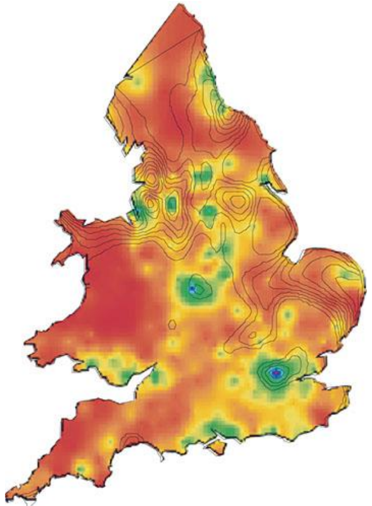
Eg Raytracing



Eg HashMaps

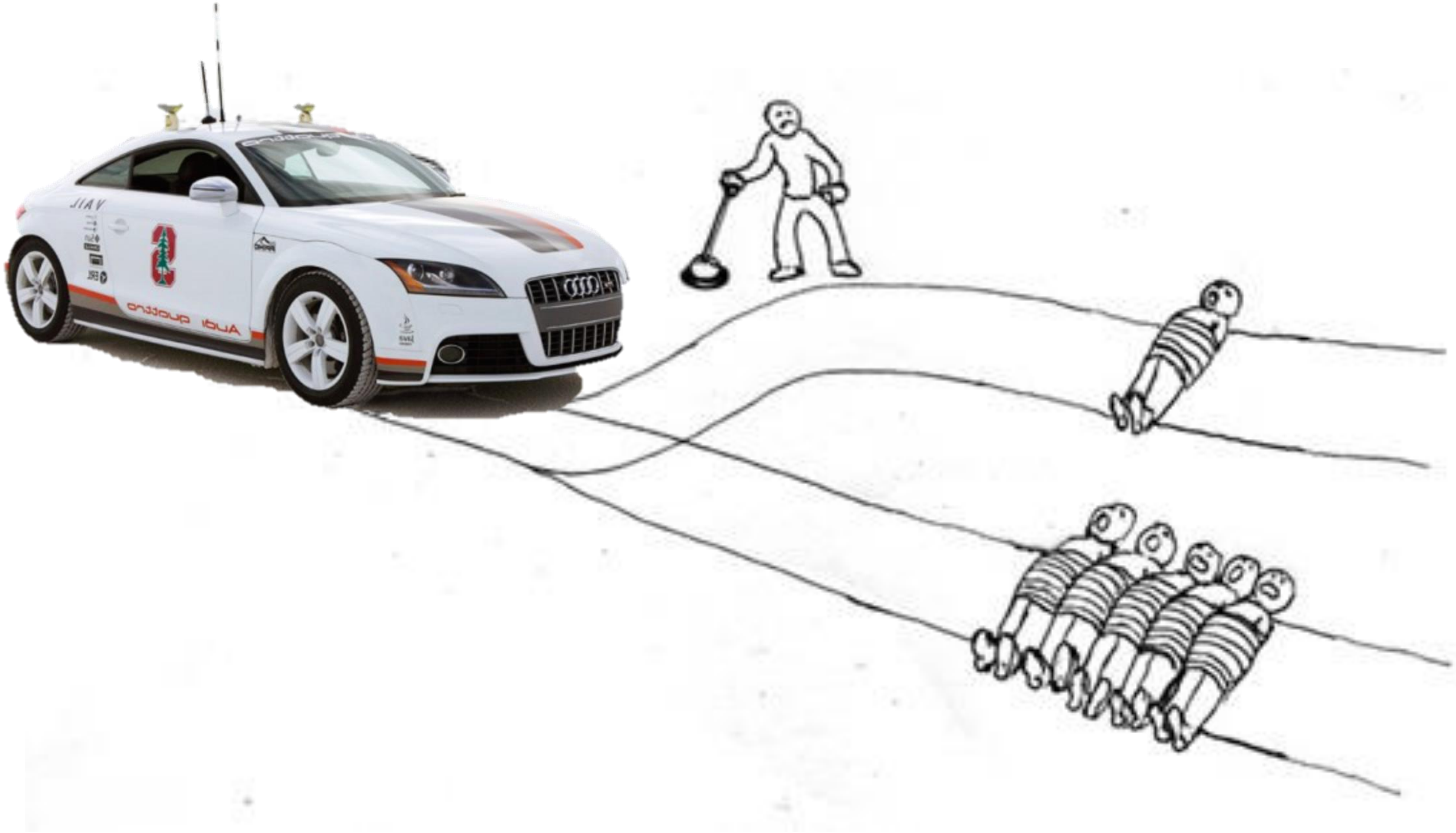


Understanding the world and building tools



Recommender Systems

Philosophy and Ethics



Most Desired Skill by PhD Students

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

"Better understanding of probability"

Learn Real Skills in CS109

Spring 2017

E

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:

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

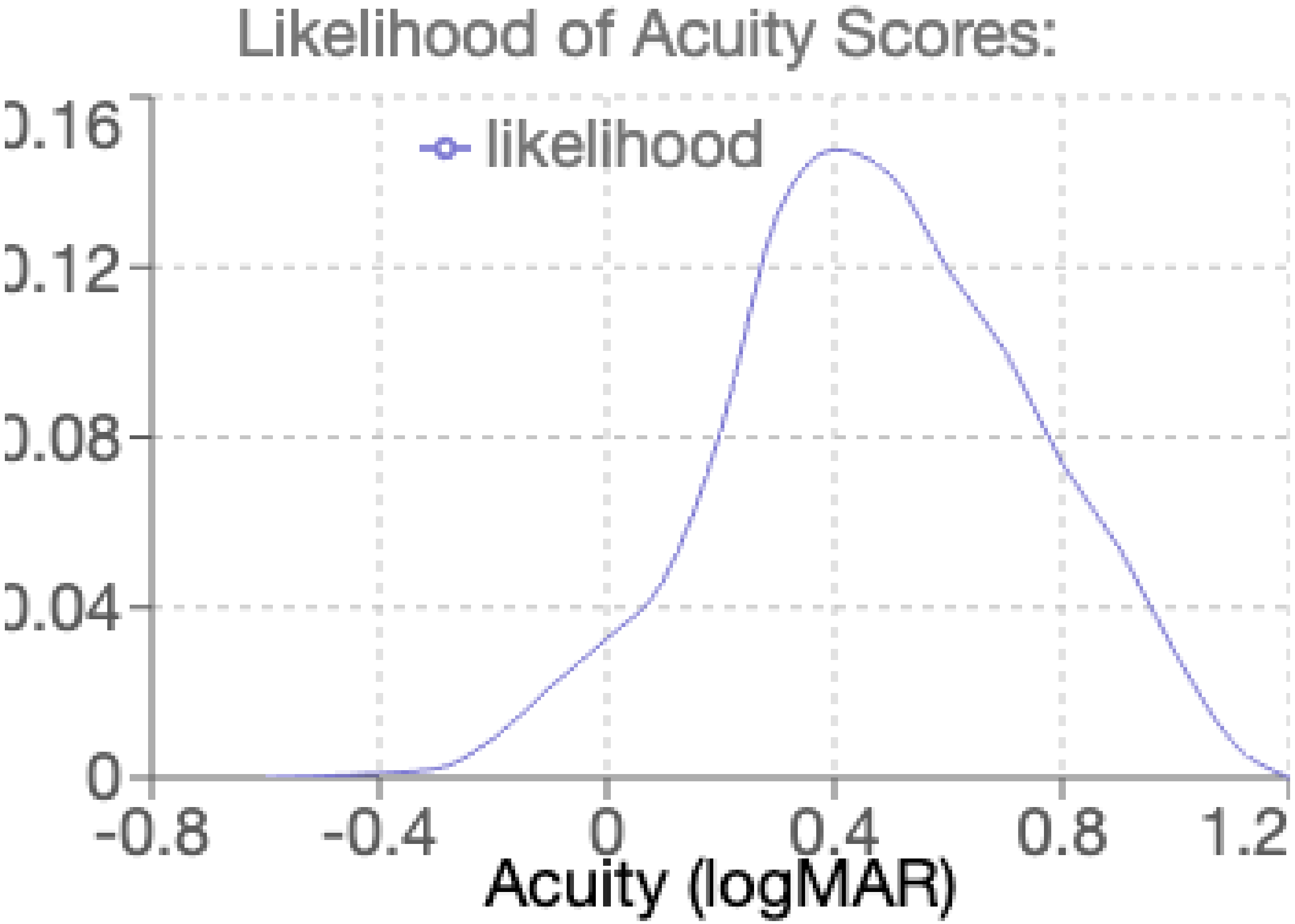
Where θ is the user's vision score and f is the font size of the letter.

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 ith letter and $f^{(i)}$ is the font size of the ith letter. Solve for any and all partial derivatives required by your answer.

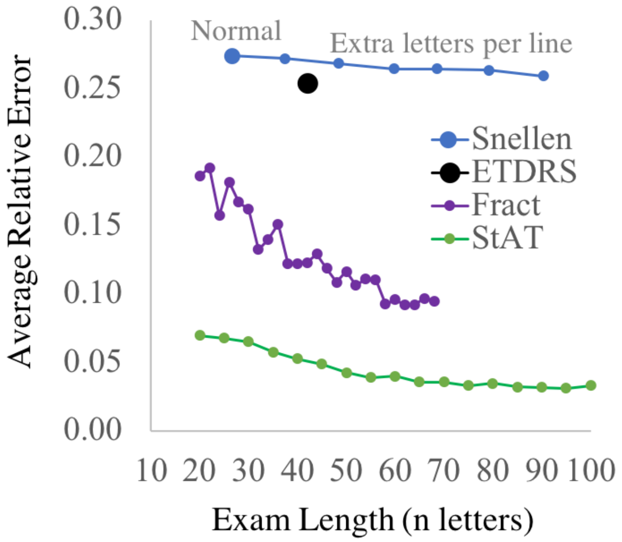
Patient sees a series of letters of different font size, and for each, answers correct or incorrect

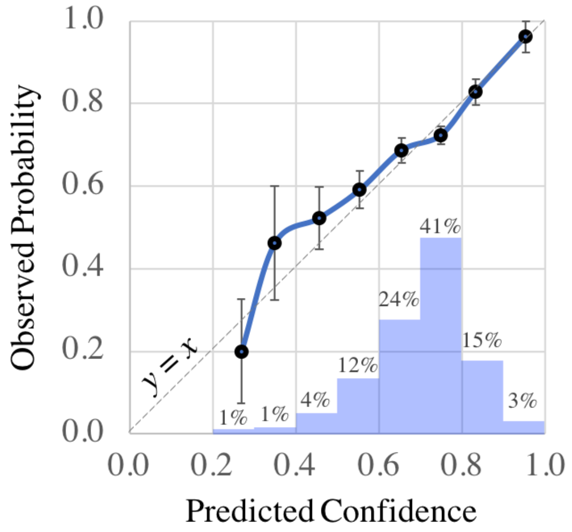
Learn Real Skills in CS109

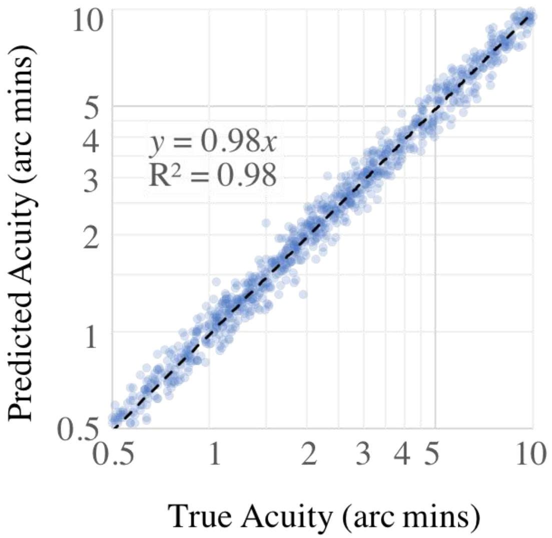
A patient has answered 20 “letter sizes” and got a few correct. What is your belief in how well they can see?



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 Lin[2]
[1]Department 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 patie... dentify the size ... ters correct. Thi...

*Equal contrib...
Copyright © 202...
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 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.

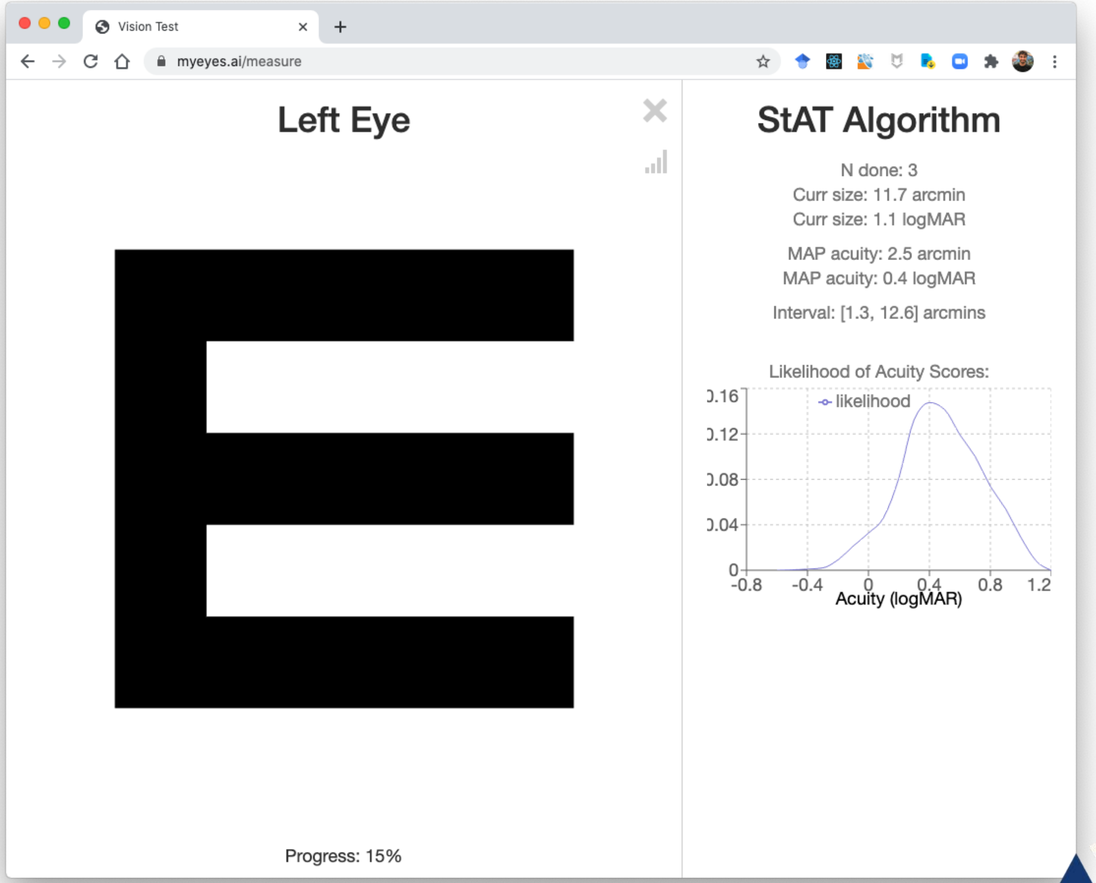
While promising, contemporary digital exams have yet to dramatically improve accuracy over traditional chart-based approaches. The current best digital exam uses a psychometric Item Response Theory (IRT) algorithm for both selecting the next letter size to query and for making a final prediction of acuity.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

...ng algorithm to adap... wn to a user. This ef... acuity belief.

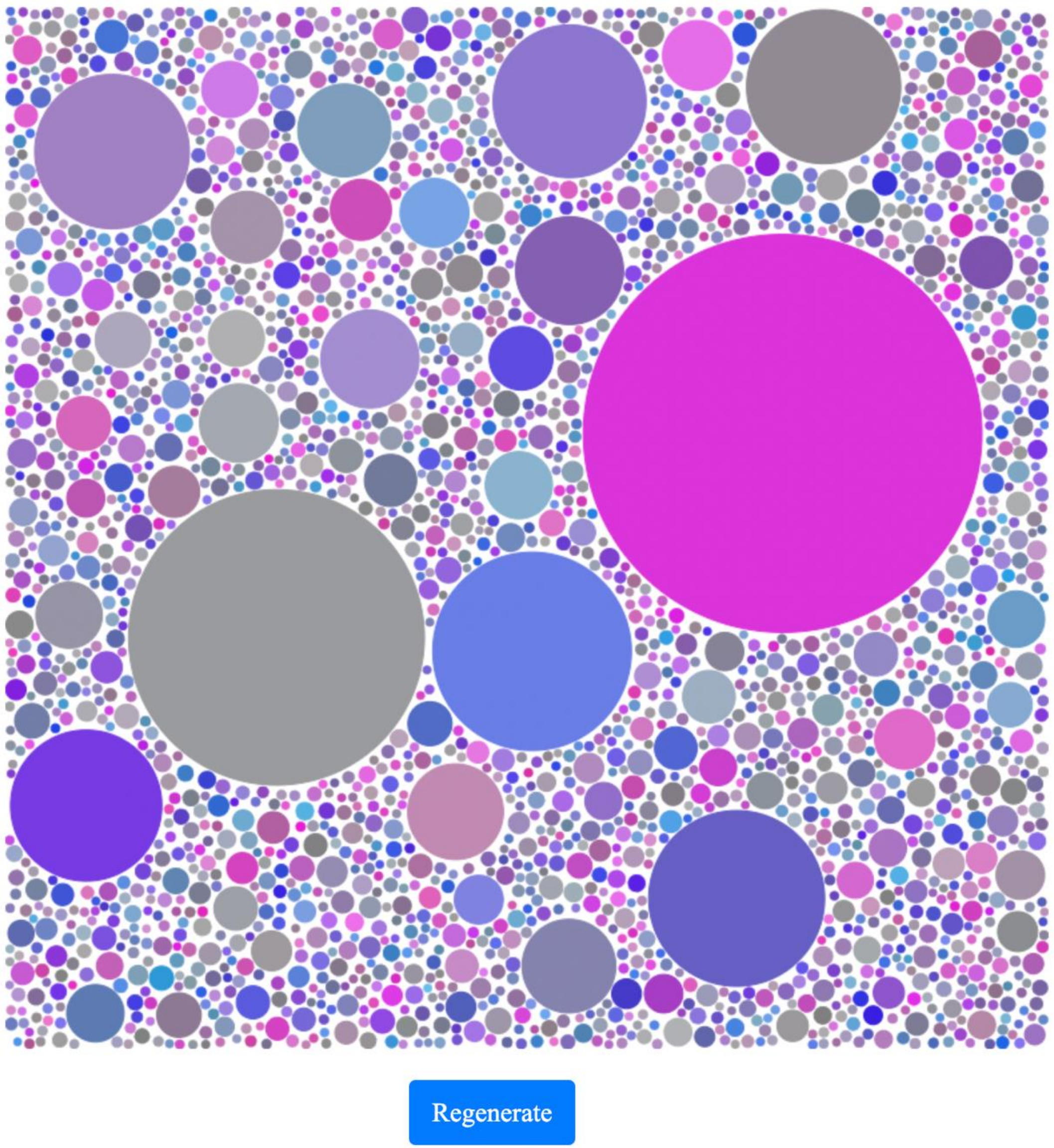
...CT, was named after ...ed. We continue in this



Science

THE LANCET

What is on a typical final?



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

Probability Fundamentals

Why learn in the time of ChatGPT?

Lets dive in…

2 min pedagogic pause.

Sample Space

- **Sample space**, S, is set of all possible outcomes of an experiment
  - Coin flip: S = {Head, Tails}
  - Flipping two coins: S = {[H, H], [H, T], [T, H], [T, T]}
  - Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}
  - # emails in a day: $S = \{x \mid x \in \mathbf{Z}, x \geq 0\}$ {non-neg. ints}
  - YouTube hrs. in day: $S = \{x \mid x \in \mathbf{R}, 0 \leq x \leq 24\}$

Event Space

- **<u>Event</u>**, E, is some subset of S $\{E \subseteq S\}$
  - Coin flip is heads: $E = \{\text{Head}\}$
  - $\geq$ 1 head on 2 coin flips: $E = \{[H, H], [H, T], [T, H]\}$
  - Roll of die is 3 or less: $E = \{1, 2, 3\}$
  - # emails in a day $\leq$ 20: $E = \{x \mid x \in \mathbf{Z}, 0 \leq x \leq 20\}$
  - Wasted day {≥ 5 YT hrs.}: $E = \{x \mid x \in \mathbf{R}, 5 \leq x \leq 24\}$

Note: When Ross uses: $\subset$, he really means: $\subseteq$

Event Space

Sample Space, S

- Coin flip
  $S = \{\text{Heads}, \text{Tails}\}$
- Flipping two coins
  $S = \{(\text{H,H}), (\text{H,T}), (\text{T,H}), (\text{T,T})\}$
- Roll of 6-sided die
  $S = \{1, 2, 3, 4, 5, 6\}$
- # emails in a day
  $S = \{x \mid x \in \mathbb{Z},\ x \geq 0\}$
- TikTok hours in a day
  $S = \{x \mid x \in \mathbb{R}, 0 \leq x \leq 24\}$

Event, E

- Flip lands heads
  $E = \{\text{Heads}\}$
- ≥ 1 head on 2 coin flips
  $E = \{(\text{H,H}), (\text{H,T}), (\text{T,H})\}$
- Roll is 3 or less:
  $E = \{1, 2, 3\}$
- Low email day (≤ 20 emails)
  $E = \{x \mid x \in \mathbb{Z},\ 0 \leq x \leq 20\}$
- Wasted day (≥ 5 TT hours):
  $E = \{x \mid x \in \mathbb{R}, 5 \leq x \leq 24\}$

What is a probability?

[suspense]

Number between 0 and 1

A number to which we ascribe meaning

$$P(E)$$

* Our belief that an event E occurs

A number to which we ascribe meaning

$$\Pr(E)$$

* Our belief that an event E occurs

What is a Probability?

How many times does it occur

The event we care about

$$\mathrm{P}(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

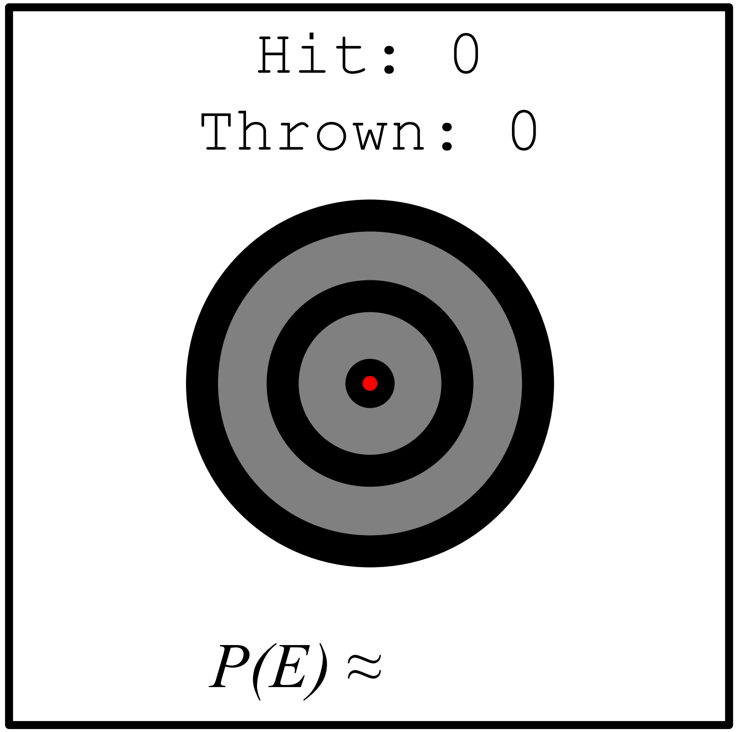
Out of (close to) infinite trials

To the course reader!

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

n is the number of trails

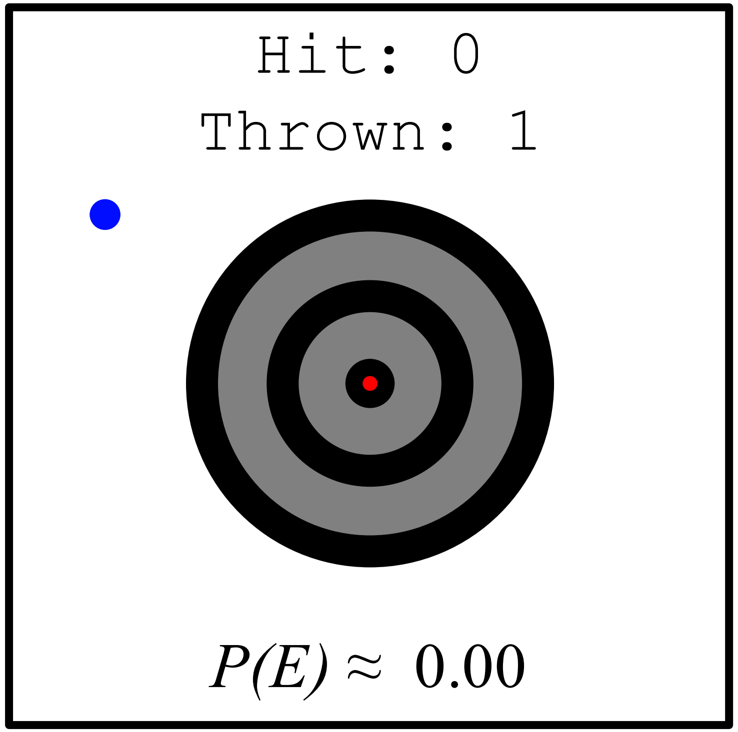


The "event" *E* is that you hit the target

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

n is the number of trails

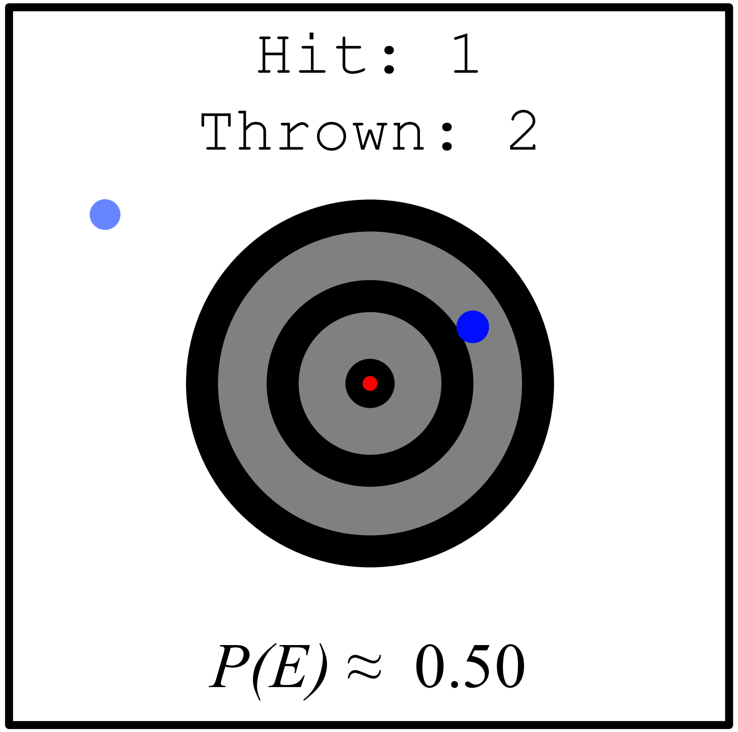


The "event" *E* is that you hit the target

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

n is the number of trails

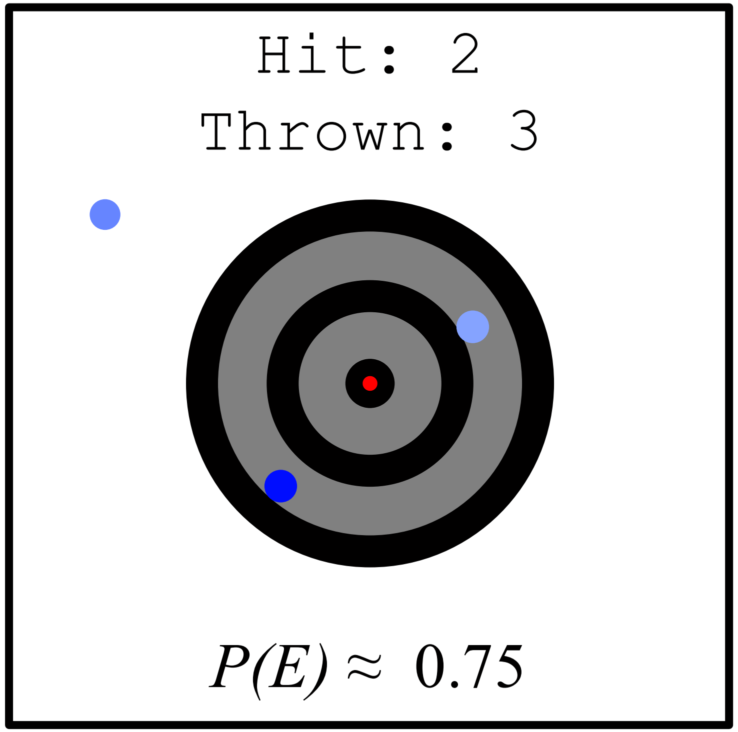


The “event” *E* is that you hit the target

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

n is the number of trails

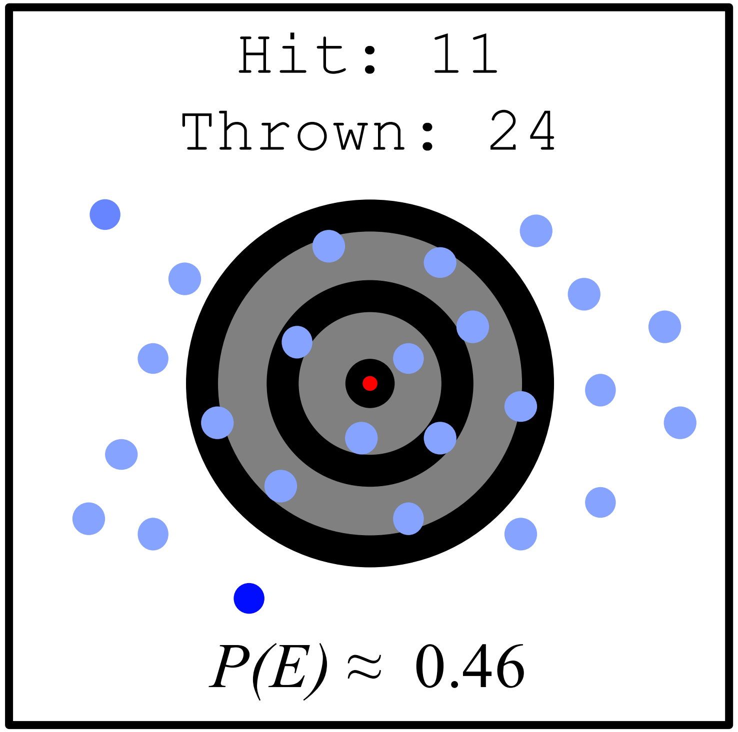


The "event" *E* is that you hit the target

What is a Probability?

$$P(E) = \lim_{n \to \infty} \frac{\text{count}(E)}{n}$$

n is the number of trails



The "event" *E* is that you hit the target

What is a Probability (in a Dataset)?

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

Dataset of weather

Trial	**Value**
1	Rainy
2	Sunny
3	Rainy
4	Cloudy
5	Rainy
6	Sunny
7	Sunny
8	Sunny
	…
10000	Cloudy

Let E be the event that it is **Sunny**

$$P(E) = \lim_{n\to\infty} \frac{n(E)}{n}$$

$$\approx \frac{\text{Count}(E)}{10000}$$

$$\approx \frac{3332}{10000} \approx 0.3332$$

Equally Likely Outcomes

Some sample spaces have **equally likely outcomes**.

- Coin flip: S = {Head, Tails}
- Flipping two coins: S = {[H, H], [H, T], [T, H], [T, T]}
- Roll of 6-sided die: S = {1, 2, 3, 4, 5, 6}

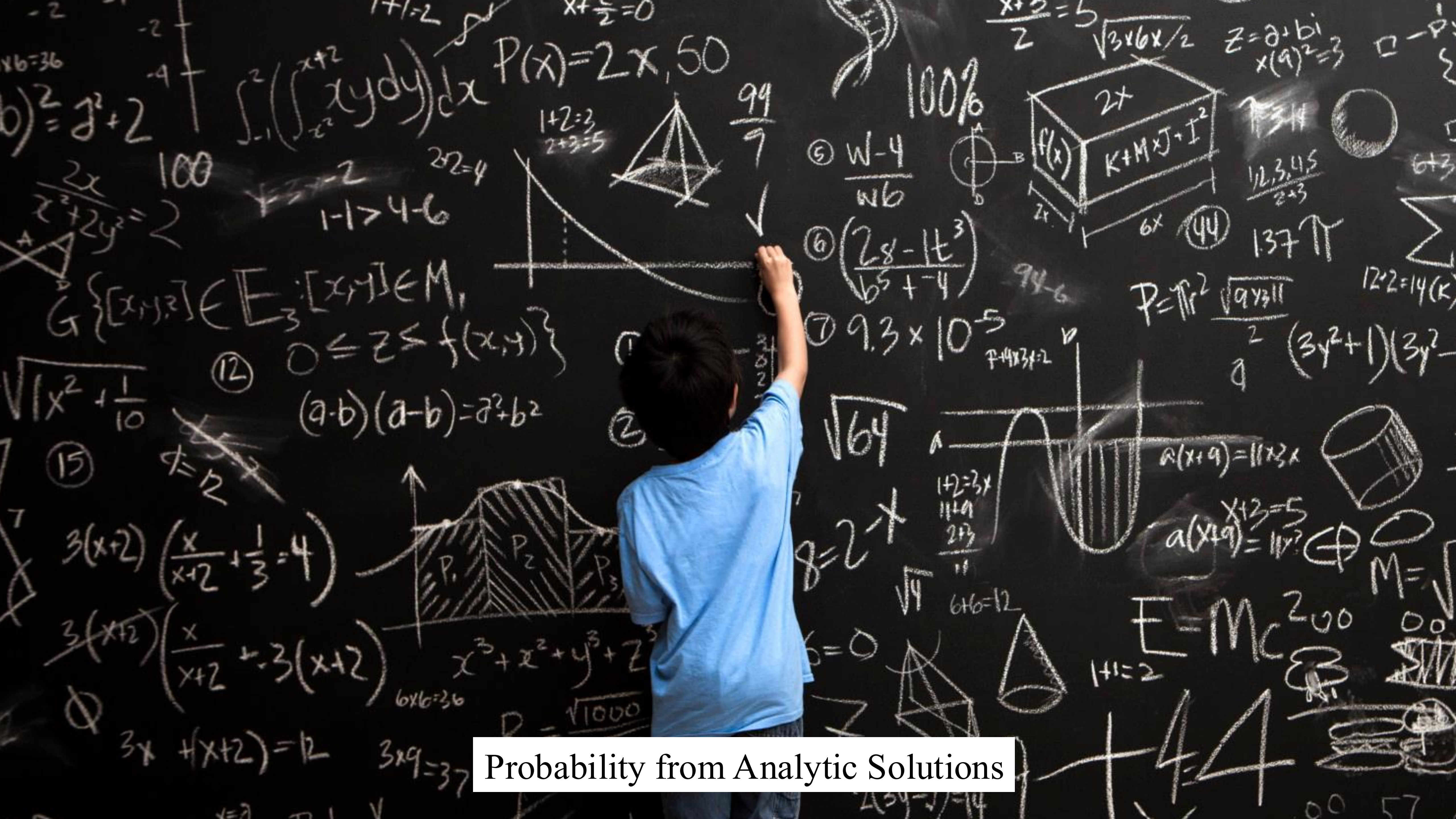
If we have equally likely outcomes, then P(Each outcome) $= \frac{1}{|S|}$

Therefore $$P(E) = \frac{\text{\# outcomes in E}}{\text{\# outcomes in S}} = \frac{|E|}{|S|}$$ {by Axiom 3}

Not Everything is Equally Likely

- Play lottery.
  - What is P(Win)?

- S = {Lose, Win}
- E = {Win}
- P(Win) = |E|/|S| = 1/2 = 50%



Probability from Analytic Solutions

The Axioms

Axioms of Probability

Recall: S = all possible outcomes. E = the event.

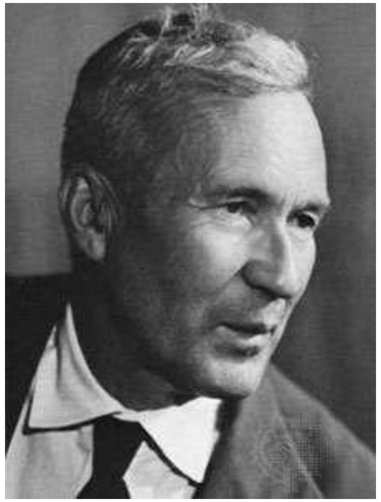
- Axiom 1: $0 \leq P(E) \leq 1$

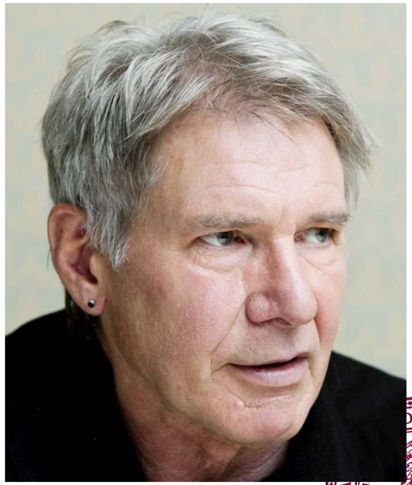
- Axiom 2: $P(S) = 1$

- Axiom 3: If events E and F are mutually exclusive:

$$P(E \cup F) = P(E) + P(F)$$

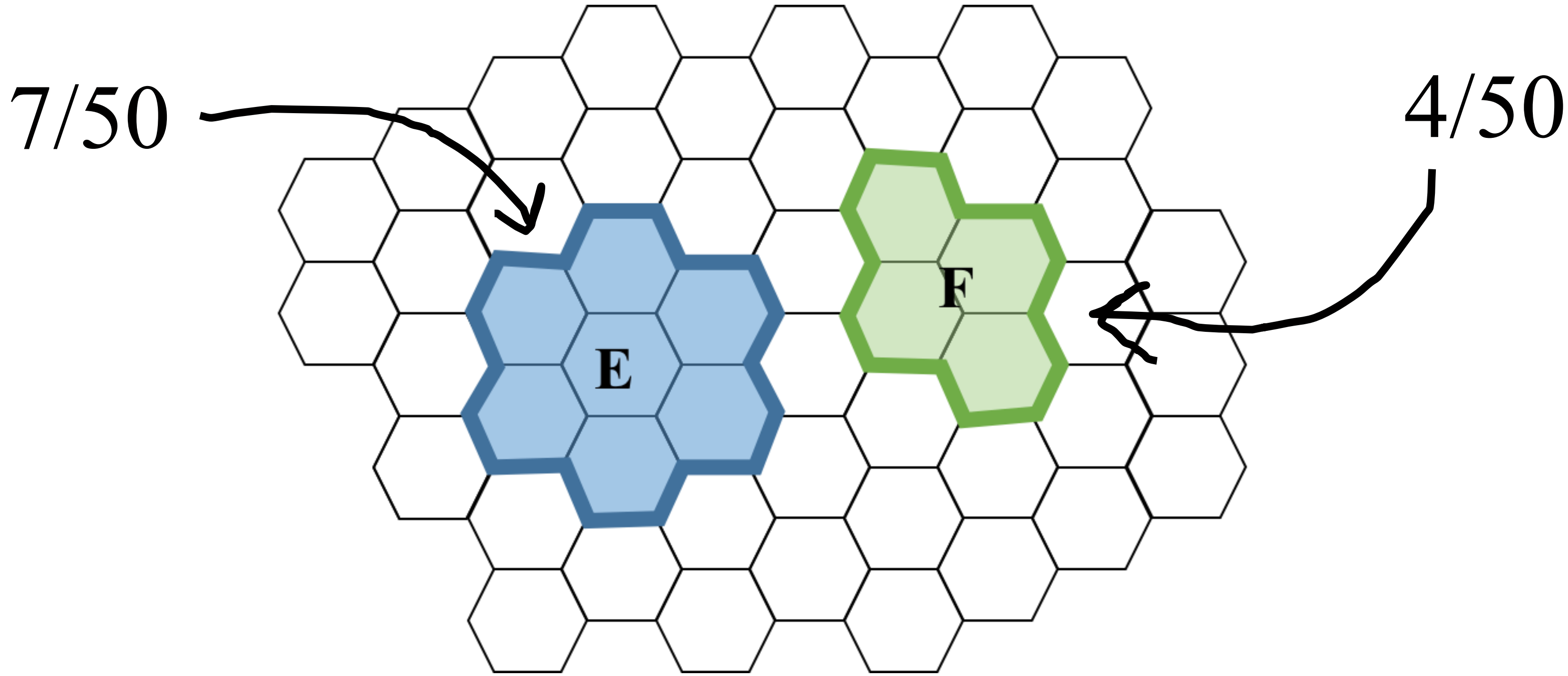
Kolmogorov





Harrison Ford

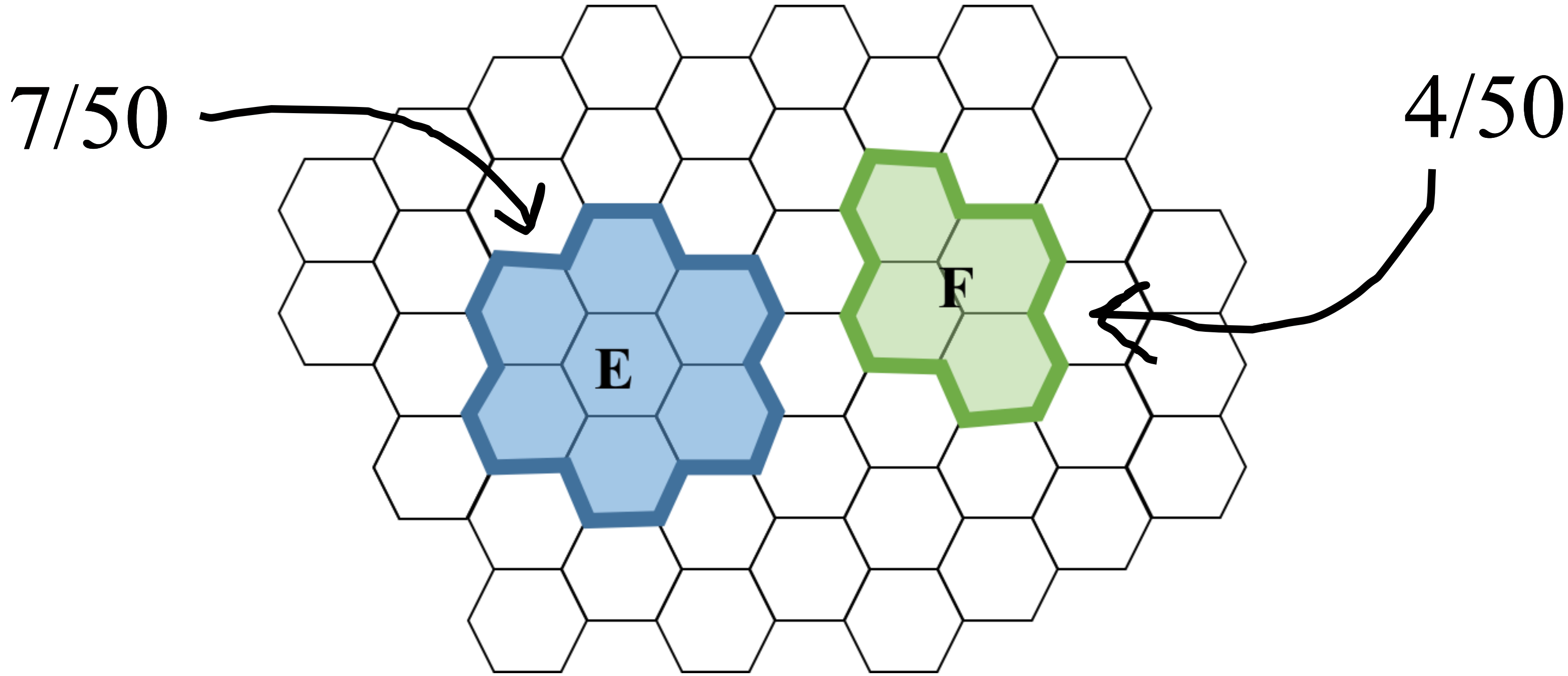
Mutually Exclusive Events



If events are mutually exclusive, probability of OR is simple:

$$P(E \cup F) = P(E) + P(F)$$

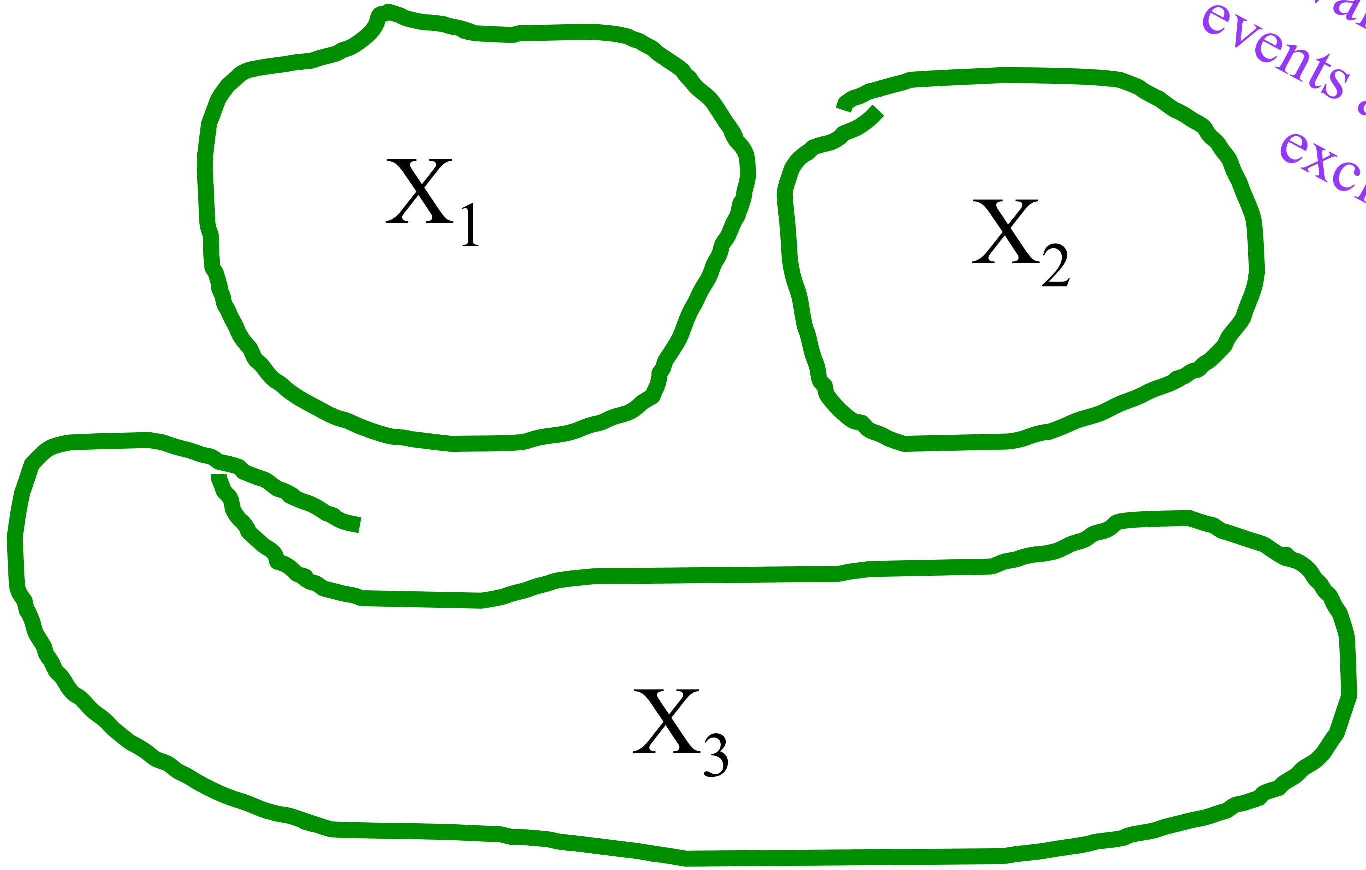
Mutually Exclusive Events



If events are mutually exclusive, probability of OR is simple:

$$P(E \cup F) = \frac{7}{50} + \frac{4}{50} = \frac{11}{50}$$

Probability of "or"



Wahoo! All my events are mutually exclusive

$$P(X_1 \cup X_2 \cup \cdots \cup X_n) = \sum_{i=1}^{n} P(X_i)$$

If events are *mutually exclusive* probability of OR is easy!

$P(E^c) = 1 - P(E)$?

$$P(E \text{ or } E^c) = P(E) + P(E^c)$$

Axiom 3. Since E and E^c are mutually exclusive

$$P(S) = P(E) + P(E^c)$$

Since everything must either be in E or E^c

$$1 = P(E) + P(E^c)$$

Axiom 2

$$P(E^c) = 1 - P(E)$$

Rearrange

Many times it is easier to calculate $P(E^C)$.

Provable Identity #1:

$$P(E^c) = 1 - P(E)$$

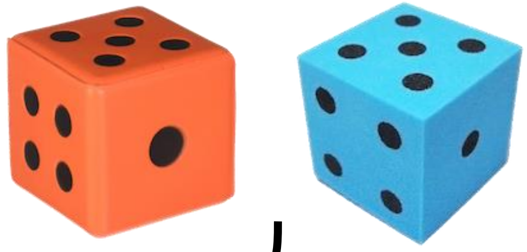
Sources of Probability: A dice story

Sum of Two Die = 7?

Roll two 6-sided dice. What is probability the sum = 7?
Let E be the event that the sum is 7

S = { [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] }

Each outcome

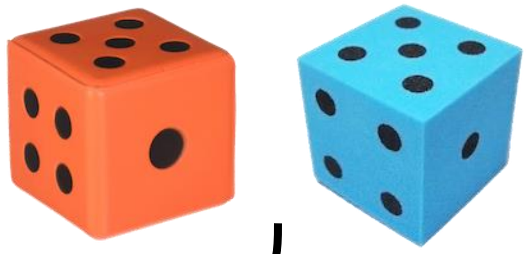


Value dice 1

Value dice 2

Sum of Two Die = 7?

Each outcome



Value dice 1

Value dice 2

Roll two 6-sidex dice. What is probability the sum = 7?
Let E be the event that the sum is 7

S = {

[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] }

E = *in blue*

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.16\bar{6}$$

Sum of Two Die = 7?

```python
import random
from tqdm import tqdm

N_TRIALS = 10000000  # getting close to infinity
TARGET_SUM = 7       # do the two dice sum to 6?

def main():
    n_events = 0
    for i in tqdm(range(N_TRIALS)):
        dice_total = run_experiment()
        if dice_total == TARGET_SUM:
            n_events += 1
    pr_e = n_events / N_TRIALS
    print(f'after {N_TRIALS} trials')
    print('P(E) ≈ ', pr_e)

def run_experiment():
    d_1 = roll_dice()
    d_2 = roll_dice()
    return d_1 + d_2

def roll_dice():
    # give me a random dice roll
    # alternatively random.randint(1, 7)
    return random.choice([1,2,3,4,5,6])

if __name__ == '__main__':
    # this starts the program in main
    main()
```

Sum of Two Die = 7?

```python
import random
from tqdm import tqdm

N_TRIALS = 10000000  # getting close to infinity
TARGET_SUM = 7       # do the two dice sum to 6?

def main():
    n_events = 0
    for i in tqdm(range(N_TRIALS)):
        dice_total = run_experiment()
        if dice_total == TARGET_SUM:
            n_events += 1
    pr_e = n_events / N_TRIALS
    print(f'after {N_TRIALS} trials')
    print('P(E) ≈ ', pr_e)

def run_experiment():
    d_1 = roll_dice()
    d_2 = roll_dice()
    return d_1 + d_2

def roll_dice():
    # give me a random dice roll
    # alternatively random.randint(1, 7)
    return random.choice([1,2,3,4,5,6])

if __name__ == '__main__':
    # this starts the program in main
    main()
```

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.16\bar{6}$$

Sum of Two Die = 7?

```python
import random
from tqdm import tqdm

N_TRIALS = 10000000  # getting close to infinity
TARGET_SUM = 7       # do the two dice sum to 6?

def main():
    n_events = 0
    for i in tqdm(range(N_TRIALS)):
        dice_total = run_experiment()
        if dice_total == TARGET_SUM:
            n_events += 1
    pr_e = n_events / N_TRIALS
    print(f'after {N_TRIALS} trials')
    print('P(E) ≈ ', pr_e)

def run_experiment():
    d_1 = roll_dice()
    d_2 = roll_dice()
    return d_1 + d_2

def roll_dice():
    # give me a random dice roll
    # alternatively random.randint(1, 7)
    return random.choice([1,2,3,4,5,6])

if __name__ == '__main__':
    # this starts the program in main
    main()
```

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.16\bar{6}$$

```
(base) ~/Desktop/win26_cs109/code/lecture1 python3.9 dice_soln.py
after 10000000 trials
P(E) =  0.1665393
```

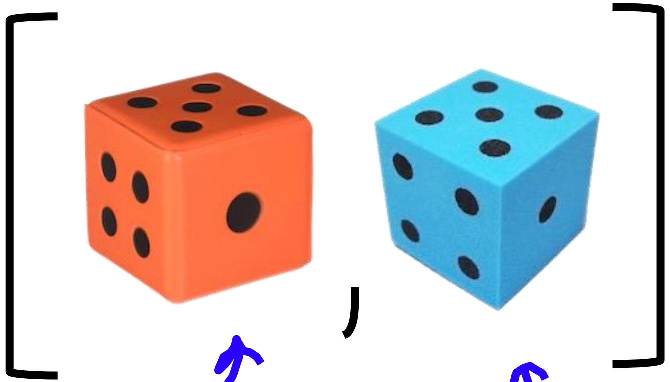
Sum of Two Die = 2?

Roll two 6-sidex dice. What is probability the sum = 2?
Let E be the event that the sum is 2

S = { [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] }

E =

Each outcome



Value dice 1

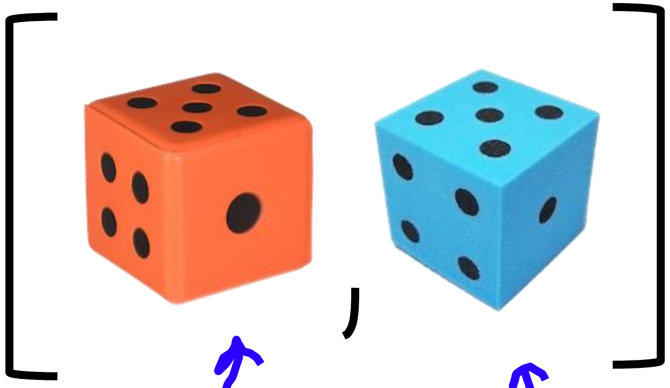
Value dice 2

Sum of Two Die = 2?

Roll two 6-sidex dice. What is probability the sum = 2?
Let E be the event that the sum is 2

S = { **[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] }

Each outcome



Value dice 1

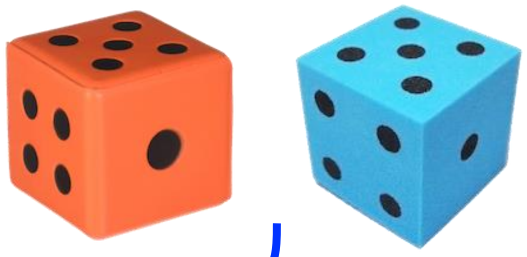
Value dice 2

E = *in red*

$$P(E) = \frac{|E|}{|S|} = \frac{1}{36} = 0.02\bar{7}$$

Other ways to make a Sample Space?

Each outcome

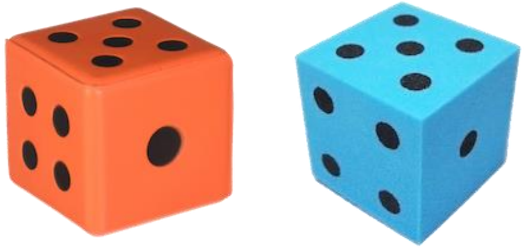


Value dice 1

Value dice 2

Sum of Two Die: Three options for the sample space

Value dice 1



Value dice 2

Think of the die as **distinct**

[5 , 5]

Value dice 1

Value dice 2

Think of the die as **indistinct**

{ 5 , 5 }

Value of a dice

Value of a dice

Just look at the sum

10

Sum of Two Die = 7? Bug: Die are Indistinct

Each outcome

Just look at the sum

Roll two 6-sidex dice. What is probability the sum = 7?
Let E be the event that the sum is 7

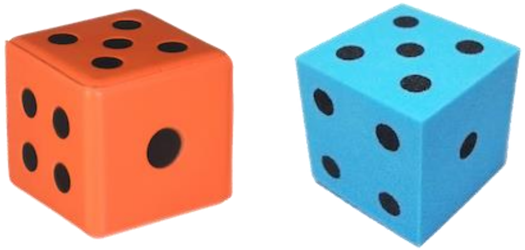
S = { 2 3 4 5 6
7 8 9 10 11 12 }

$$P(E) = \frac{|E|}{|S|} = \frac{1}{11} = 0.0\overline{9}$$

E = *in red*

Sum of Two Die: Three options for the sample space

Value
dice 1



Value
dice 2

Think of the die as **distinct**

[5 , 5]

Value
dice 1

Value
dice 2

Think of the die as **indistinct**

{ 5 , 5 }

Value of
a dice

Value of
a dice

Just look at the sum

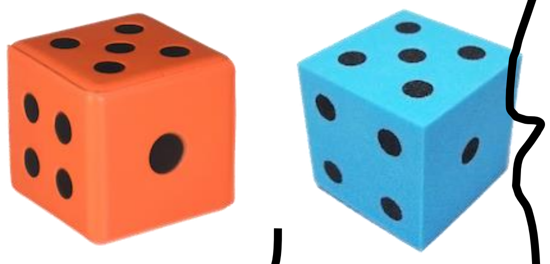
10

Sum of Two Die = 7? Bug: Die are Indistinct

Roll two 6-sidex dice. What is P(sum = 7)?

S = { {1,1} {1,2} {1,3} {1,4} {1,5} **{1,6}**
{2,2} {2,3} {2,4} **{2,5}** {2,6}
{3,3} **{3,4}** {3,5} {3,6}
{4,4} {4,5} {4,6}
{5,5} {5,6}
{6,6} }

Each outcome



Value dice 1

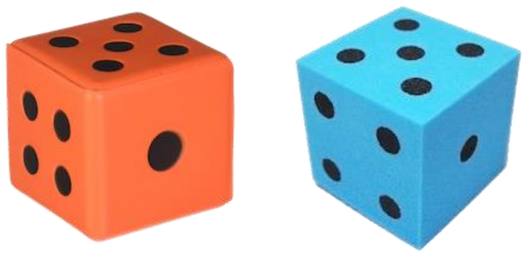
Value dice 2

E = *in blue*

$$P(E) = \frac{|E|}{|S|} = \frac{3}{20} = 0.15?$$

Sum of Two Die: Three options for the sample space

Value
dice 1



Value
dice 2

Think of the die as **distinct**

[5, 5]

Value
dice 1

Value
dice 2

Think of the die as **indistinct**

{5, 5}

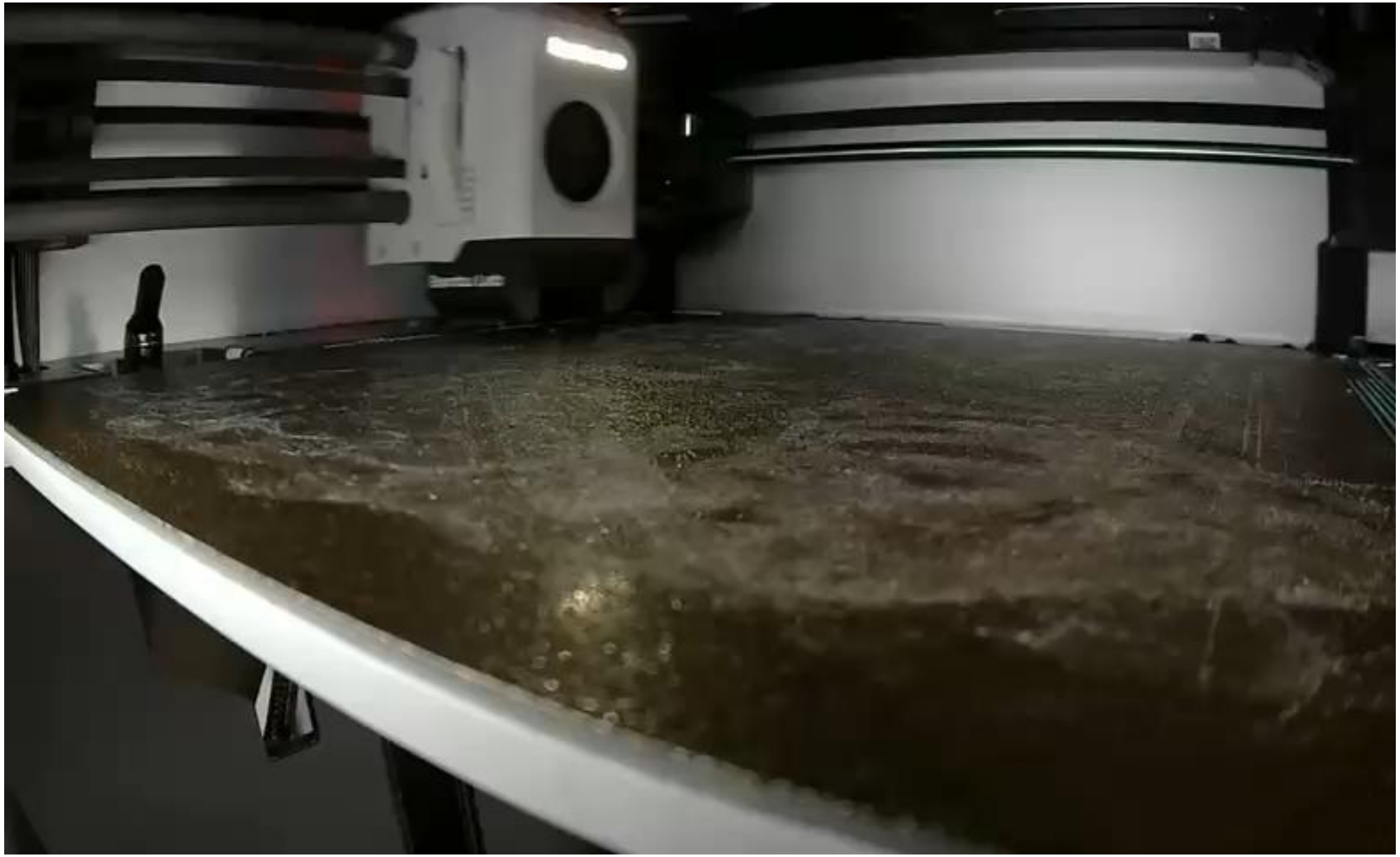
Value of
a dice

Value of
a dice

Just look at the sum

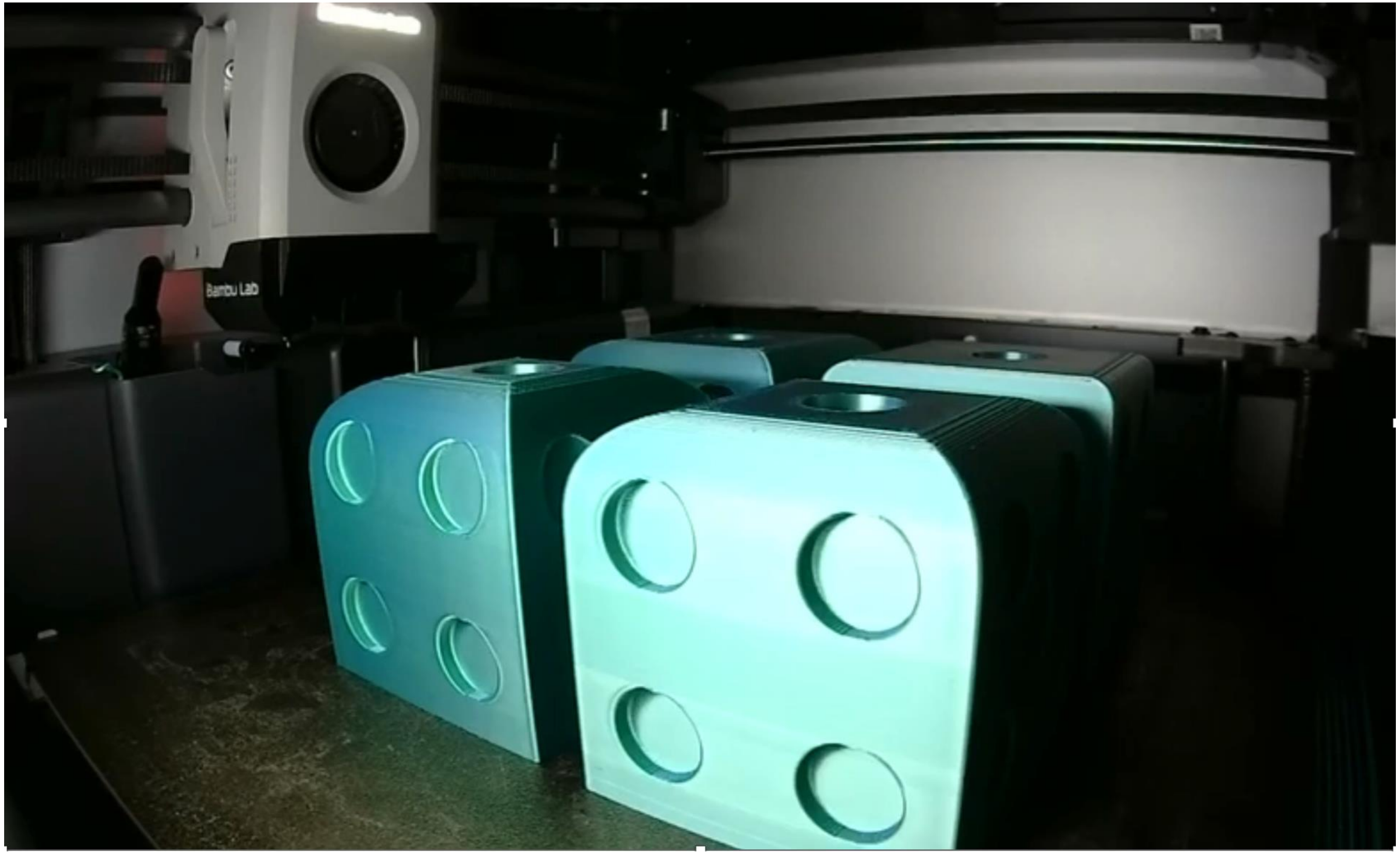
10

Sum of Two of These Die = 7?



We printed this for CS109

Sum of Two of These Die = 7?



We printed this for CS109


I'm going on an adventure!

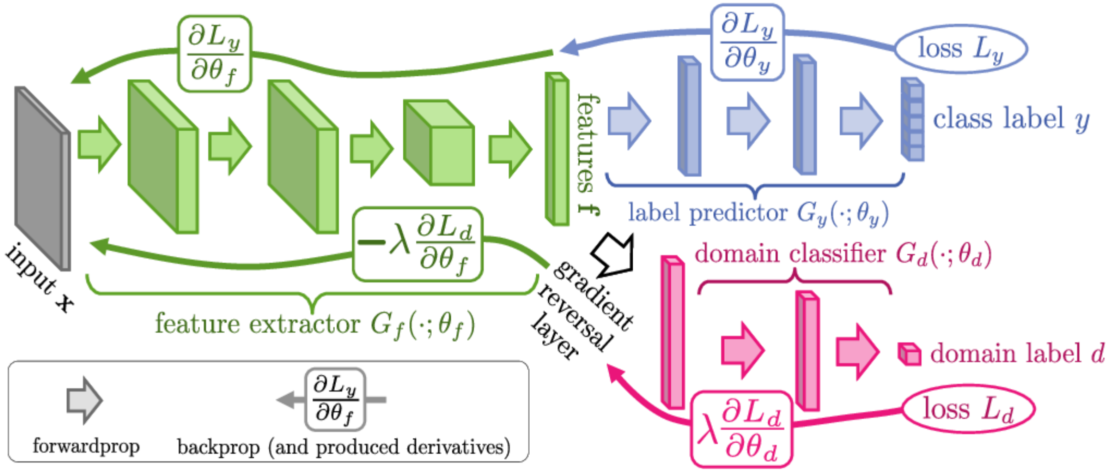
Before we go....

A tale of Research by CS109 Students

Fair AI with Adversarial Network

221 Citations

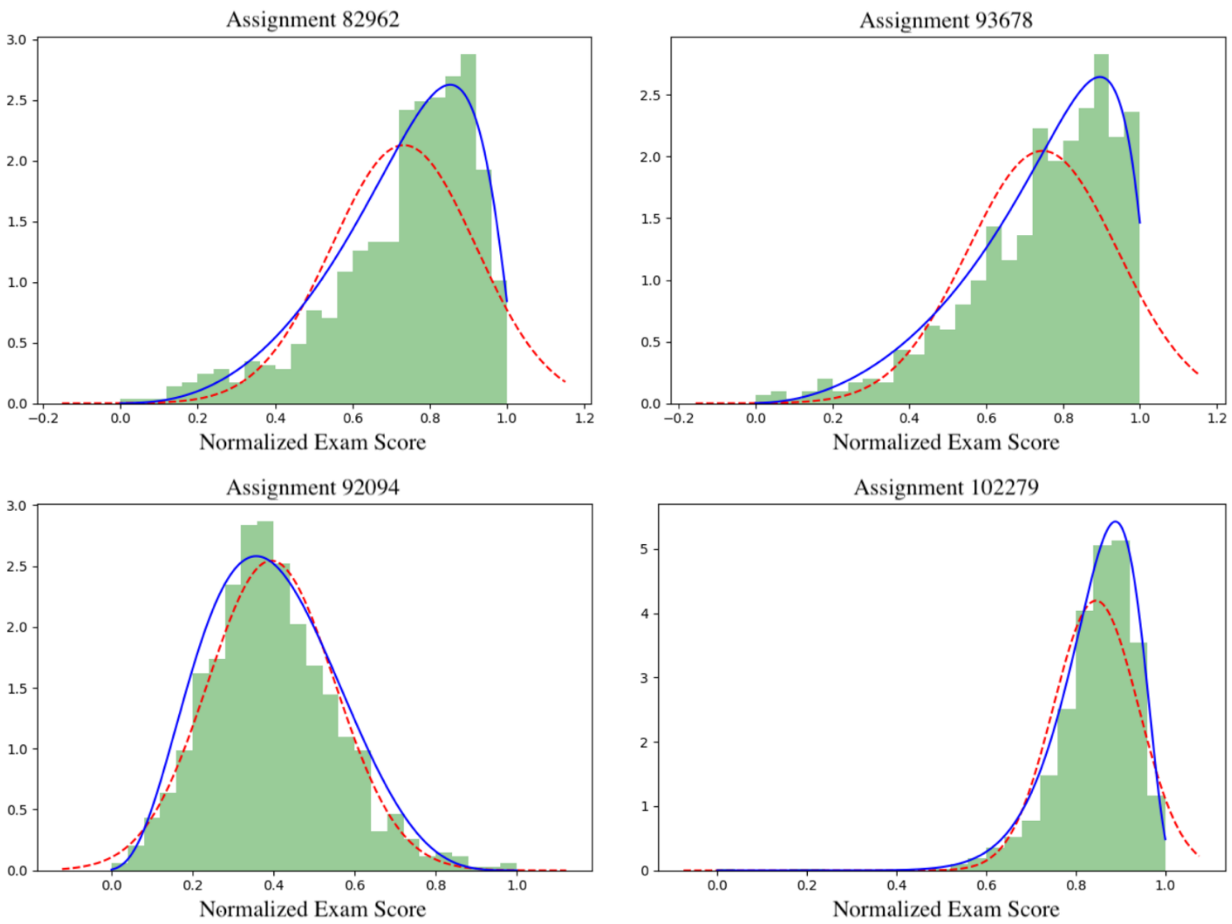
With undergrads Christina Wadsworth and Francesca Vera

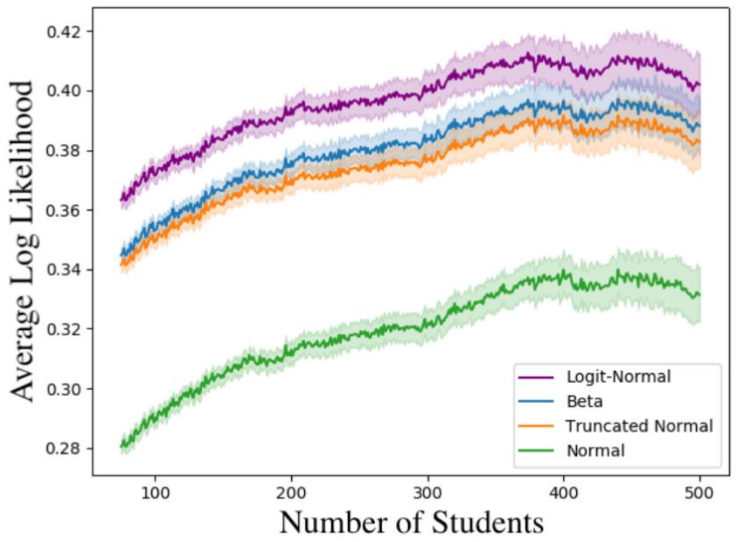


Model	Accuracy	FP Gap	FN Gap
COMPAS Scores (Our Test Set)	0.68	0.17	0.22
Our Recidivism Model	0.70	0.15	0.27
Our Chosen Adversarial Model	0.70	**0.01**	**0.02**
Bechavod et al. AVD Penalizers (2017)	0.65	**0.02**	**0.04**
Bechavod et al. SD Penalizers (2017)	0.66	**0.02**	**0.03**
Bechavod et al. Vanilla 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

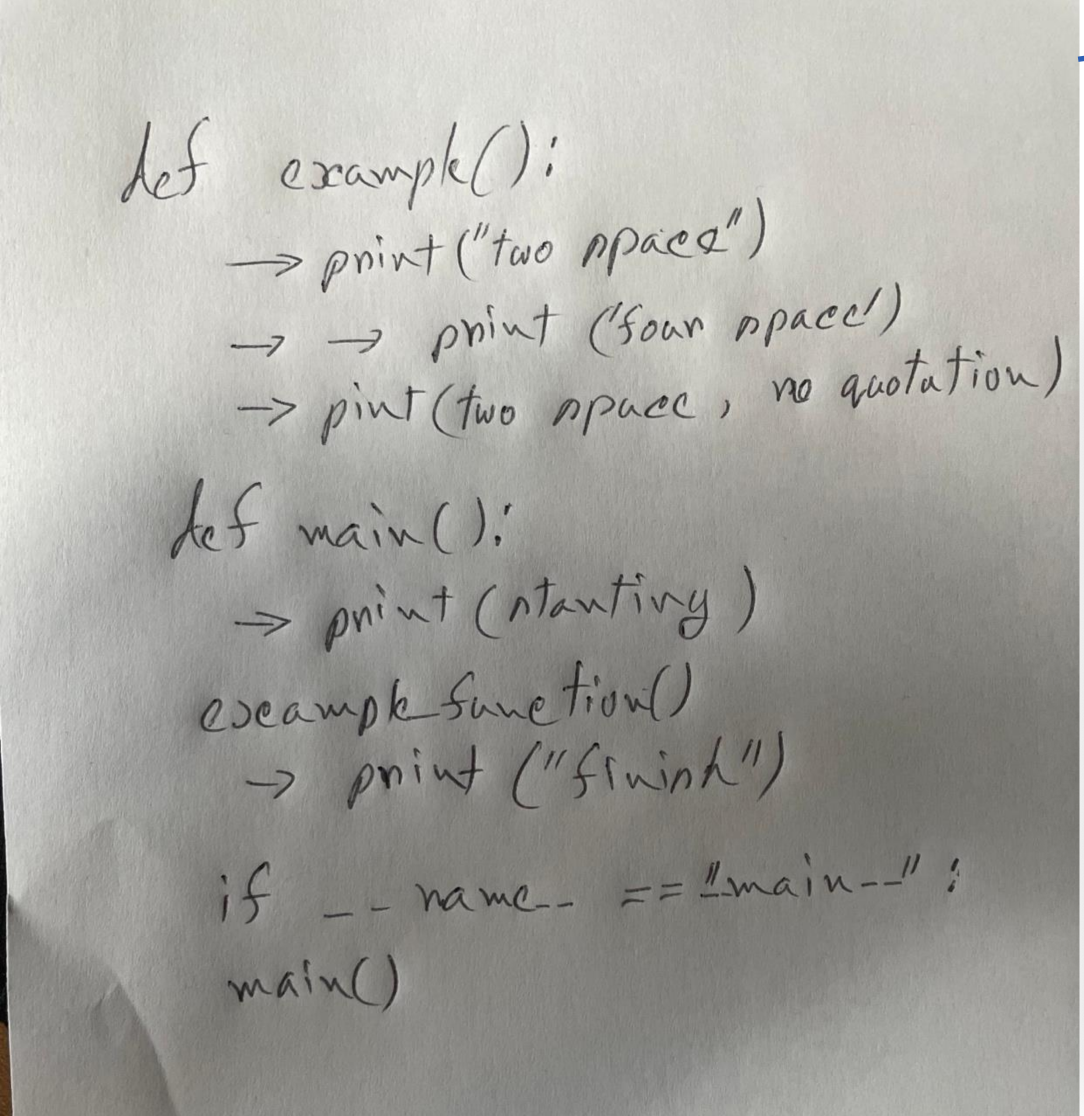




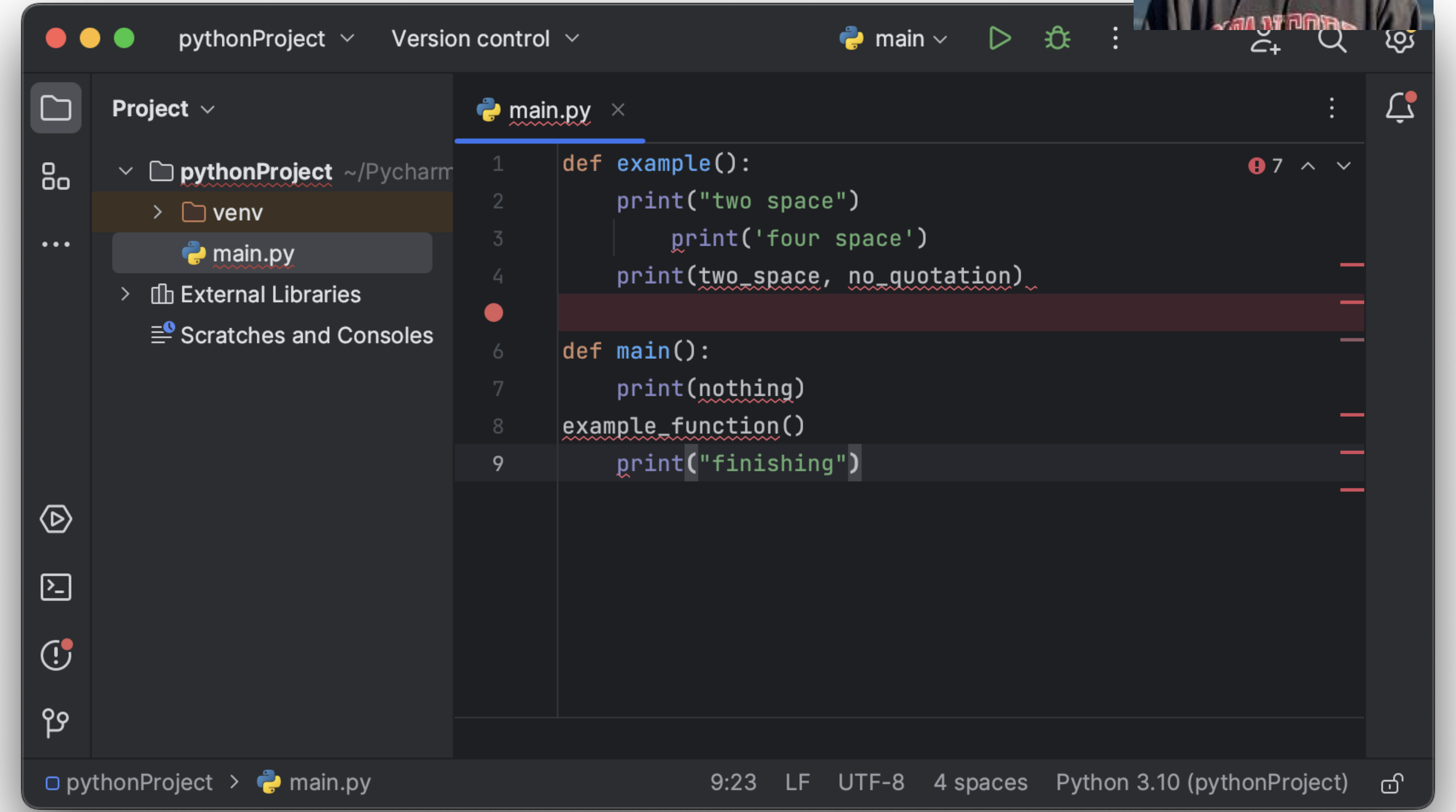
Noah Arthurs, CS109 Student

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

Analyzed 4,000 Gradescope Courses



Transcribed handwritten code

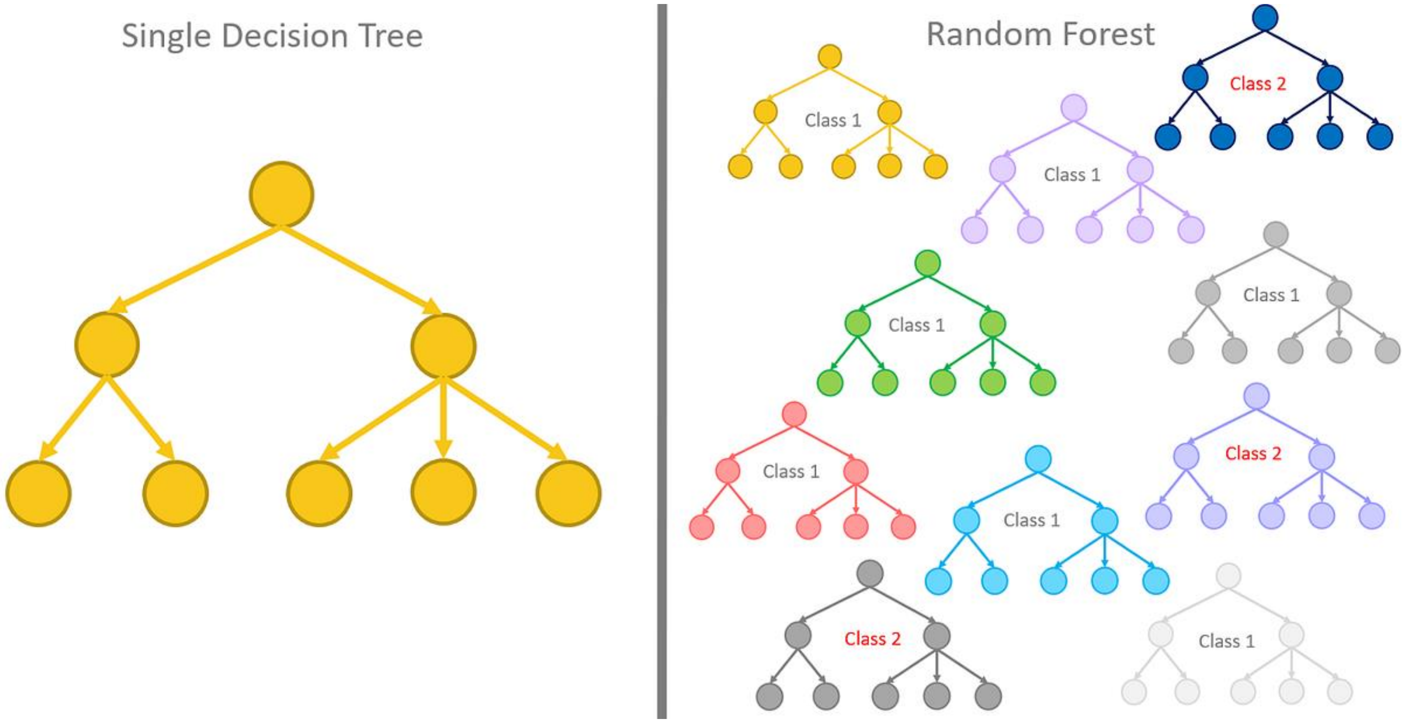


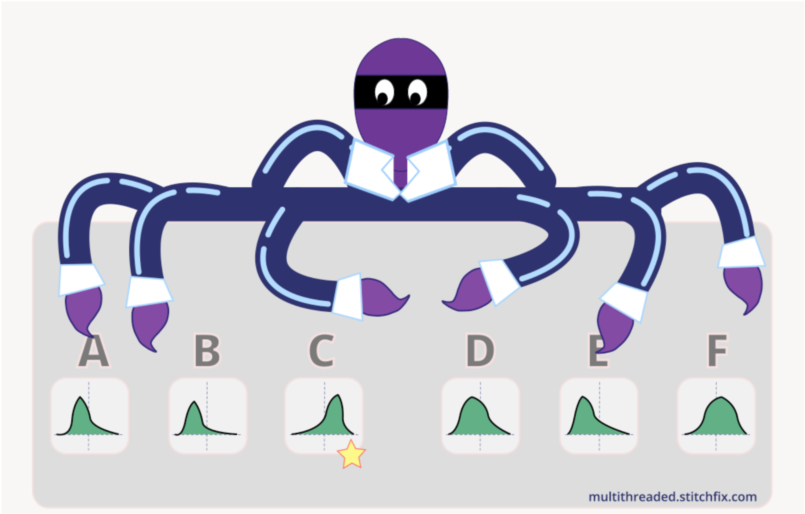
Math question: Can you relate pixel position to intended indentation using probability?

Undergraduate from CS109 Fall 23

Won Best Undergraduate Paper for L@S

Faster Forest Training Using Multi-Armed Bandits





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


I'm going on an adventure!