

CS109: Probability for Computer Scientists

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



But first...
who am I?

Professor Chris Piech

Teaching at Stanford

CS106A

Programming
Methodologies

LAST: SPR 2024

CS109

Probability for Computer
Scientists

LAST: WIN 2025

8,000+ students over 10 years

CS398

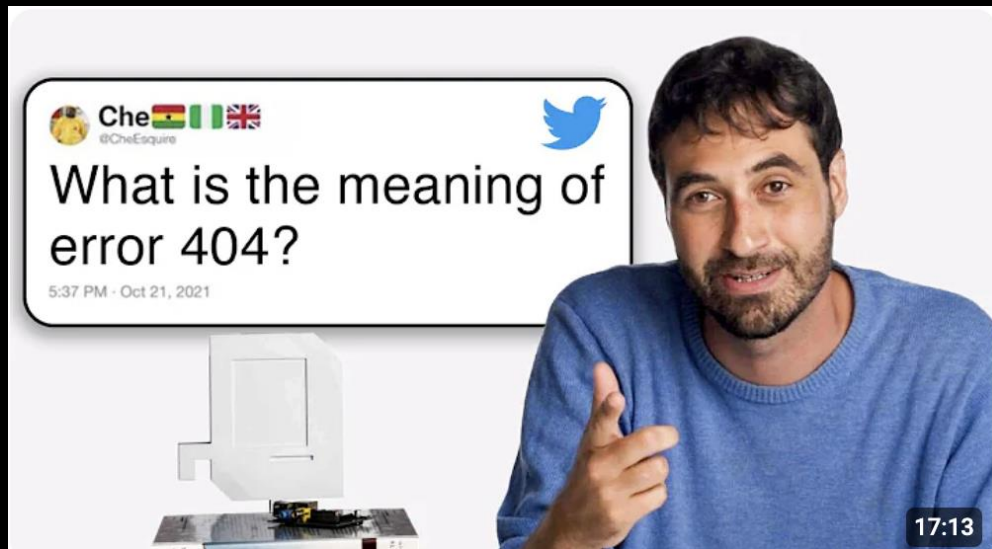
Computational Education

LAST: SPR 2022

CS106B

Programming
Abstractions

LAST: FALL 2016



Code in Place: the
course with the most
teachers



From Three Places!



Grew up in Nairobi, Kuala Lumpur before Stanford!



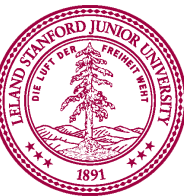
Long History in CS109



I took the first CS109 back when I looked like this



Been teaching it since 2014



Stanford University

Computer Science Department

Stanford AI

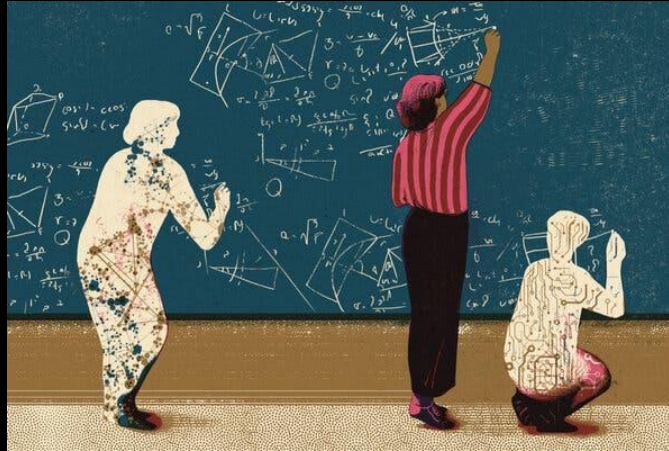
Piech Lab



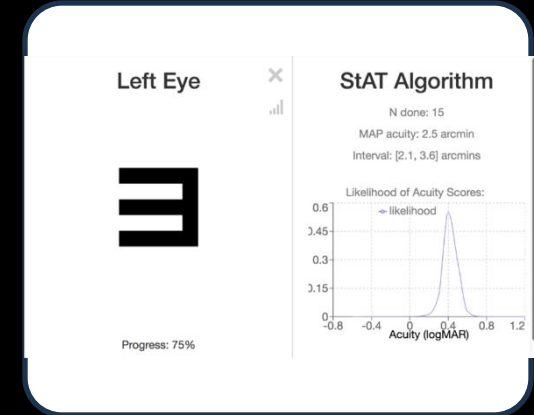
Professor Chris Piech



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



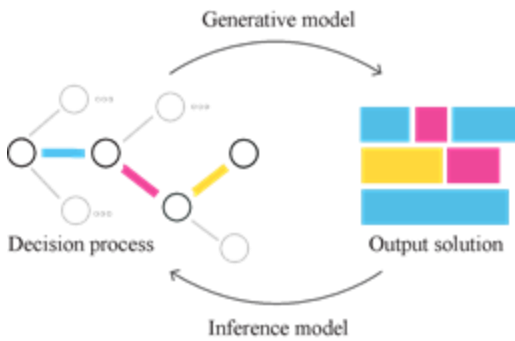
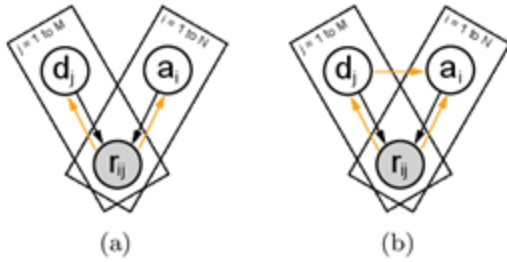
Co-Invented Generative Grading, used in Code.org, featured in the New York Times



Co-Created Stanford Acuity Test, Most Accurate Eye Exam!



Computational Education (and medicine too)



THIS IS MATH

$$\begin{array}{r} 1 \text{ to } \\ 243 \\ - 87 \\ \hline 156 \end{array}$$

THIS IS MATH ON COMMON CORE

ANY QUESTIONS?

AI generated feedback

Students evaluate the feedback

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

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

Traduza esta frase

Is this house for sale?

esta casa está à venda?



Play to Grade

demo_student_2 Not Graded

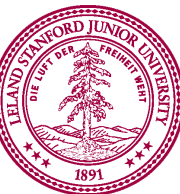
Assignment: Breakout

Mouse movement	4/4	▼
Brick drawing	5/5	▼
Paddle drawing	4/4	▼
Ball drawing	4/4	▼
Constants	0/2	▼
Wall bouncing	3/3	▼
Paddle bouncing	2/2	▼
Paddle skewering	0/1	▲
Ball does not become skewered on paddle (1) <input type="checkbox"/>		
New life	4/4	▼

Total Score 37/40

Video Code Demo

Best of STEM award for AI Tutor with Code.org

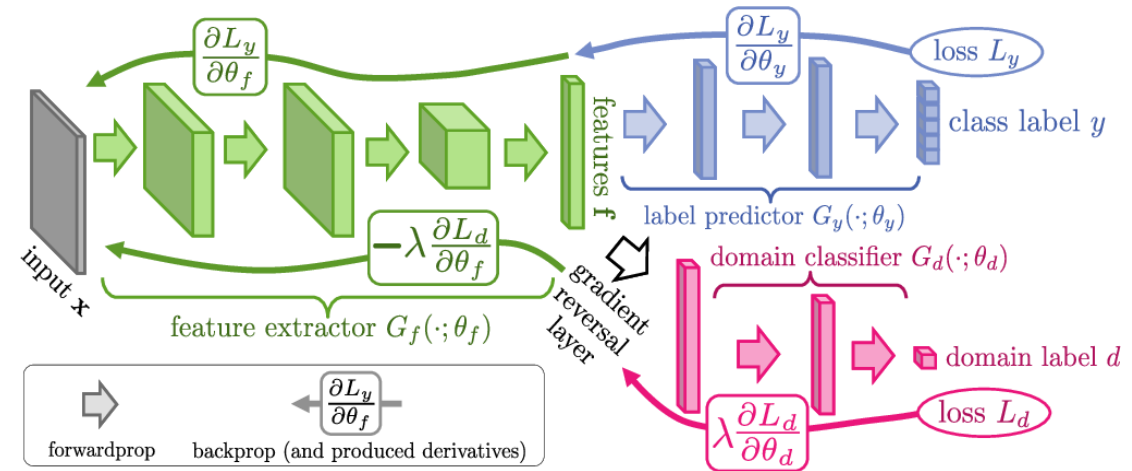


Research by CS109
Students

Fair AI with Adversarial Network

221 Citations

With undergrads Christina Wadsworth and Francesca Vera

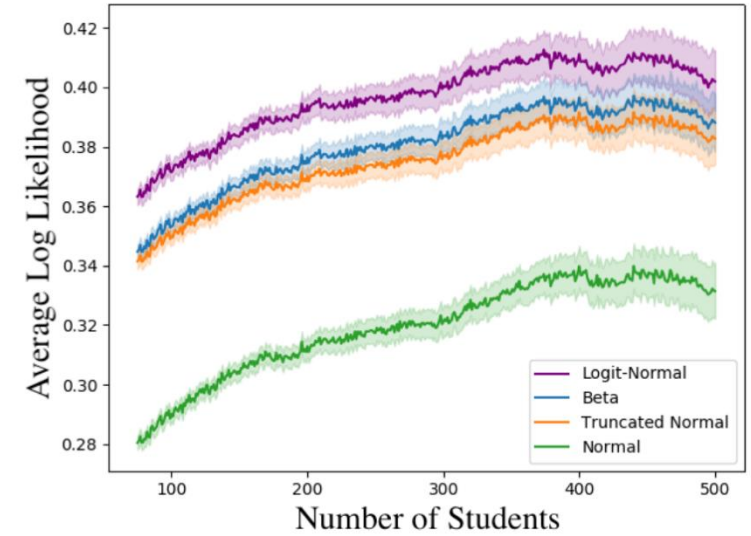
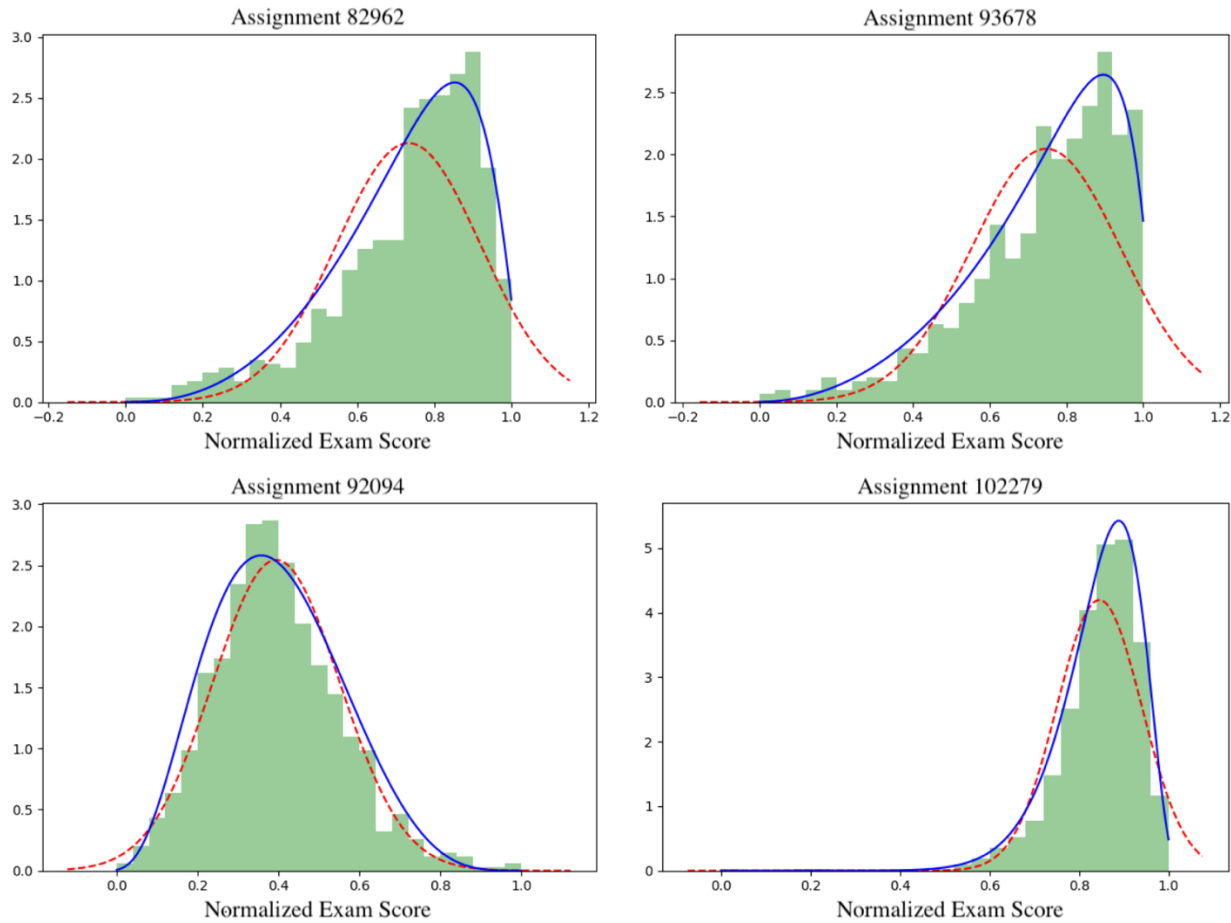


MODEL	ACCURACY	FP GAP	FN GAP
COMPAS SCORES (OUR TEST SET)	0.68	0.17	0.22
OUR RECIDIVISM MODEL	0.70	0.15	0.27
OUR CHOSEN ADVERSARIAL MODEL	0.70	0.01	0.02
BEHAVOD ET AL. AVD PENALIZERS (2017)	0.65	0.02	0.04
BEHAVOD ET AL. SD PENALIZERS (2017)	0.66	0.02	0.03
BEHAVOD ET AL. VANILLA REGULARIZED (2017)	0.67	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?



Grades are Not Normal



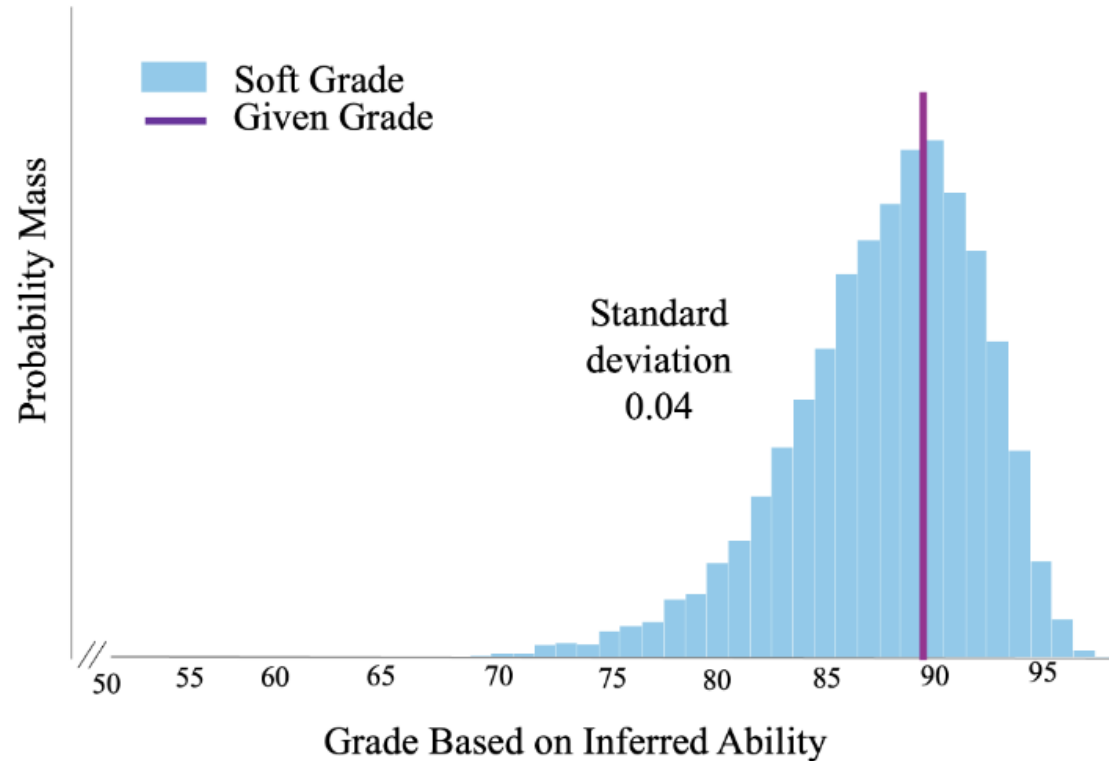
Noah Arthurs, CS109 Student



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

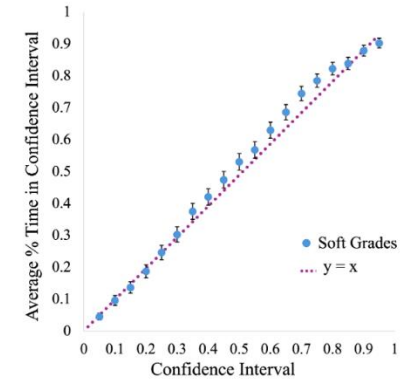


Soft Grades



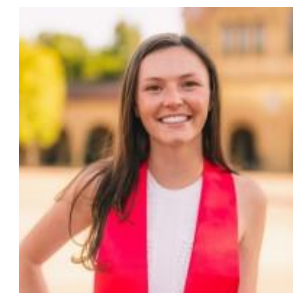
$$\sigma_{\text{posterior}}^2 = \left(\frac{1}{\sigma_{\text{prior}}^2} + \sum_{j=1}^k \frac{1}{\epsilon_j^2} \right)^{-1}$$

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

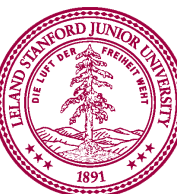


	Exact	Likelihood ±1	±3	Expected Calibration Error (ECE)
Oracle	0.119 ± 0.007	0.324 ± 0.013	0.612 ± 0.016	0.018
Soft Grades	0.118 ± 0.007	0.323 ± 0.014	0.605 ± 0.016	0.020
FMN Baseline	0.066 ± 0.007	0.192 ± 0.019	0.407 ± 0.027	0.131

Math question: Can you represent uncertainty around grades given in classes like the ones at Stanford?



See also DPO for grading



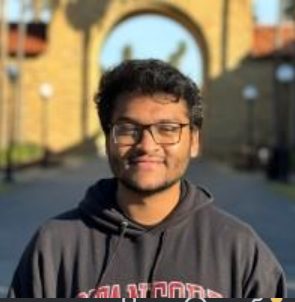
```
def example():  
    → print("two space")  
    → → print('four space')  
    → pint(two space, no quotation)
```

```
def main():  
    → print(rinting)
```

```
example_function()  
    → print("finishing")
```

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

Transcribed handwritten code

A screenshot of a code editor window. The title bar shows 'pythonProject' and 'Version control'. The left sidebar shows a project tree with 'pythonProject', 'venv', and 'main.py'. The main editor area shows the following Python code:

```
1 def example():  
2     print("two space")  
3     print('four space')  
4     print(two space, no quotation)  
6 def main():  
7     print(rinting)  
8     example_function()  
9     print("finishing")
```

The code is color-coded and has line numbers. The status bar at the bottom shows 'pythonProject > main.py', '9:23', 'LF', 'UTF-8', '4 spaces', and 'Python 3.10 (pythonProject)'.

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

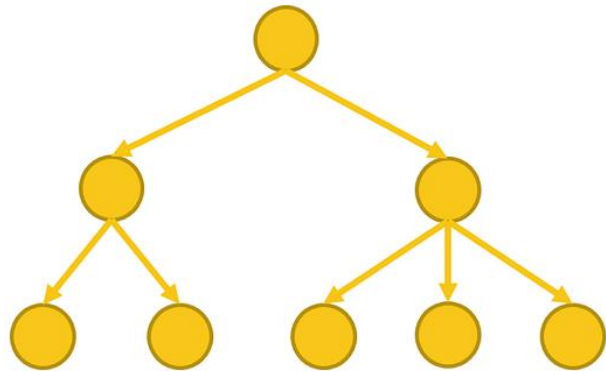


Undergraduate from CS109 Fall 23

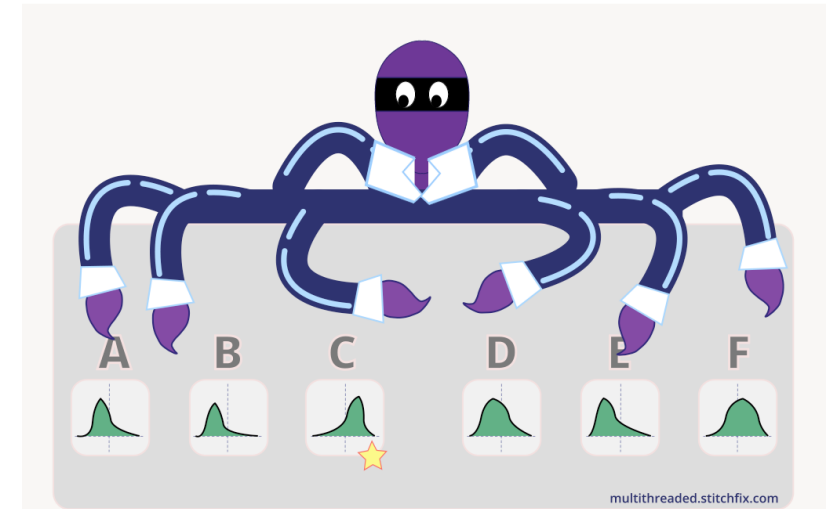
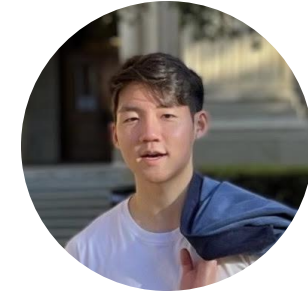
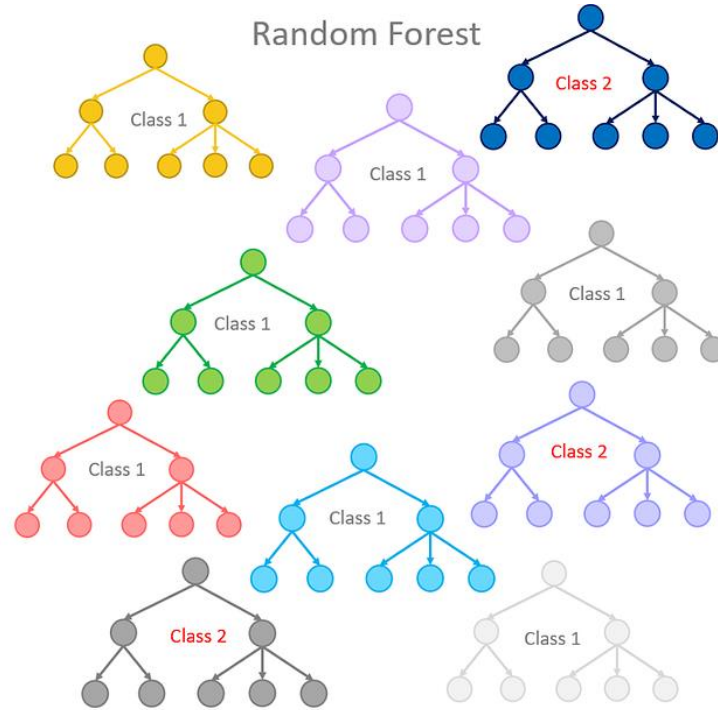
Won Best Undergraduate Paper for L@S

Faster Forest Training Using Multi-Armed Bandits

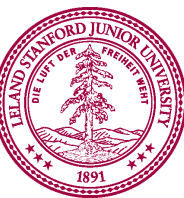
Single Decision Tree



Random Forest



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



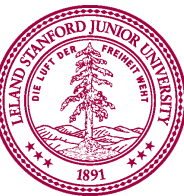
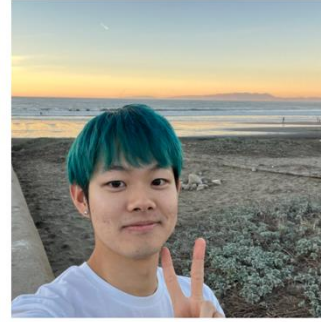
So many things to love in this world



Most amazing family



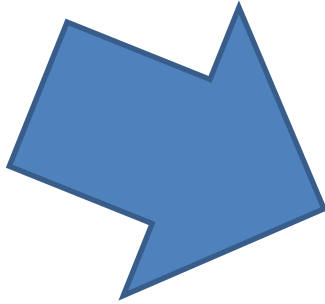
Fantastic Teaching Team



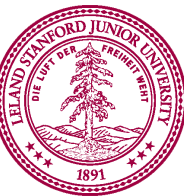
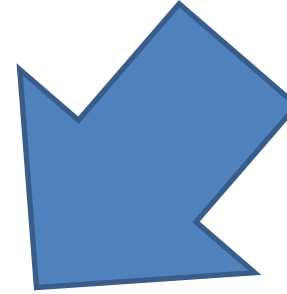
Course mechanics

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

Essential Information



cs109.stanford.edu



Are you in the right place?

Prerequisites

What you really need:

CS106B/X (important, coreq ok):

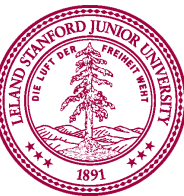
- Recursion
- Hash Tables
- Binary Trees
- Programming

CS103 (not necessary):

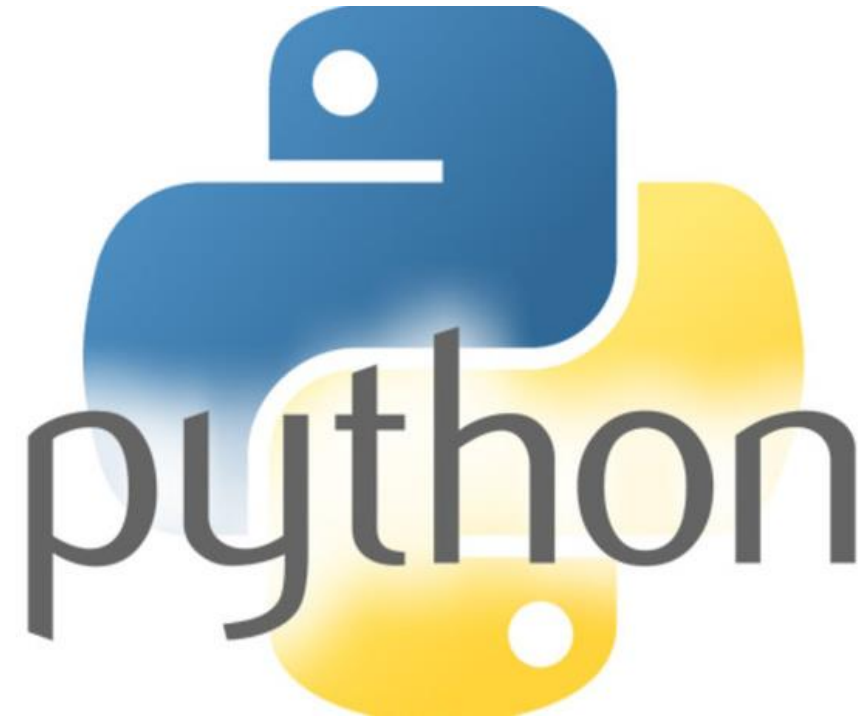
- Proof techniques (induction)
- Set theory
- Math maturity

Math 21 or equivalent (important, coreq ok)

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

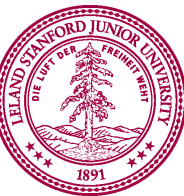


Coding in CS109



Review session on Friday

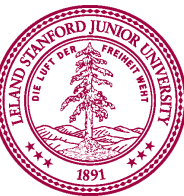
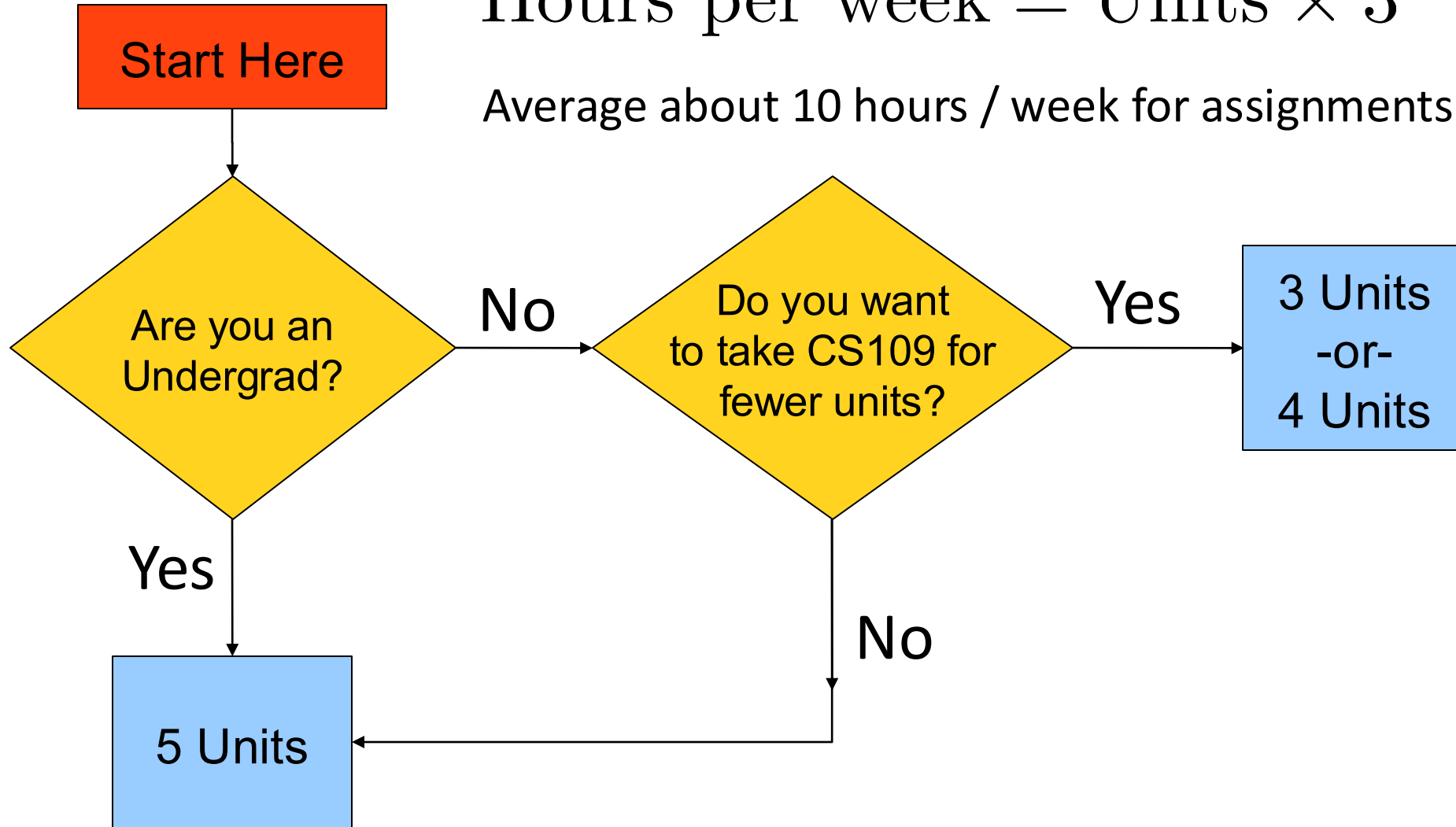
Piech, CS109, Stanford University



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, Oct 28th, 7pm

30%

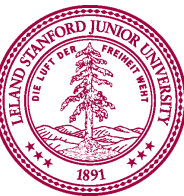
Final

3 hour exam, Dec 9th, 8:30am

20%

Participation

Section, PEP



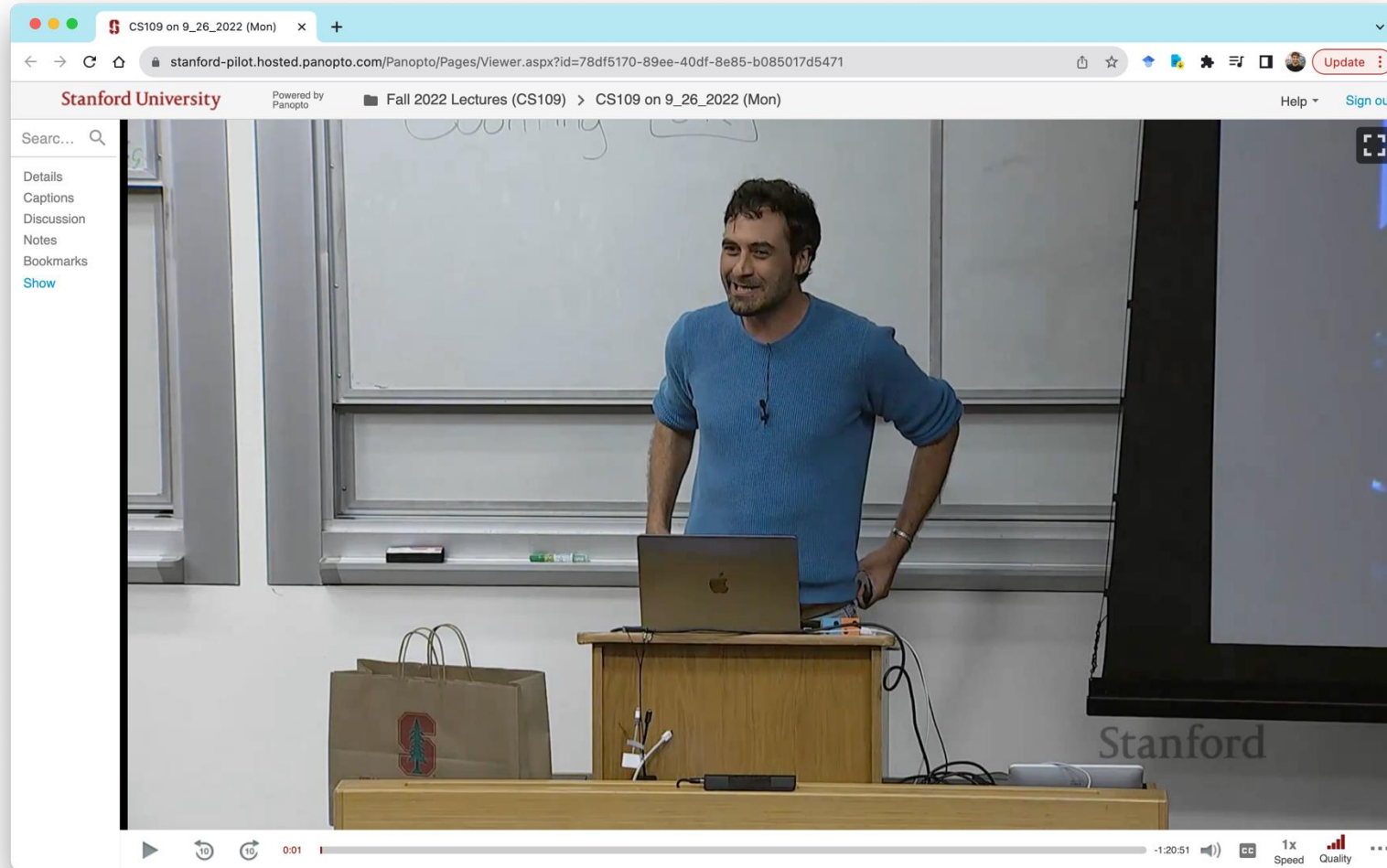
Personal Exam Prep (PEP)

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

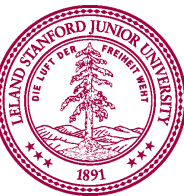
Song of the Quarter

[coming on Wednesday]

Is Class Online?



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



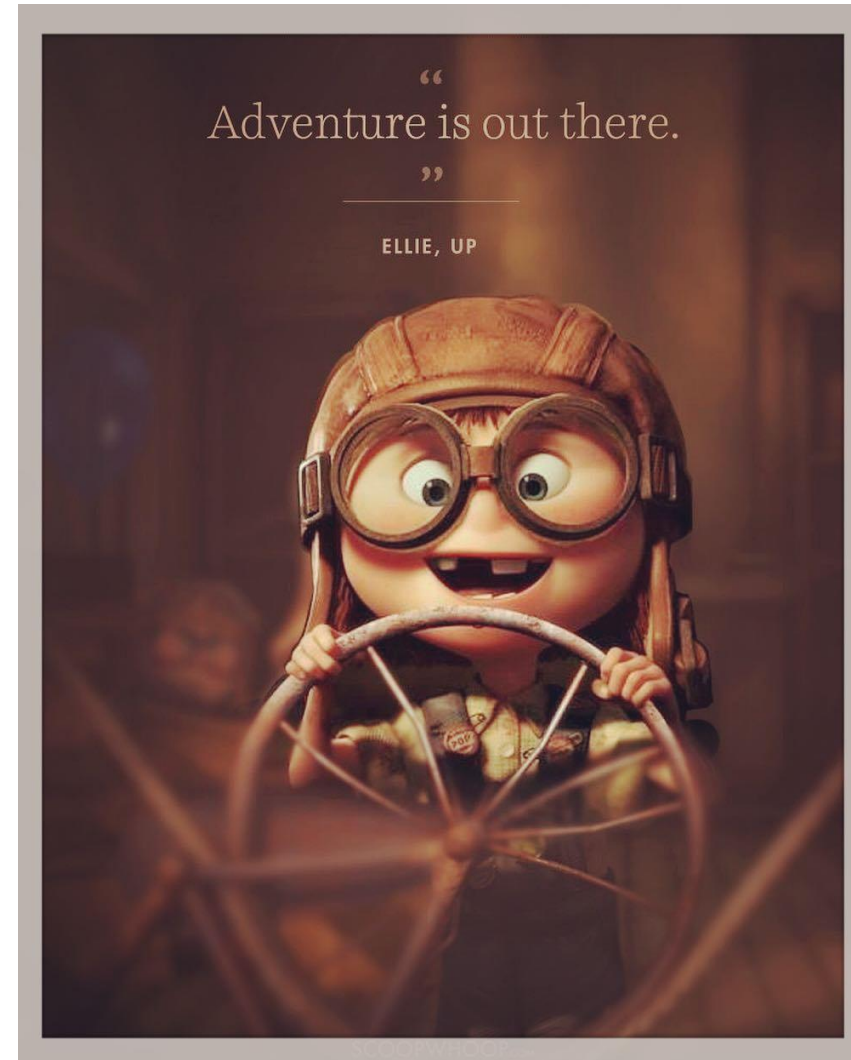
Come to Class

Section assumes you
are caught up to
Monday's lecture

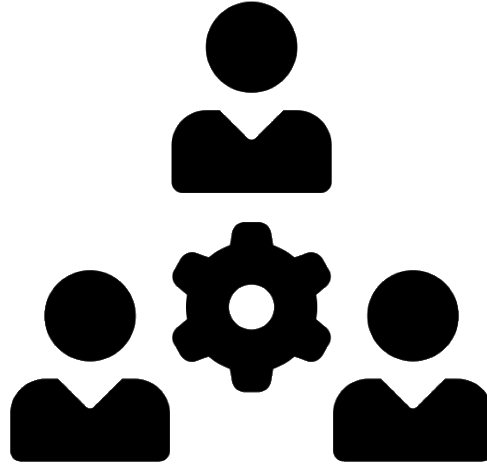
"I met my husband in
your CS109 lectures"
- Student from 3 years
ago

Attending lecture
strongly correlates
with better grades

We have fun.



Ask questions



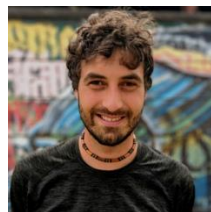
Q&A forum
All announcements



“Working” office hours
start on Saturday



Email cs109@cs.stanford.edu



Chris has 1:1 CS109 office hours (not for psets)

Chris Wrote a 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

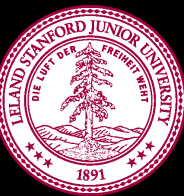


CS109
Department of Computer Science
Stanford University
December 2020
V 0.1.0.4


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

Piech, CS109, Stanford University



ACE Companion Course



The screenshot shows a Google Docs document with a colorful illustration of a forest scene at the top. The text on the page provides details about the CS ACE Program Application for Fall 2025, including its purpose, requirements, and deadlines.

CS ACE Program Application - Fall 2025

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.

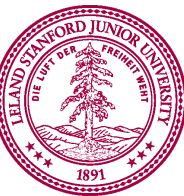
The priority deadline is 11:59 PM Friday September 19, 2025.

The **FINAL** deadline for ACE applications is Friday, September 26, 2025, by 11:59 PM. Applicants will be notified of their application result on a rolling basis.

ACE course enrollment is limited to undergraduate students. Priority will be given to undergraduates pursuing degrees in the School of Engineering.



<https://forms.gle/ayGGXCsXrh1tgaVp9>



CS109 | Syllabus

web.stanford.edu/class/cs109/handouts/syllabus.html

CS109 Course Resources Problem Sets Lecture Section Schedule

Syllabus

Schedule

Honor Code

Office Hours

Course Reader

Python Review

Latex Cheat Sheet

Lecture Videos

Syllabus


A DAY AGO

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

Teaching Team

Professor: Chris Piech
✉ piech@cs
🏠 Durand 305

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



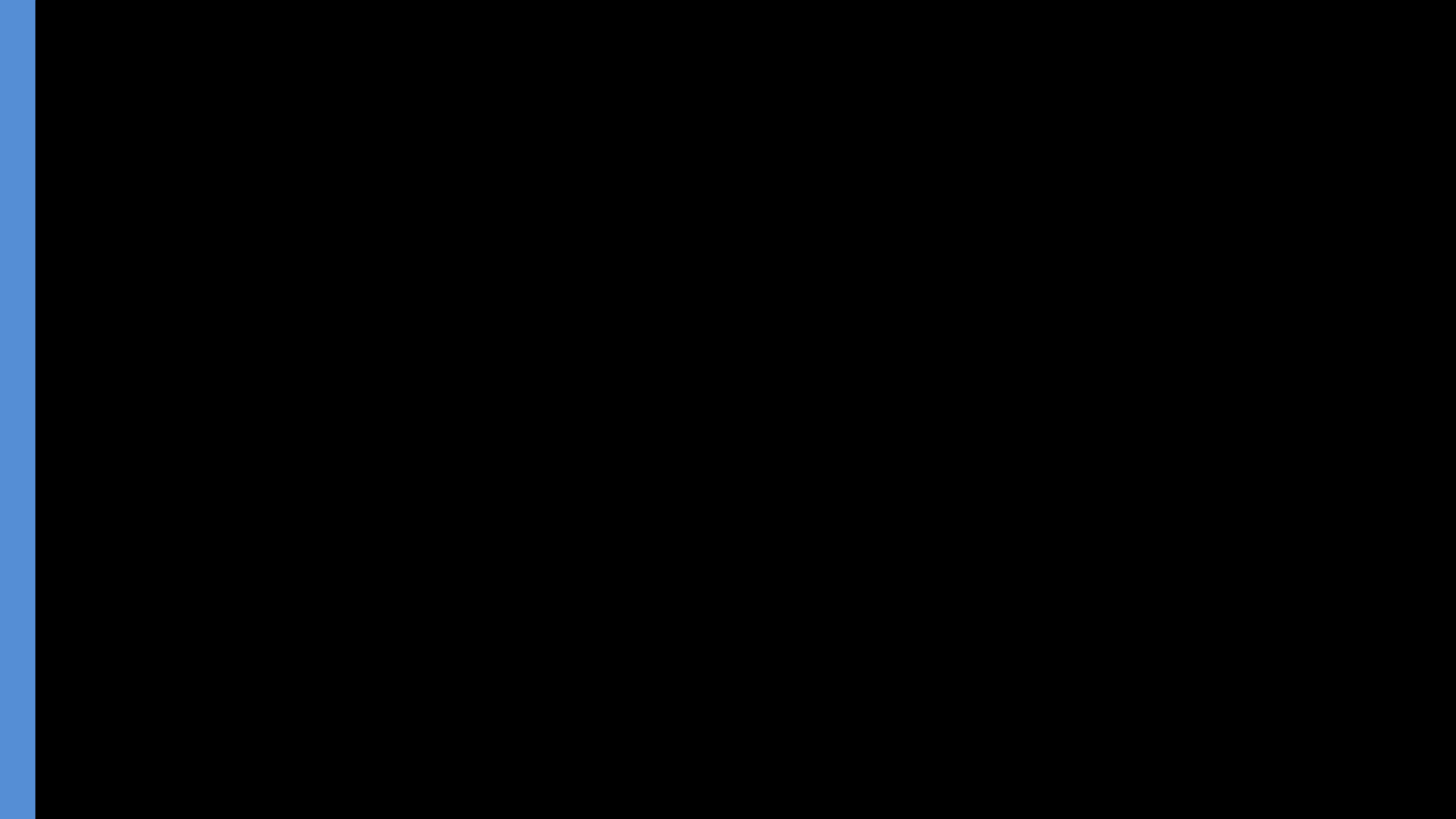
<https://web.stanford.edu/class/cs109/handouts/syllabus.html>

Teaching Team

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

There is always a better way!

- Improved curricula
- Under the hood of an LLM



Story of Modern AI

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

Brief History

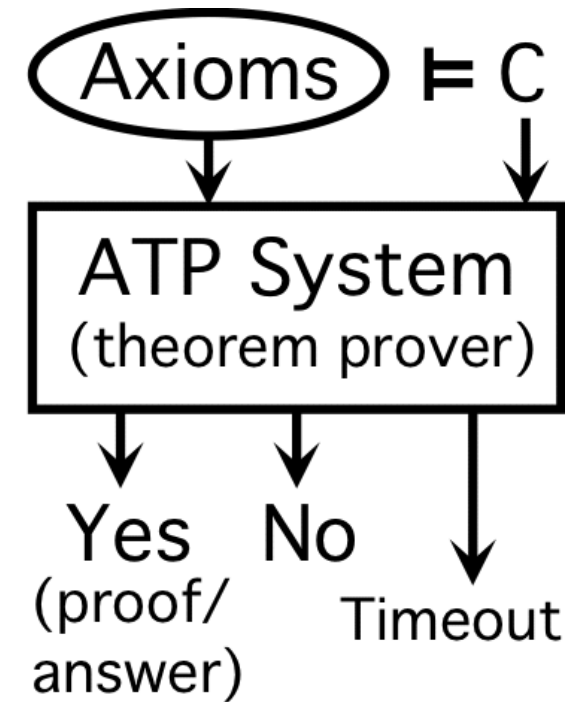


Early Optimism 1950s

1952

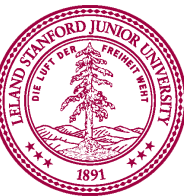


1955



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.

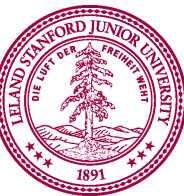


(Russian)



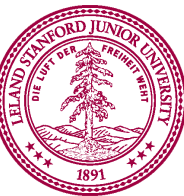
The vodka is good but the meat is rotten.

The world is too complex



BRACE YOURSELVES

WINTER IS COMING



Something is going on in the world of AI

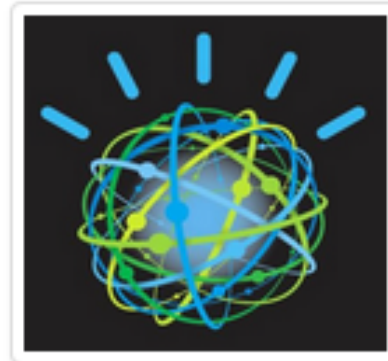
Big Milestones Part 1



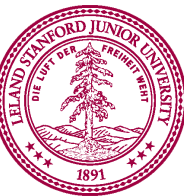
1997 Deep Blue



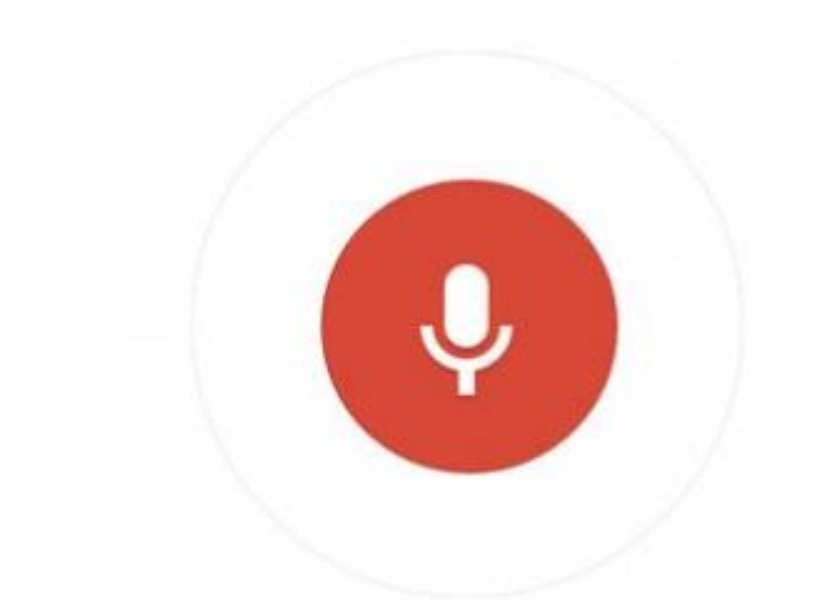
2005 Stanley



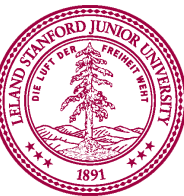
2011 Watson



I was told speech was 30 years out



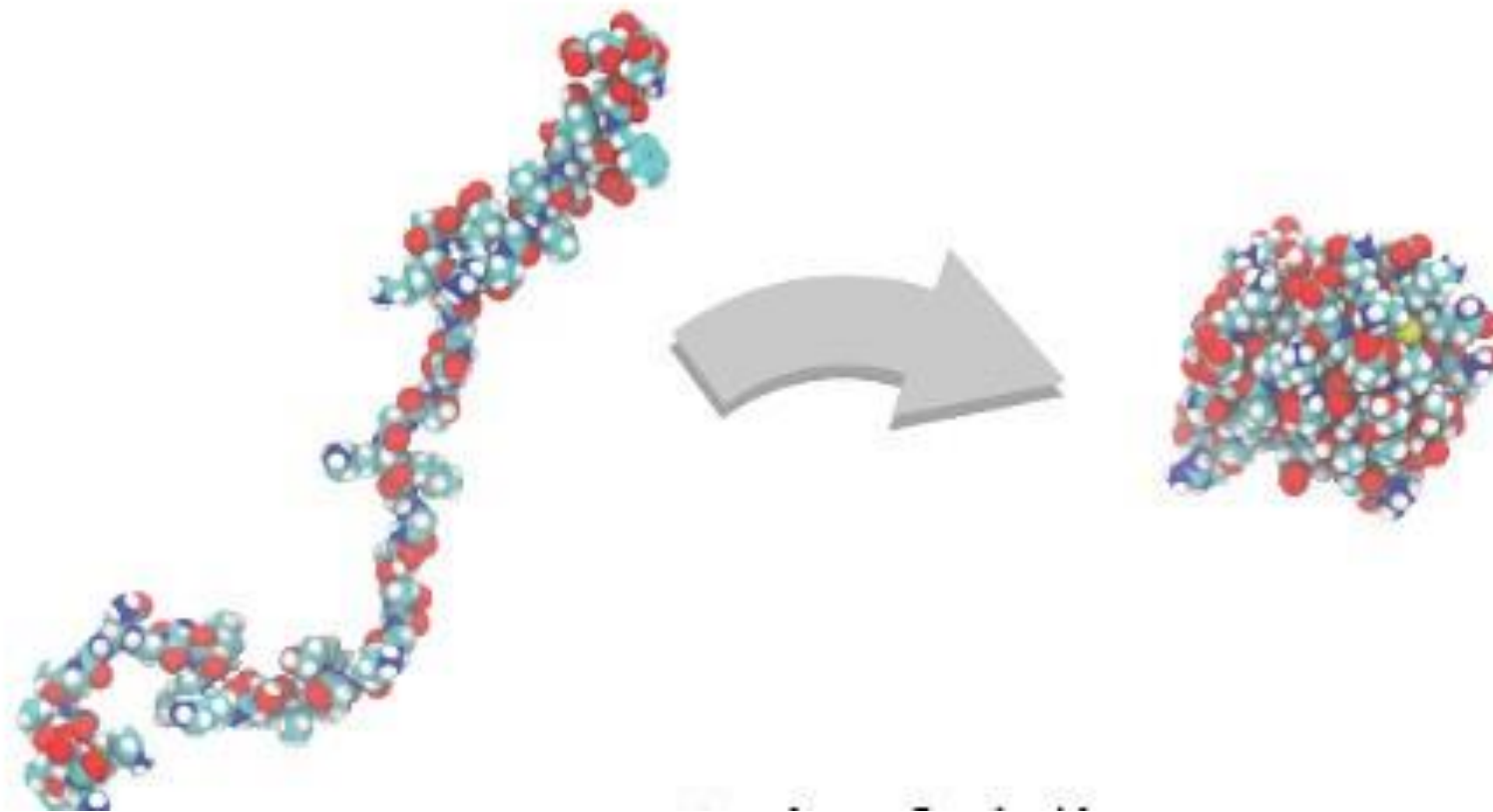
Almost perfect...



The last remaining board game



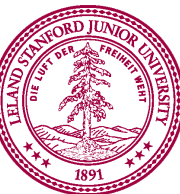
Protein Folding



Directions From A to B

The screenshot shows a Google Maps interface with the following elements:

- Address Bar:** 2600 Columbia St to Mexico City, Mexico
- Search Bar:** google.com/maps/dir/2600+Columbia+St,+Palo+Alto,+CA+94304,+USA/Mexico+City,+CDMX,+Mexico/@28.1116947,-119.664044...
- Left Sidebar:**
 - Home (2600 Columbia St)
 - Mexico City, Mexico
 - Options
 - Send directions to your phone
 - Route 1: 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.", "Your destination is in a different time zone." Details link.
 - Route 2: via México 15D, 38 hr, 2,207 miles.
 - Route 3: via I-5 S and I-10 E, 38 hr, 2,405 miles.
 - Explore Mexico City (Restaurants, Hotels, Gas stations, Parking Lots, etc.)
- Map:** Shows a blue route from San Francisco, CA to Mexico City, Mexico, passing through Los Angeles, San Diego, Phoenix, Tucson, Ciudad Juárez, and Monterrey.
- Map Controls:** Layers, Google logo, Map data ©2022 Google, INEGI, United States, Terms, Privacy, Send feedback, 200 mi scale.



3:31



Google Translate



English



Ukrainian



ENGLISH



Please translate this into Ukrainian.
Thank you



Camera



Conversation



Transcribe

UKRAINIAN



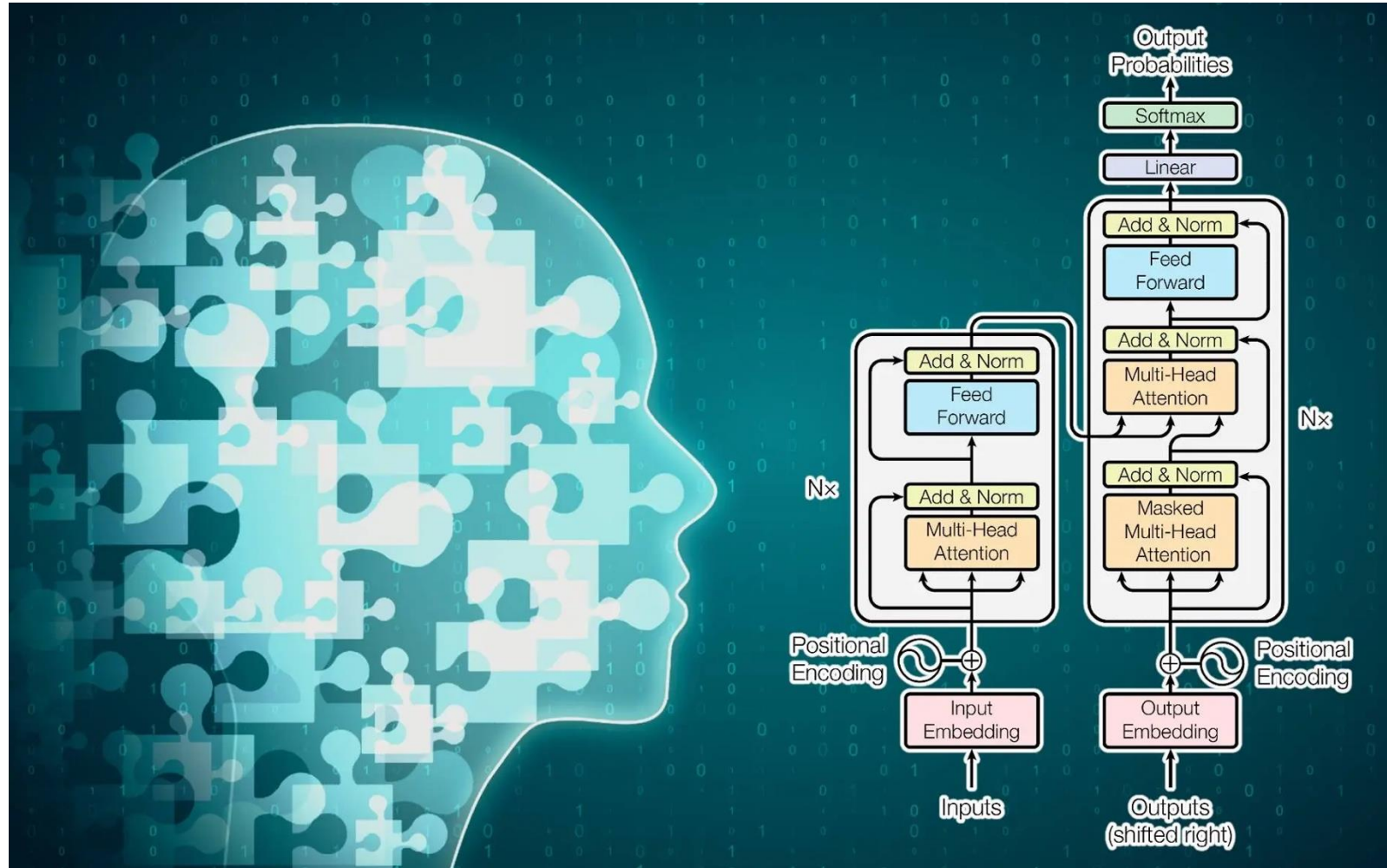
Будь ласка, переведіть це

Self Driving Cars



And then in 2022, everything changed, again

AI that (seems) to understand language



DallE Images made by Freya (2 years old)



What is going on?

[suspense]

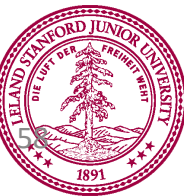
Focus on one problem

Computer Vision



Chihuahua or muffin?

Piech, CS109, Stanford University



Can you do it?

Chihuahua or Muffin?



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



```
class DigitDetector:
```

```
    def detect(raw_image):
```

```
        # Return a 0 or 1 based on the pixels of the image
```

```
        # TODO
```



Perhaps there is an insight?

Two Great Ideas

1. Artificial Neurons

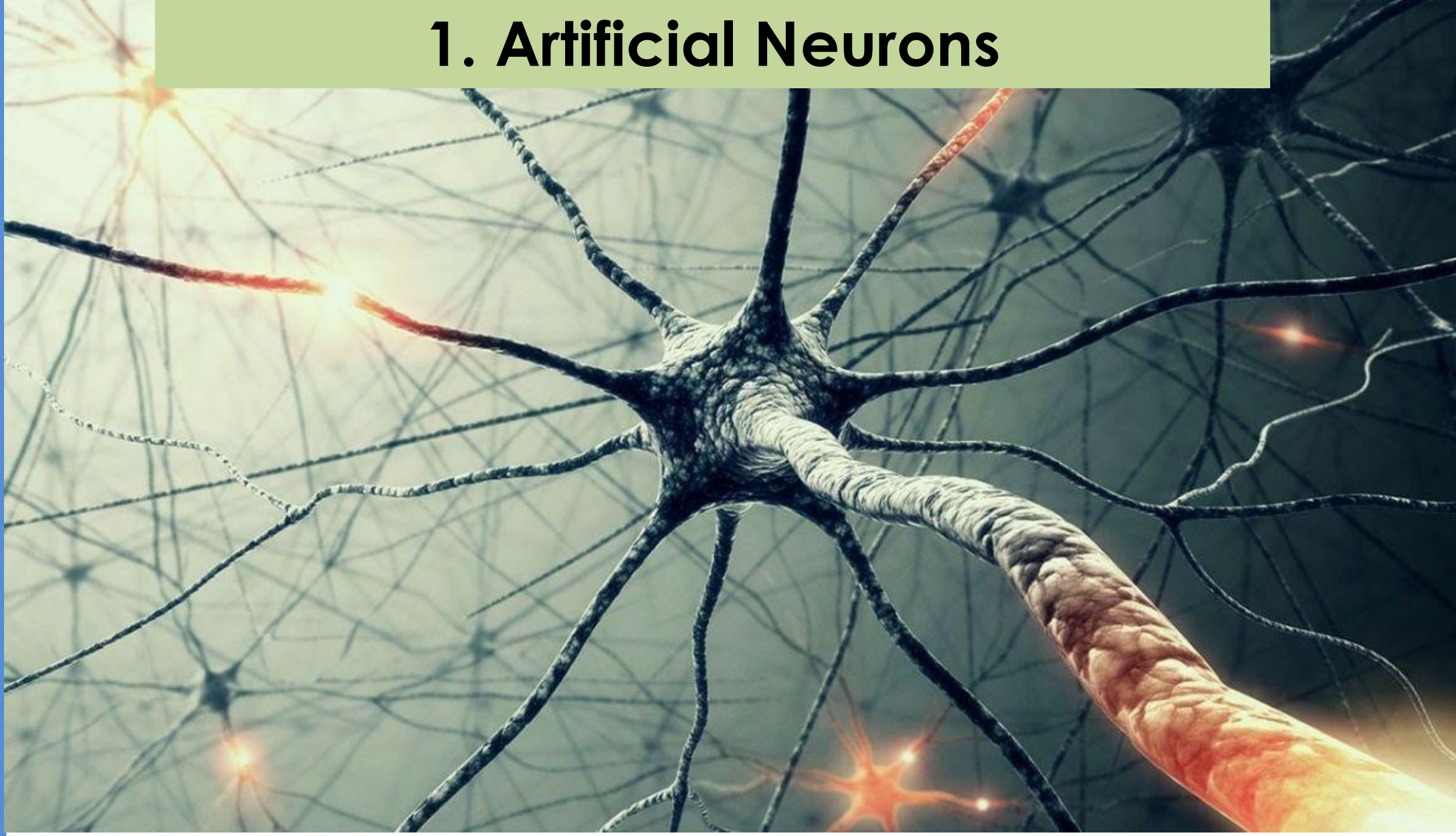
2. Learn by Example

Two Great Ideas

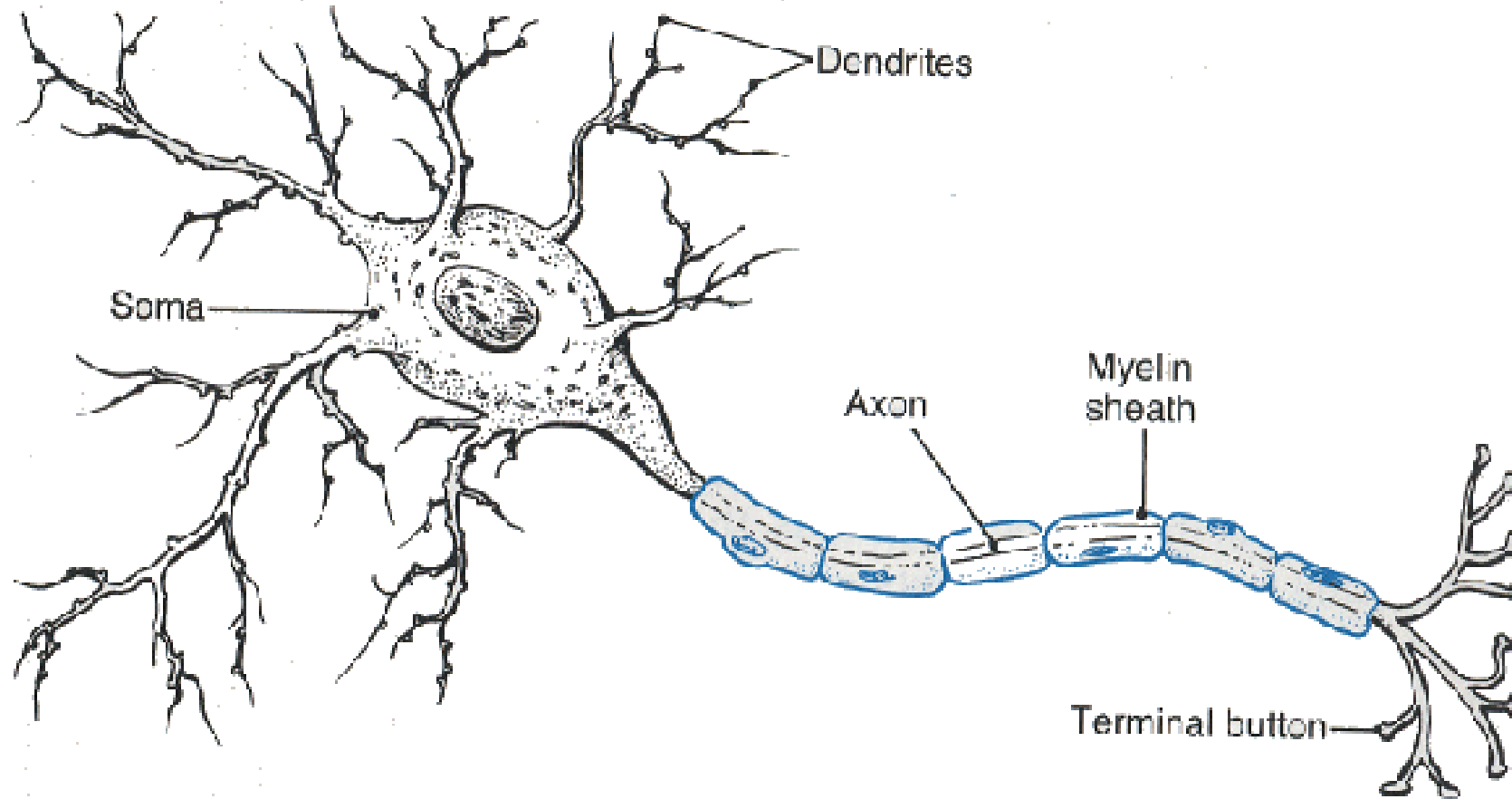
1. Artificial Neurons

2. Learn by Example

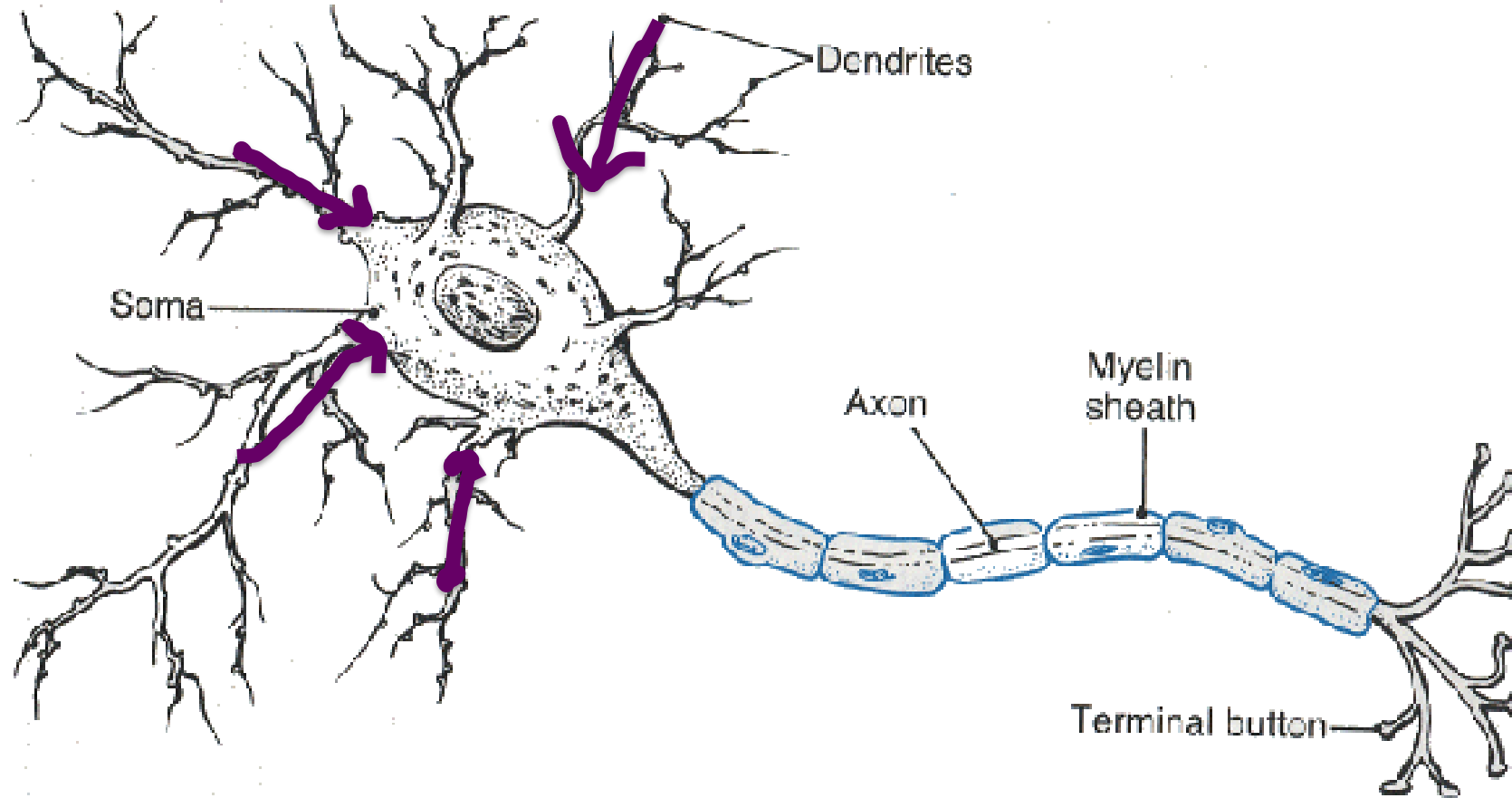
1. Artificial Neurons



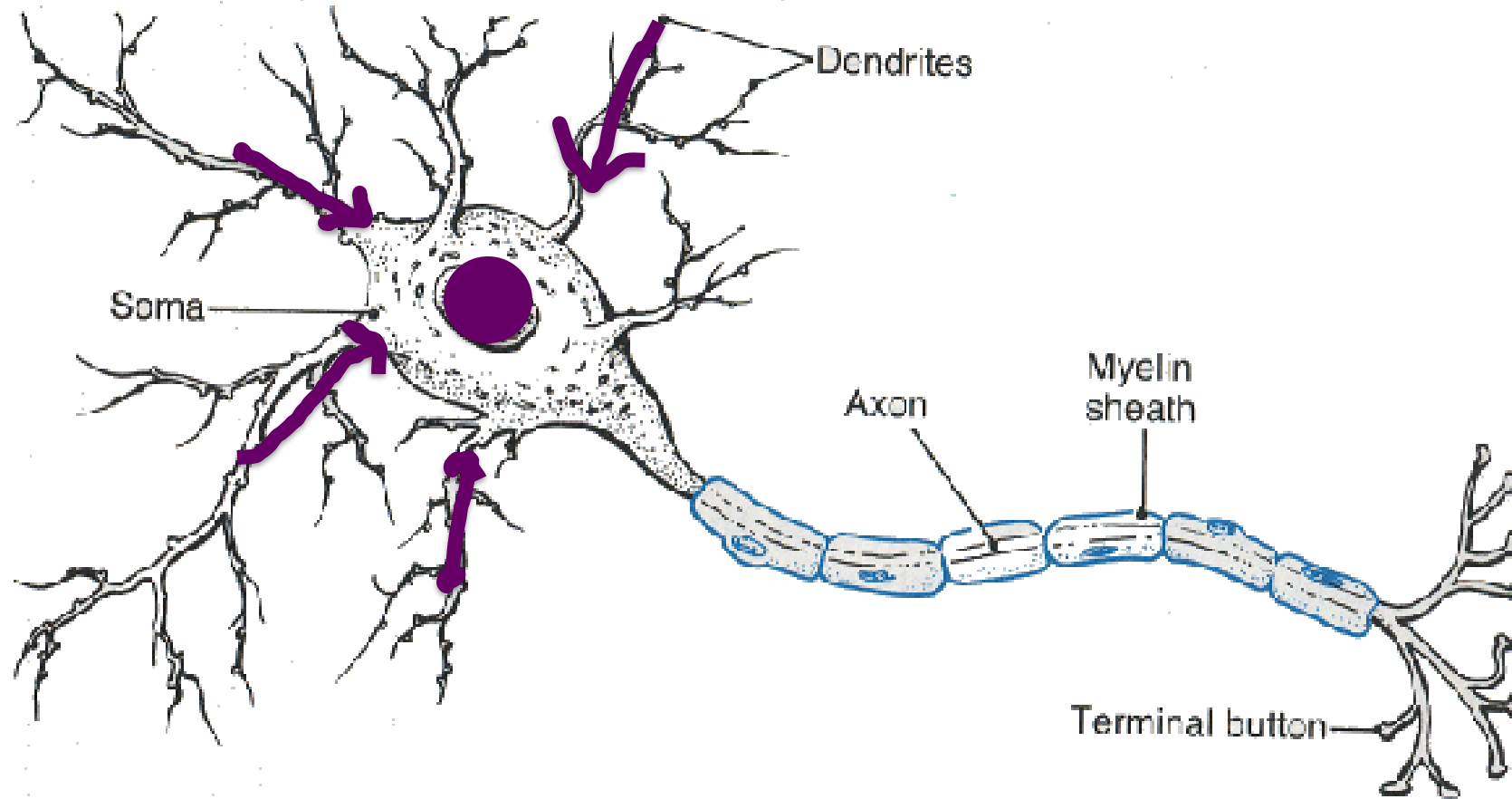
Neuron



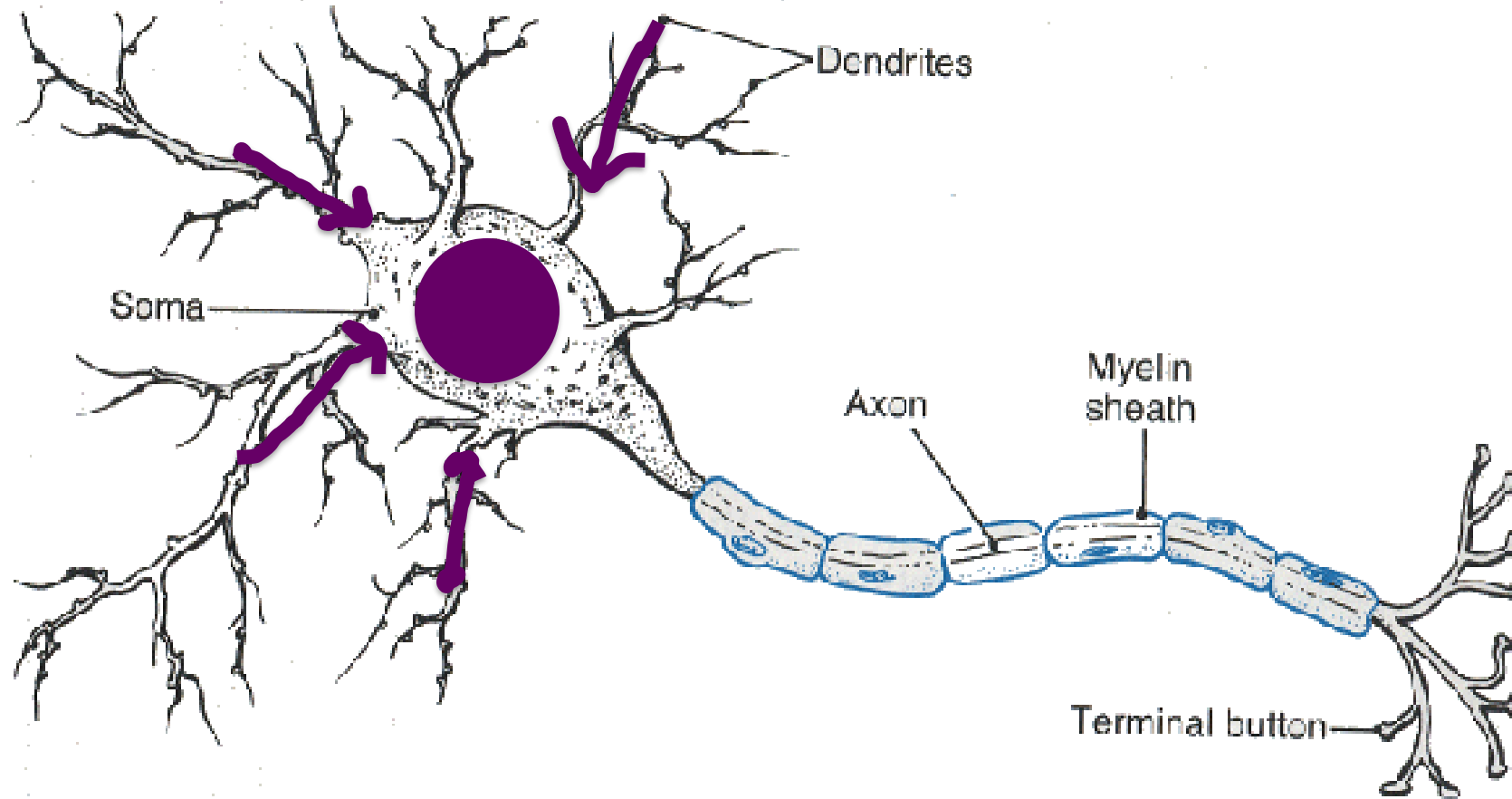
Neuron



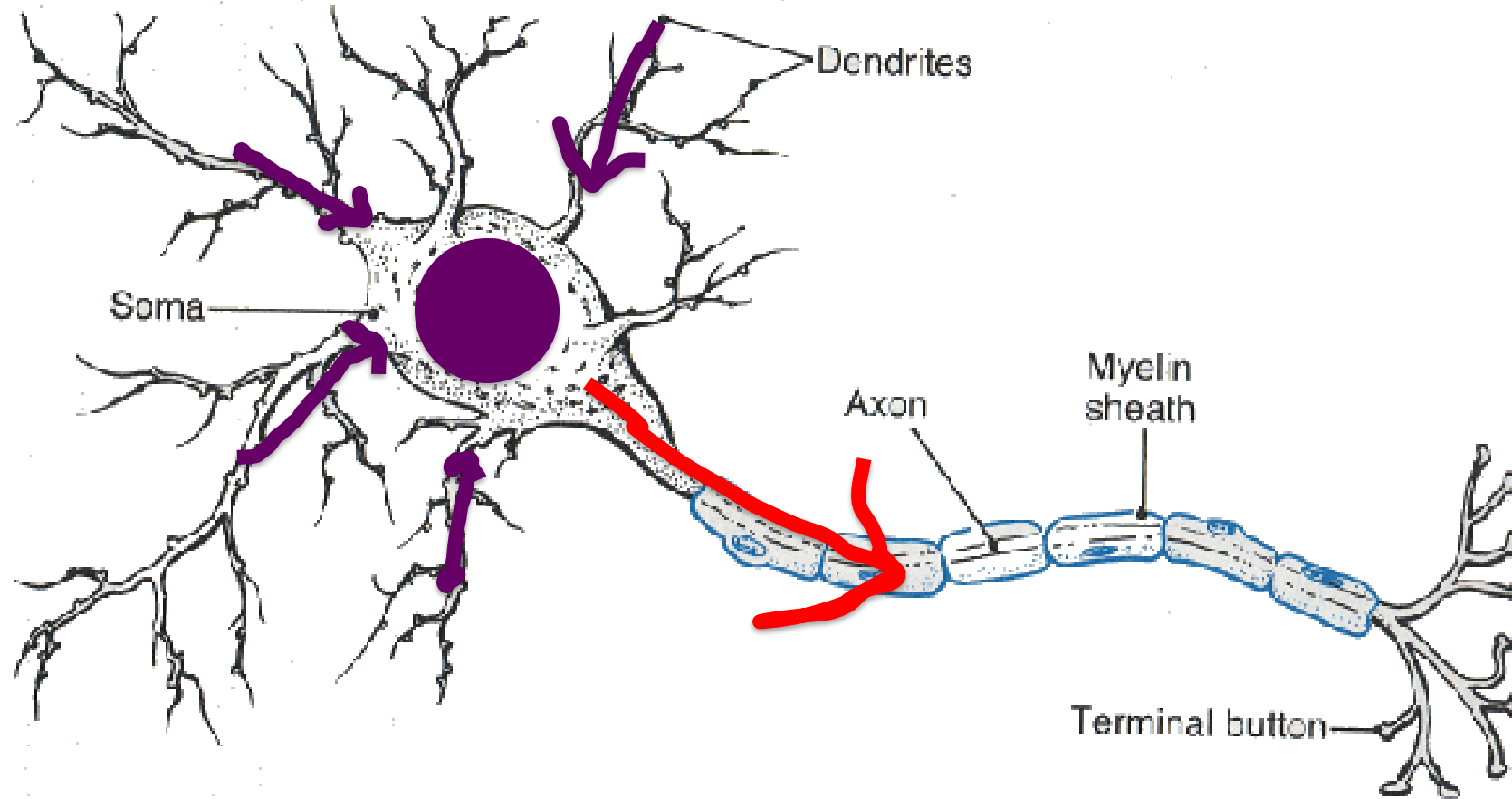
Neuron



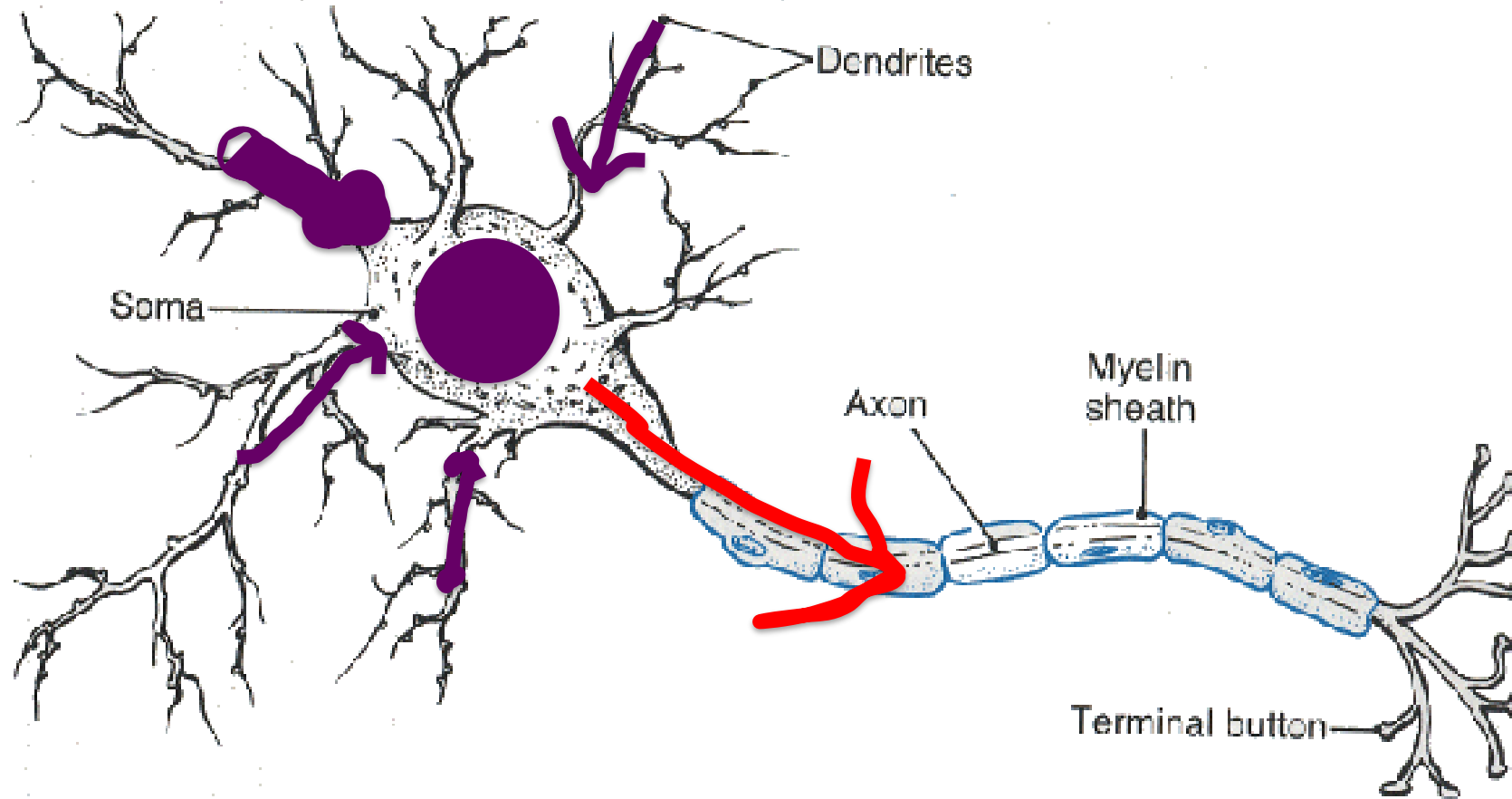
Neuron



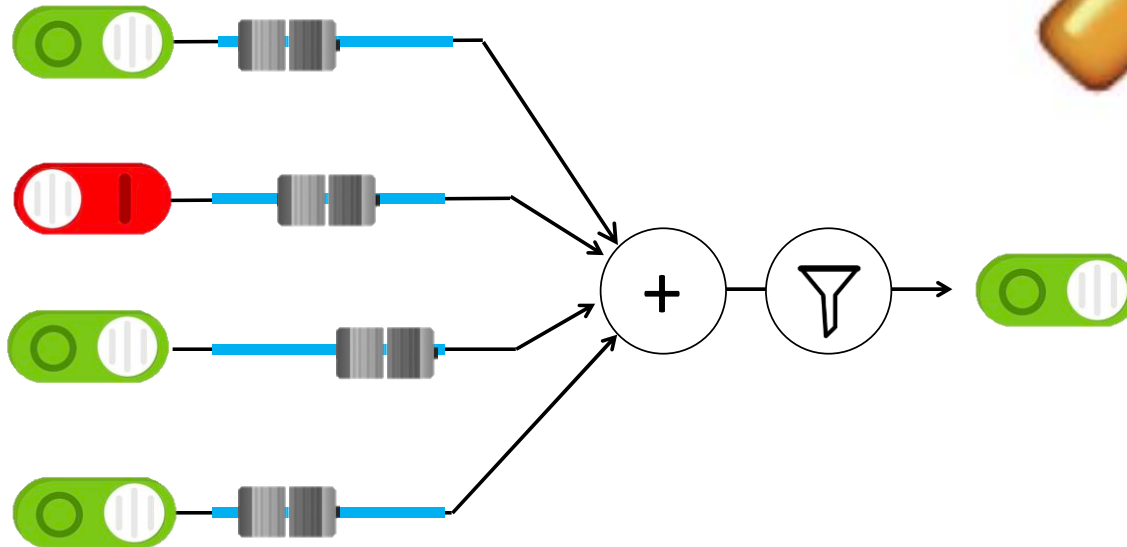
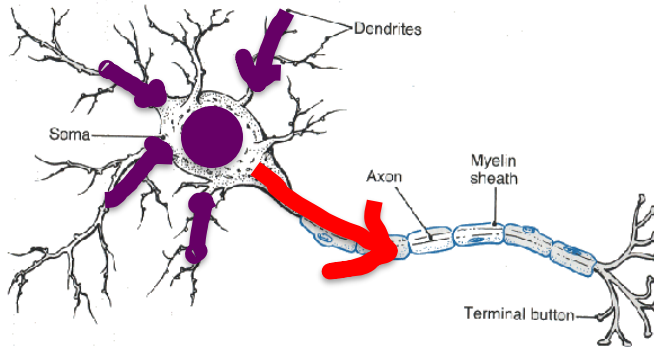
Neuron



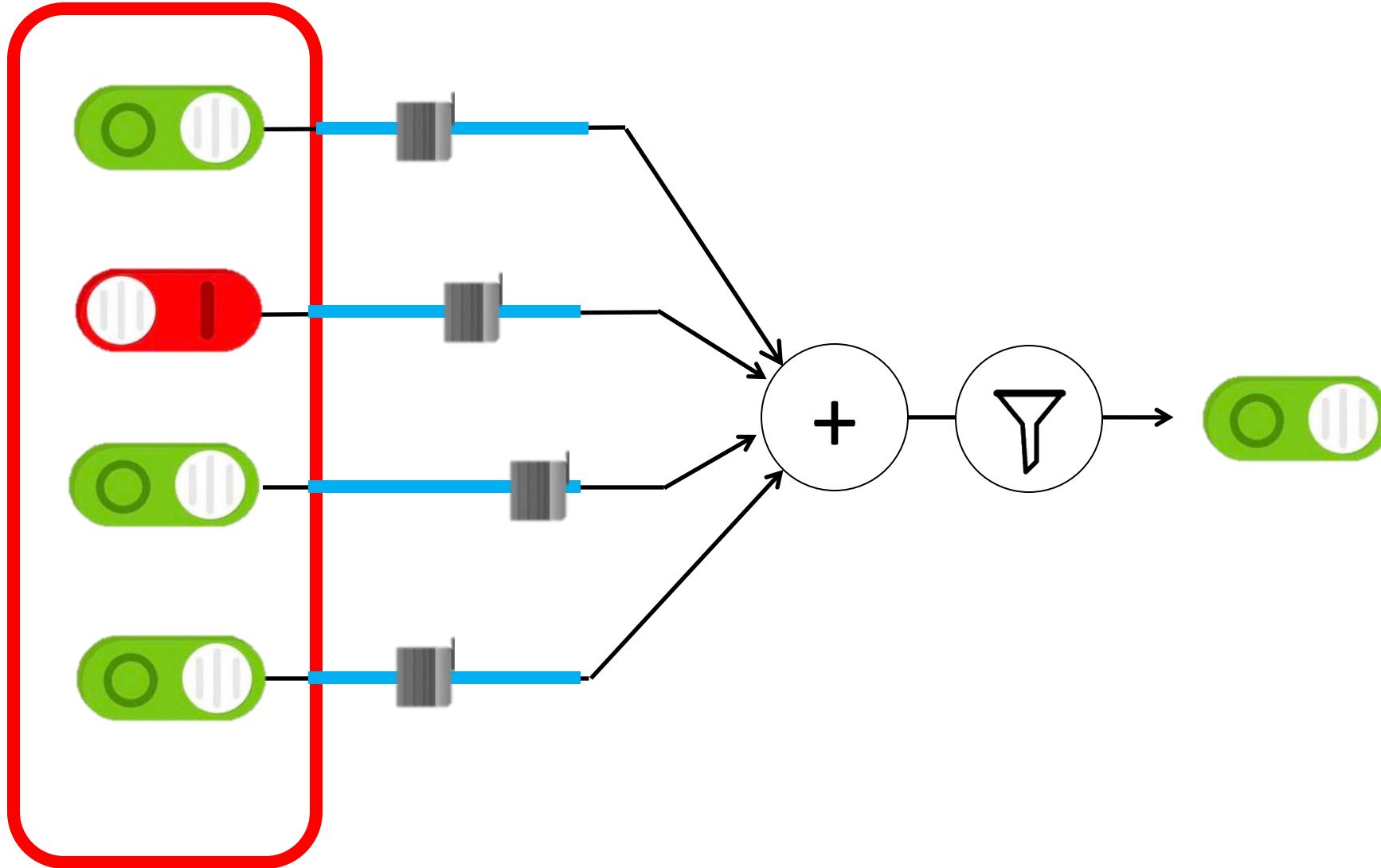
Some Inputs are More Important



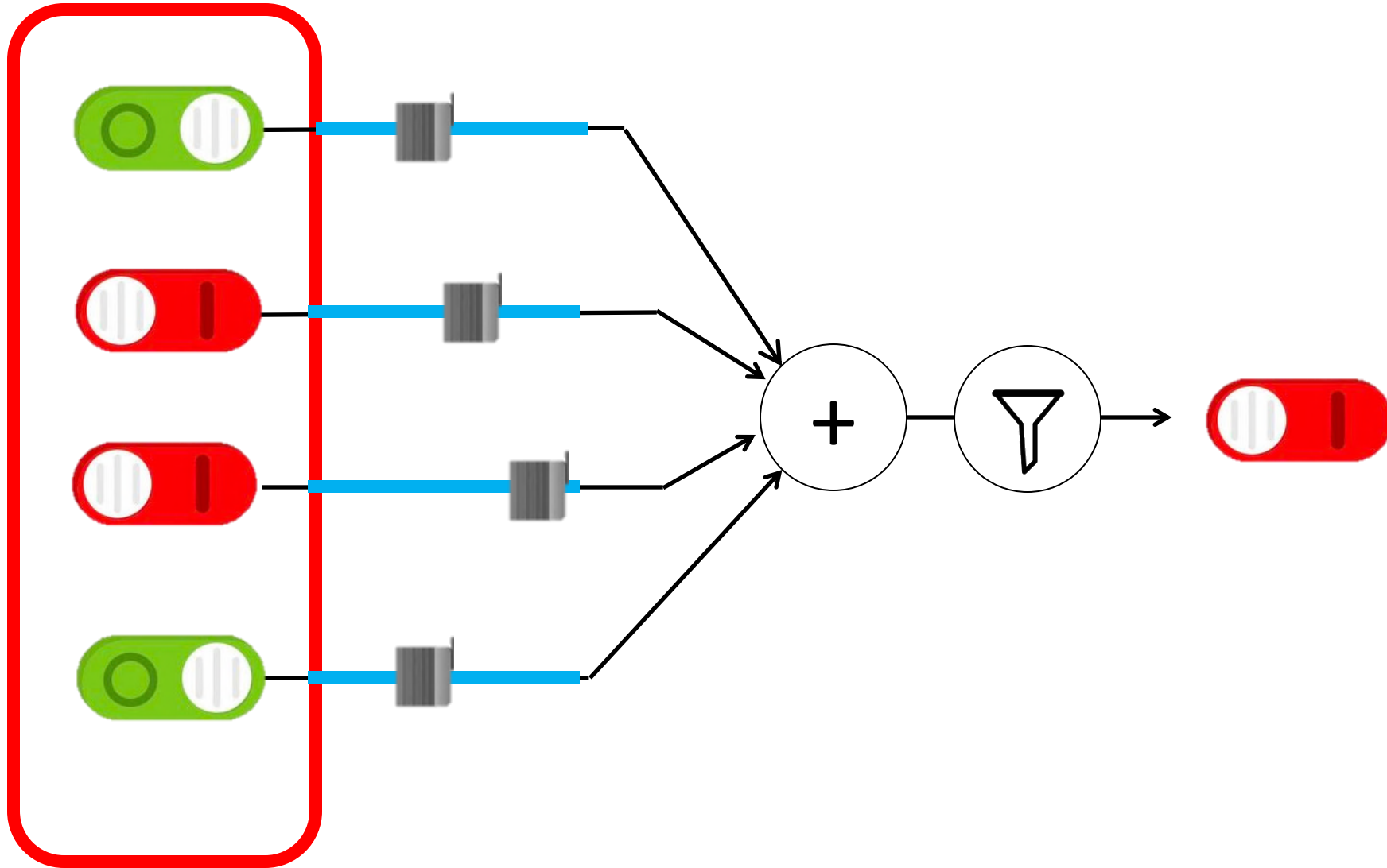
Artificial Neuron



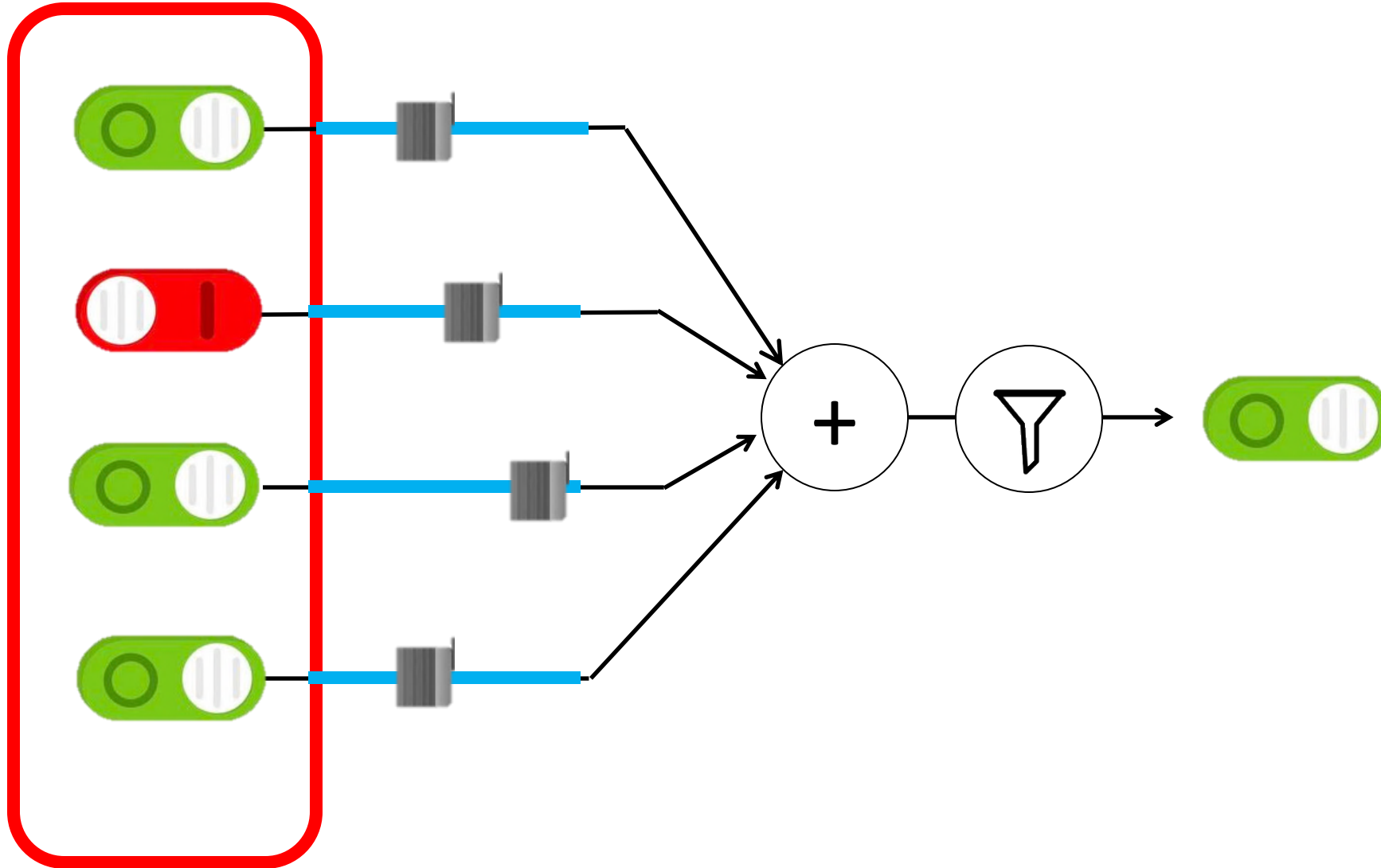
Inputs



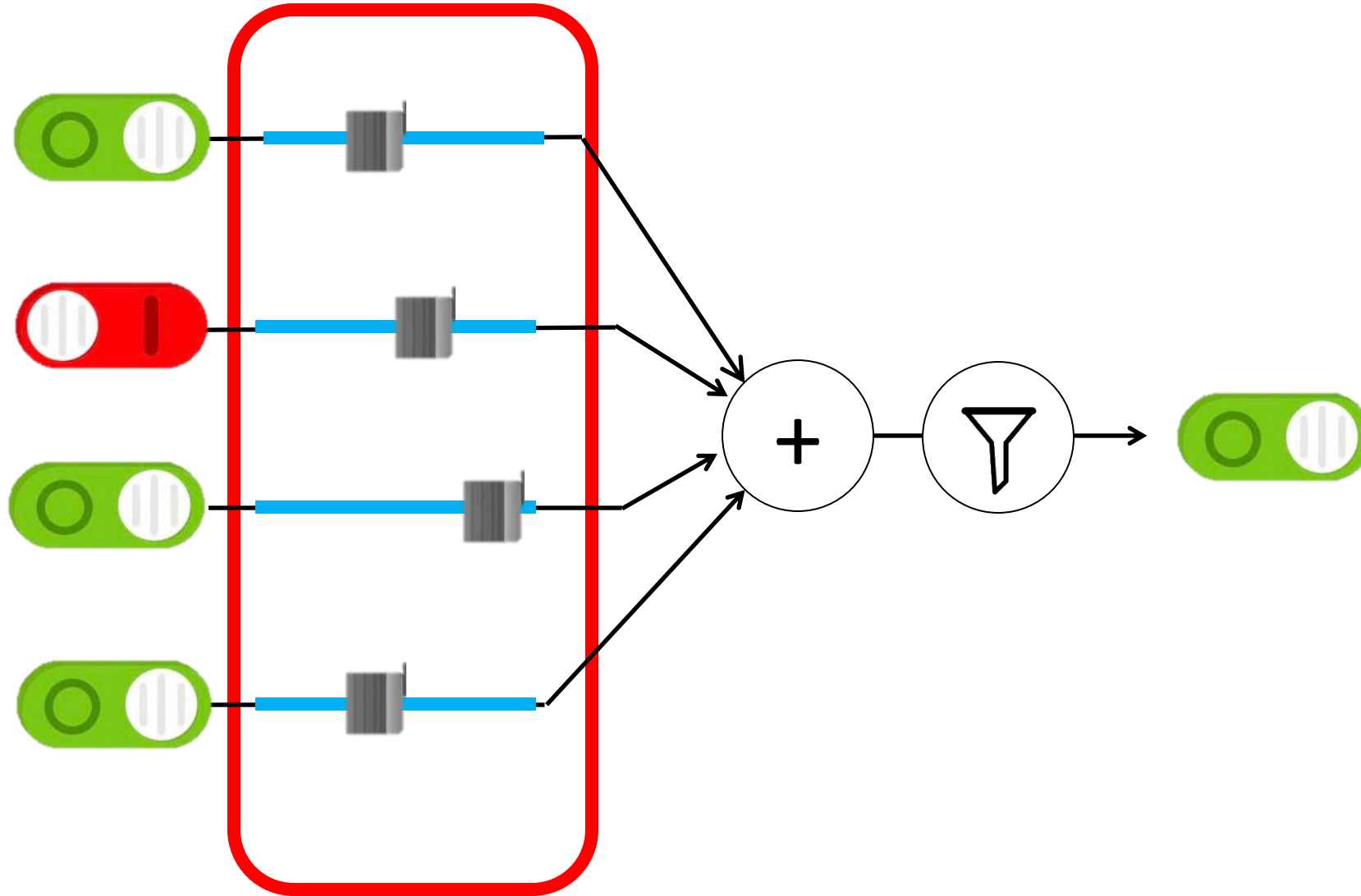
Inputs



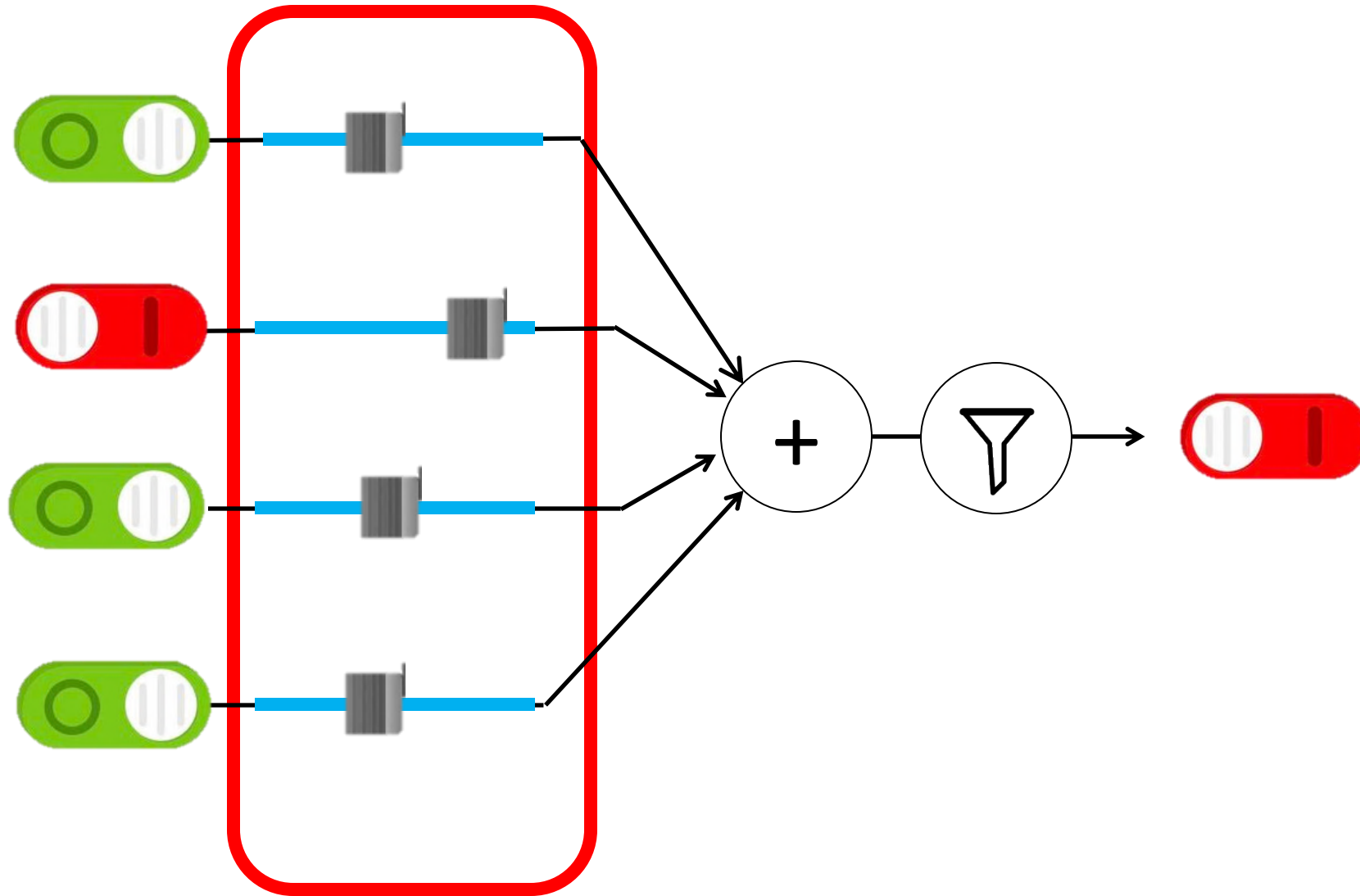
Inputs



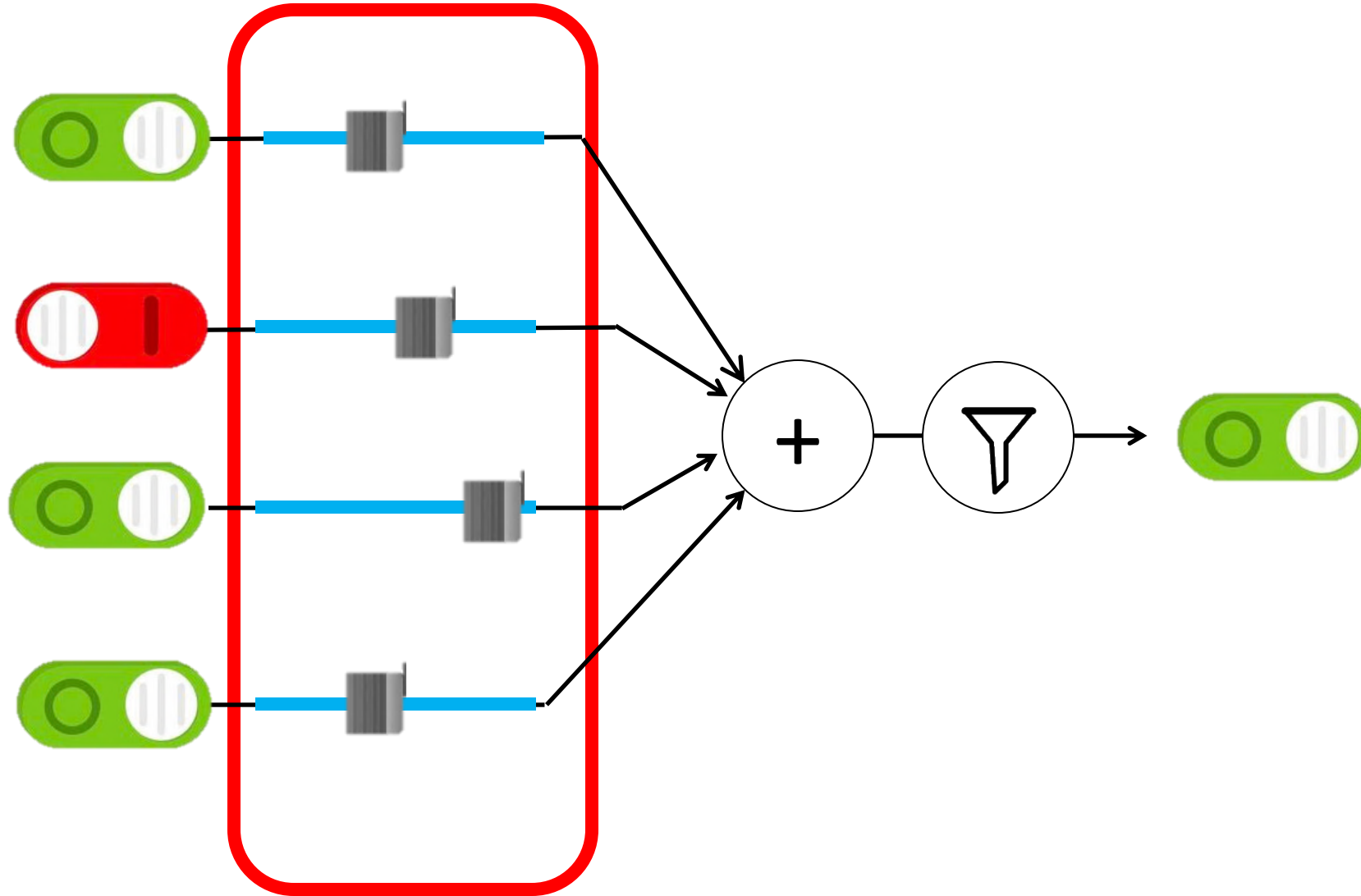
Weights



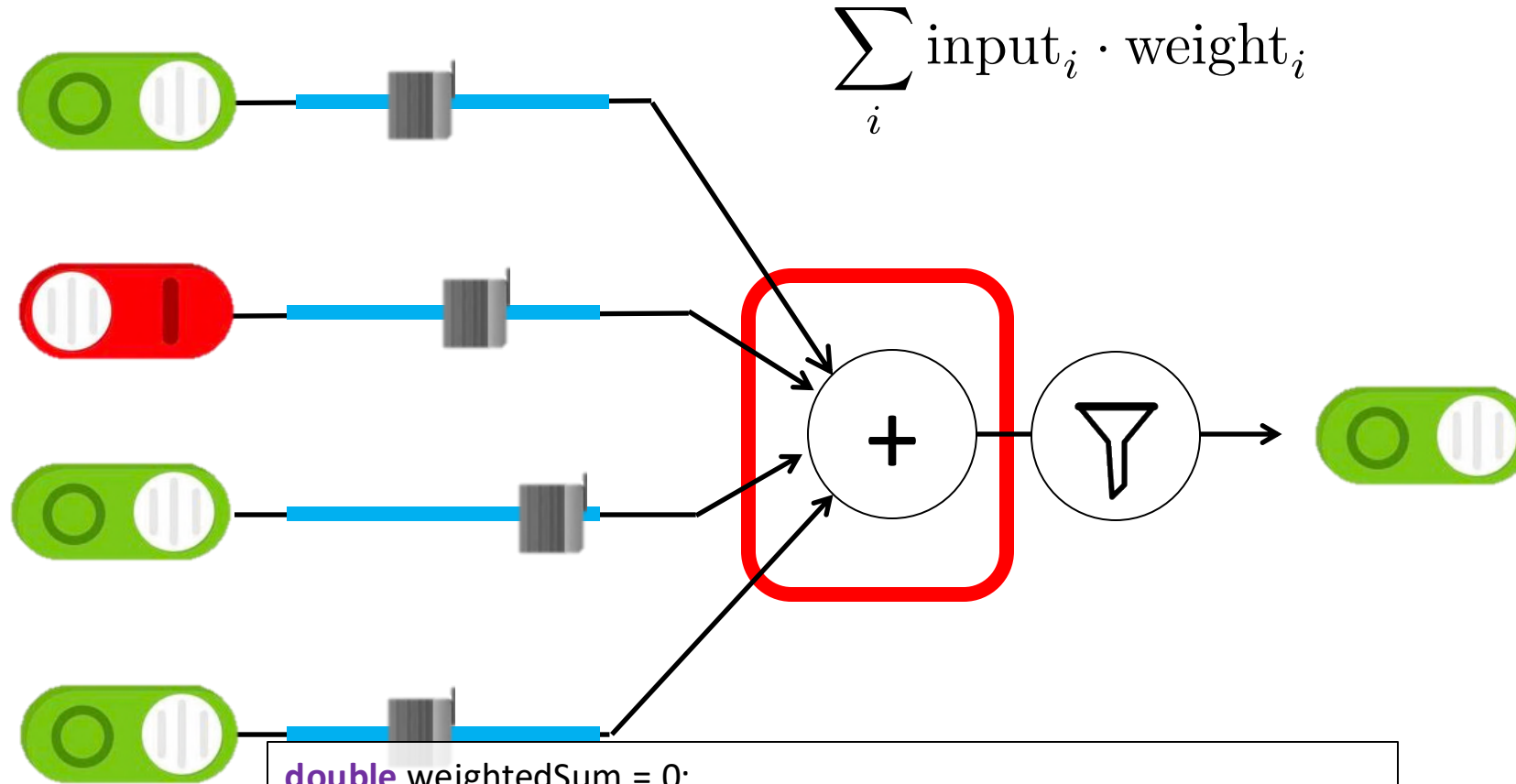
Weights



Weights



Weighted Sum

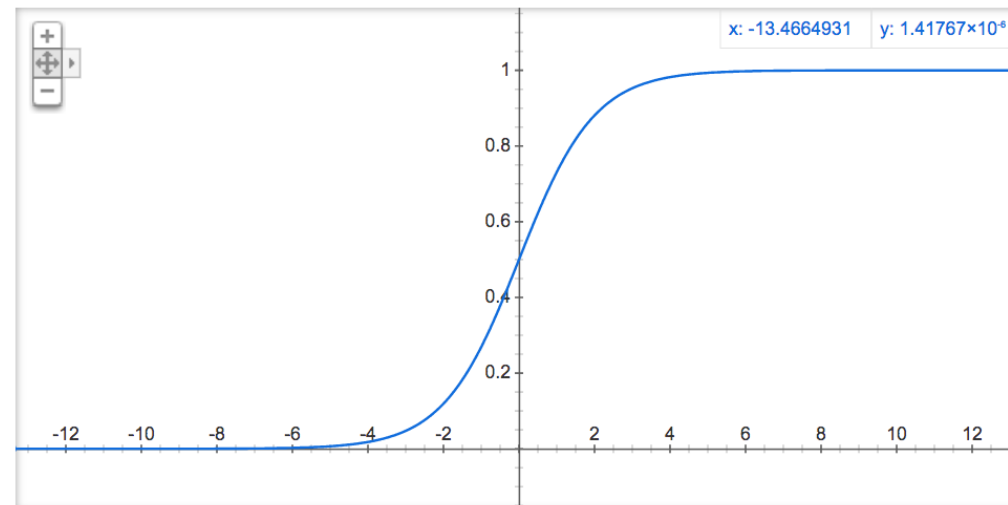
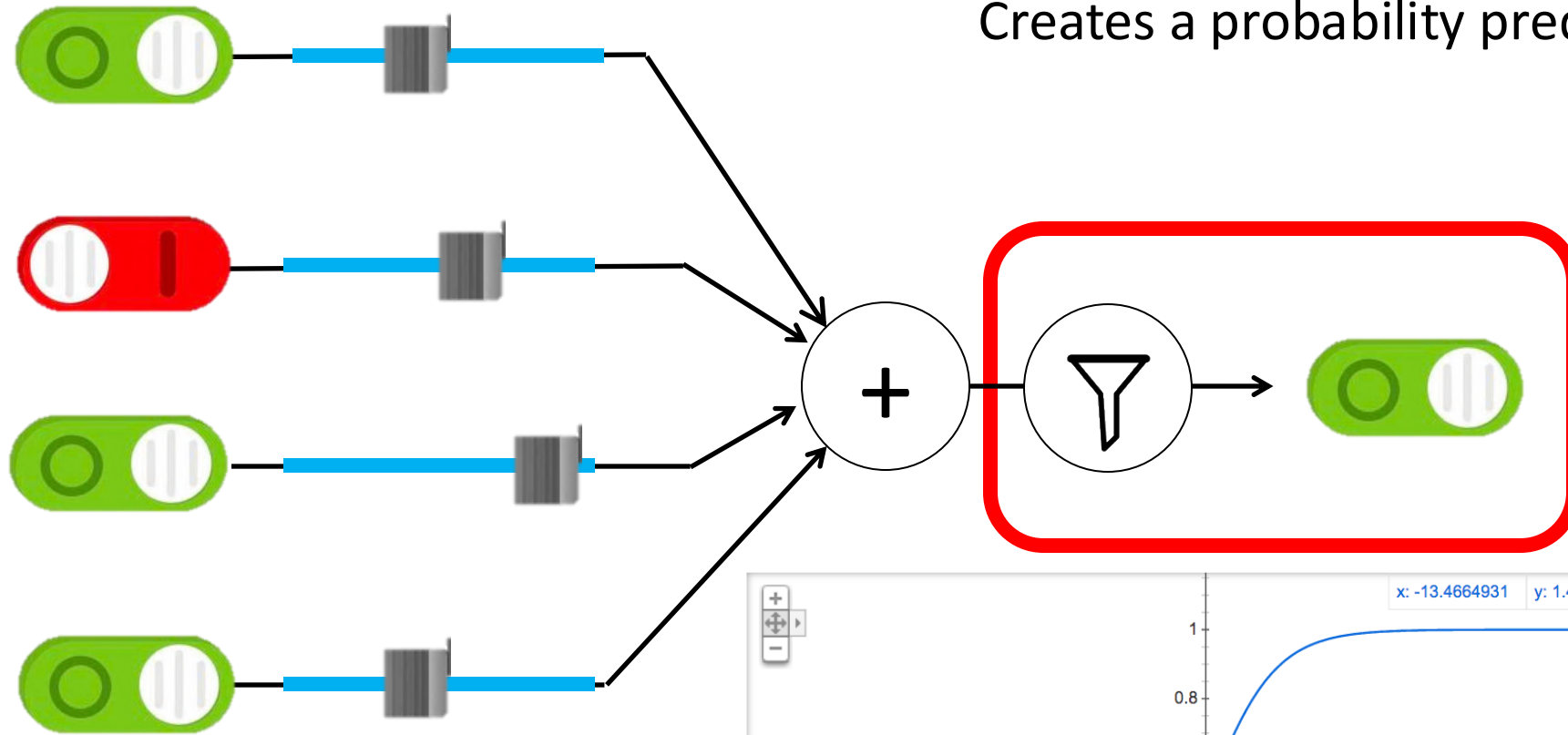


```
double weightedSum = 0;  
weightedSum += input0 * weight0;  
weightedSum += input1 * weight1;  
weightedSum += input2 * weight2;  
weightedSum += input3 * weight3;
```



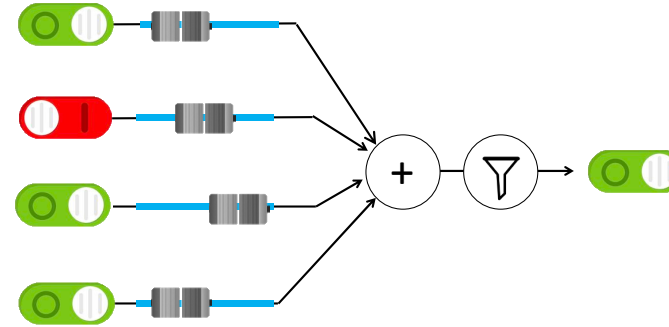
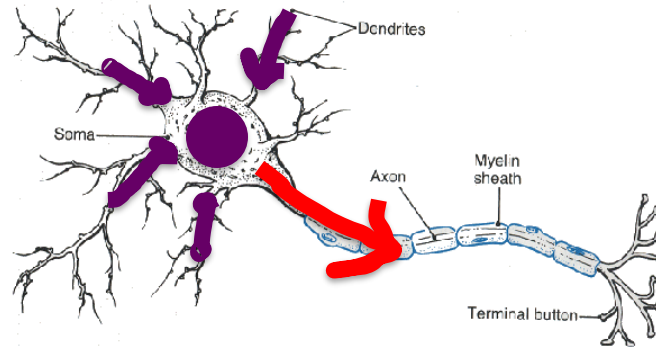
Filter and Output

Creates a probability prediction

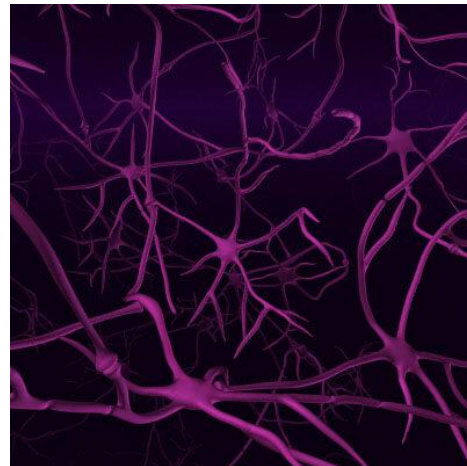


Biological Basis for Neural Networks

- A neuron



- Your brain

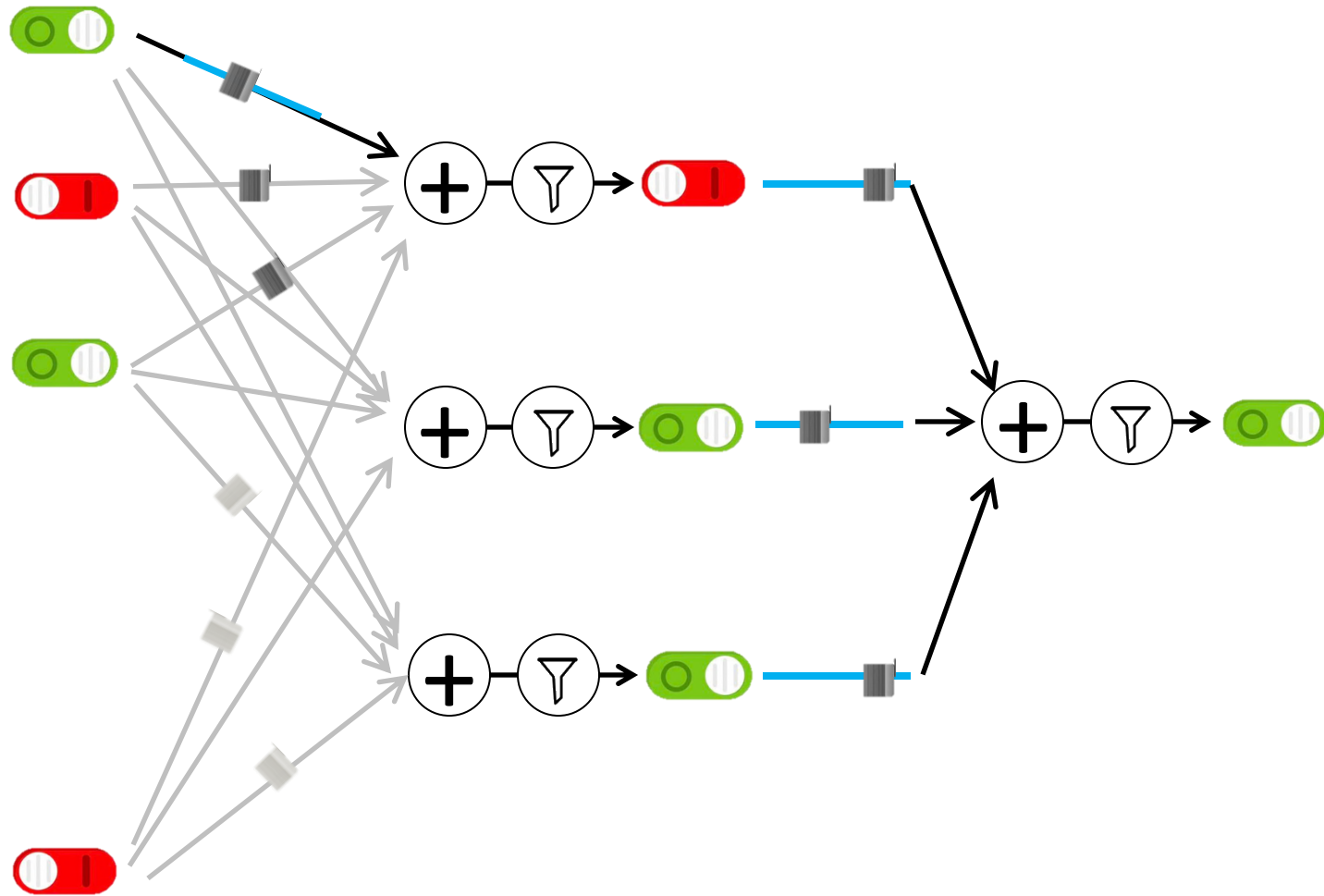


???

Actually, it's probably someone else's brain

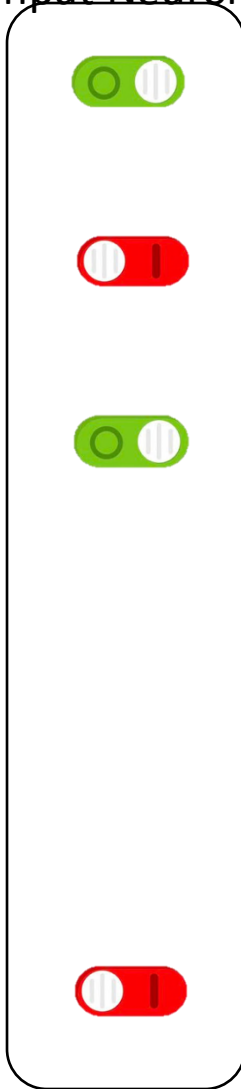


Put Many Together

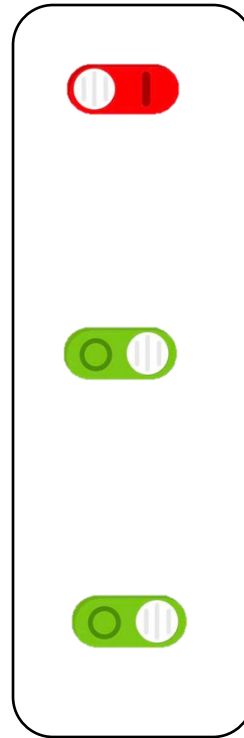


Put Many Together

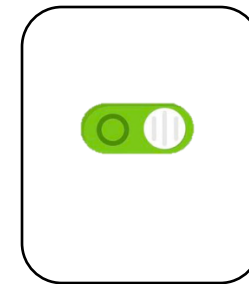
Input Neurons



Hidden Neurons



Output Neurons

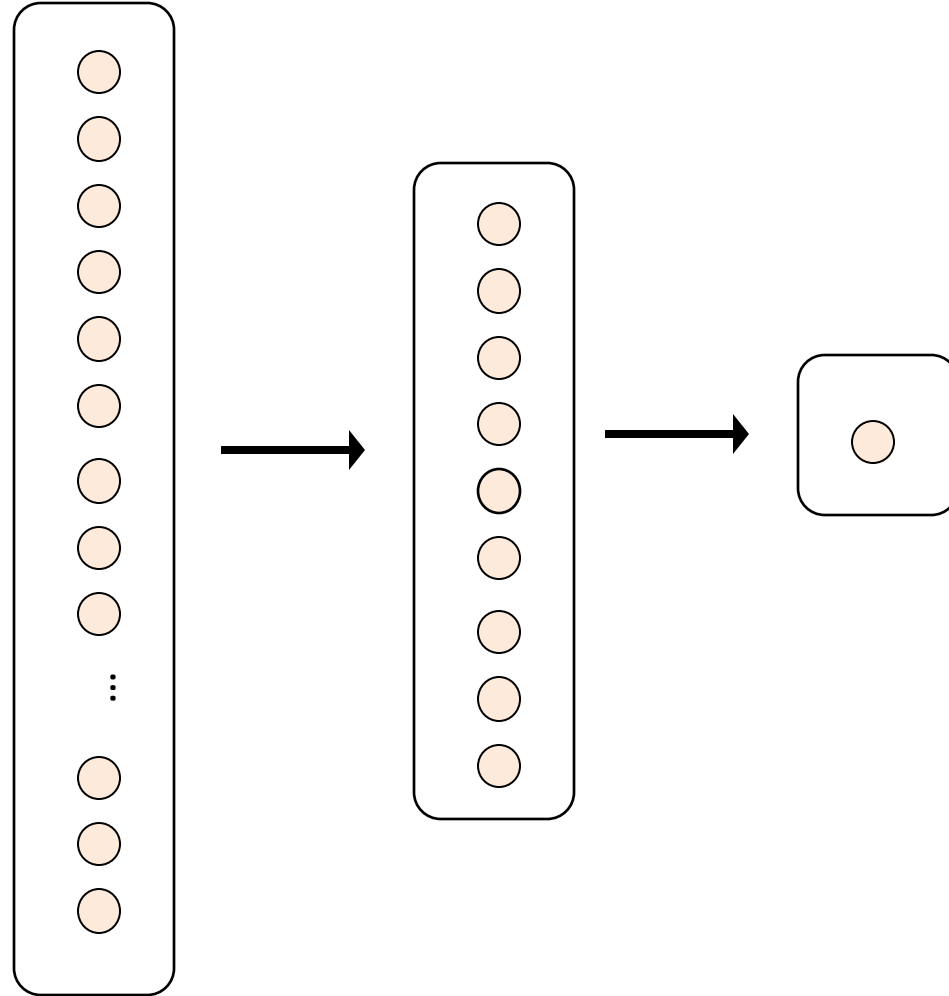


Making a Prediction

Input Neurons

Hidden Neurons

Output Neurons

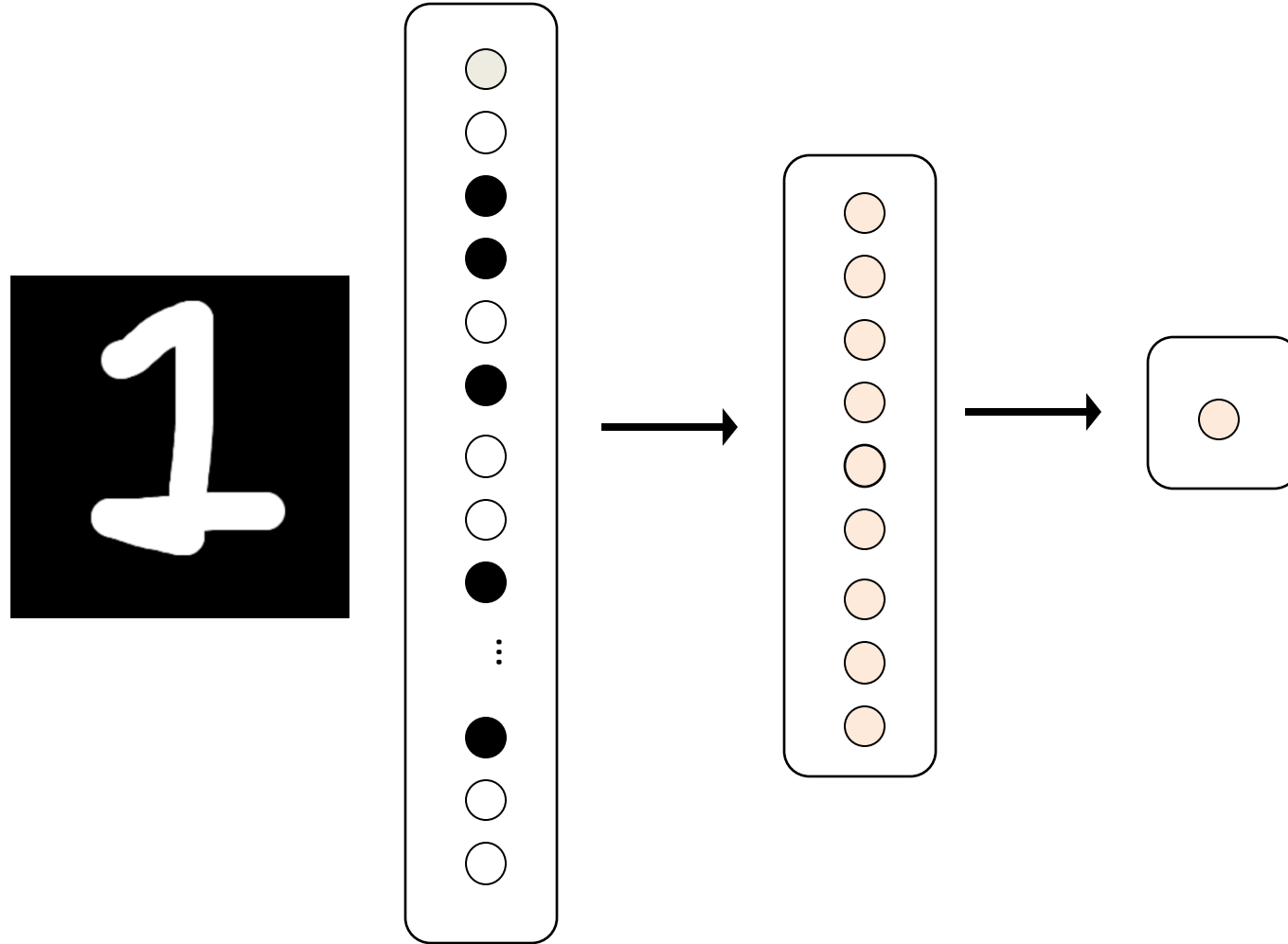


Making a Prediction

Input Neurons

Hidden Neurons

Output Neurons

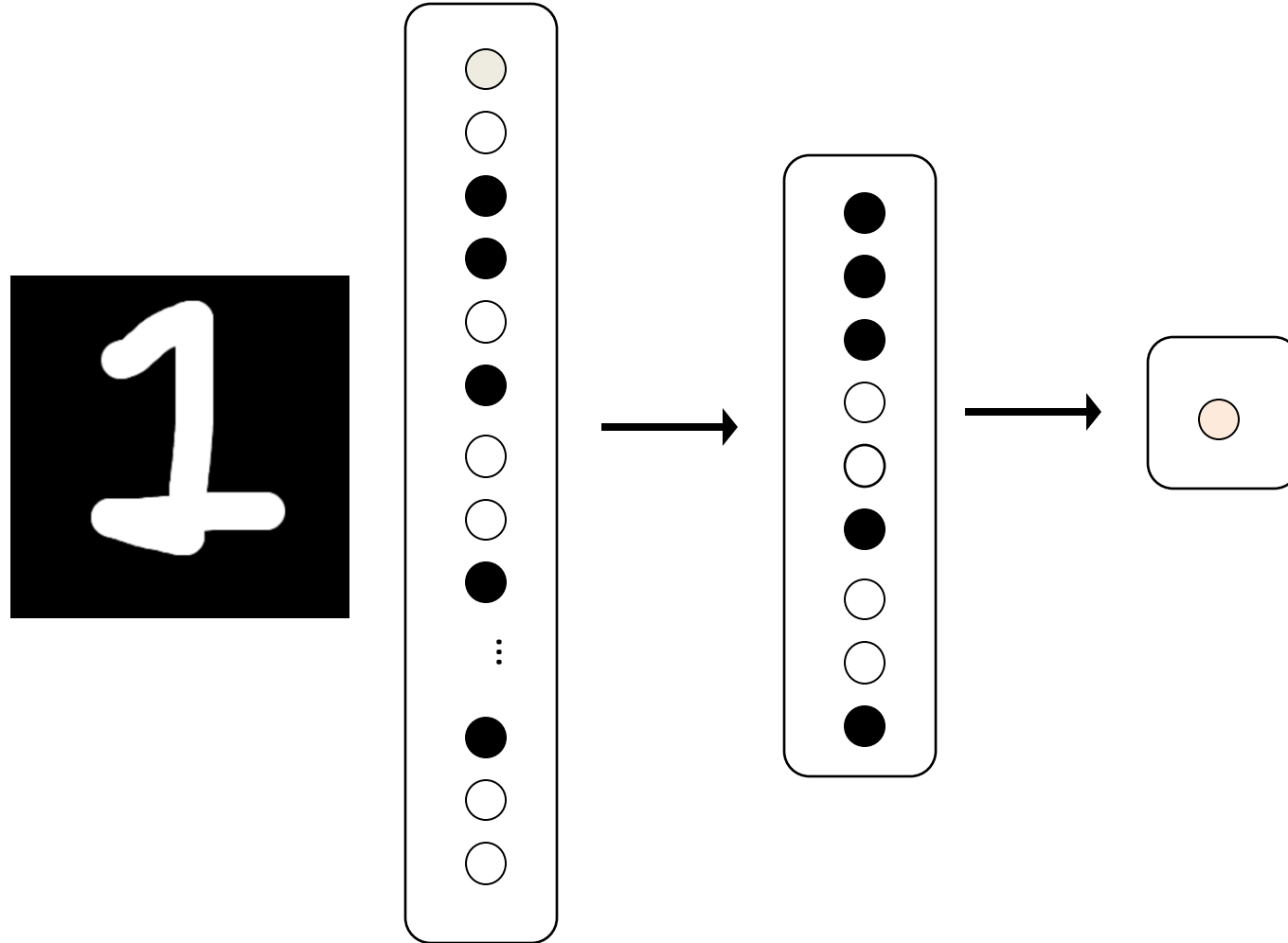


Making a Prediction

Input Neurons

Hidden Neurons

Output Neurons

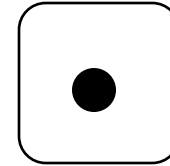
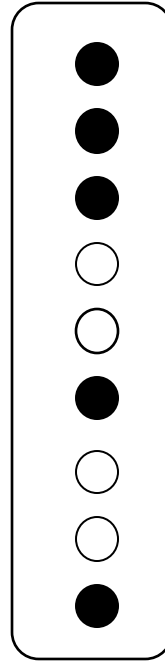
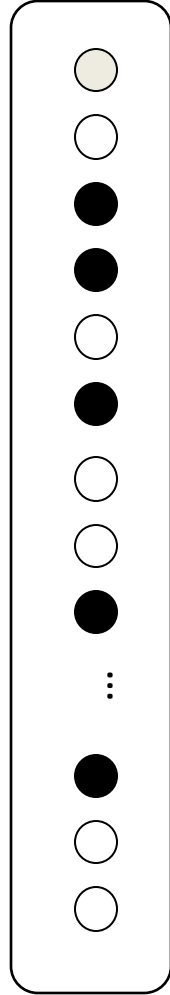


Making a Prediction

Input Neurons

Hidden Neurons

Output Neurons

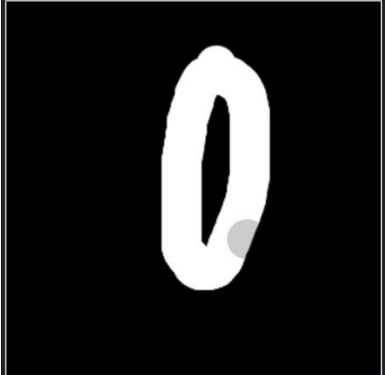




I think that is a
picture of a
one!



Demonstration

Draw your number here



X  

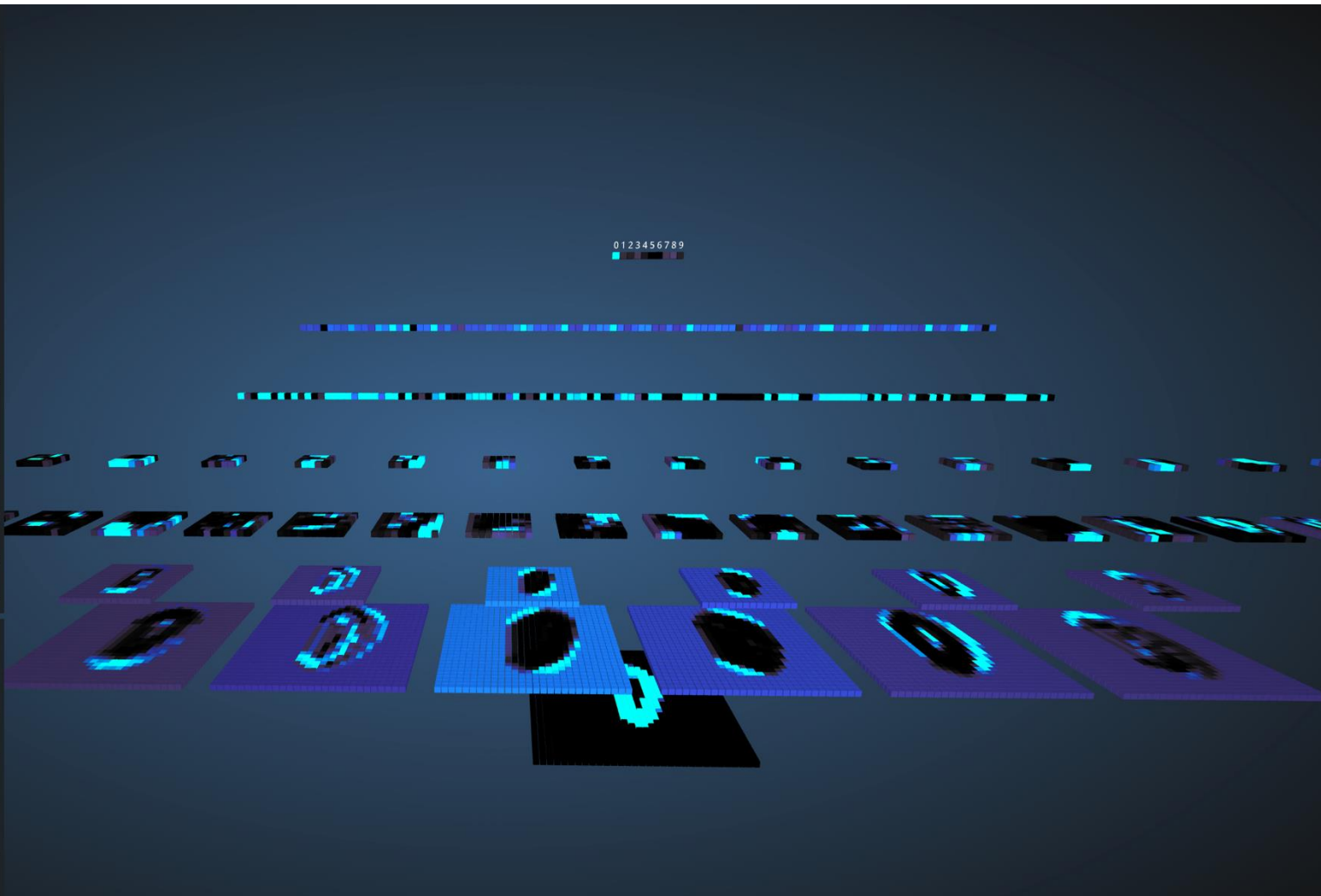
Downsampled drawing:

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




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




Interpret the Output as Prediction

Draw your number here



0 1 2 3 4 5 6 7 8 9



X [Pencil] [Eraser]

Downsampled drawing:

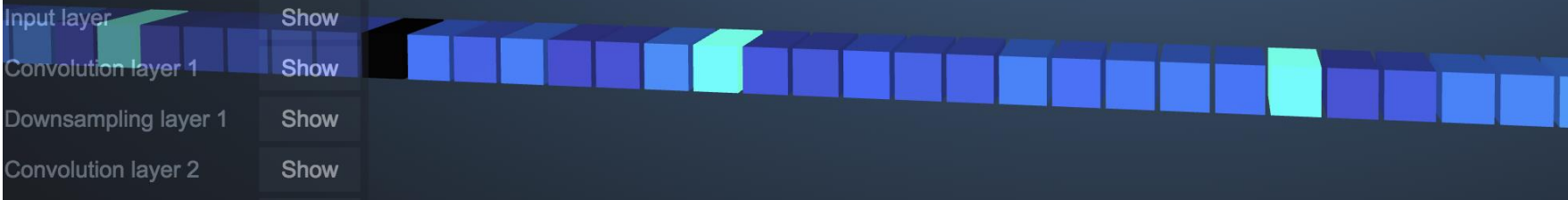
First guess: 3

Second guess: 3

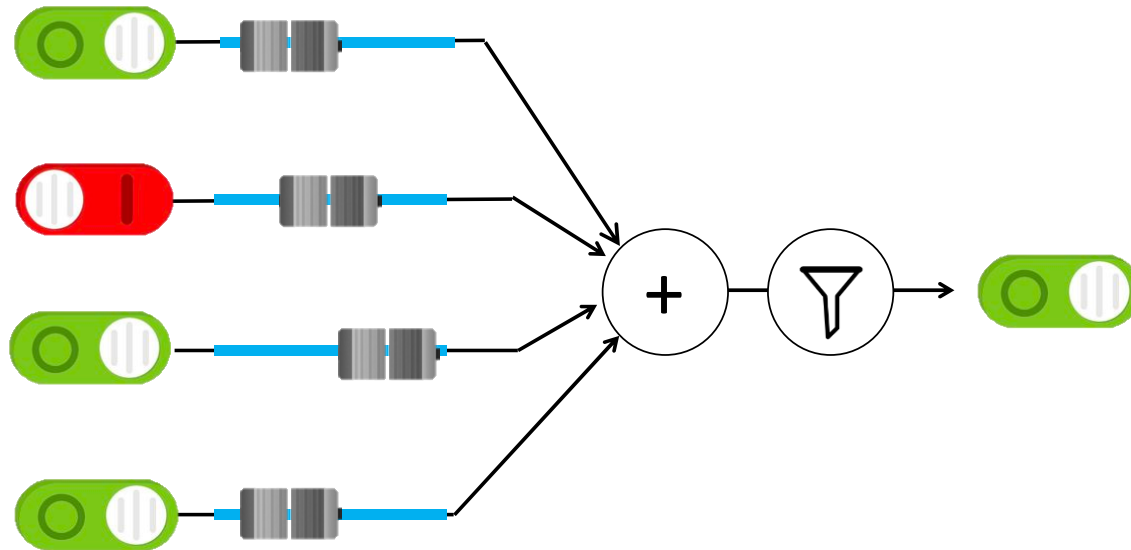
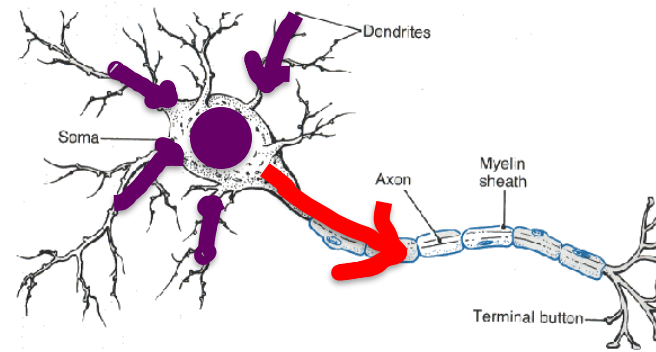
8

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

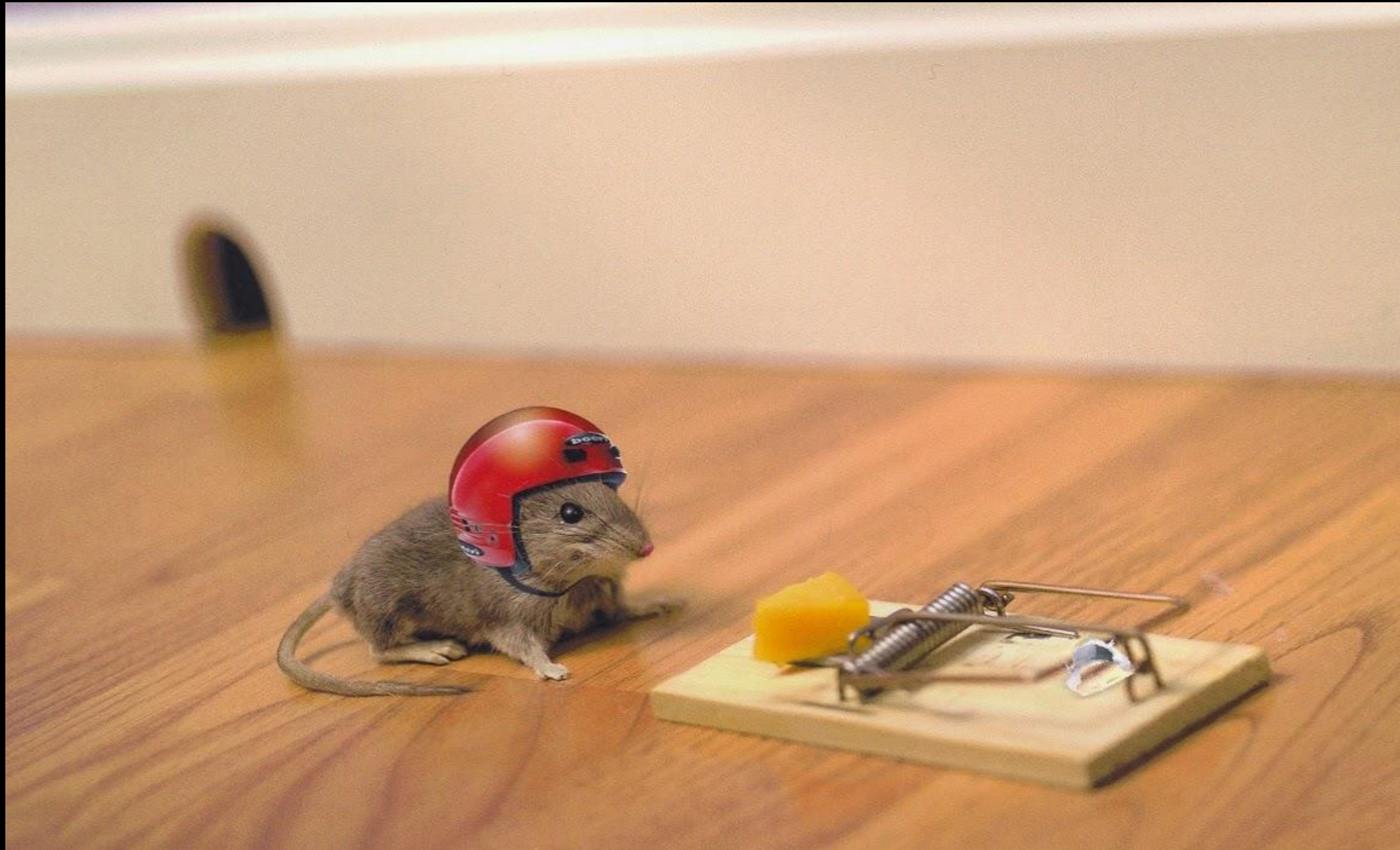
2. Learn by Example

Two Great Ideas

1. Artificial Neurons

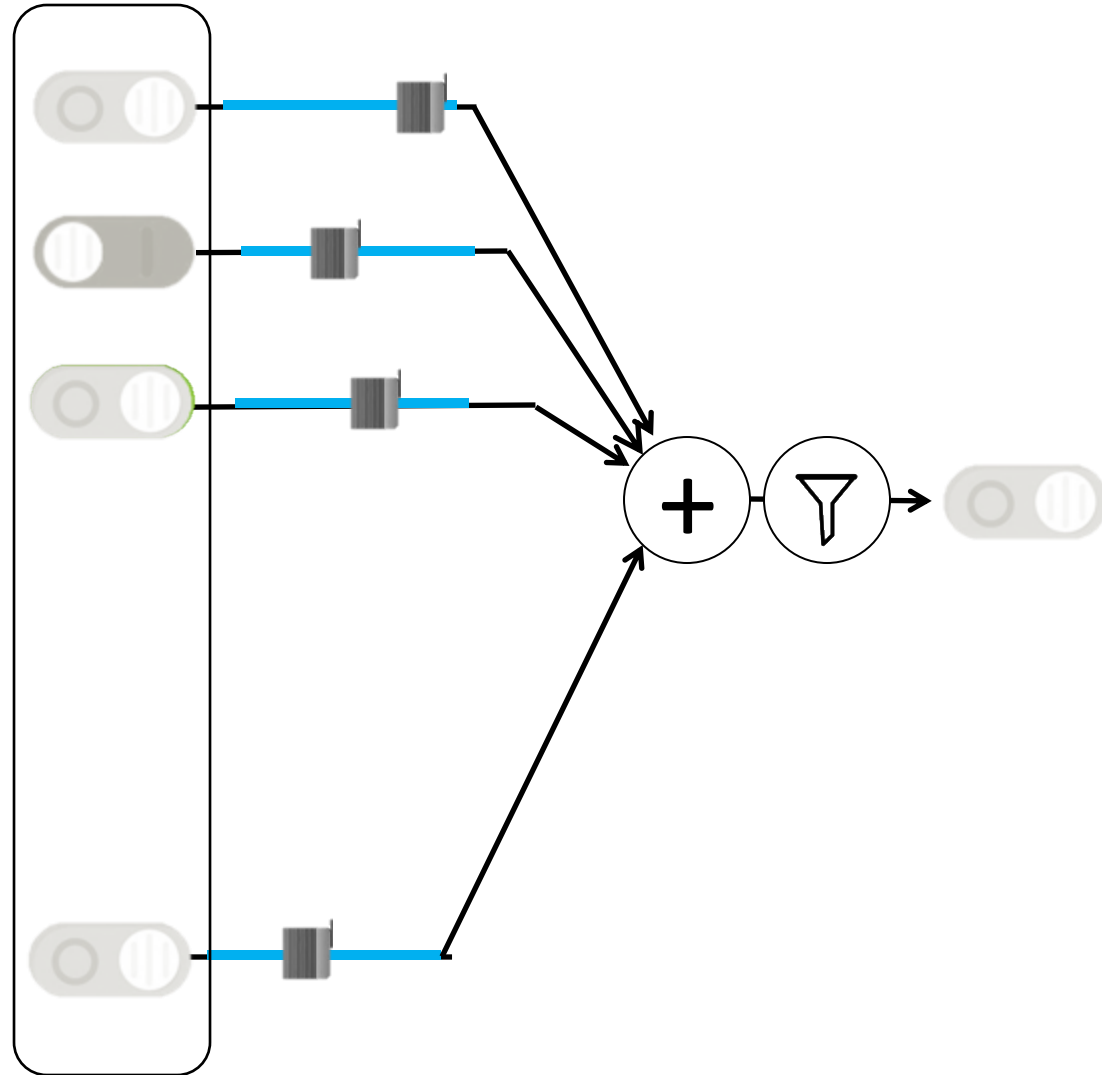
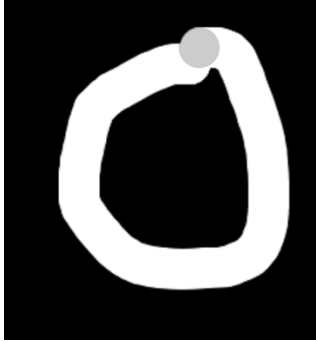
2. Learn by Example

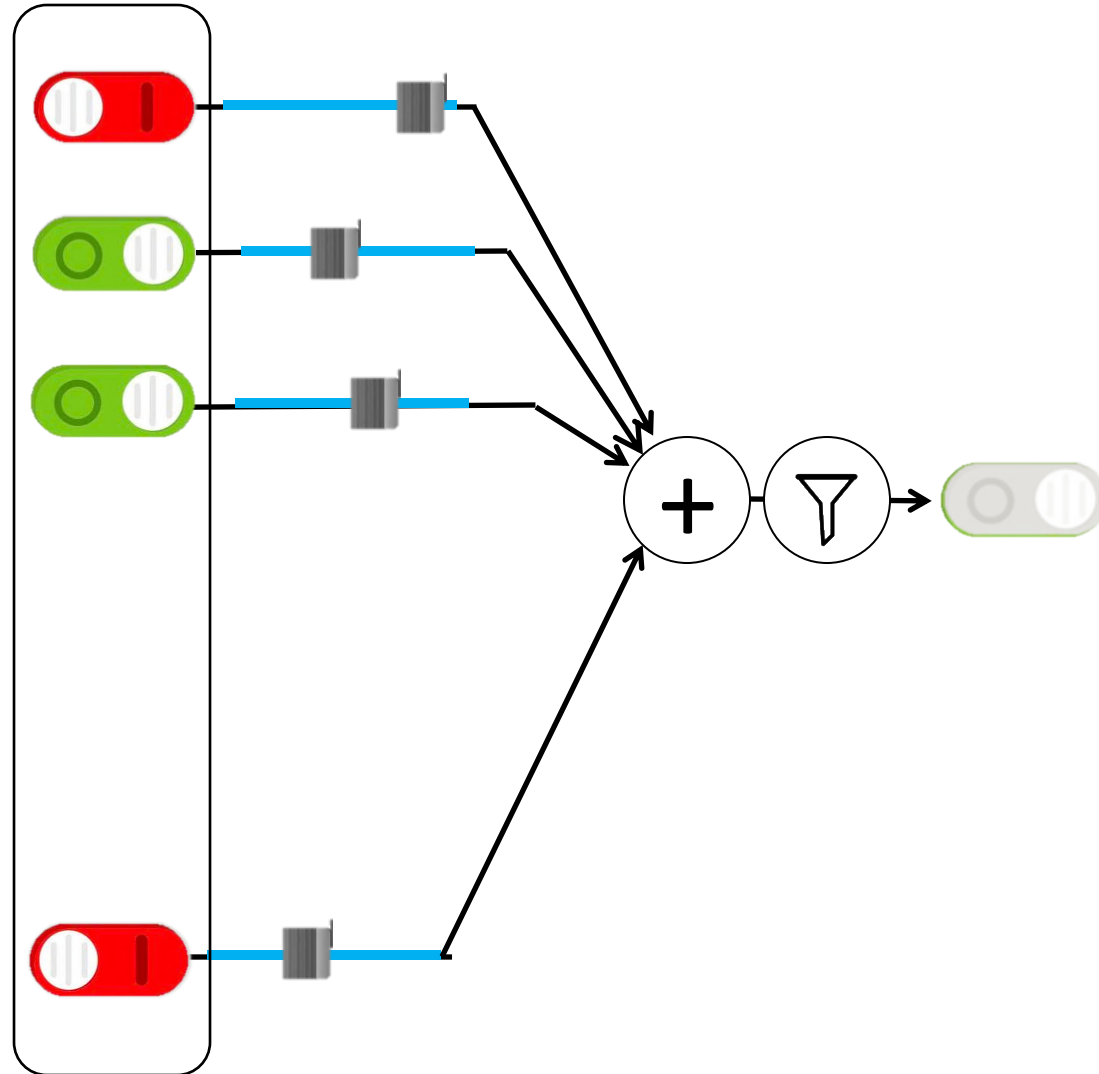
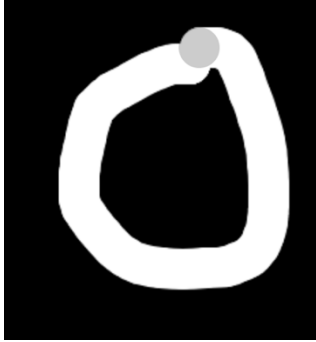
2. Learn From Experience

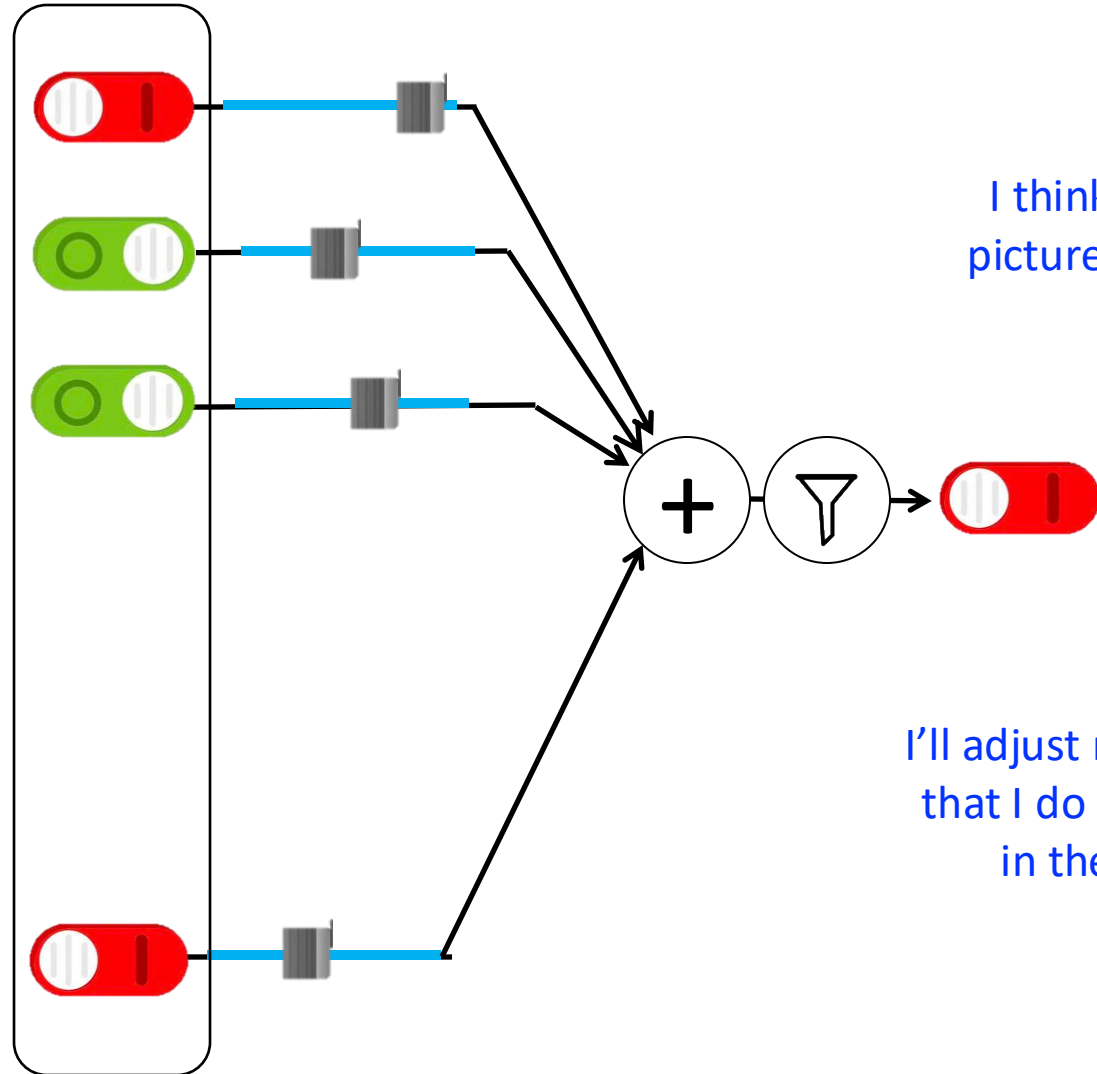
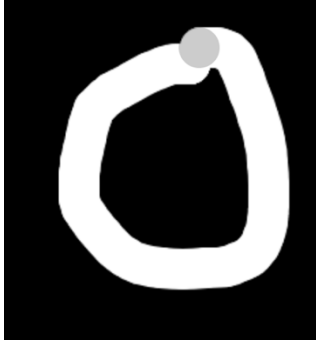


Learn by Example







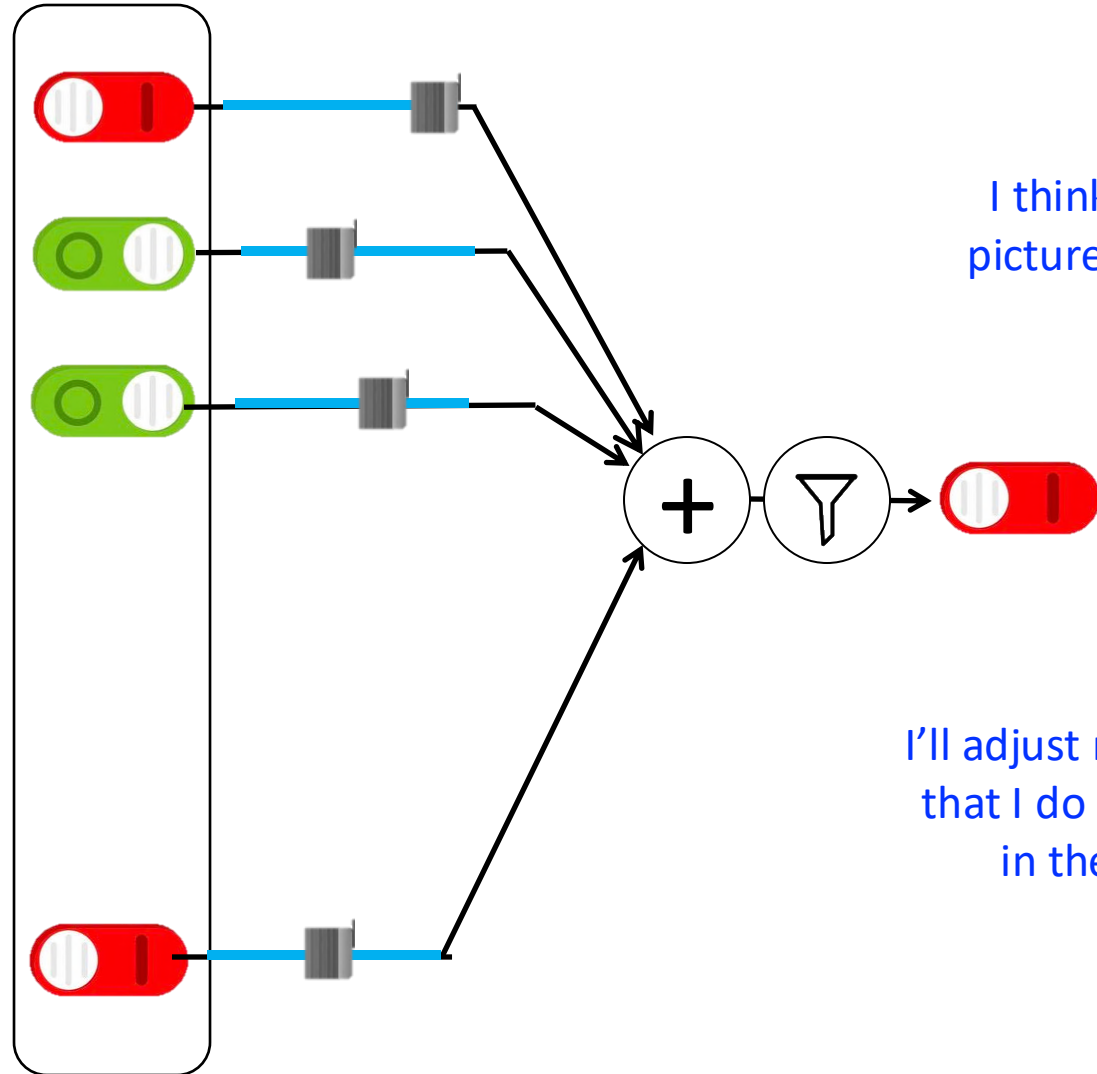
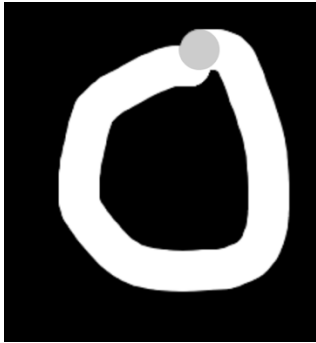


I think that is a picture of a **One!**

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

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



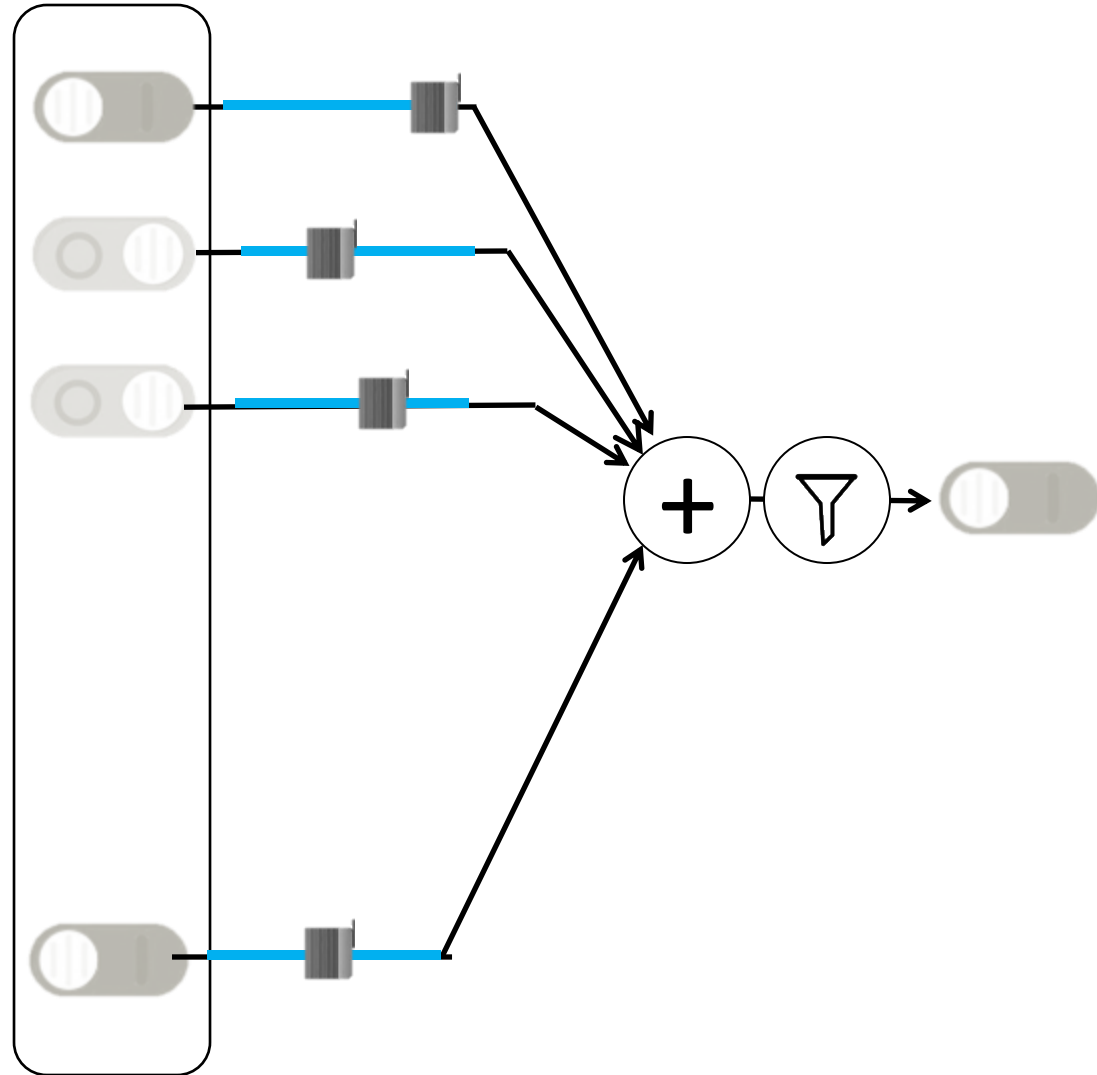


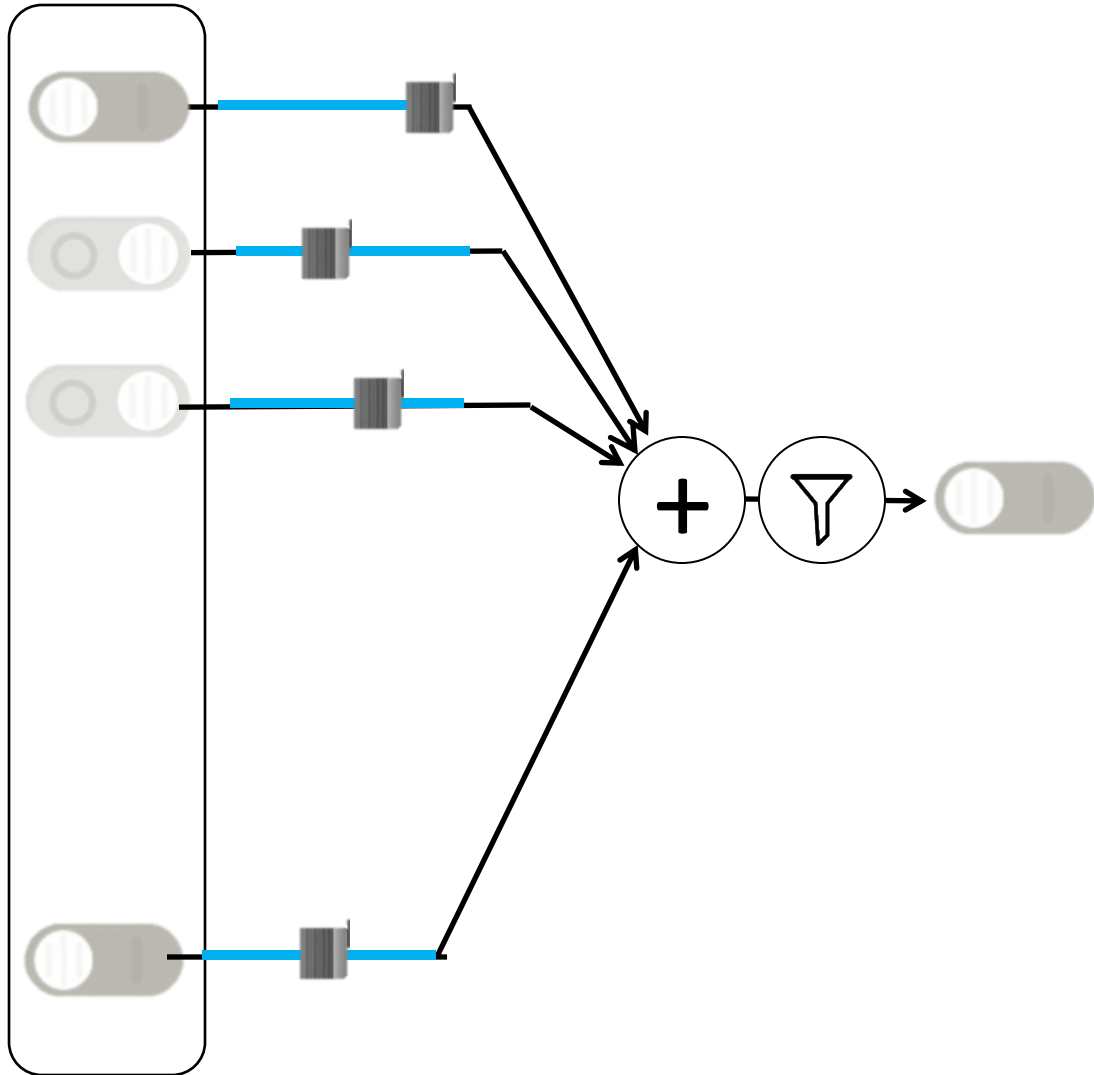
I think that is a picture of a **One!**

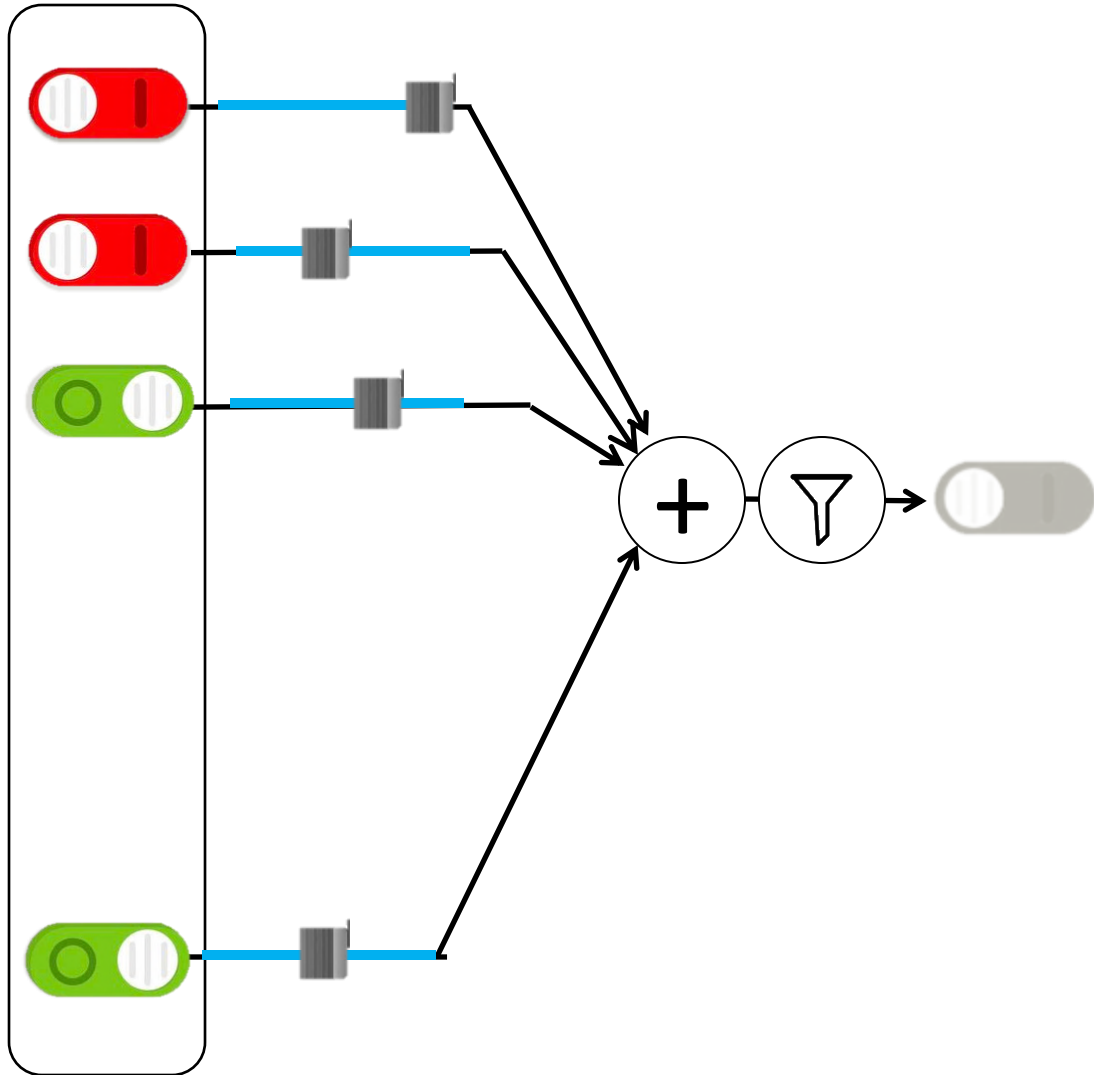
What do you mean it's actually a **Zero?**

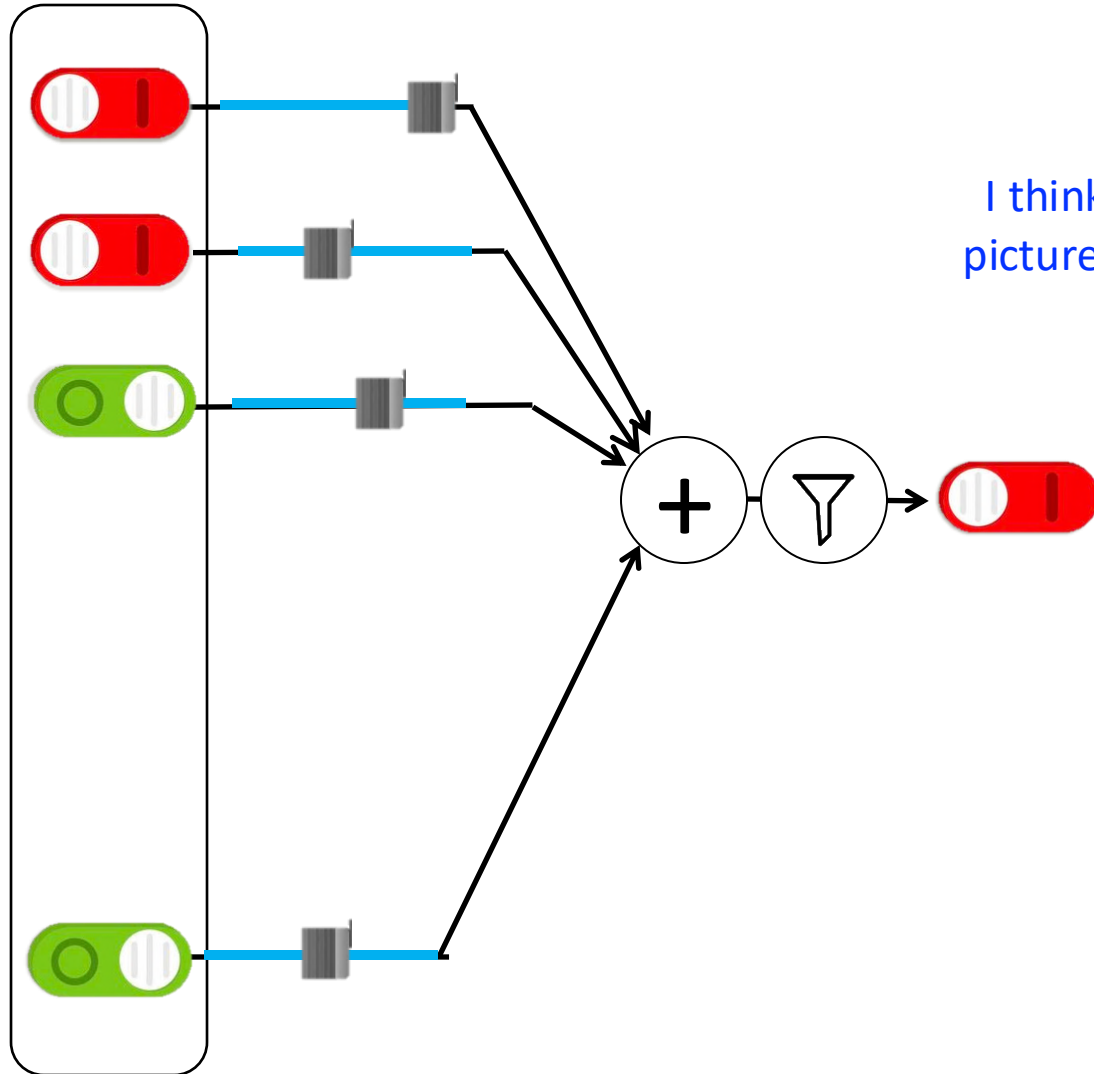
I'll adjust my sliders so that I do a better job in the future







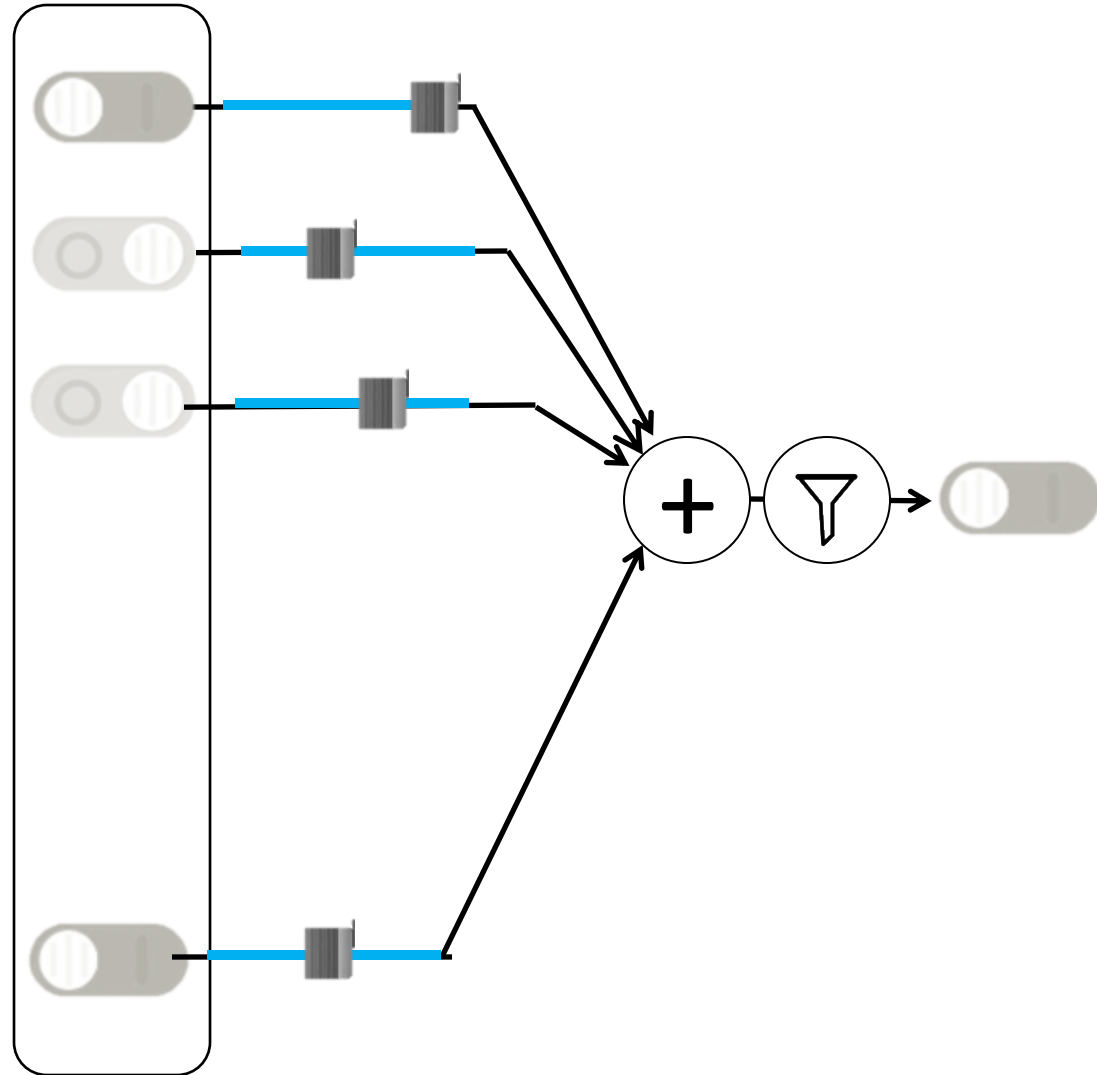


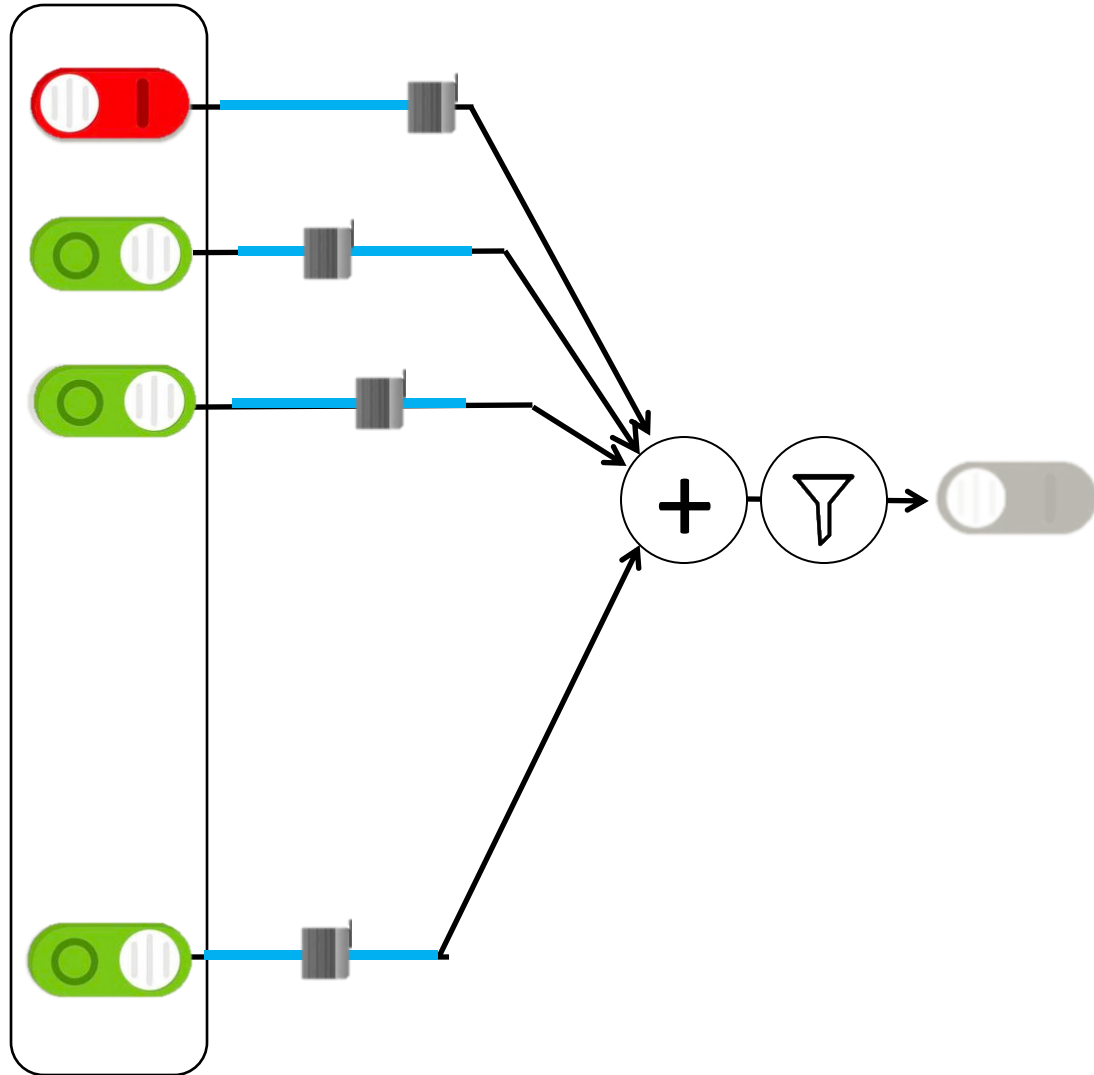


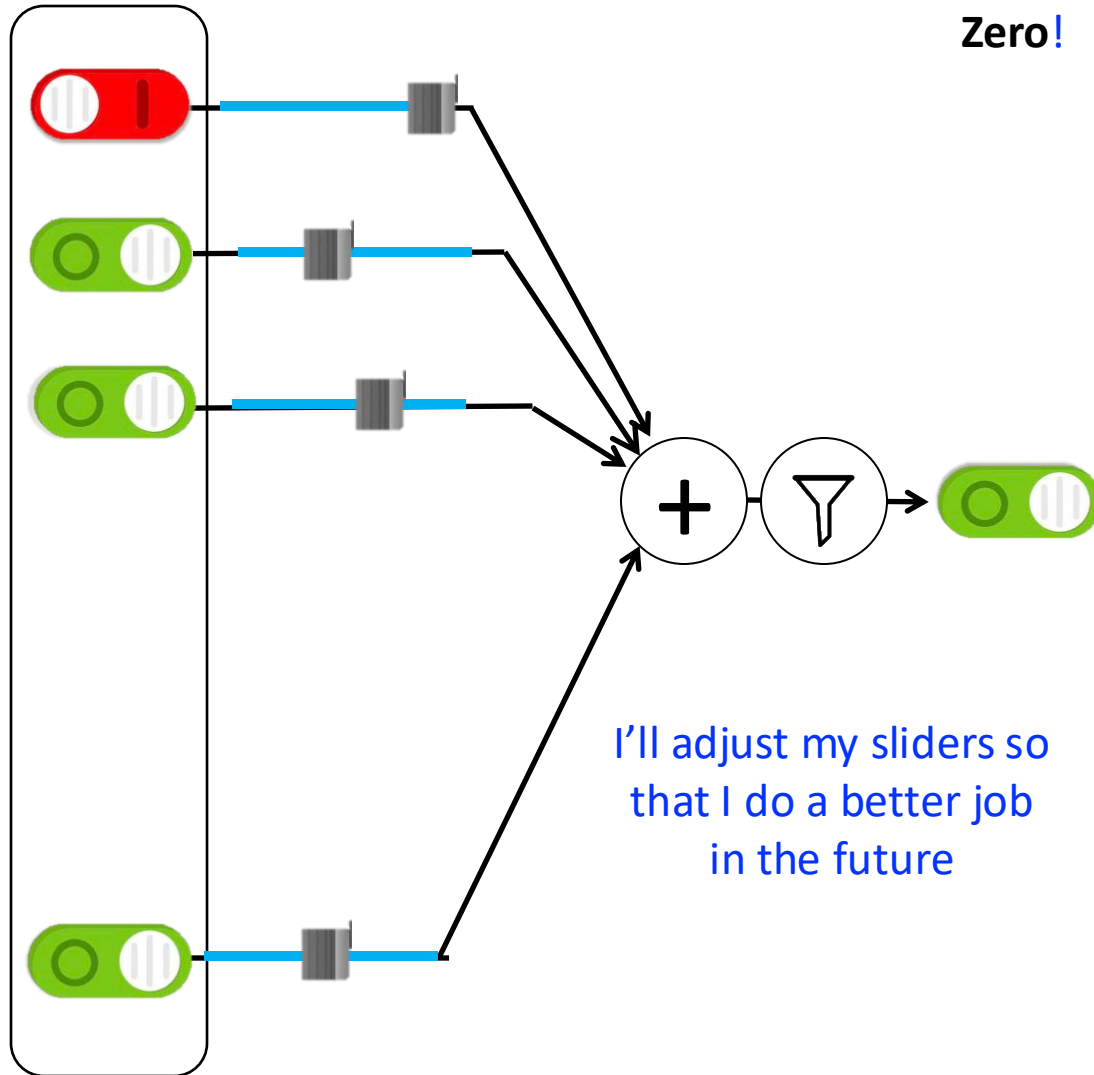
I think that is a picture of a **One!**

Wahoo I got it right!







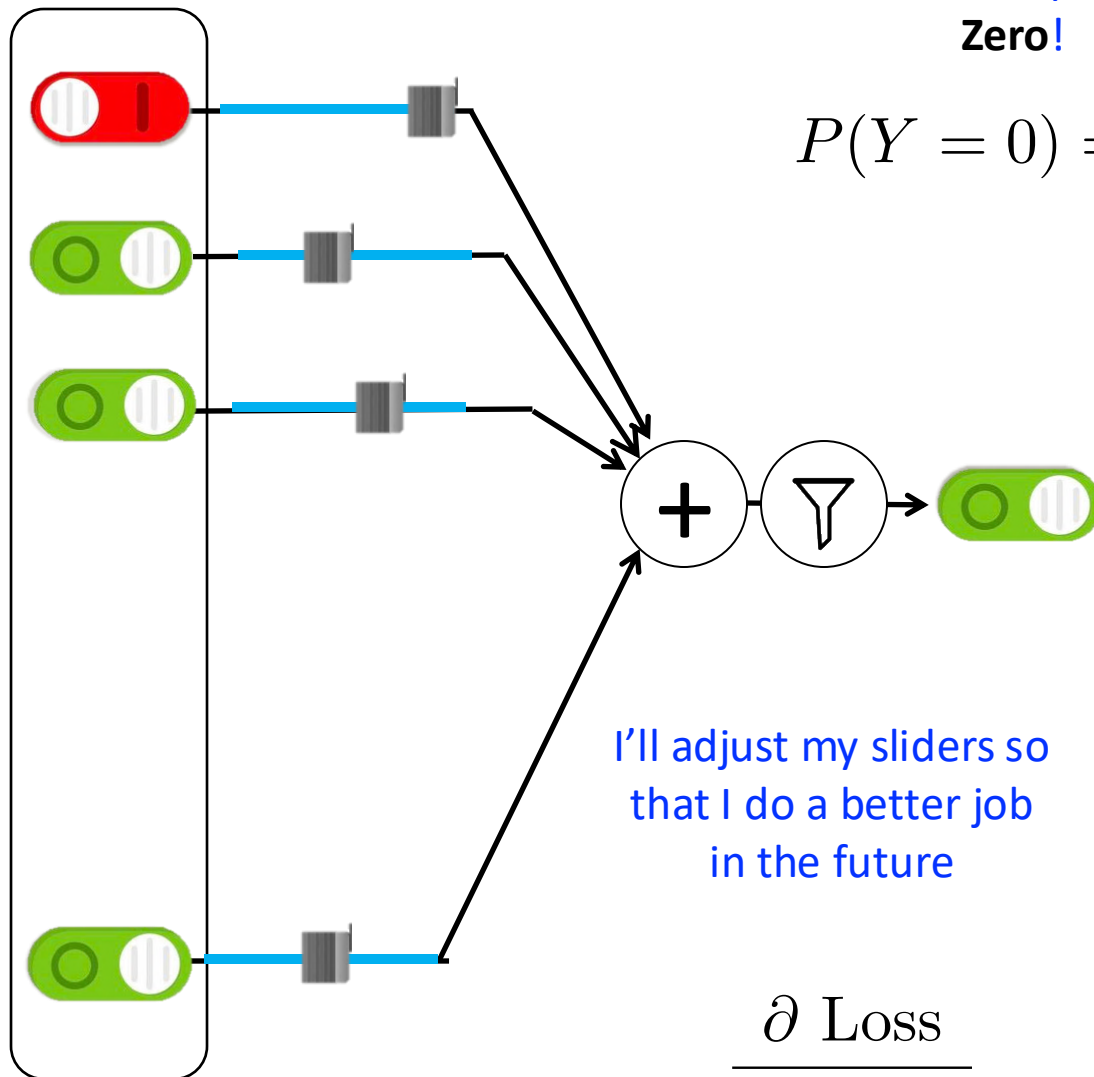


I think that is a picture of a
Zero!

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**

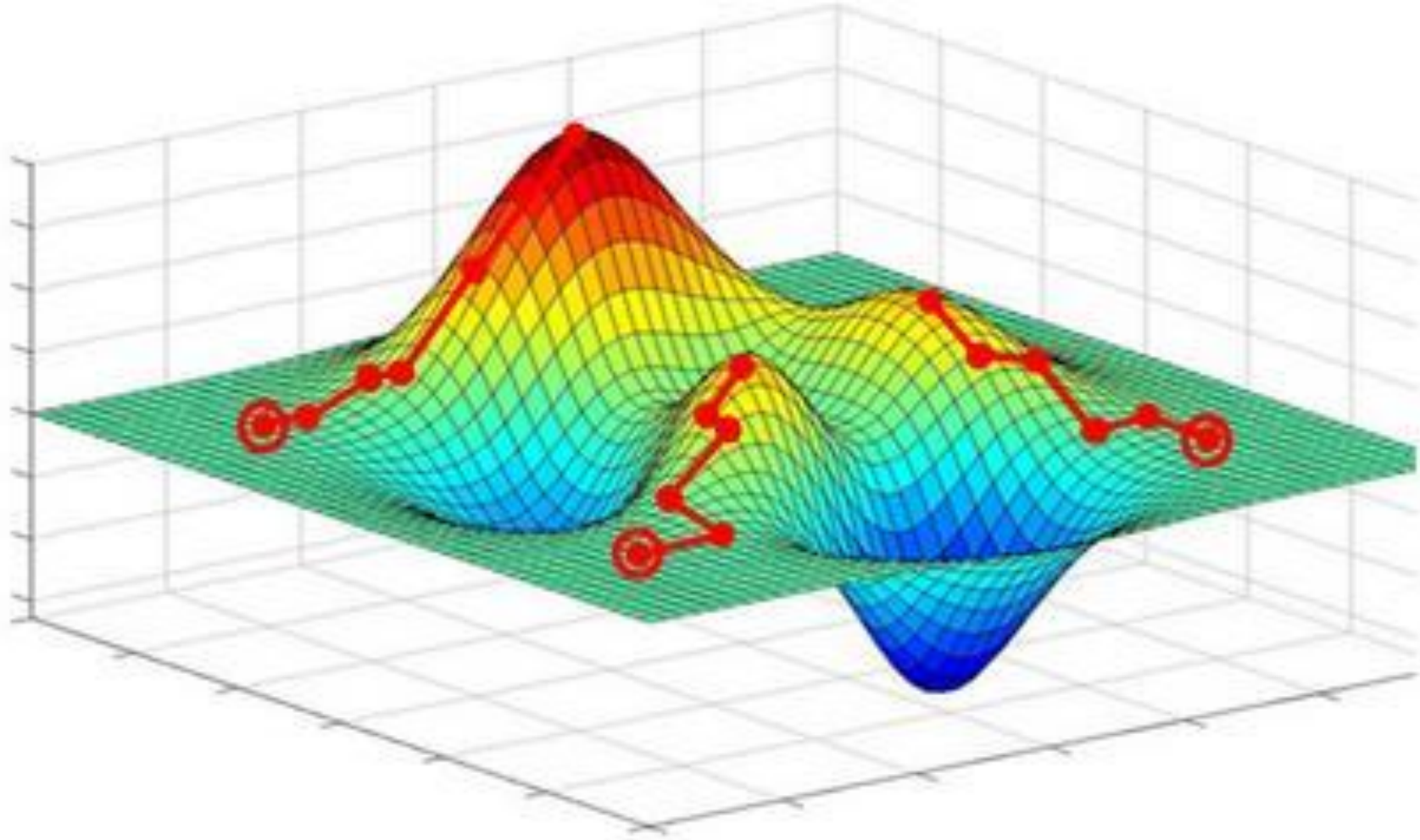
$$\text{Loss} = 1$$

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

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

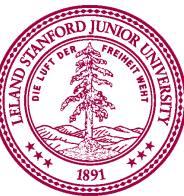


Gradient Descent



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

Piech, CS109, Stanford University



Gradient of Probability

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

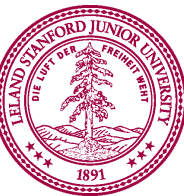
$$\hat{y} = \sigma \left(\sum_{j=0}^{m_h} \mathbf{h}_j \theta_j^{(\hat{y})} \right)$$

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

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

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

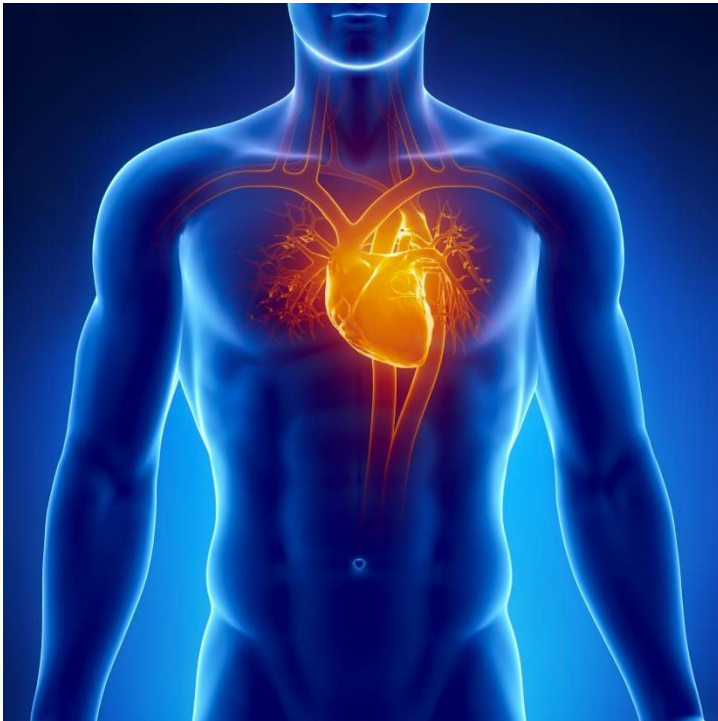
You will be able to do this.



Where you will be by the end of class

CS109: Theory Class focused on Applications

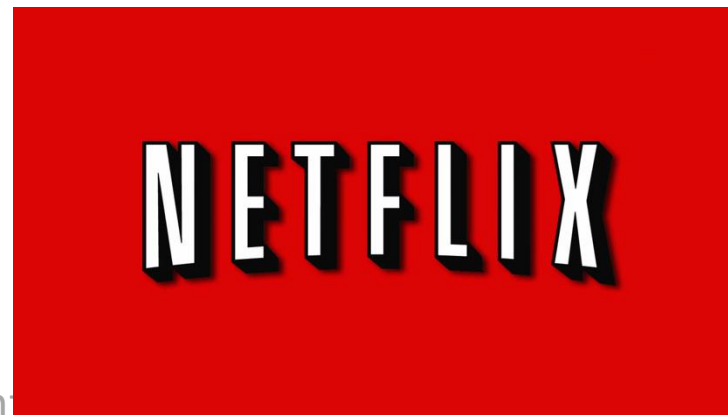
Heart



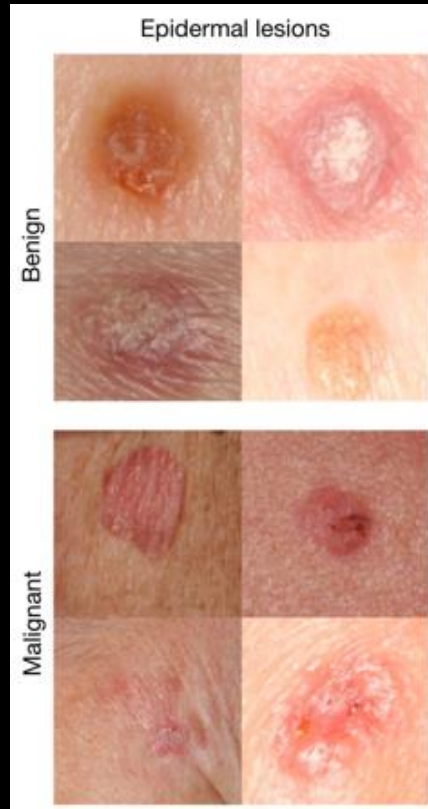
Ancestry



Netflix



Where is this Useful?



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

Developed recently, at Stanford.

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

What about Generative AI?

Who Invented Generative AI for Images?



Deep Unsupervised Learning using Nonequilibrium Thermodynamics by Jascha Sohl Dickstein

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



Deep Unsupervised Learning using Nonequilibrium Thermodynamics

Jascha Sohl-Dickstein
Stanford University

JASCHA@STANFORD.EDU

Eric A. Weiss
University of California, Berkeley

EAWEISS@BERKELEY.EDU

Niru Maheswaranathan
Stanford University

NIRUM@STANFORD.EDU

Surya Ganguli
Stanford University

SGANGULI@STANFORD.EDU

Abstract

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

1. Introduction

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

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

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

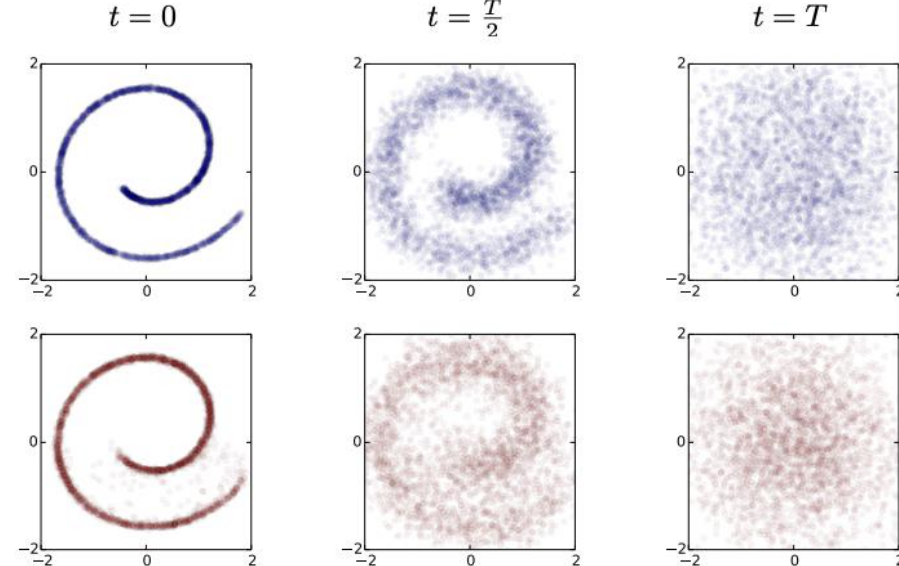
1.1. Diffusion probabilistic models

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

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

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

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



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

Will teach this at the end of the quarter

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

End of Story

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

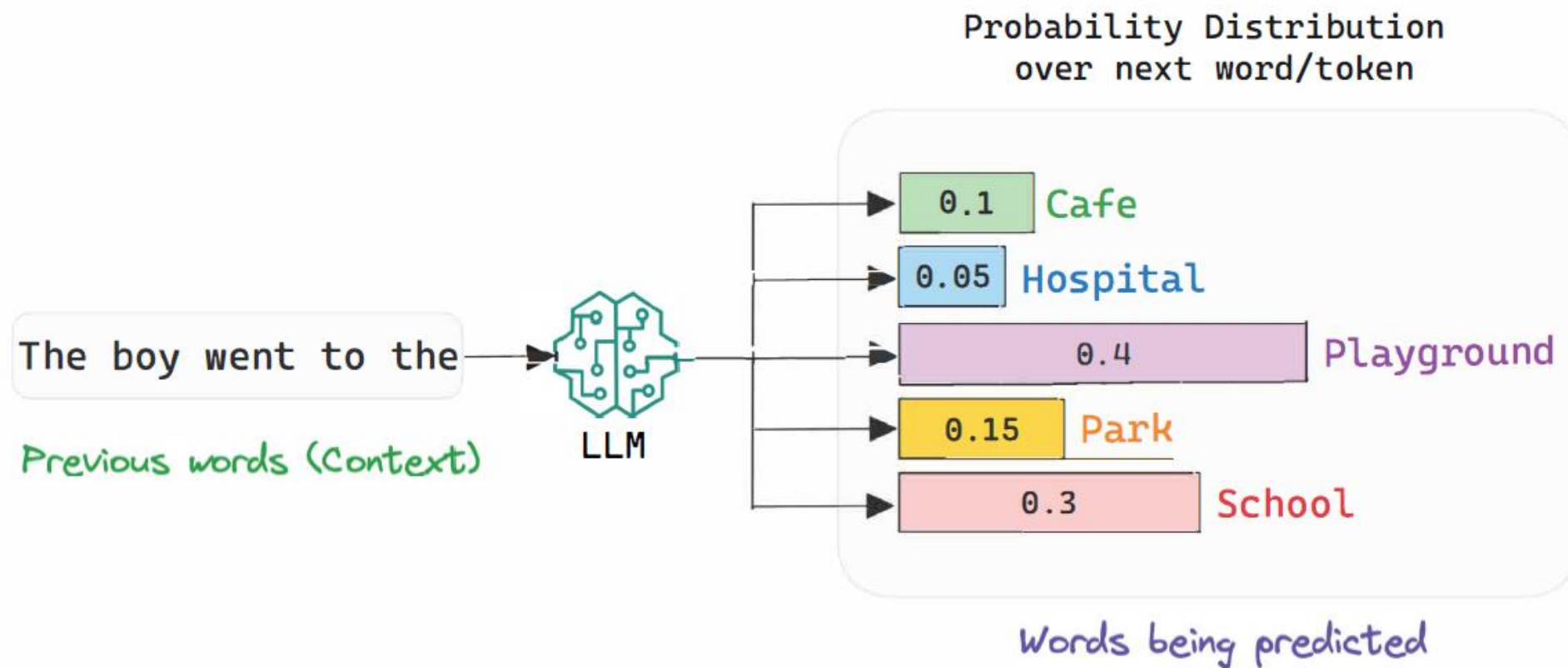
Probability is more than just machine learning

Abundance of important problems



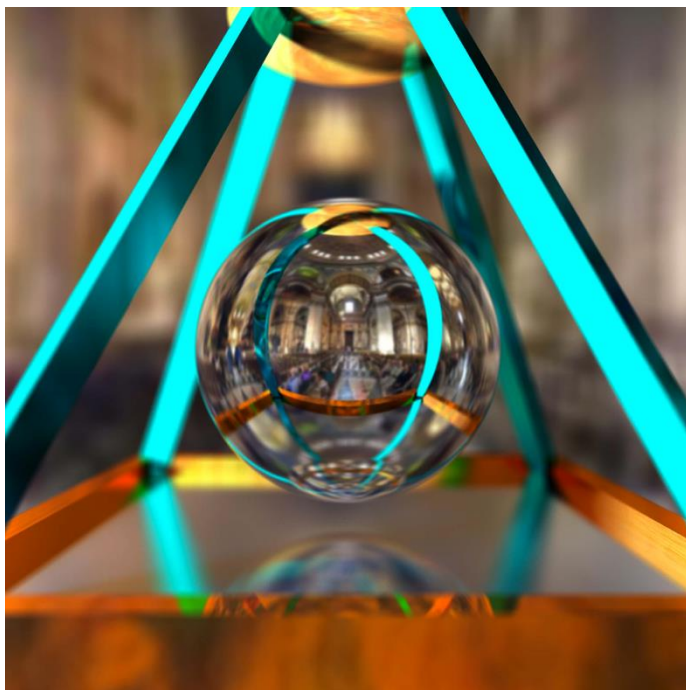
Many Open Problems

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

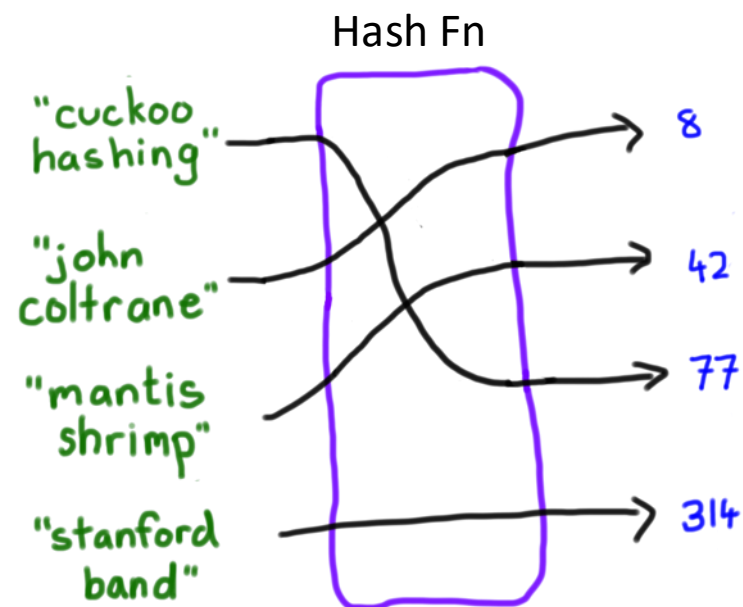


Algorithms and Probability

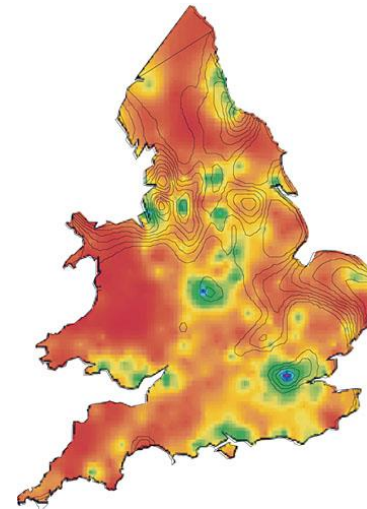
Eg Raytracing



Eg HashMaps



Understanding the world and building tools



Recommender Systems

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

Your 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%)**

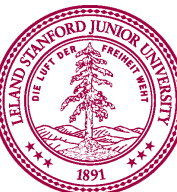
In Stock.
Ships from and sold by **Amazon.com**. Gift-wrap available.

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](#)

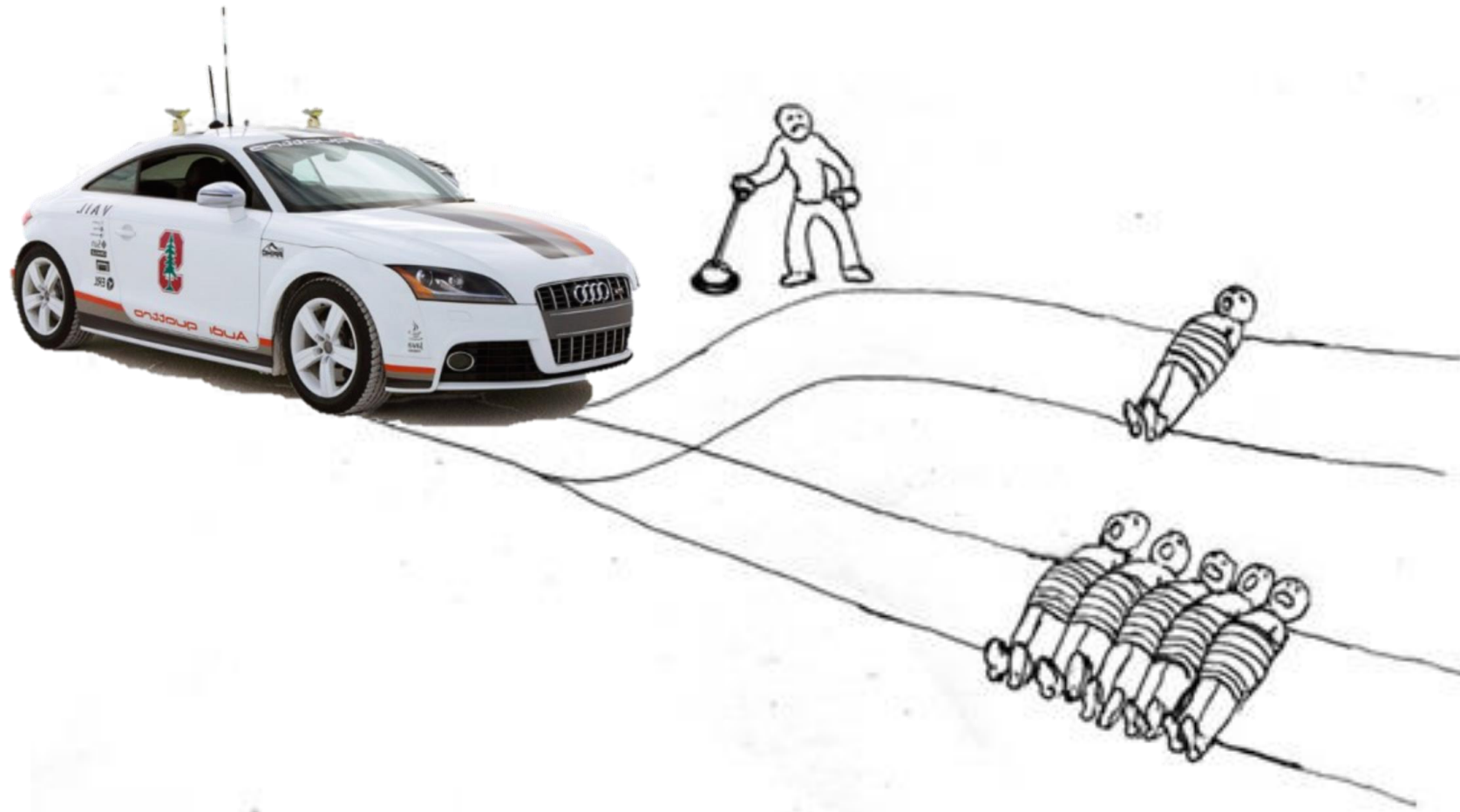
Customers Who Bought This Item Also Bought

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



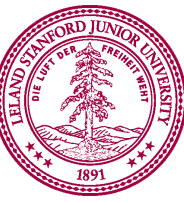
Most Desired Skill by PhD Students

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

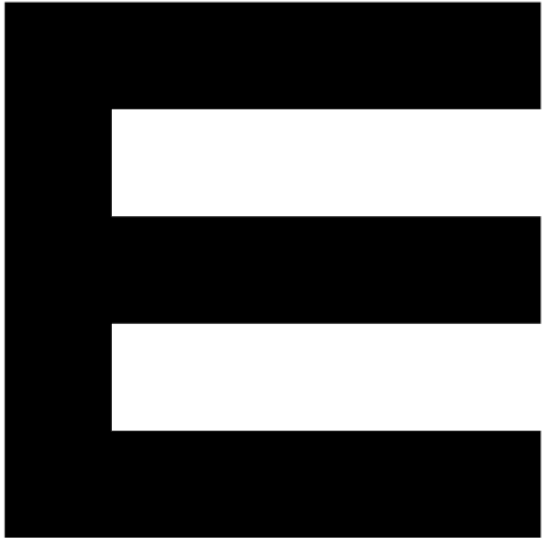
“Better understanding of probability”



Learn Real Skills in CS109



Spring 2017



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

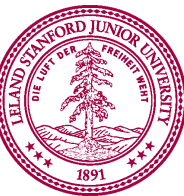
You decide that the vision tests given by eye doctors could have more precise results if we used an approach inspired by logistic regression. In a vision test a user looks at a letter with a particular font size and either correctly guesses the letter or incorrectly guesses the letter.

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.

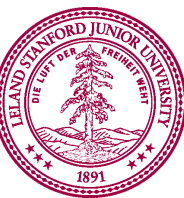
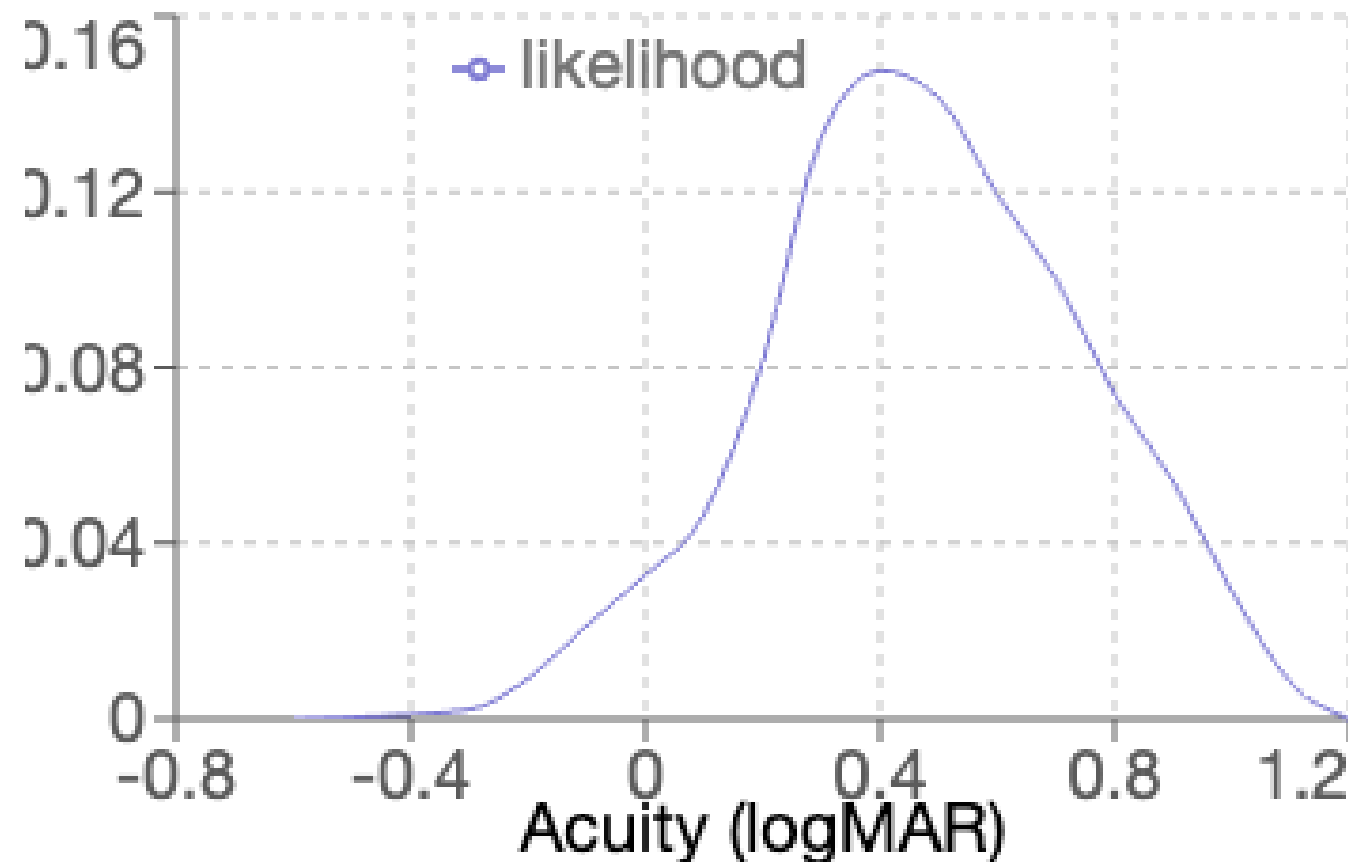
Explain how you could estimate a user's vision score (θ) based on their 20 responses $(f^{(1)}, y^{(1)}) \dots (f^{(20)}, y^{(20)})$, where $y^{(i)}$ is an indicator variable for whether the user correctly identified the i th letter and $f^{(i)}$ is the font size of the i th letter. Solve for any and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



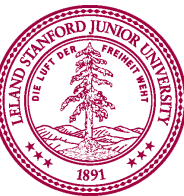
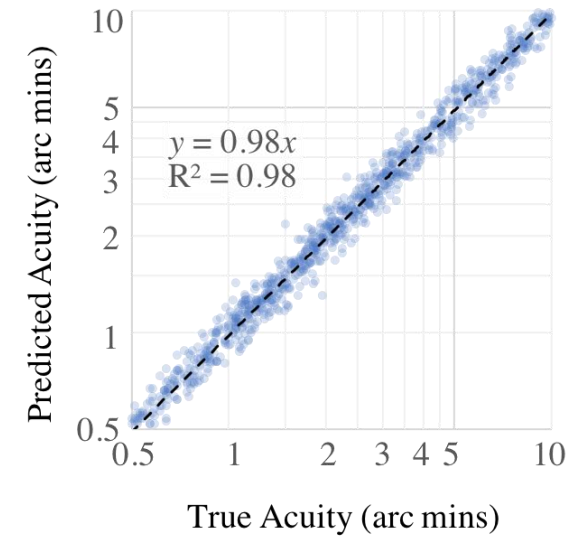
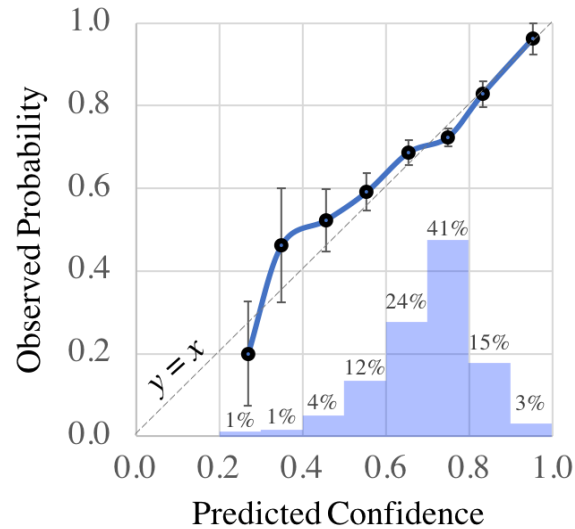
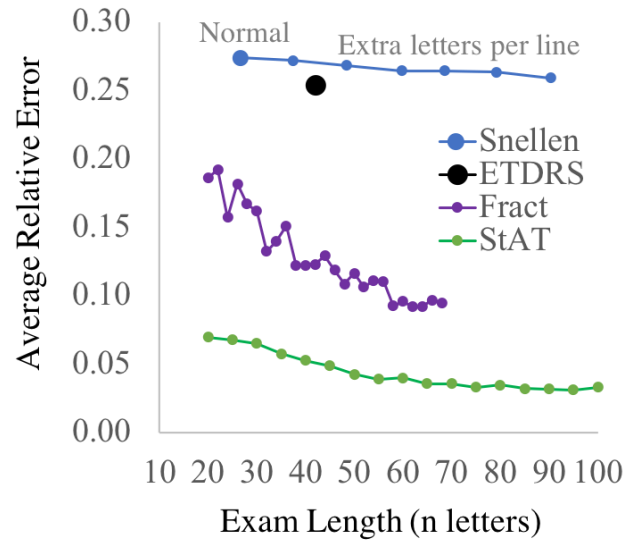
Learn Real Skills in CS109

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

Likelihood of Acuity Scores:



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

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.

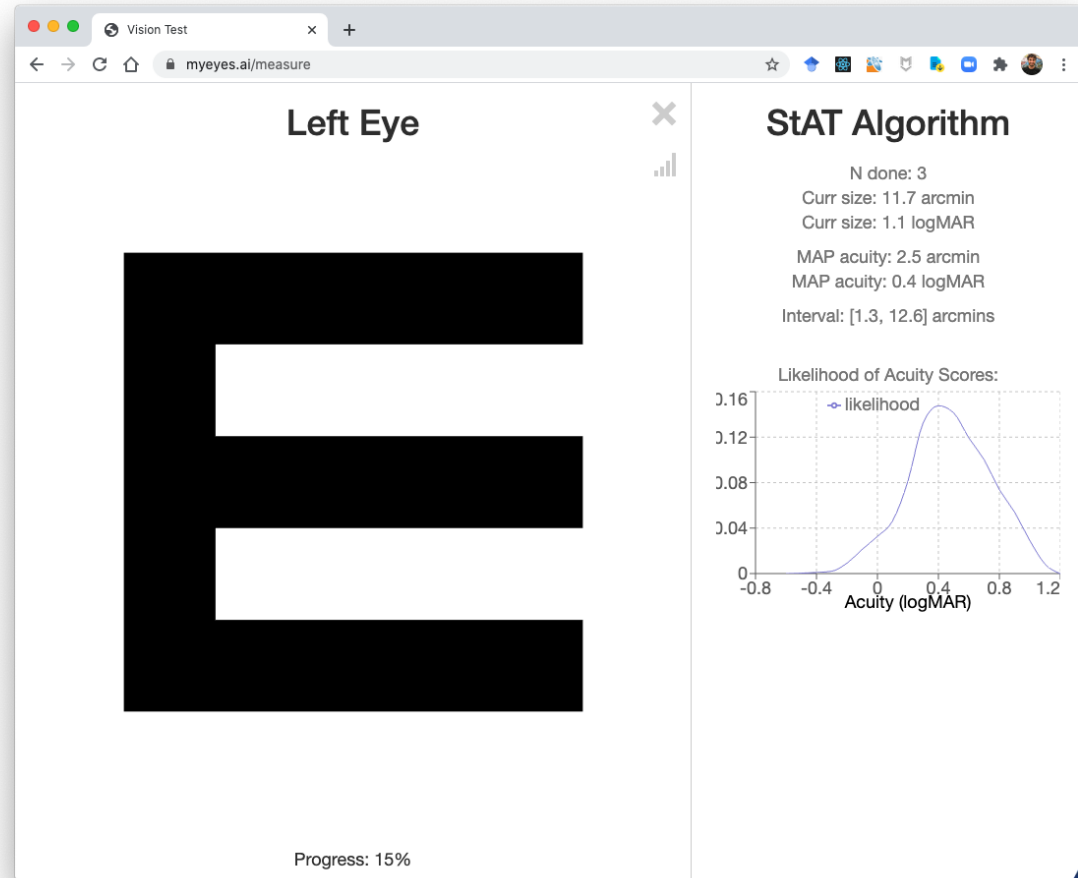
While promising, contemporary digital exams have yet to dramatically improve accuracy over traditional chart-based approaches. The current best digital exam uses a psychometric Item Response Theory (IRT) algorithm for both selecting the next letter size to query and for making a final prediction of acuity.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

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

StACT was named after Ed. We continue in this



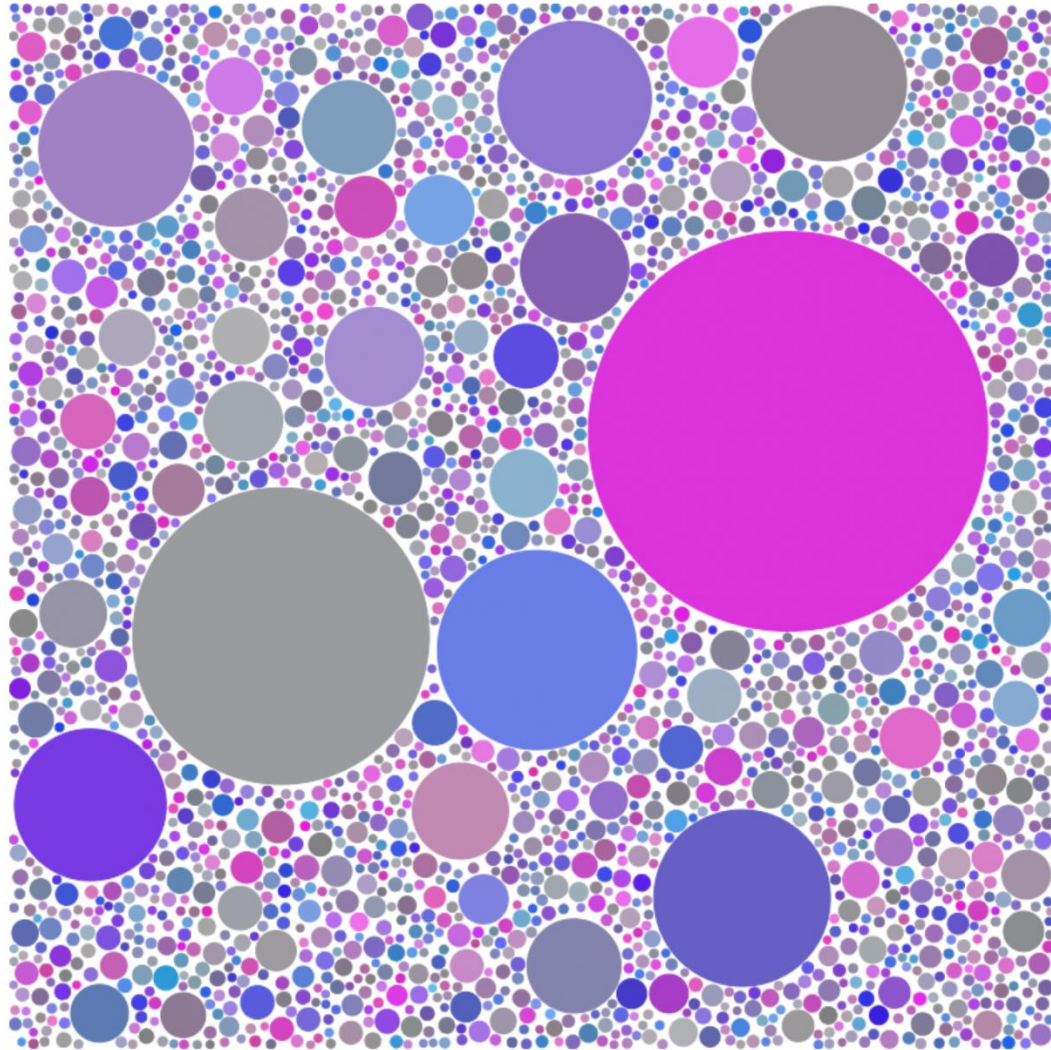
Science

THE LANCET



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

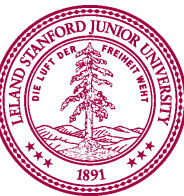
What is on a typical final?



Regenerate

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

https://chrspiach.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

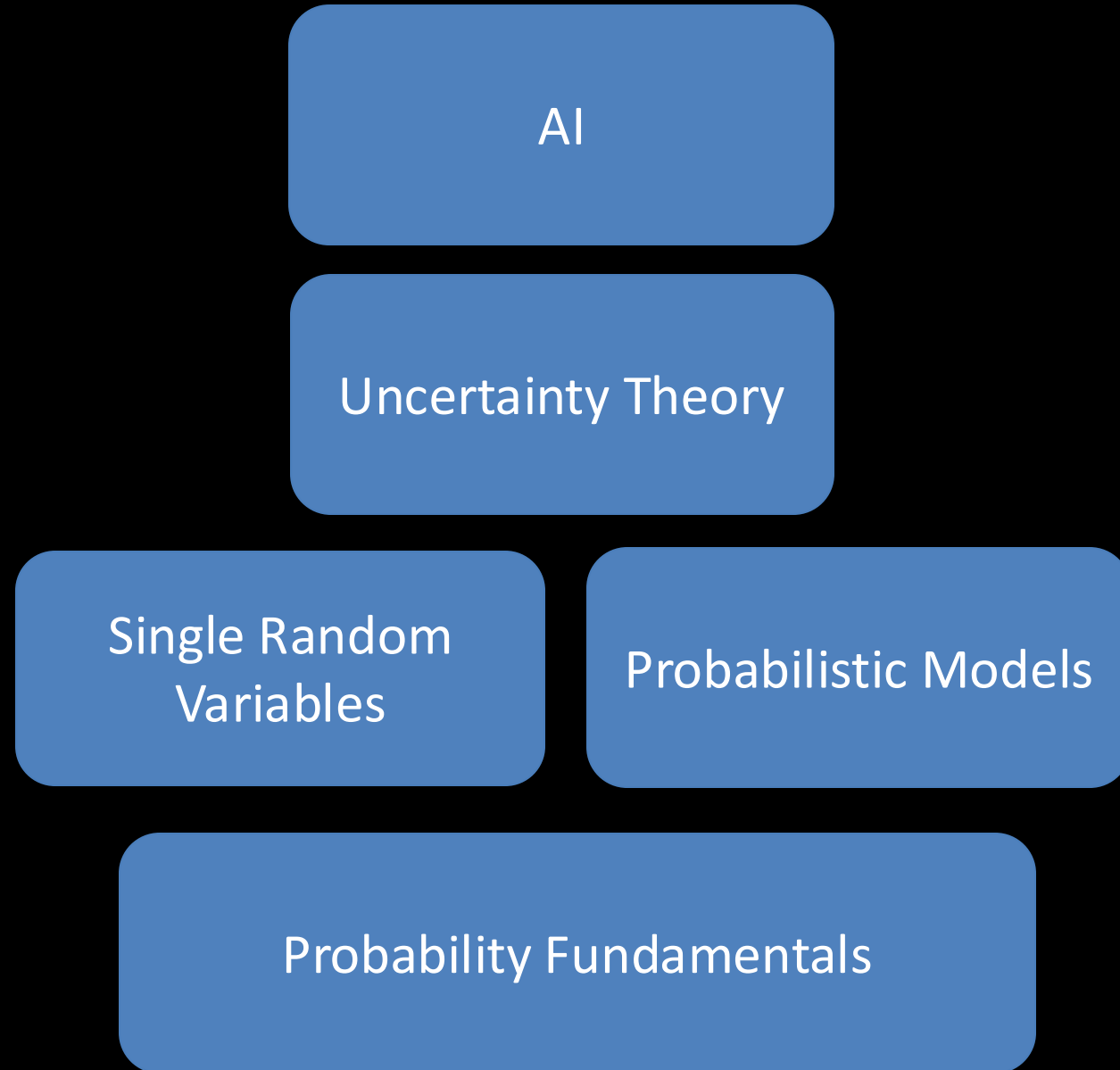
Delayed gratification

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.



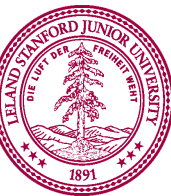
Why learn in the time of ChatGPT?

Lets dive in...

2 min pedagogic pause.

Sample Space

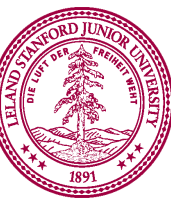
- **Sample space**, S , is set of all possible outcomes of an experiment
 - Coin flip: $S = \{\text{Head, Tails}\}$
 - Flipping two coins: $S = \{[H, H], [H, T], [T, H], [T, T]\}$
 - Roll of 6-sided die: $S = \{1, 2, 3, 4, 5, 6\}$
 - # emails in a day: $S = \{x \mid x \in \mathbf{Z}, x \geq 0\}$ {non-neg.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



Event Space

- **Event**, E , is some subset of S $\{E \subseteq S\}$
 - Coin flip is heads: $E = \{\text{Head}\}$
 - ≥ 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 ≤ 20 : $E = \{x \mid x \in \mathbf{Z}, 0 \leq x \leq 20\}$
 - Wasted day $\{\geq 5 \text{ YT hrs.}\}$: $E = \{x \mid x \in \mathbf{R}, 5 \leq x \leq 24\}$

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



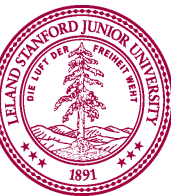
Event Space

Sample Space, S

- Coin flip
 $S = \{\text{Heads, Tails}\}$
- 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
 $S = \{x \mid x \in \mathbb{Z}, x \geq 0\}$
- 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
 $E = \{(H,H), (H,T), (T,H)\}$
- 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

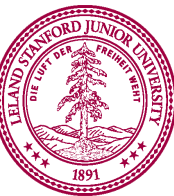


A number to which we ascribe meaning



$\text{Pr}(E)$

* Our belief that an event E occurs



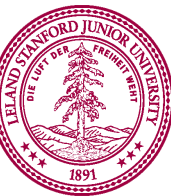
What is a Probability?

The event we
care about

How many times does it
occur

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

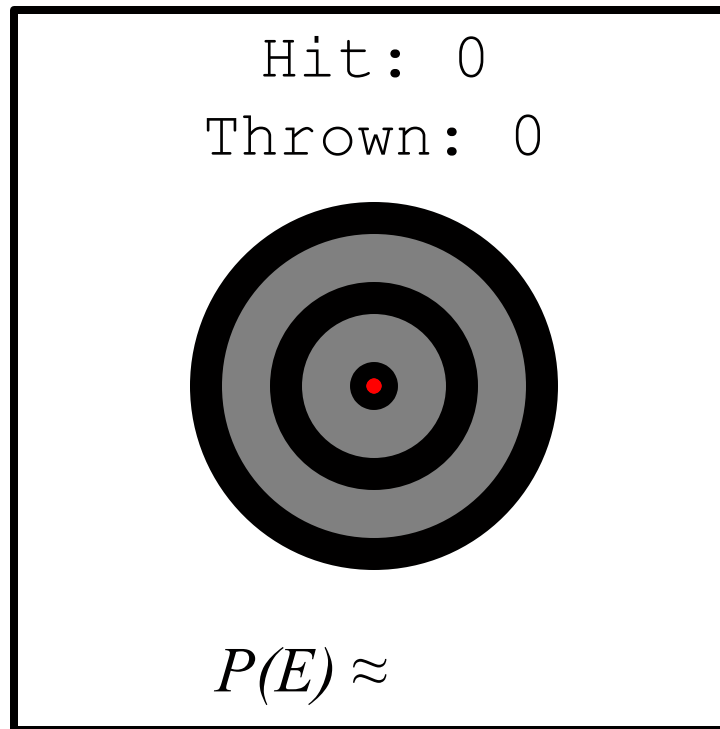
Out of (close to) infinite
trials



What is a Probability?

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

n is the number
of trials

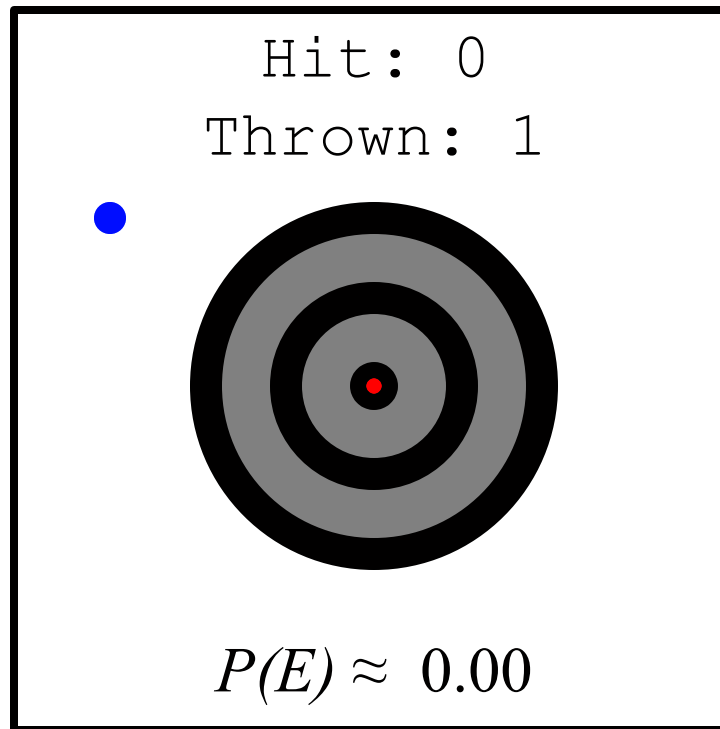


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

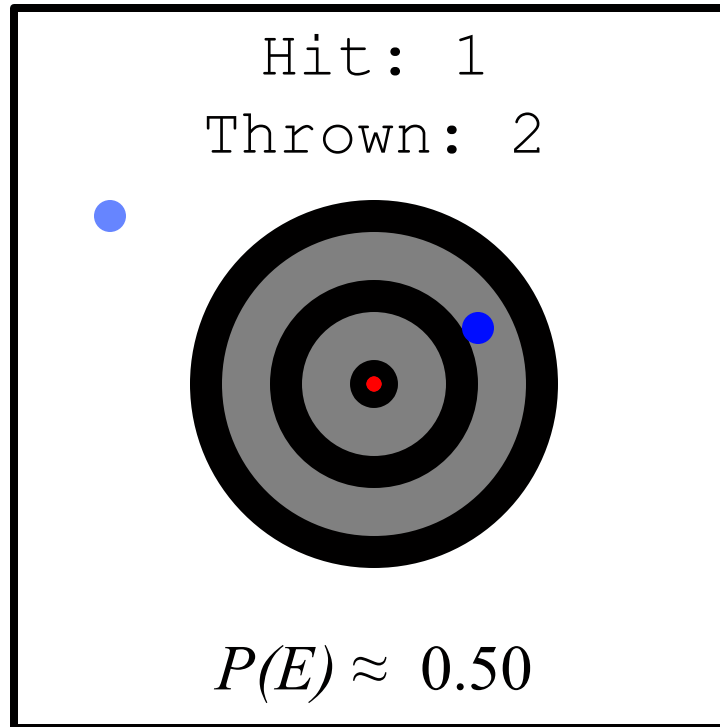


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

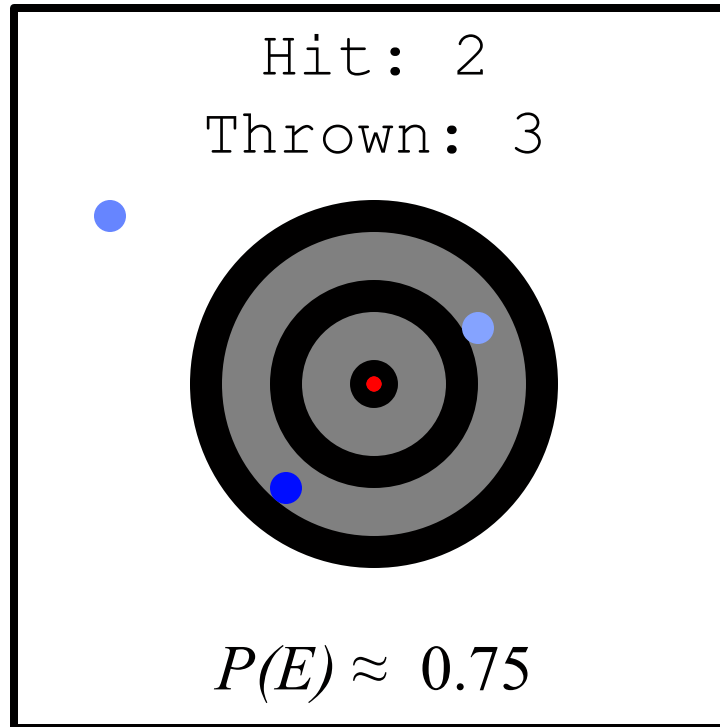


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials

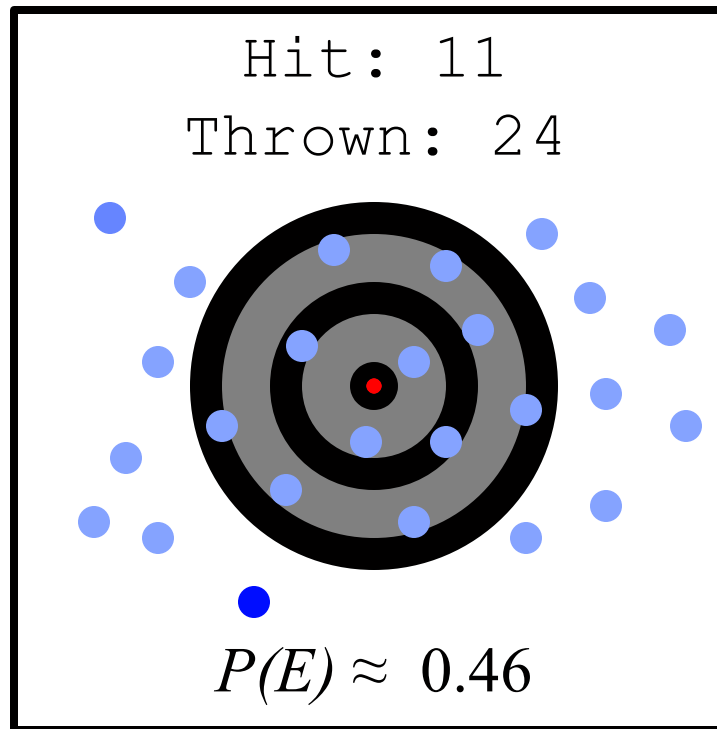


The “event” E
is that you hit
the target

What is a Probability?

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

n is the number
of trials



The “event” E
is that you hit
the target

What is a Probability (in a Dataset)?

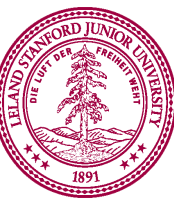
$$P(E) = \lim_{n \rightarrow \infty} \frac{n(E)}{n}$$

Dataset of weather

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

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

$$\begin{aligned} P(E) &= \lim_{n \rightarrow \infty} \frac{n(E)}{n} \\ &\approx \frac{\text{Count}(E)}{10000} \\ &\approx \frac{3332}{10000} \approx 0.3332 \end{aligned}$$



Equally Likely Outcomes

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

- Coin flip: $S = \{\text{Head, Tails}\}$
- 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 $P(\text{Each outcome}) = \frac{1}{|S|}$

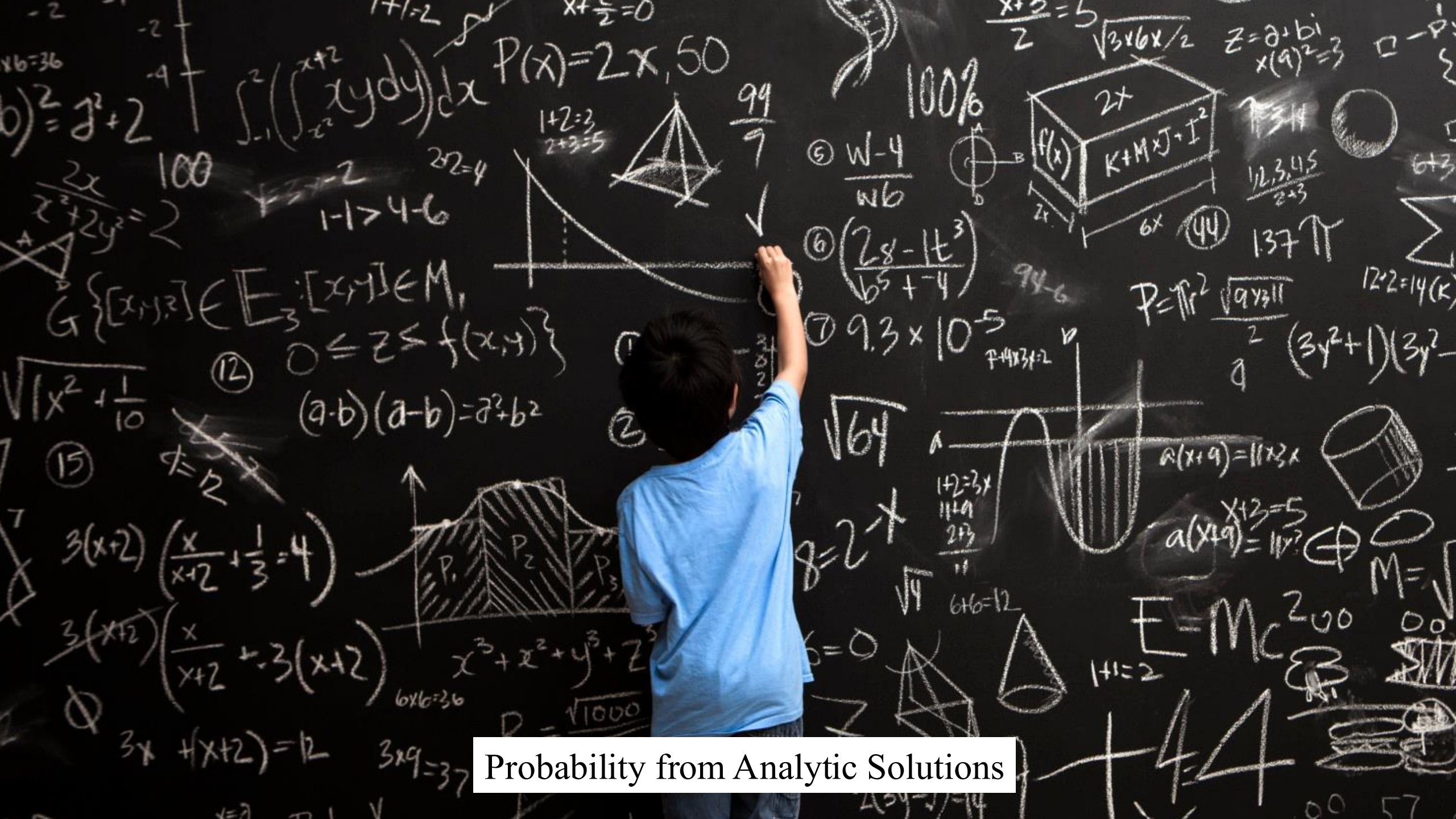
Therefore $P(E) = \frac{\# \text{ outcomes in } E}{\# \text{ outcomes in } S} = \frac{|E|}{|S|}$ {by Axiom 3}

Not Everything is Equally Likely

- Play lottery.
 - What is $P(\text{Win})$?
-

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





Probability from Analytic Solutions

The Axioms



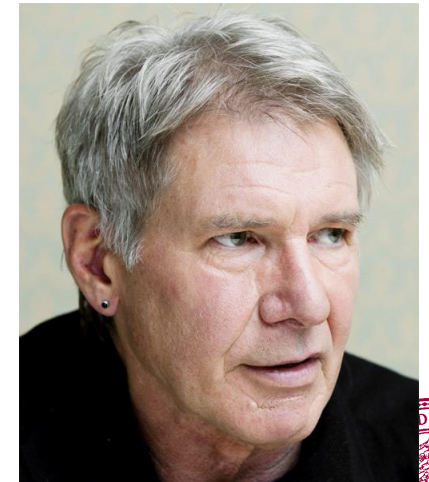
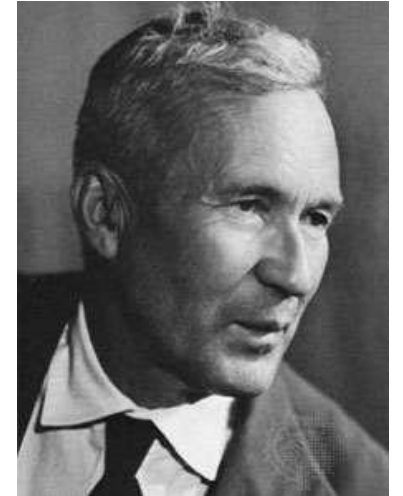
Axioms of Probability

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:

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

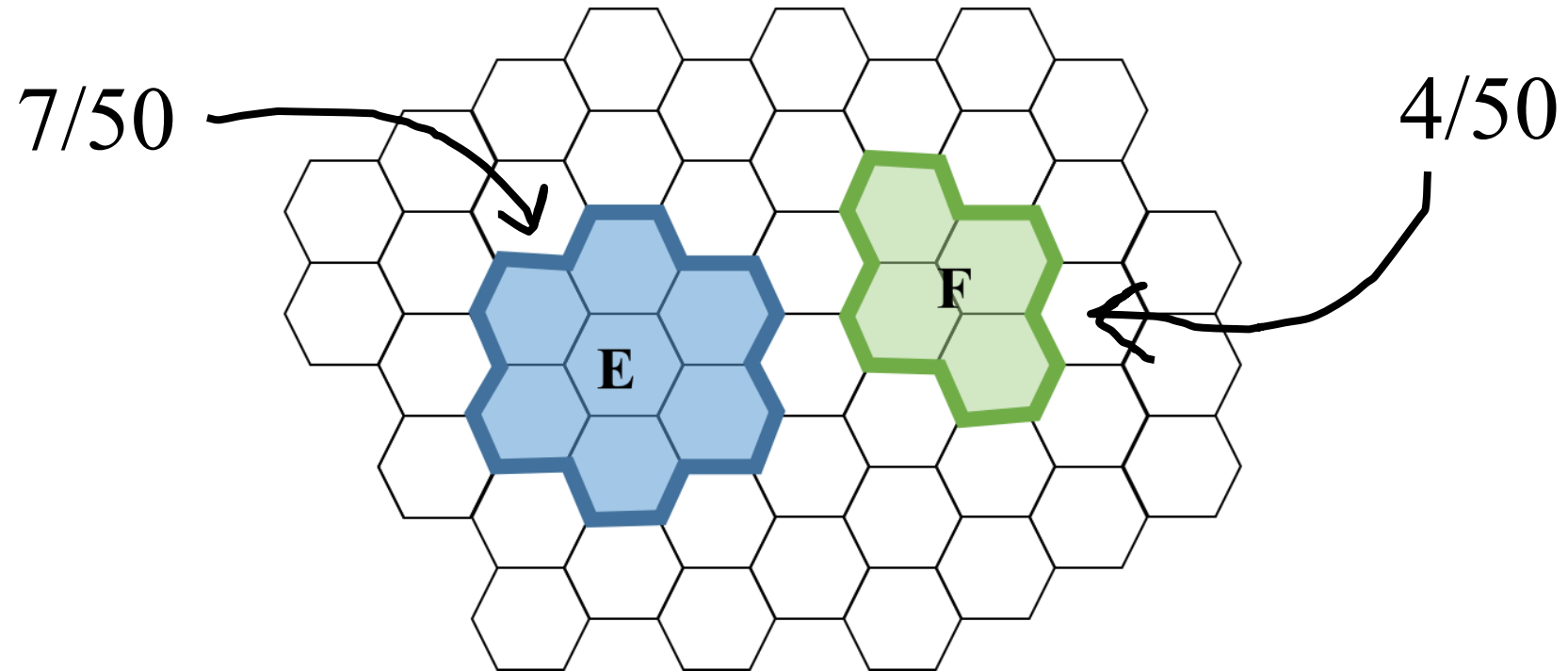
Kolmogorov



Harrison Ford



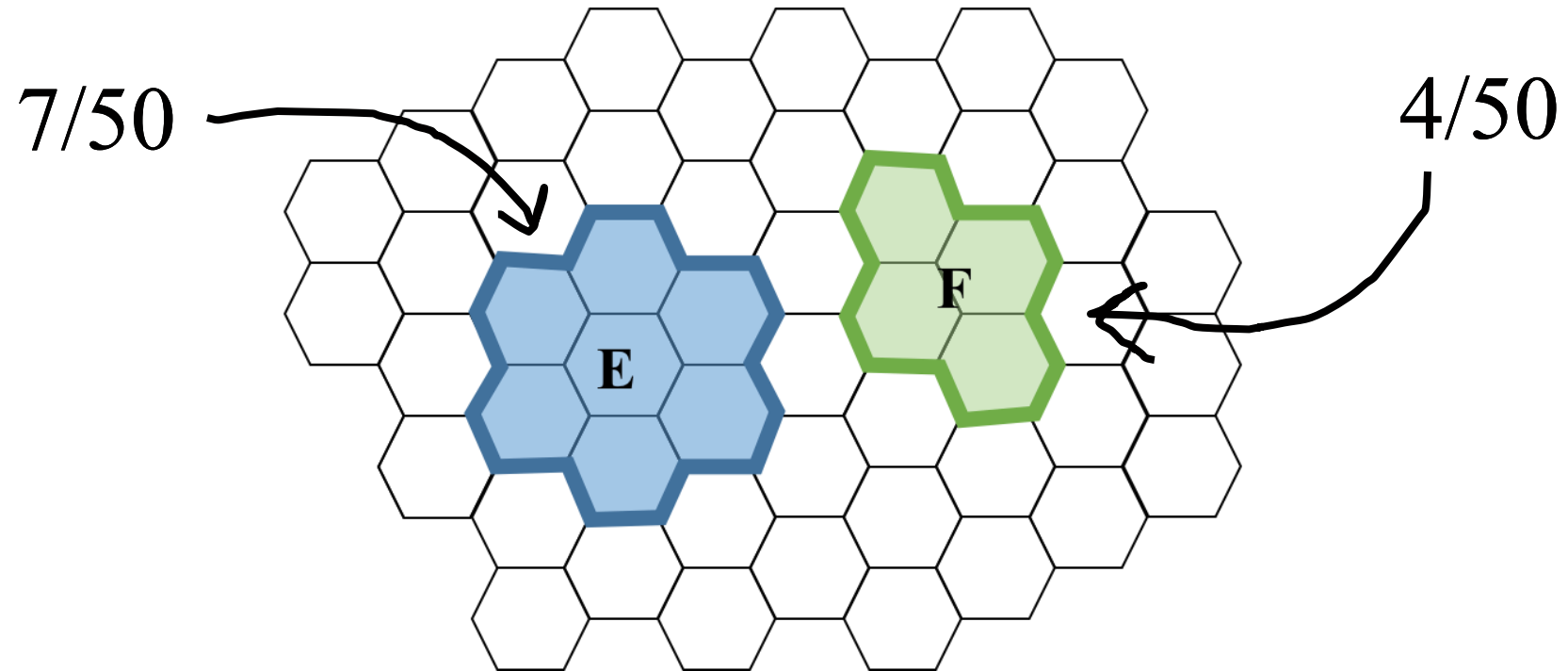
Mutually Exclusive Events



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

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

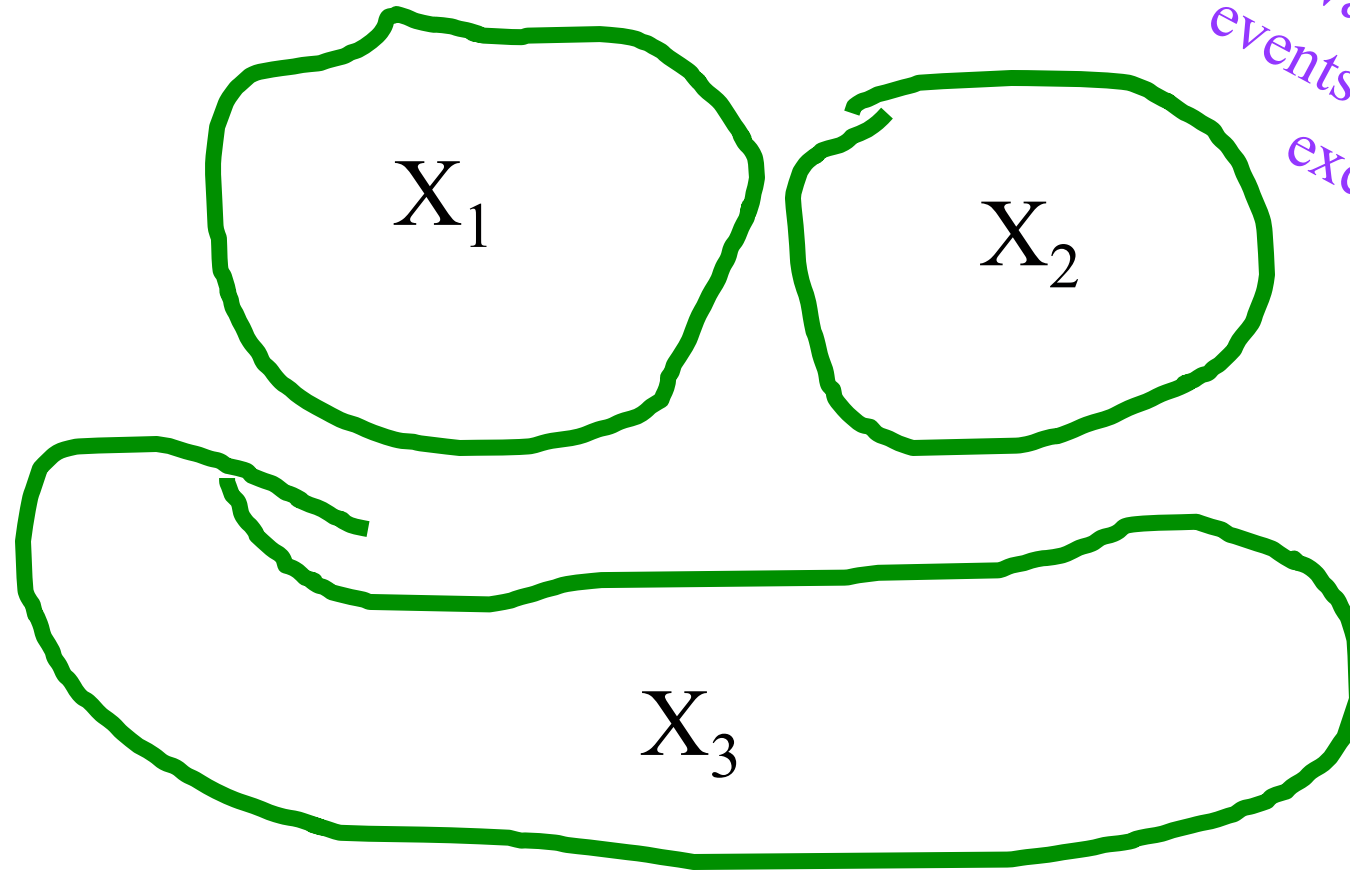
Mutually Exclusive Events



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

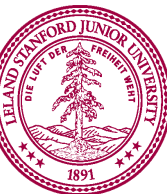
$$P(E \cup F) = \frac{7}{50} + \frac{4}{50} = \frac{11}{50}$$

Probability of "or"



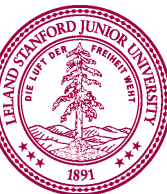
Wahoo! All my
events are mutually
exclusive

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





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

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

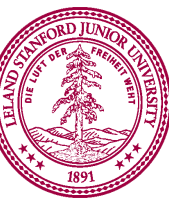
Since everything must either be in E or E^c

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

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?

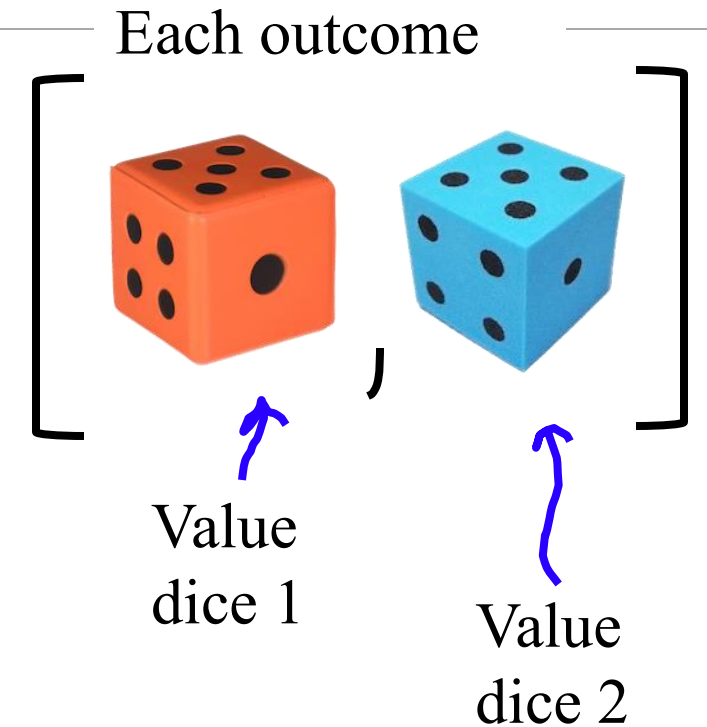
Roll two 6-sided dice. What is probability the sum = 7?

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]

$\}$

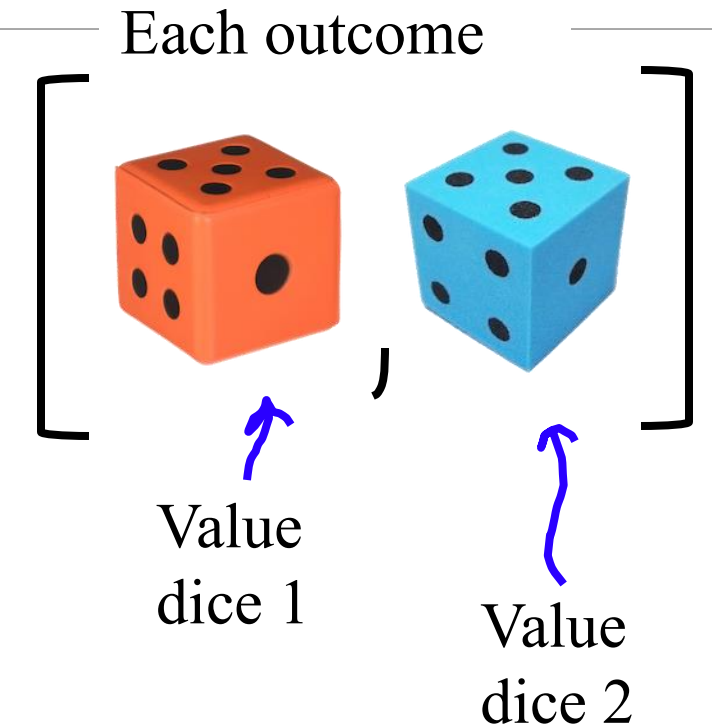


Sum of Two Die = 7?

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] }



E = *in blue*

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.1\overline{6}$$

Sum of Two Die = 7?

```
1  import random
2  from tqdm import tqdm
3
4  N_TRIALS = 10000000 # getting close to infinity
5  TARGET_SUM = 7      # do the two dice sum to 6?
6
7  def main():
8      n_events = 0
9      for i in tqdm(range(N_TRIALS)):
10         dice_total = run_experiment()
11         if dice_total == TARGET_SUM:
12             n_events += 1
13     pr_e = n_events / N_TRIALS
14     print(f'after {N_TRIALS} trials')
15     print('P(E) ≈ ', pr_e)
16
17  def run_experiment():
18     d_1 = roll_dice()
19     d_2 = roll_dice()
20     return d_1 + d_2
21
22  def roll_dice():
23     # give me a random dice roll
24     # alternatively random.randint(1, 7)
25     return random.choice([1,2,3,4,5,6])
26
27  if __name__ == '__main__':
28     # this starts the program in main
29     main()
```

$$P(E) = \frac{|E|}{|S|} = \frac{6}{36} = 0.1\overline{6}$$

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL

```
piech@Chriss-MBP-2 3 % python dice_soln.py
after 10000000 trials
P(E) = 0.1666913
-
```

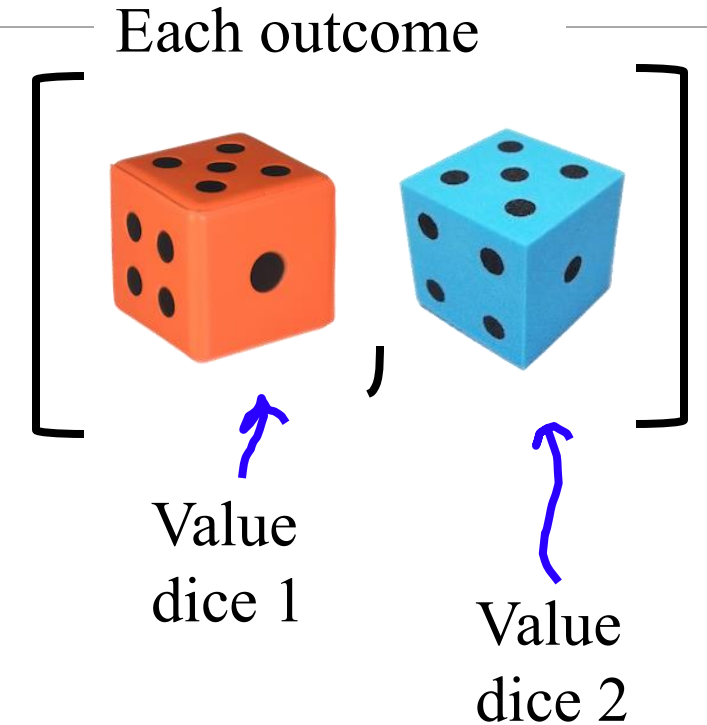
Sum of Two Die = 2?

Roll two 6-sided dice. What is probability the sum = 2?

Let E be the event that the sum is 2

$$S = \{ \begin{array}{cccccc} [1,1] & [1,2] & [1,3] & [1,4] & [1,5] & [1,6] \\ [2,1] & [2,2] & [2,3] & [2,4] & [2,5] & [2,6] \\ [3,1] & [3,2] & [3,3] & [3,4] & [3,5] & [3,6] \\ [4,1] & [4,2] & [4,3] & [4,4] & [4,5] & [4,6] \\ [5,1] & [5,2] & [5,3] & [5,4] & [5,5] & [5,6] \\ [6,1] & [6,2] & [6,3] & [6,4] & [6,5] & [6,6] \end{array} \}$$

E =



Sum of Two Die = 2?

Roll two 6-sided dice. What is probability the sum = 2?

Let E be the event that the sum is 2

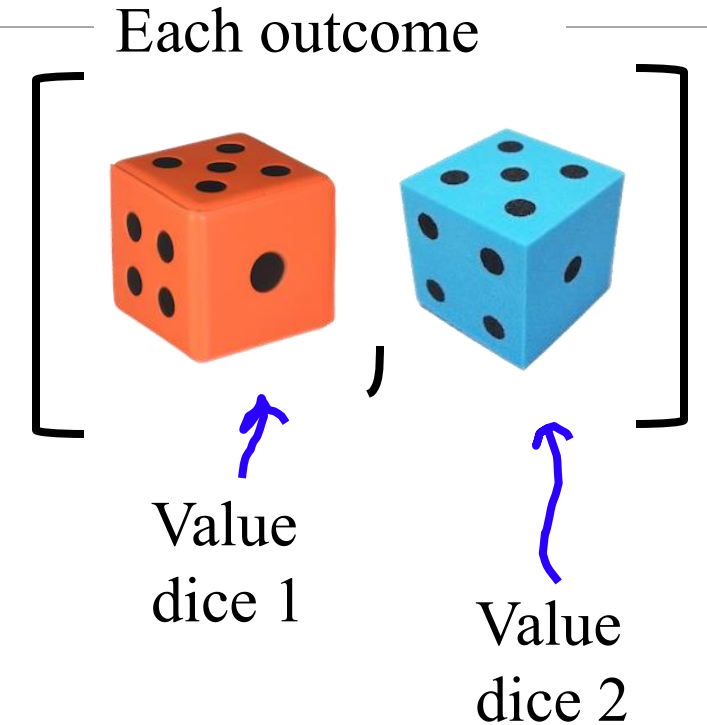
$S = \{$

[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]

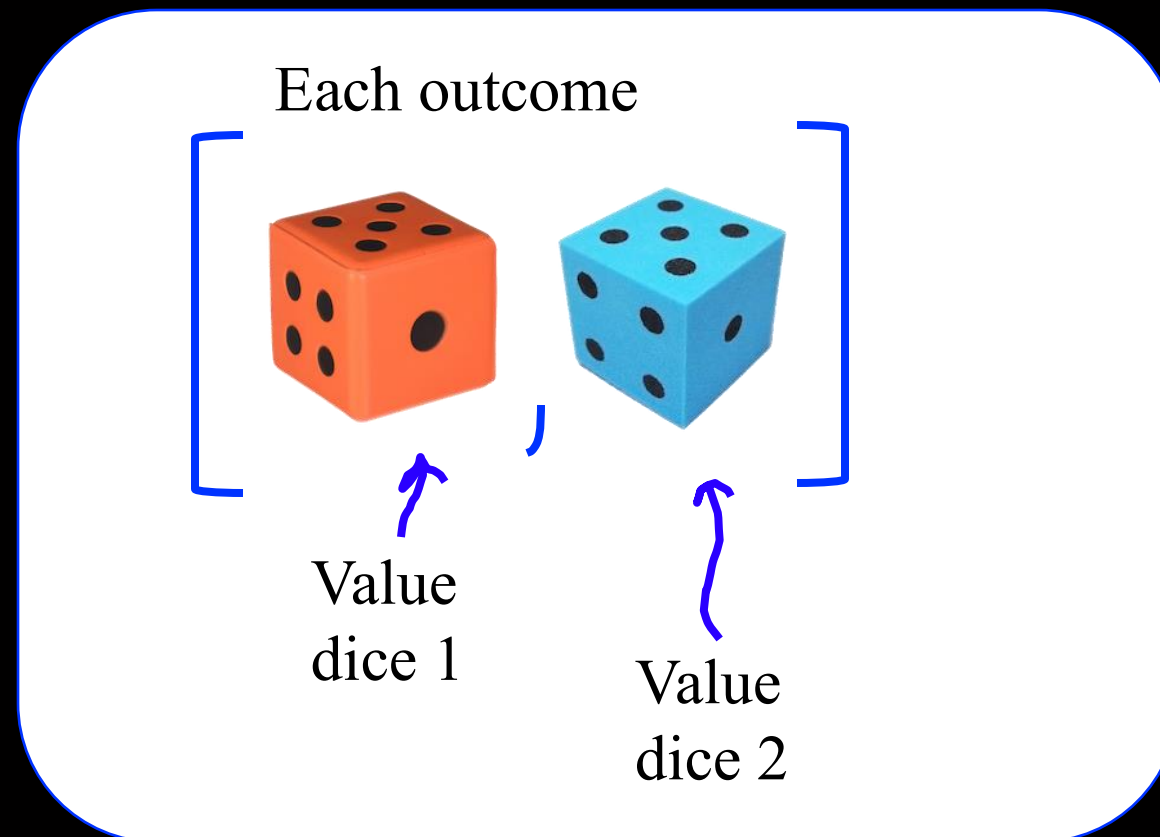
$\}$

E = *in red*

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



Other ways to make a Sample Space?



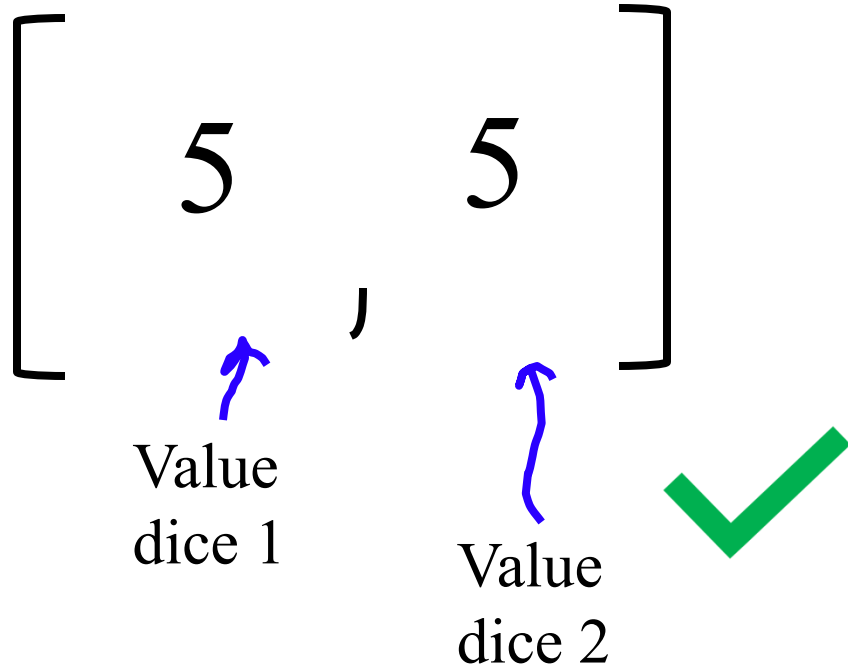
Sum of Two Die: Three options for the sample space

Value
dice 1

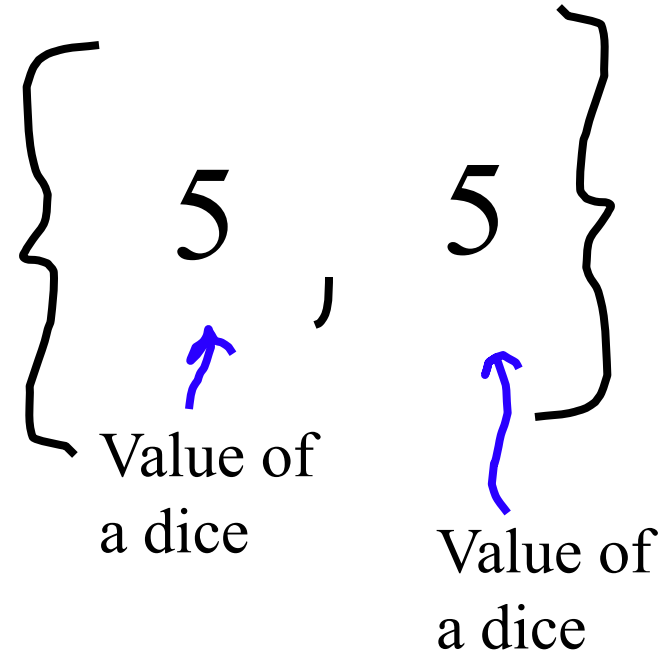


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**



Just look at the sum

10

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

Each outcome

Roll two 6-sided dice. What is probability the sum = 7?

Just look at the sum

Let E be the event that the sum is 7

$S = \{$

	2	3	4	5	6
7	8	9	10	11	12

$\}$

E = *in red*

$$P(E) = \frac{|E|}{|S|} = \frac{1}{11} = 0.\overline{09}$$

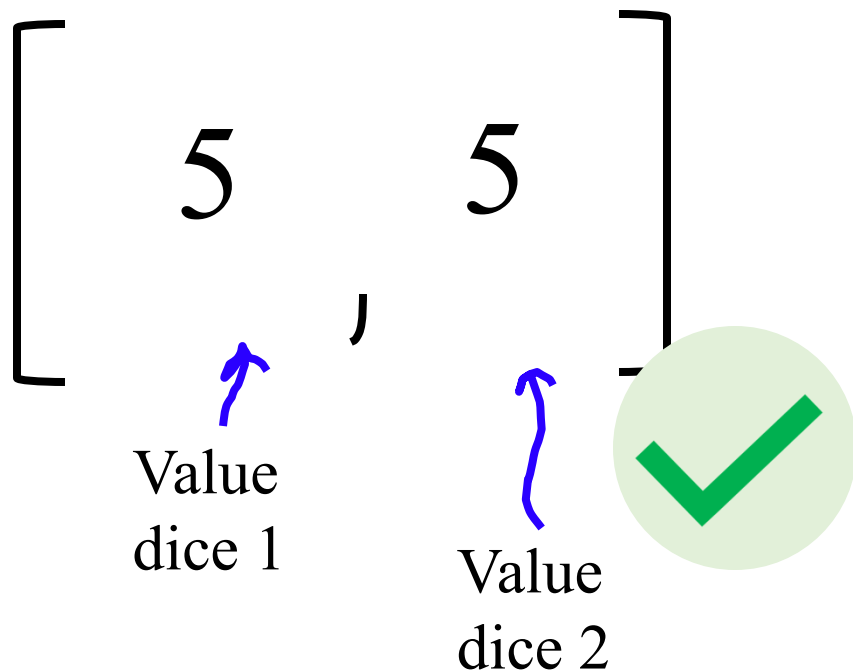
Sum of Two Die: Three options for the sample space

Value
dice 1

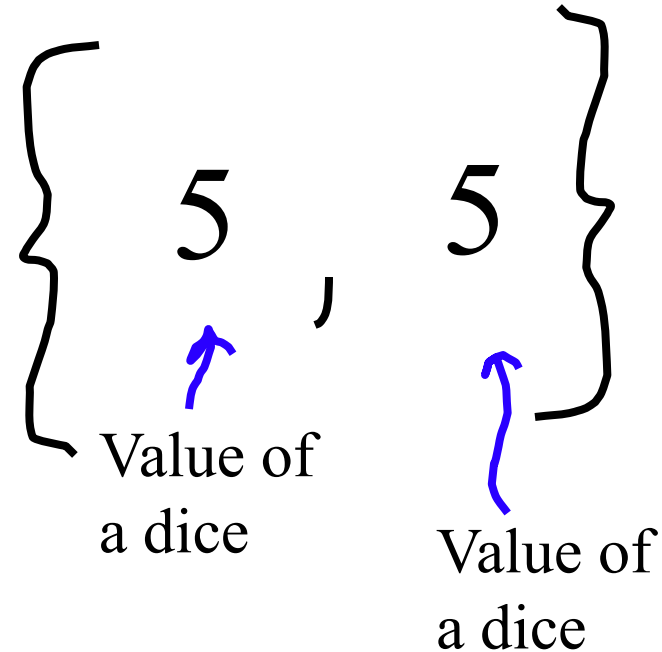


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**



Just look at the sum

10



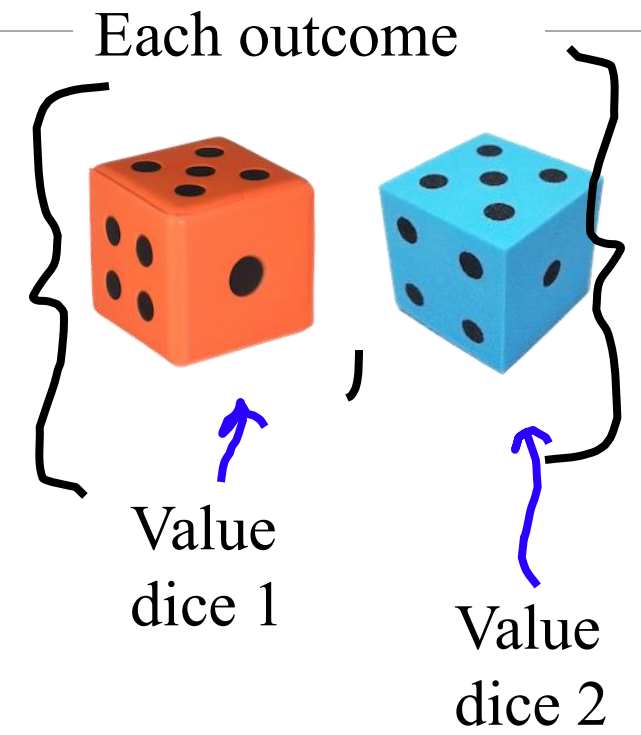
Sum of Two Die = 7? Bug: Die are Indistinct

Roll two 6-sided dice. What is $P(\text{sum} = 7)$?

$S = \{$

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

$\}$



$E = \textit{in blue}$

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

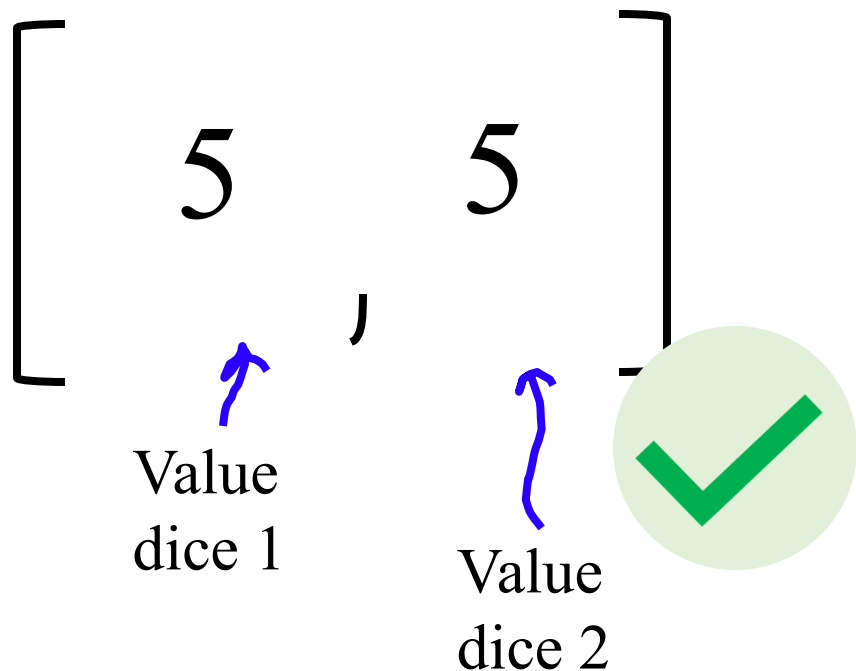
Sum of Two Die: Three options for the sample space

Value
dice 1

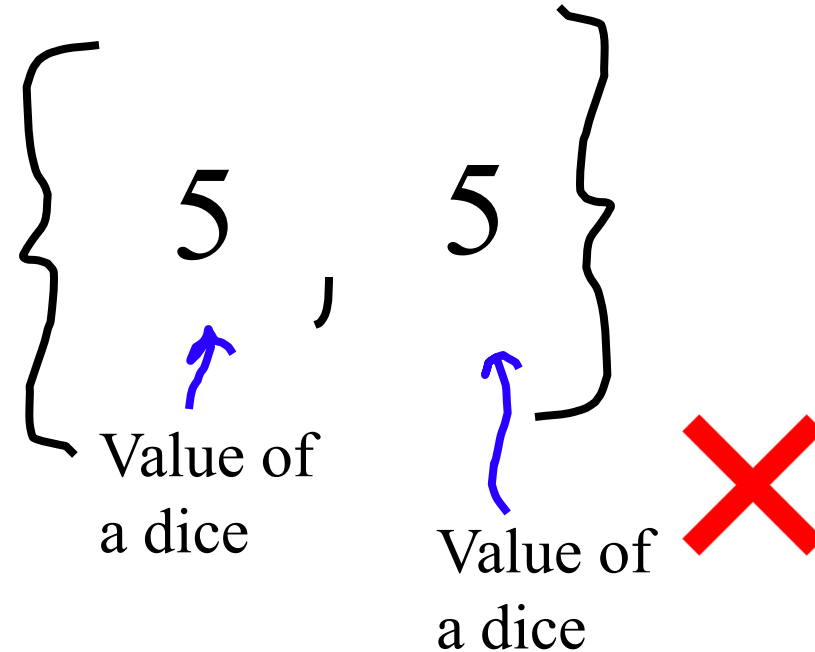


Value
dice 2

Think of the die as **distinct**



Think of the die as **indistinct**

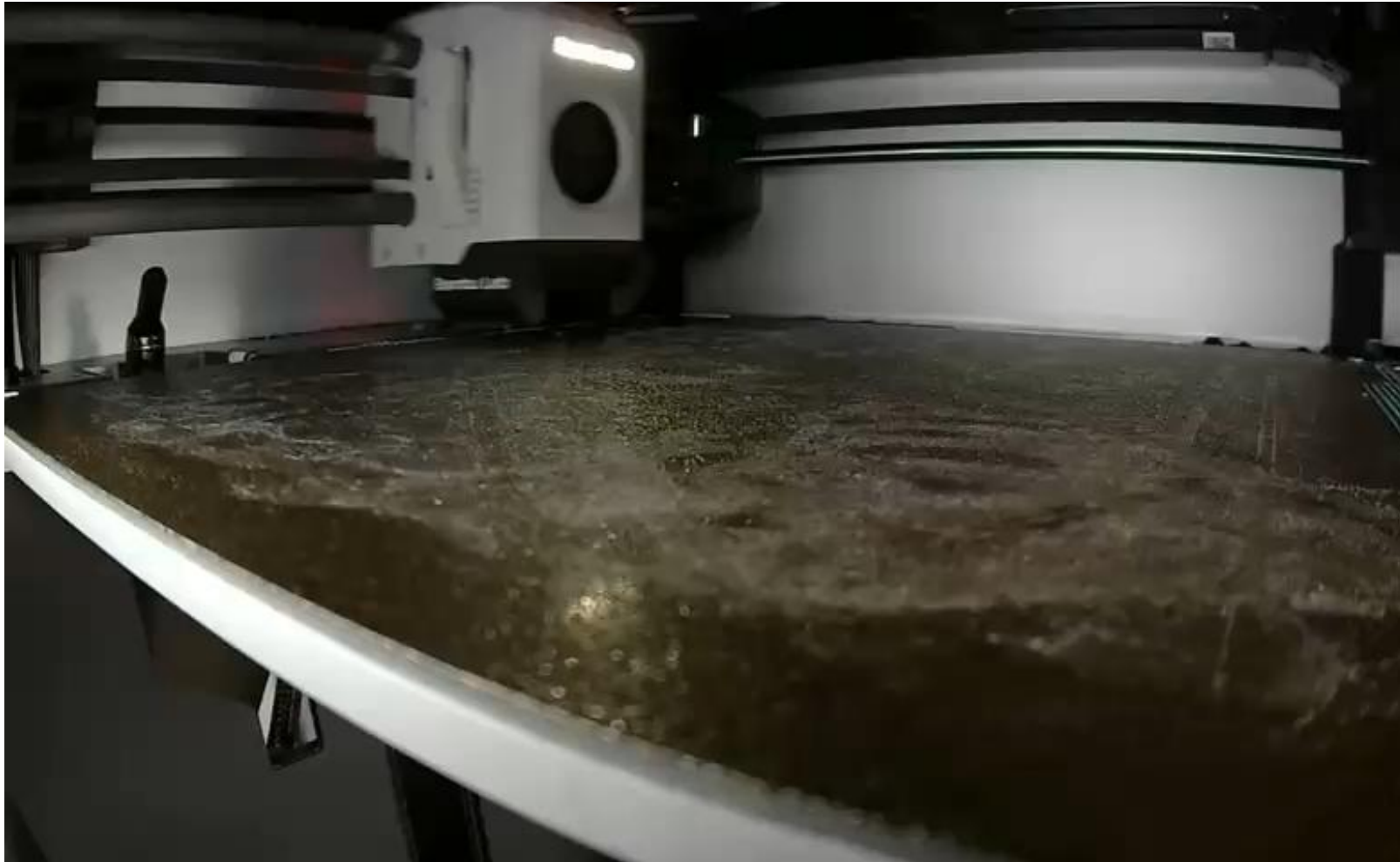


Just look at the sum

10

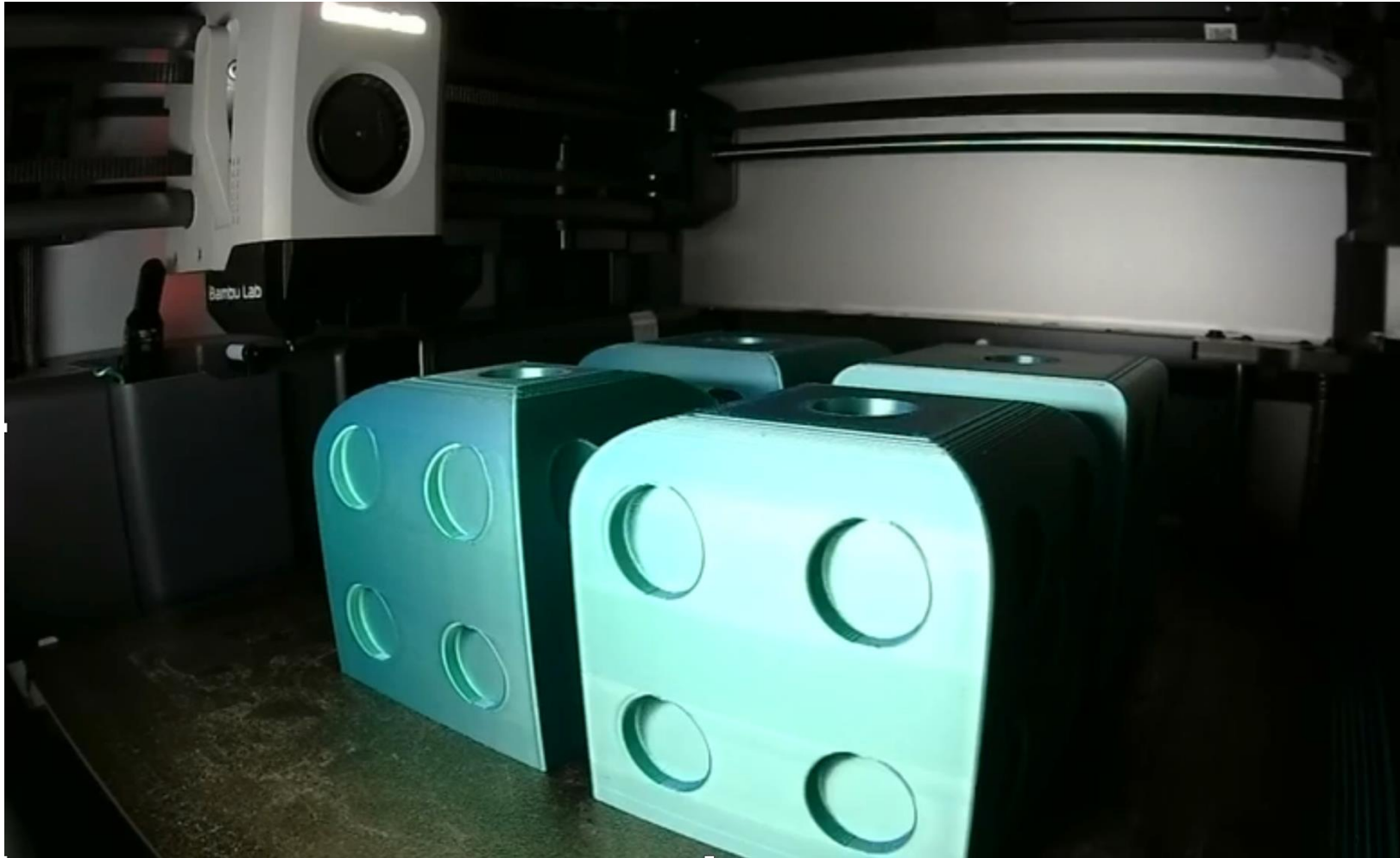


Sum of Two of These Die = 7?



I printed this last week

Sum of Two of These Die = 7?



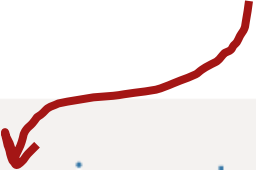
I printed this last week

Does an LLM understand
dice probability?



New API

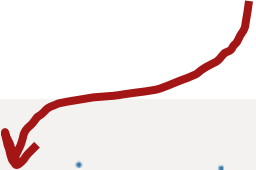
Llama-3.3-70B



```
from llm import string_pr
pr = string_pr("Hello, world")
print(pr)
```

Challenge for Tomorrow

Llama-3.3-70B



```
from llm import string_pr
pr = string_pr("Hello, world")
print(pr)
```

If you used a prompt like this:

```
f"Dice simulator output. Sum of two random 6 sided dice:
{outcome}"
```

Could you test an LLMs understanding of probability?

Challenge for Tomorrow

Llama-3.3-70B

```
from llm import string_pr
pr = string_pr("Hello, world")
print(pr)
```

If you used a prompt like this:

```
f"Dice simulator output. Sum of two random 6 sided dice:
{outcome}"
```

Could you test an LLMs understanding of probability?

Outcome	string_pr
2	3.0×10^{-30}
	⋮
7	9.9×10^{-29}
	⋮
12	2.4×10^{-30}

