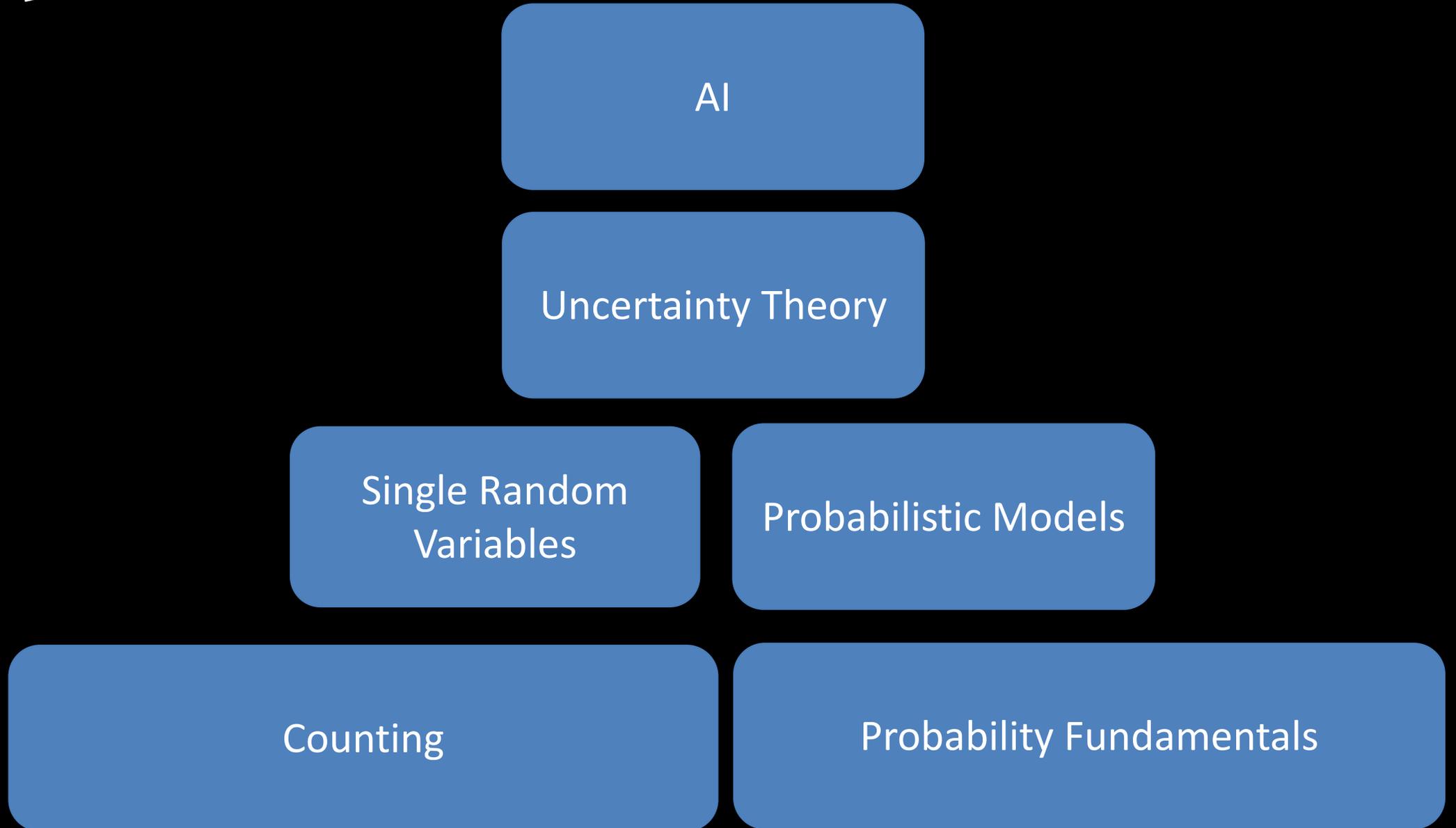


Future of Probability

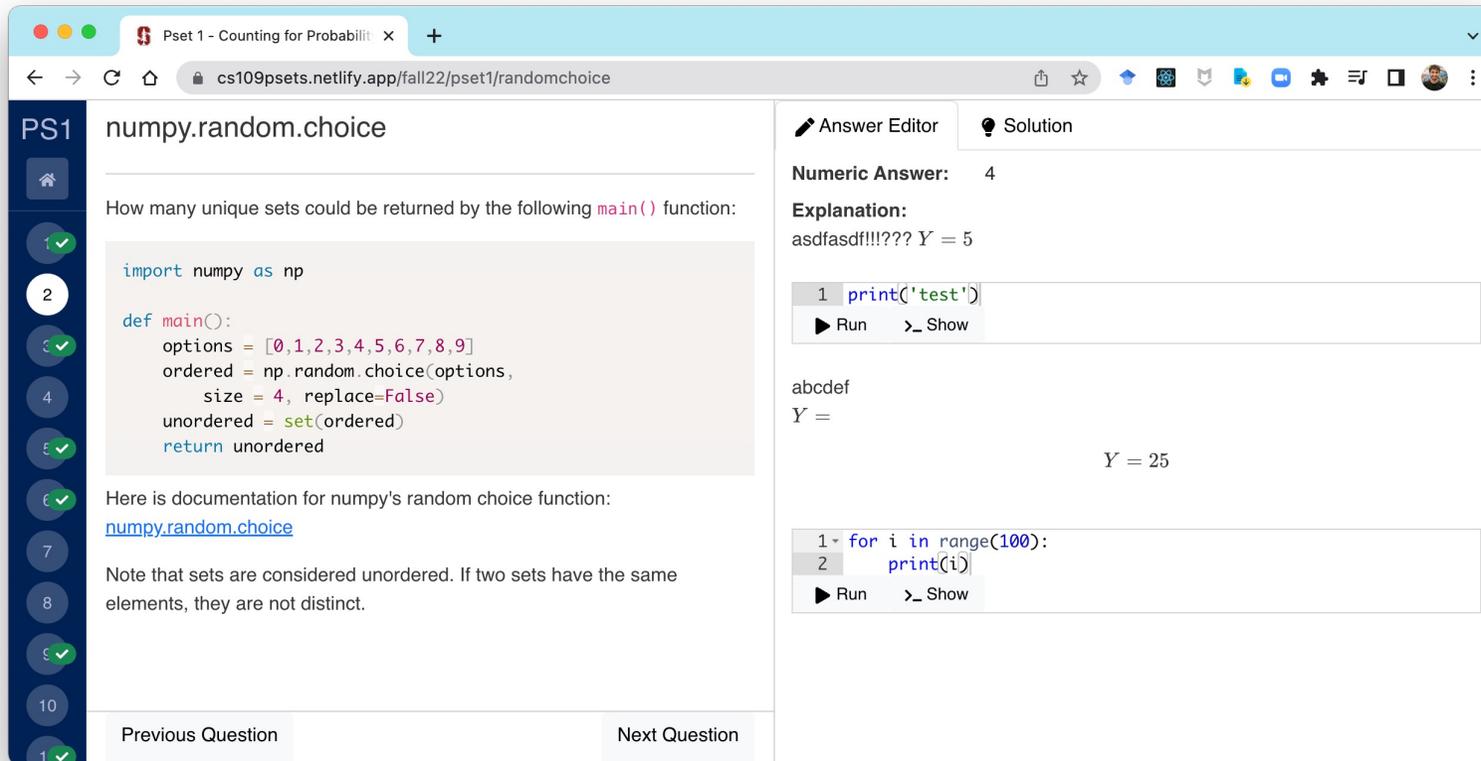
Chris Piech



Learning Goal: Abundance of important problems



Chose k examples from a dataset of lots of student work



The screenshot shows a web browser window with the URL `cs109psets.netlify.app/fall22/pset1/randomchoice`. The page title is "PS1 numpy.random.choice". The main content area contains the following text:

How many unique sets could be returned by the following `main()` function:

```
import numpy as np

def main():
    options = [0,1,2,3,4,5,6,7,8,9]
    ordered = np.random.choice(options,
                               size = 4, replace=False)
    unordered = set(ordered)
    return unordered
```

Here is documentation for numpy's random choice function: [numpy.random.choice](#)

Note that sets are considered unordered. If two sets have the same elements, they are not distinct.

Navigation buttons: Previous Question, Next Question

The right sidebar contains an "Answer Editor" and a "Solution" section. The "Answer Editor" shows a numeric answer of 4 and an explanation: "asdfasdf!!!! Y = 5". Below this is a code editor with the following code:

```
1 print('test')
```

The "Solution" section shows the output of the code: "abcdef" and "Y = 25". Below this is another code editor with the following code:

```
1 for i in range(100):
2   print(i)
```

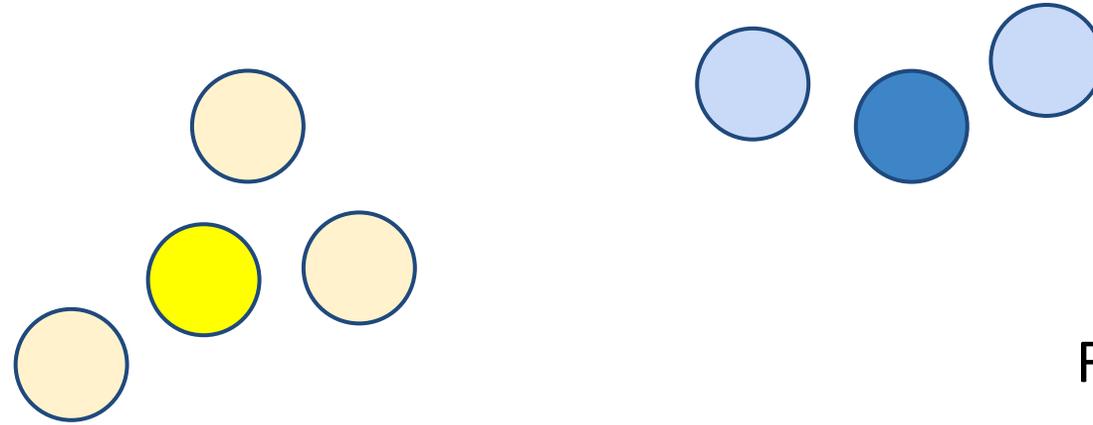
300 solutions

Find the 10 solutions which are most representative



K Medoids: A Classic Algorithm

Chose the k nodes such that the sum of minimized distances is as small as possible



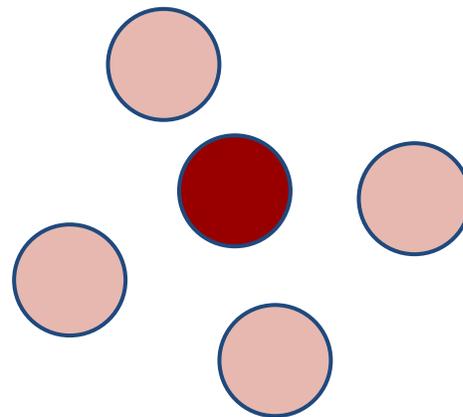
For any two points you can query the distance between them. May be non-Euclidean

Before

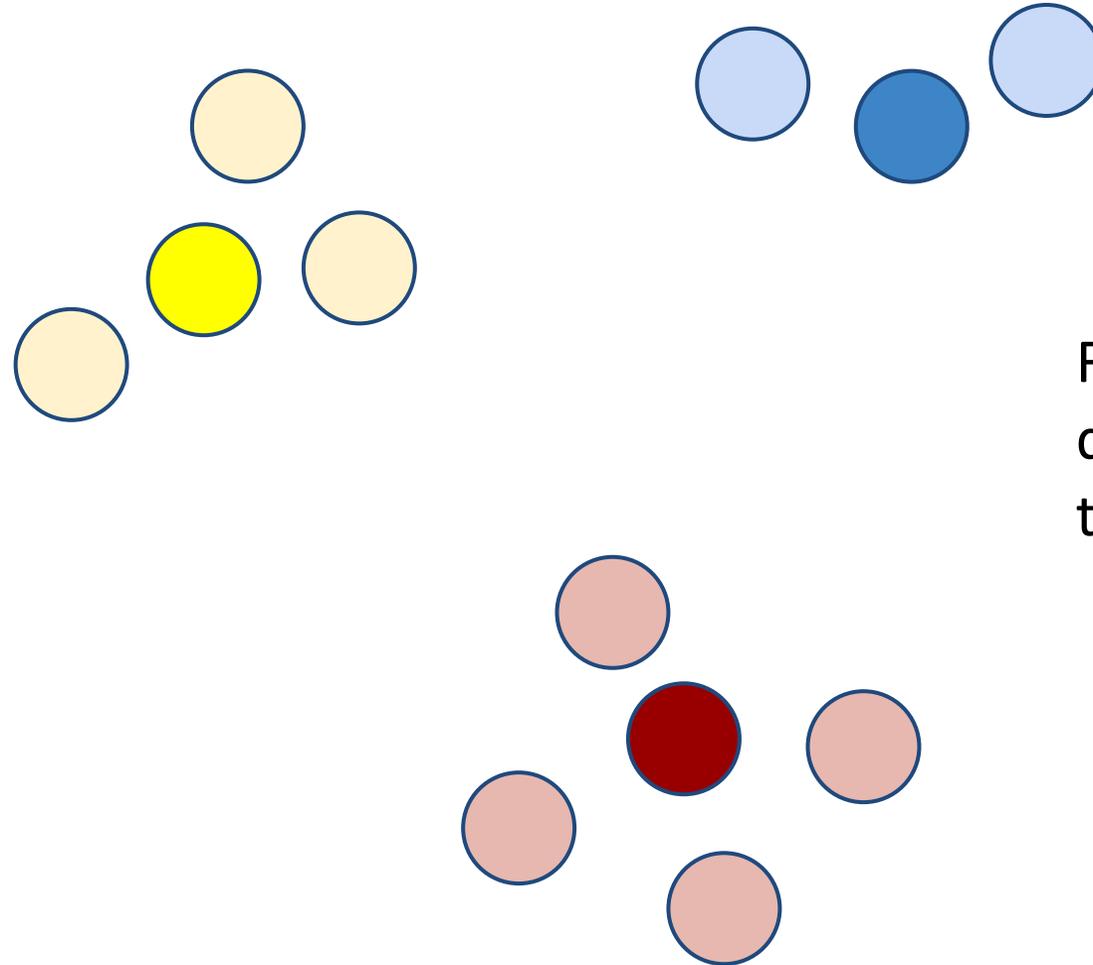
After

$$O(n^2)$$

$$O(n \log n)$$



Step 1: Chose the node closest to the rest (on average)



For any two points you can query the distance between them. May be non-Euclidean



The inner loop can be thought of as Thompson Sampling



3, 4



10, 11, 9, 8



12, 11, 10, 14



2, 0



Allowed us to revisit several core algorithms

- *k*-Medoids
 - **Mo Tiwari, ... Chris Piech.** "*BanditPAM: Almost Linear Time *k*-medoids Clustering via Multi-Armed Bandits*", NeurIPS 2020
- Random Forests
 - **Mo Tiwari, ... Chris Piech.** "*MABSplit: Faster Forest Training Using Multi-Armed Bandits*", NeurIPS 2022
- Maximum Inner Product Search
 - **Mo Tiwari, ... Chris Piech.** "Faster Inner Product Search in High Dimensions", Submitted to AAAI 2023 (favorable reviews)



Today

Digital Future of Probability

Themes to look out for:

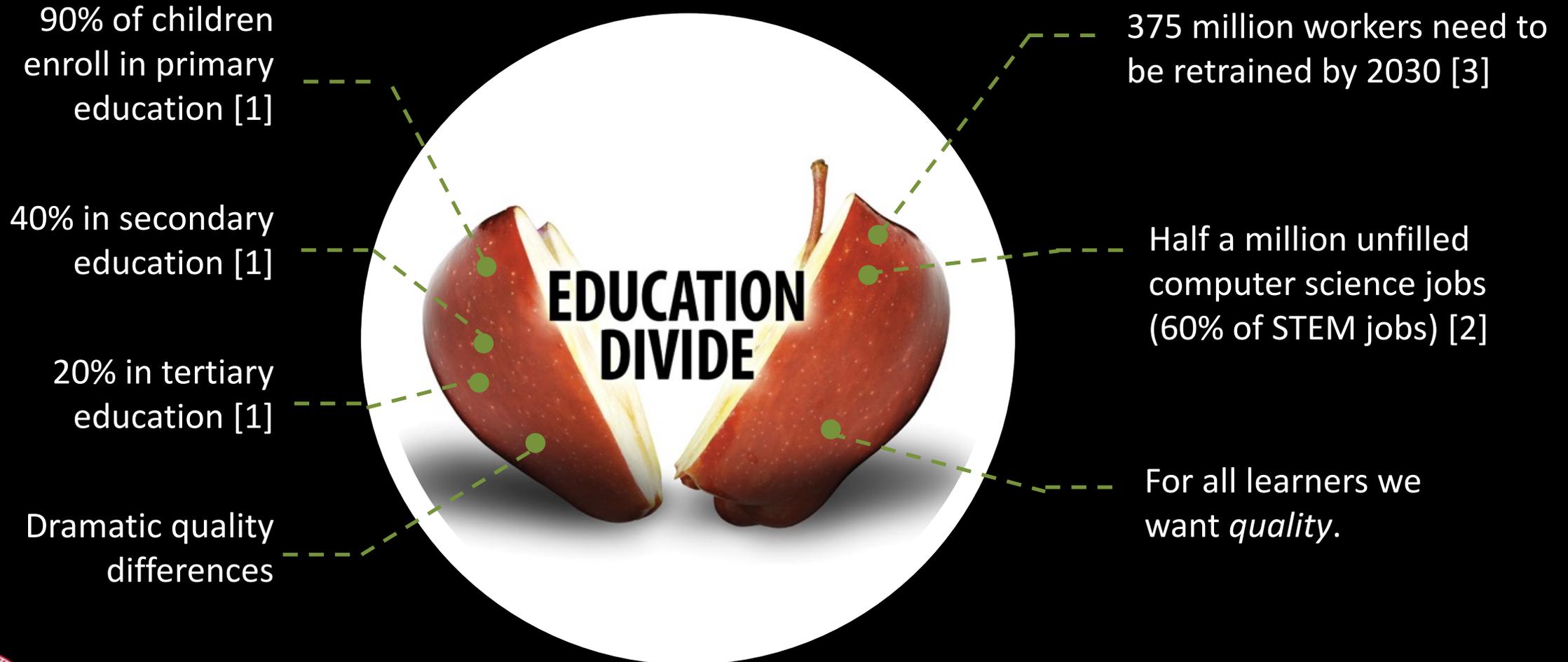
Probabilistic Modelling vs Deep Learning
Distributions vs Point Estimates

Let me tell you a story about one particular problem.

It will give you a sense of what it takes to do research in **applied probability**, and our desire to solve the problems, will lead us to deep **theoretical** challenges in modern AI.

Application -> Theory

Quality Education Gap



[1] *World development indicators 2015*. World Bank Publications, 2015.

[2] USA Bureau of Labor Statistics Employment Projections , 2016.

[3] Jobs lost, jobs gained. McKinsey Global Institute, 2017.



Smart Phone Access

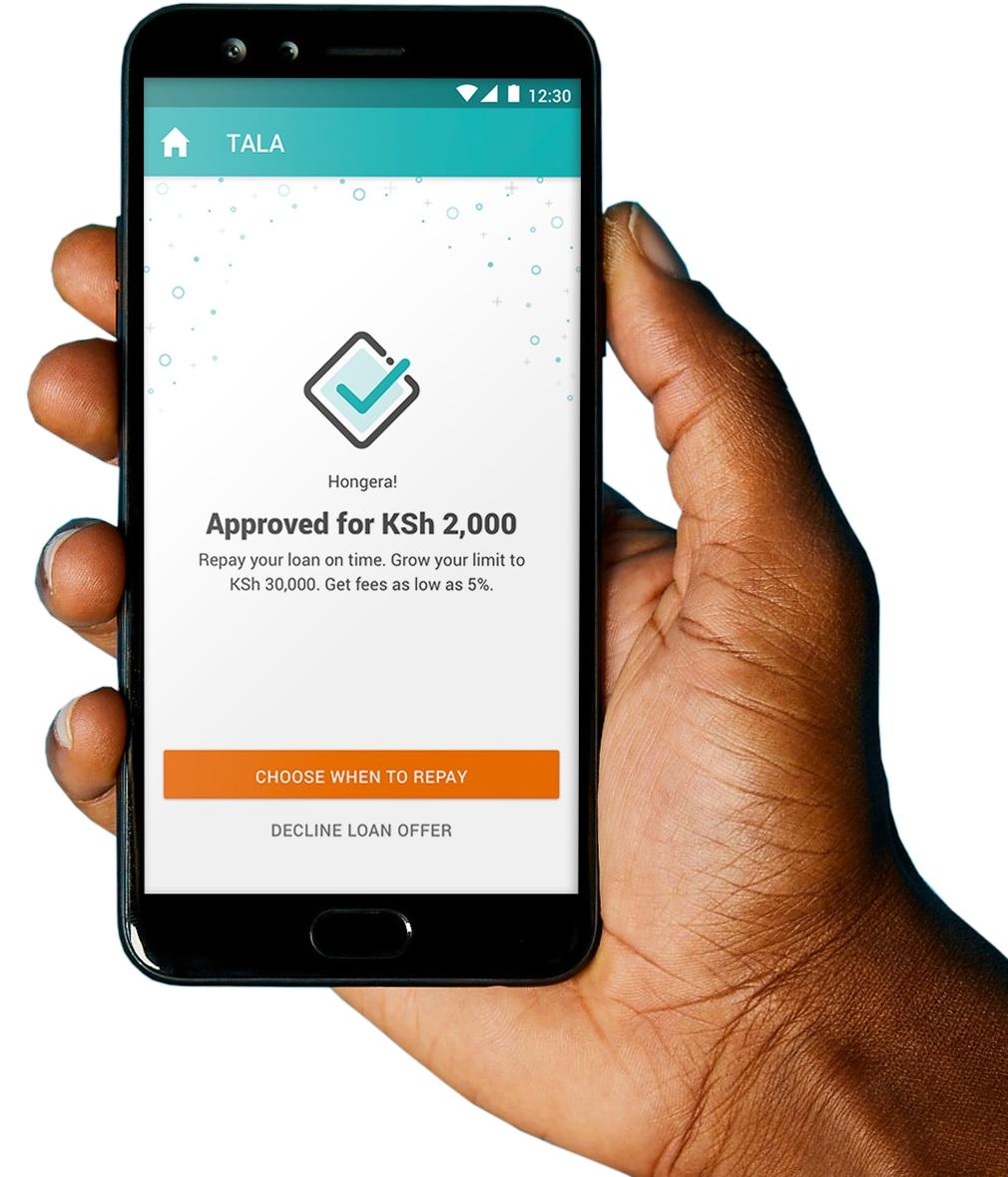
Advanced Economies



Emerging Economies

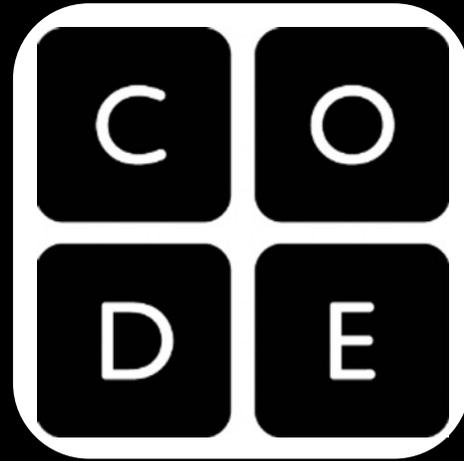


- Smartphone
- Mobile
- No phone

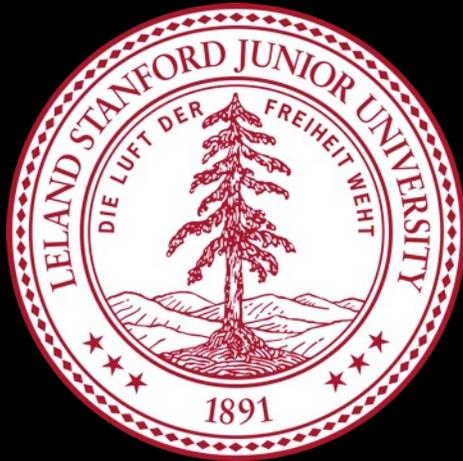


Unprecedented Data

coursera



edX

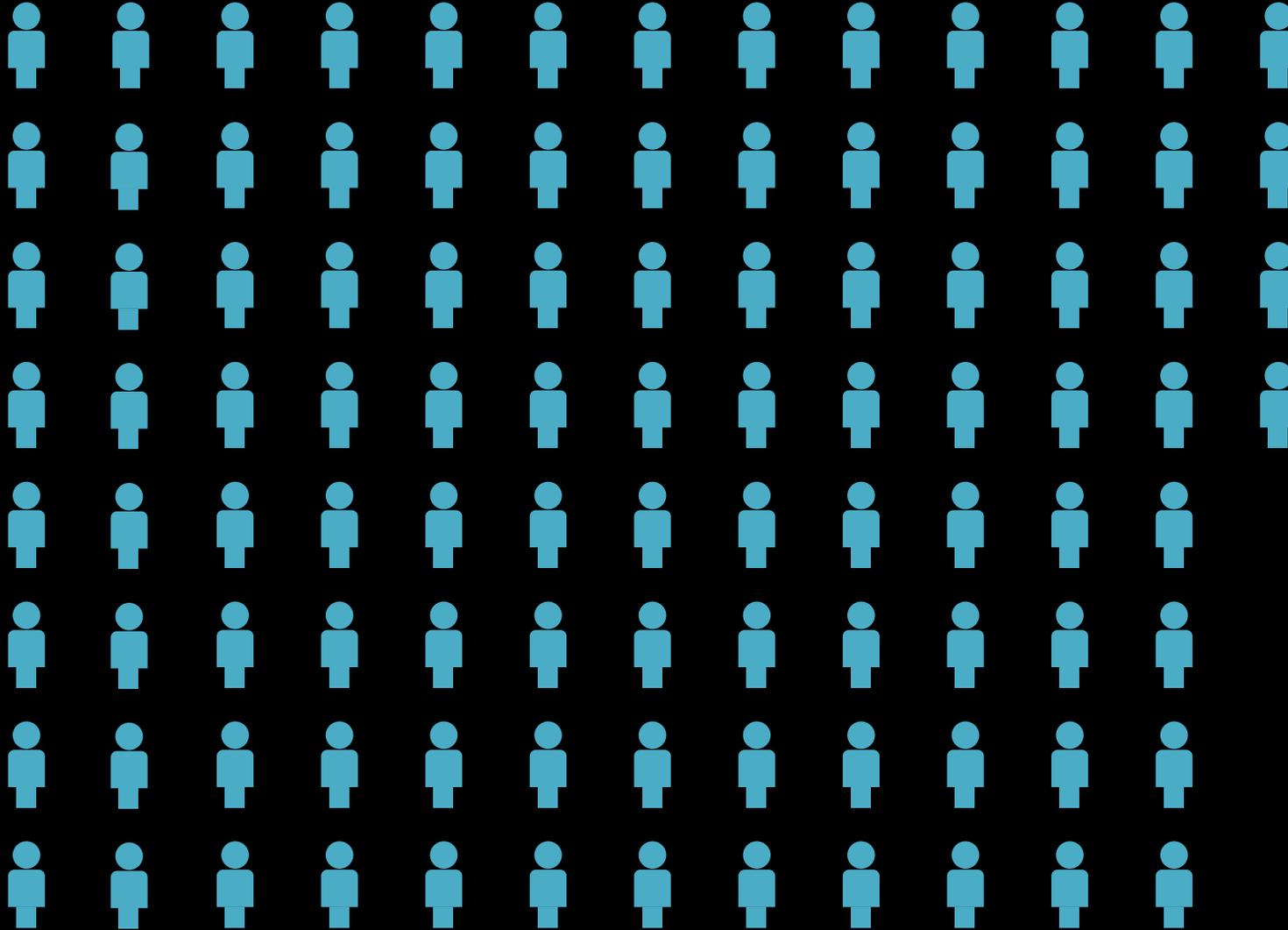


**U
UDACITY**

Over 50 million learners



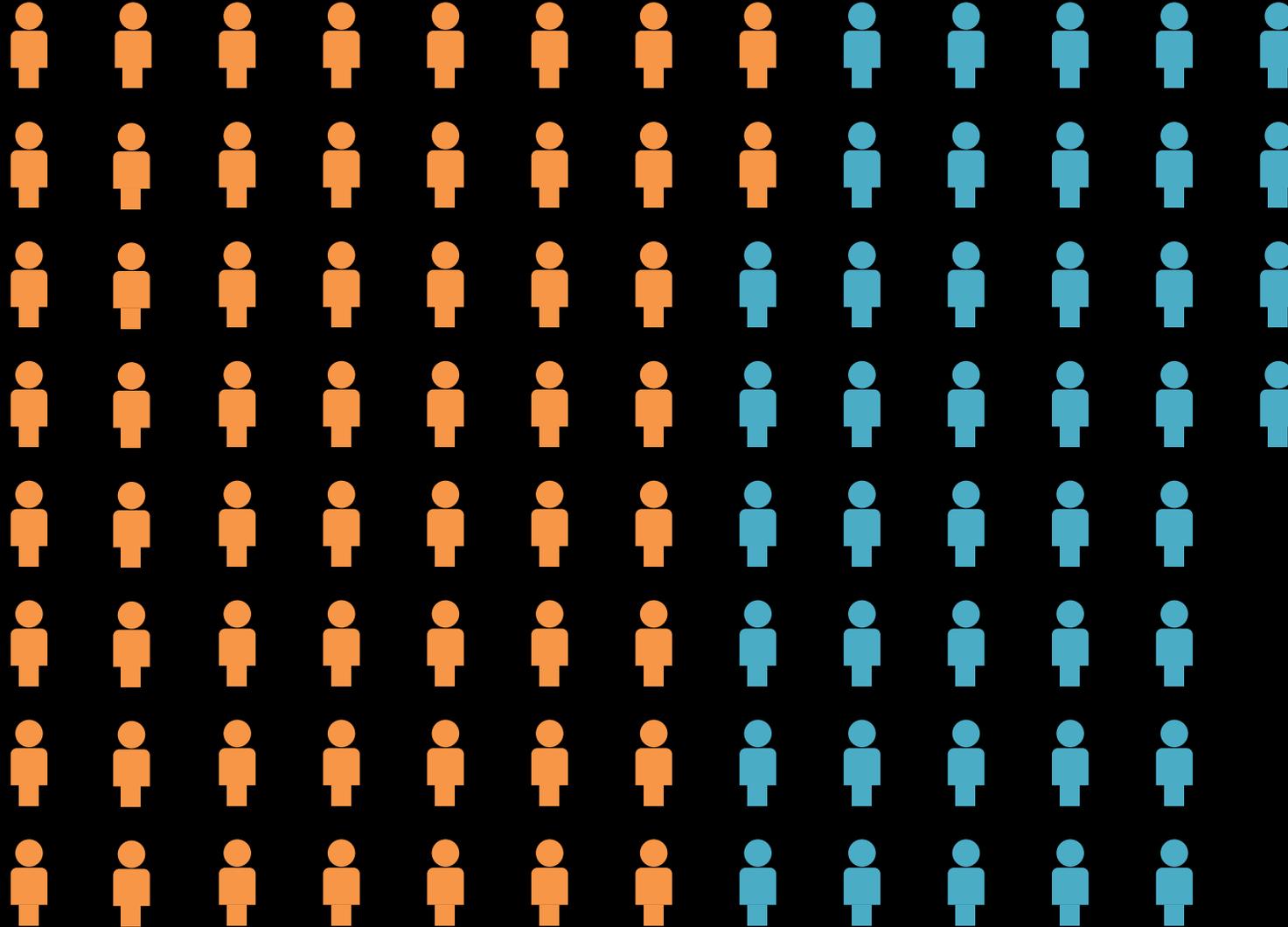
US K-12 Students



= 500,000 learners



Code.org Students



= 500,000 learners



Code.org by the numbers

1,234,127 teachers

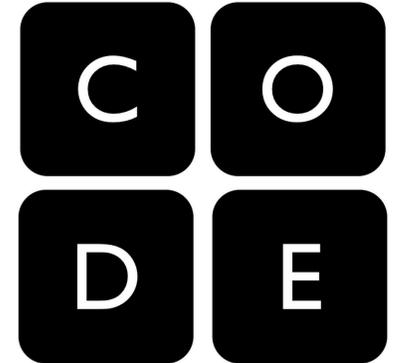
42M unique enrolled students

Used in **180+** countries

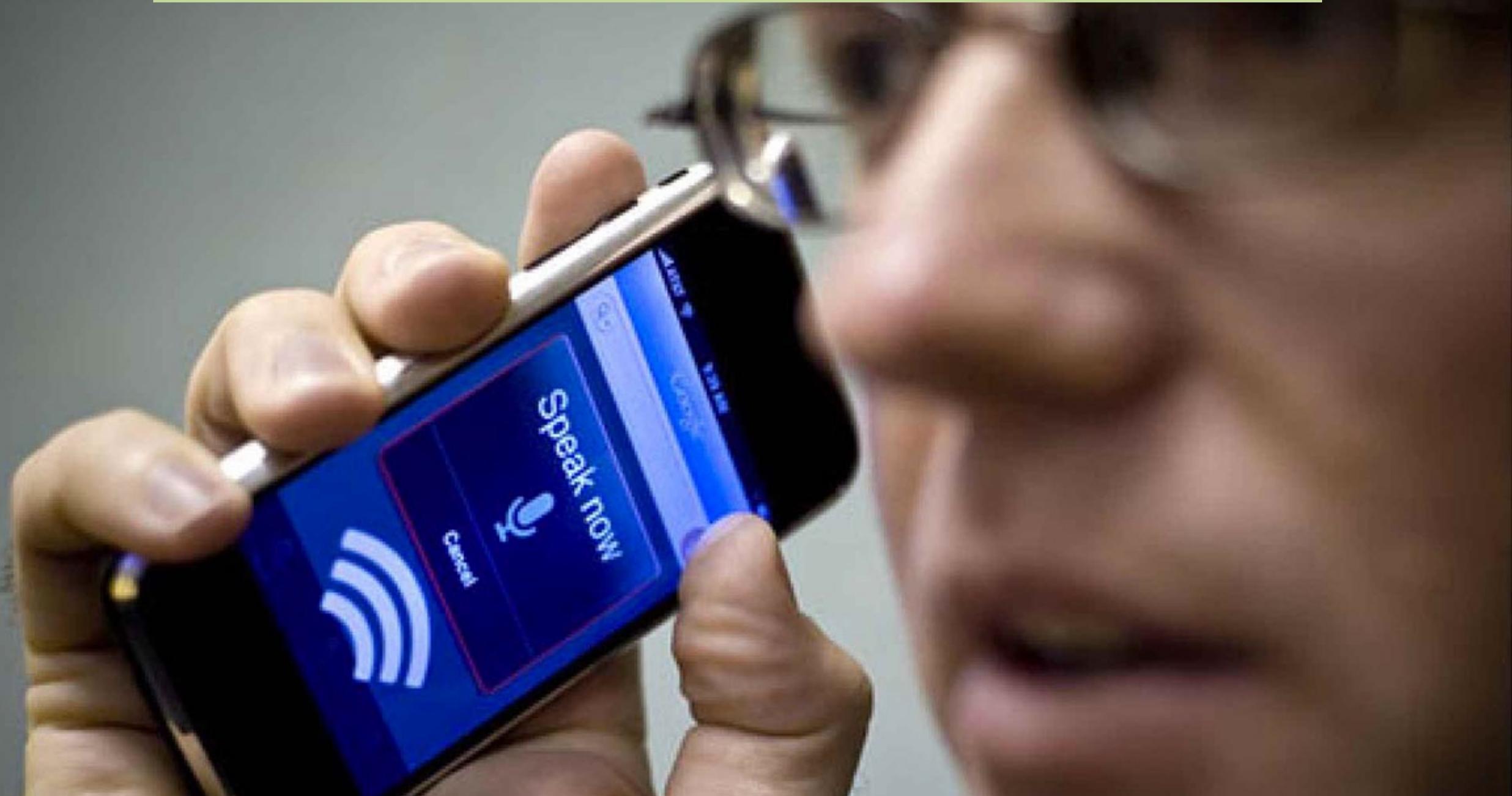
832M hour of code sessions

4 papers publish with our lab

50M K12 students in the US



Speech Recognition



Clear
Societal Need

Scale
education

Grand Challenge
in Education

Autonomously support
education by better
understanding students.

Online
assignments

Deep
Learning

New
Datasets of
Learning

AI
Renaissance



Feedback is Labor Intensive



Online classes have not solved the feedback problem [1].



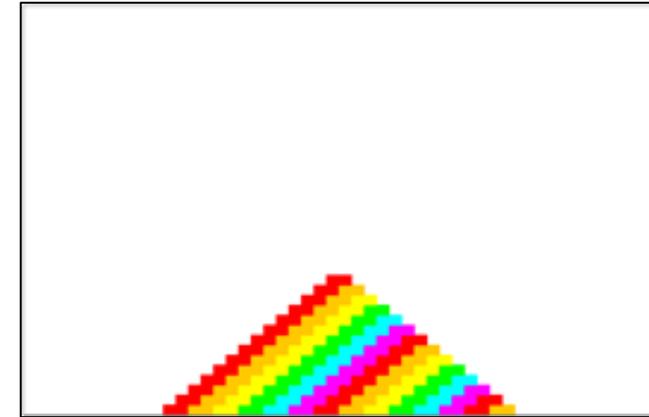
[1] *Deconstructing Disengagement. Analysing learner subpopulations in MOOCs.* Kizilcec, Piech, Schneider. Over 600 citations since 2013



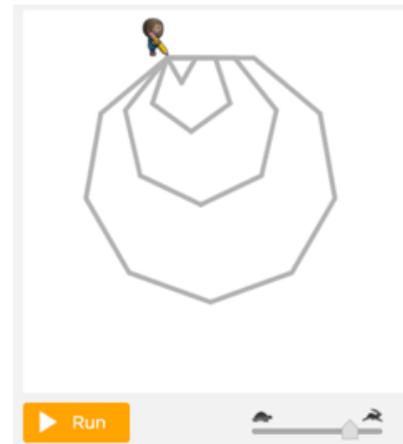
Many domains of student work

$$\begin{aligned} & 9 \times 6 \\ &= (10 - \boxed{}) \times 6 \\ &= 10 \times 6 - \boxed{} \times 6 \\ &= 60 - \boxed{} \\ &= \boxed{} \end{aligned}$$

```
when run
repeat until [acorn]
do
  if path ahead
  do
    turn left
  else
    move forward
```



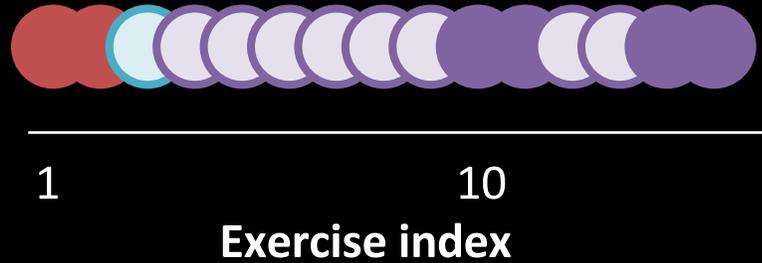
Why did the original pilgrims come to America?



Chapter 0: Always start simple

First deep learning for education

KHAN
Student



Exercise Type:

Answer:



Solving for x-intercept



Correct



Solving for y-intercept



Incorrect



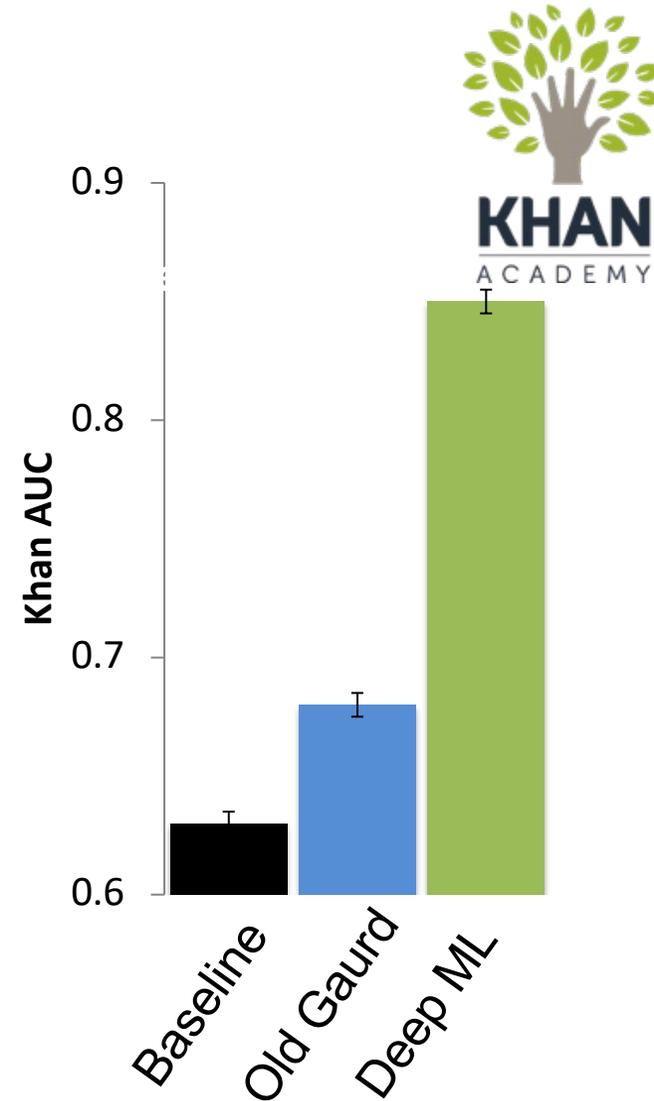
Graphing linear equations



Square roots



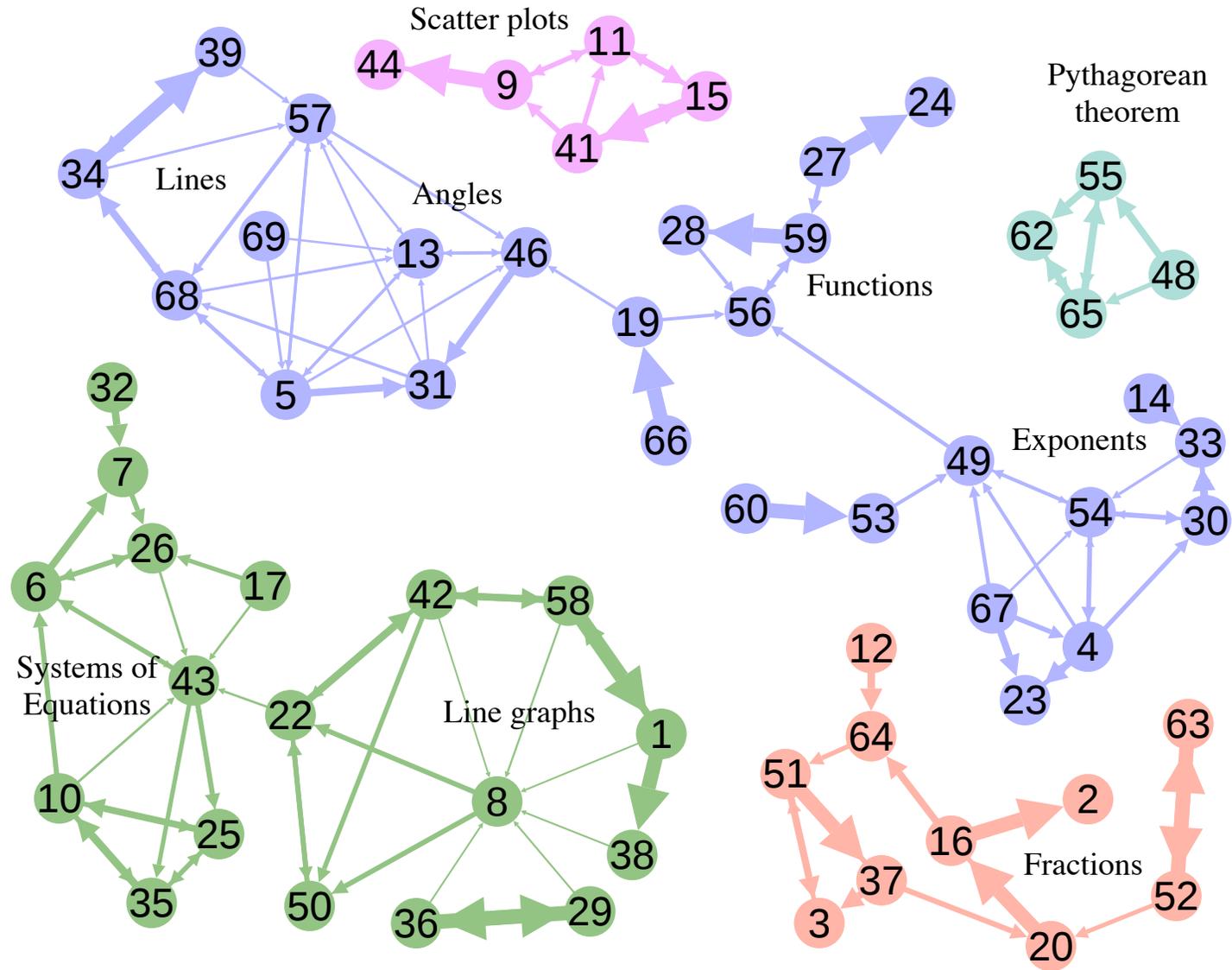
Slope of a line



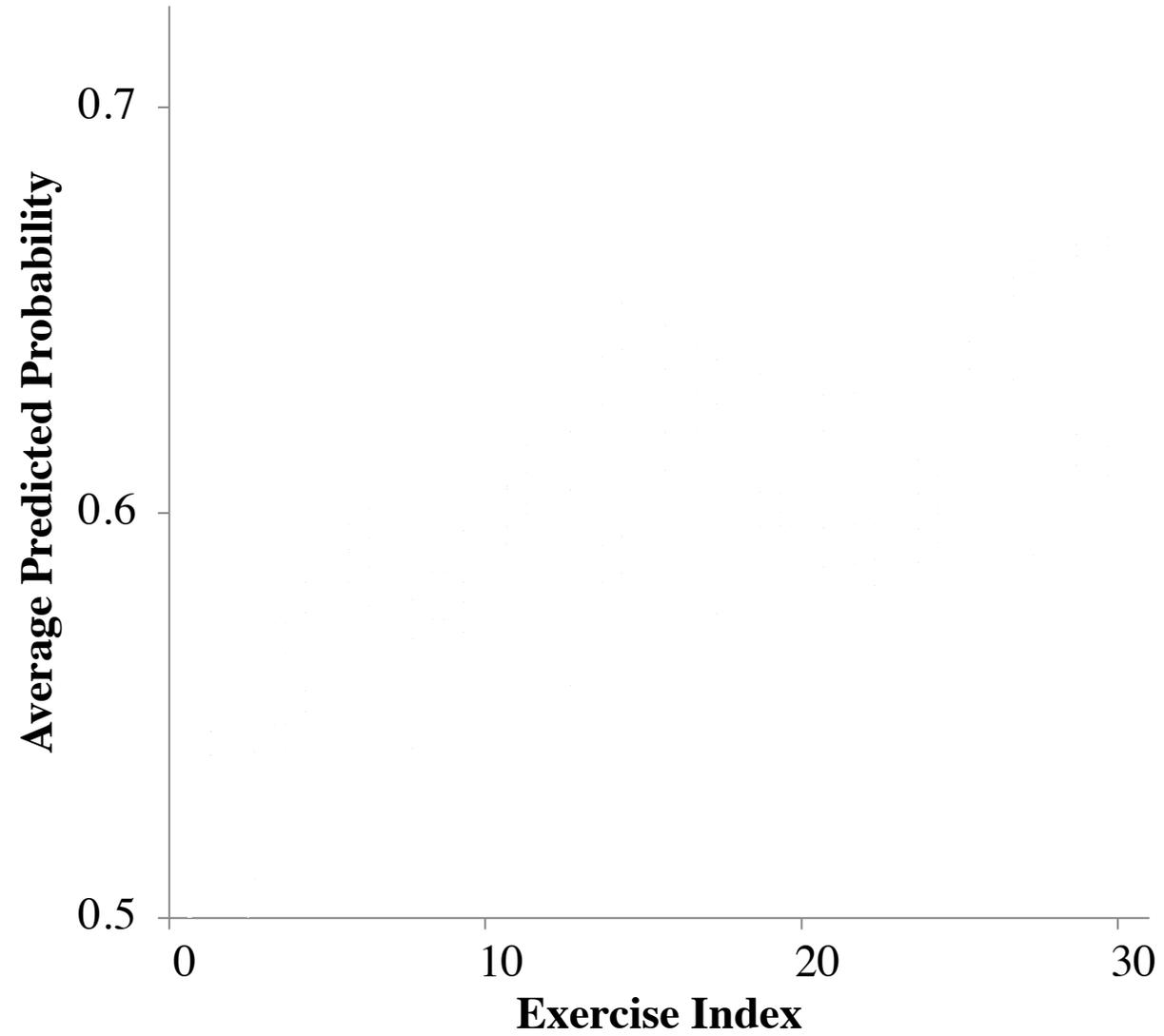
Old Problem



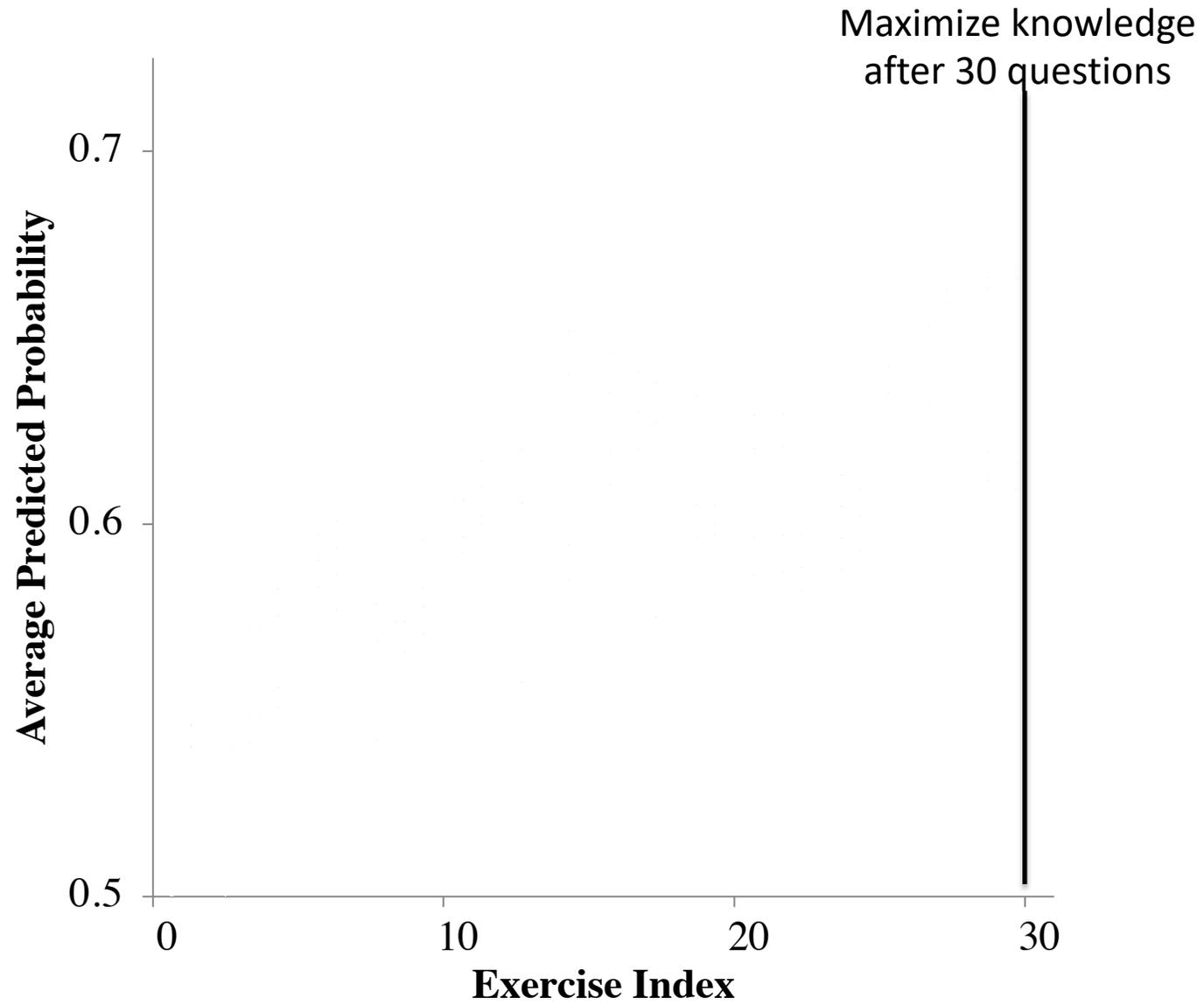
Learns Concept Relationships



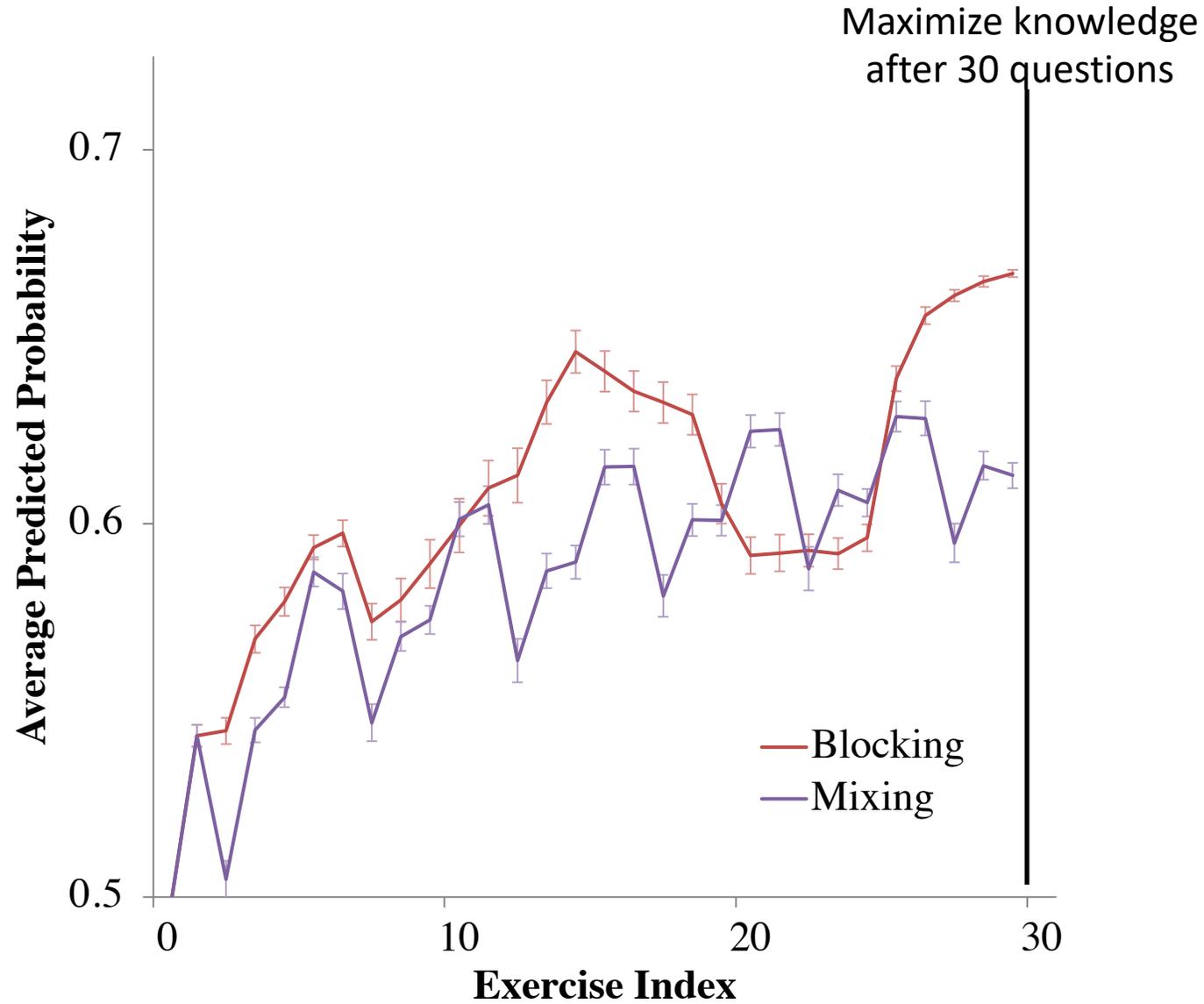
Optimal Teaching



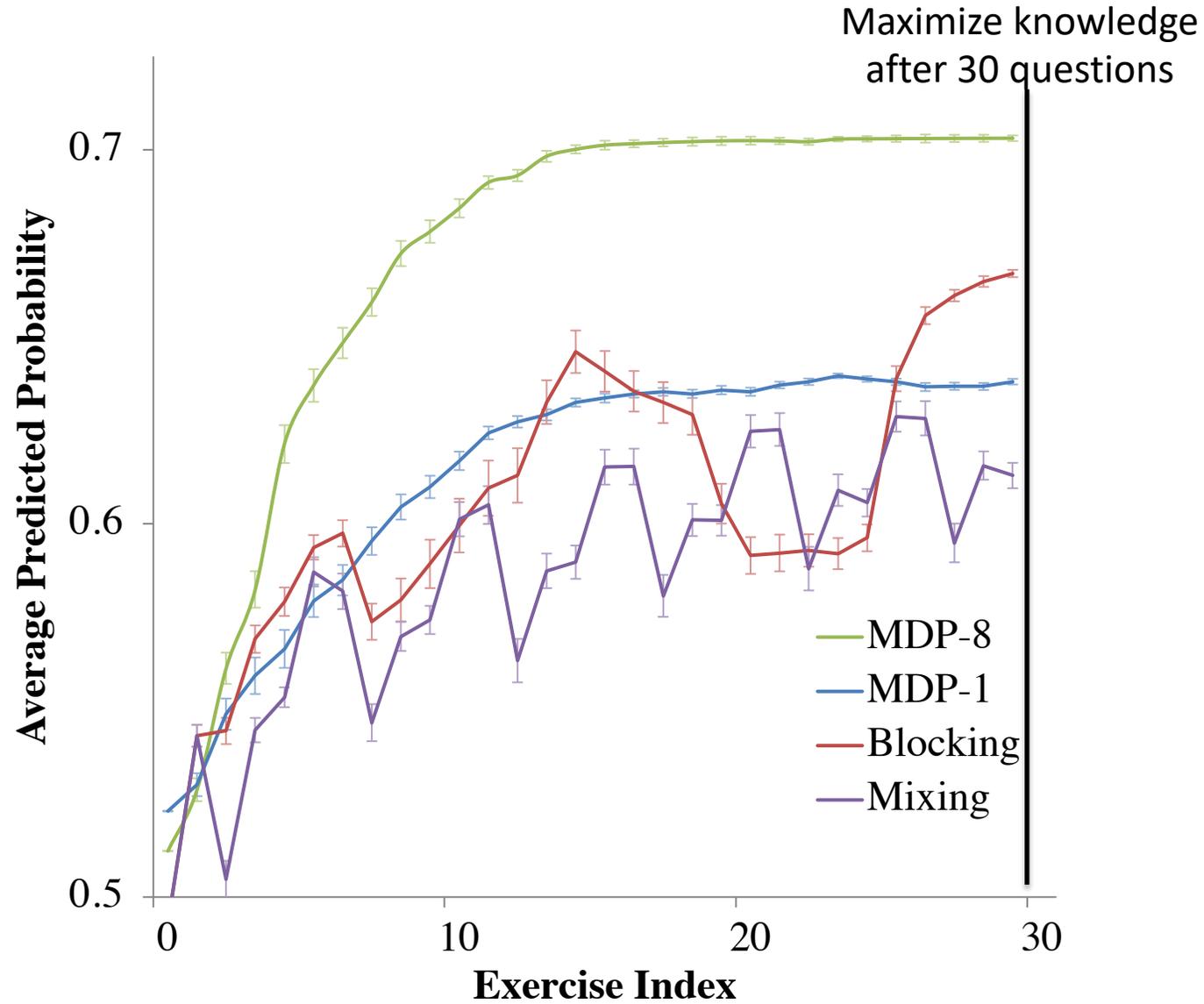
Optimal Teaching



Optimal Teaching



Optimal Teaching



We truly would rather move beyond
correct / incorrect



Some domains are very hard

Introduction to Python Upgrade to Pro

script.py

```
1 my_name = "Codecademy"  
2 print("Hello and welcome " + my_name + "!"  
3
```

File "script.py", line 3

SyntaxError: unexpected EOF while parsing

Hi! I'm Elsa of Arendelle. Help me create a single line.

Not quite. You have to use a block you aren't using yet.

Less

Blocks workspace: 3 / 2 blocks Start Over Show Code

- move forward by 100 pixels
- turn right by 90 degrees
- turn left by 90 degrees
- when run
 - turn left by 90 degrees
 - turn left by 90 degrees



Can you understand this code?

Top Secret

Chris

Not Secure

Question Solution

Instructions

- If there are many moves, focus on the first one
- Random code strategy is for when the student seems to be trying things randomly
- Lookout for students who don't get nesting or pre/post conditions. Often extra blocks in a body is an indication that they don't get that the post of the loop has to match the precondition

Question



```
import code.org.*;

public class MySoln {
    public void run() {
        move(50);
        for(int i=0; i<4; i++){
            if(frontIsClear()) {
                turnLeft(90);
            }
            for(int j=0; j<i; i++){
                move(i * 20);
                turnRight(120);
                move(10);
            }
        }
    }
}
```

Student 0

Label Console

✓ Num Done: 8273

Strategy

- Beeper Boundry (most people do this)
- Triangle Strategy
- Recursive Strategy

Looping

- Correct use of looping
- Doesn't use a while
- Doesn't have correct stop condition

Stanford TAs label 800 submissions

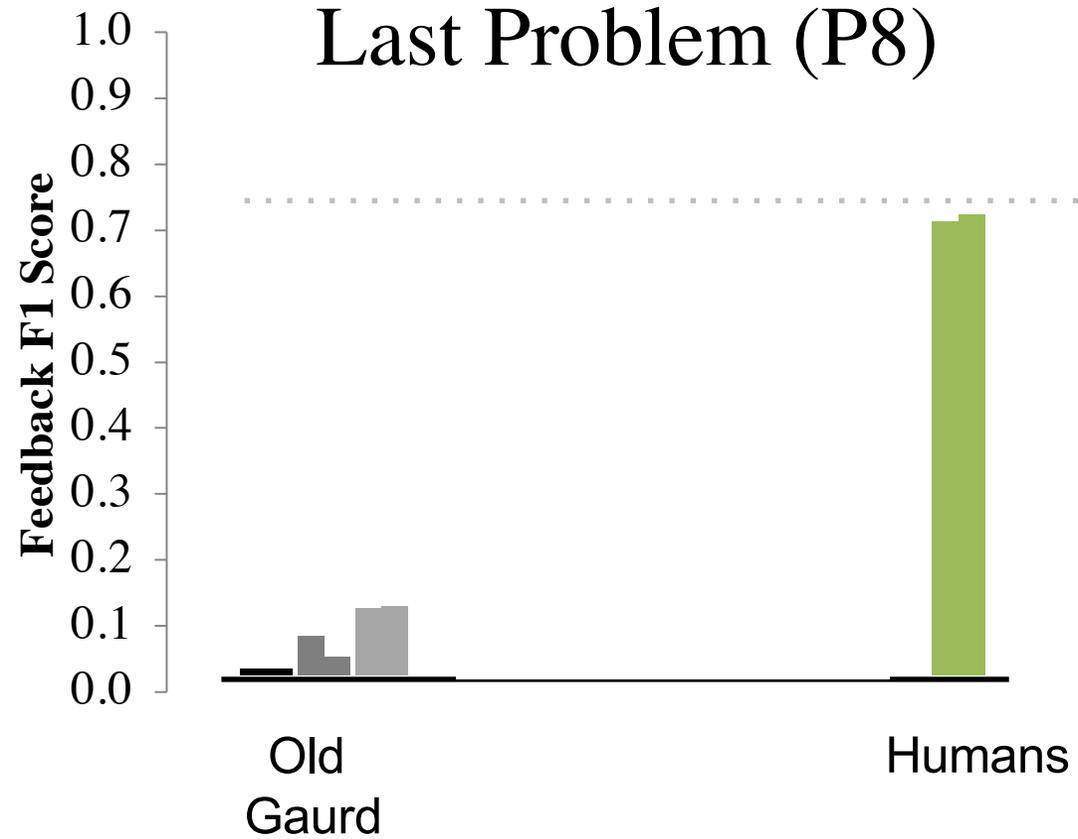
- Loop post condition doesn't match precondition
- Repetition of bodies

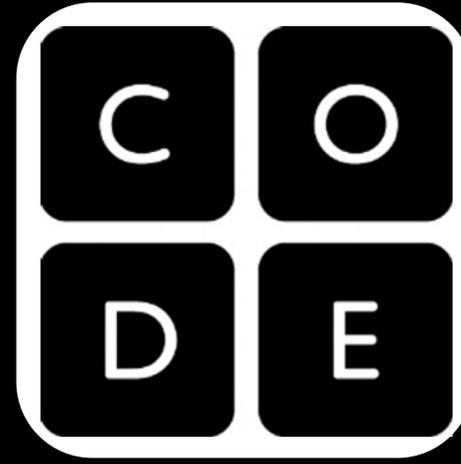
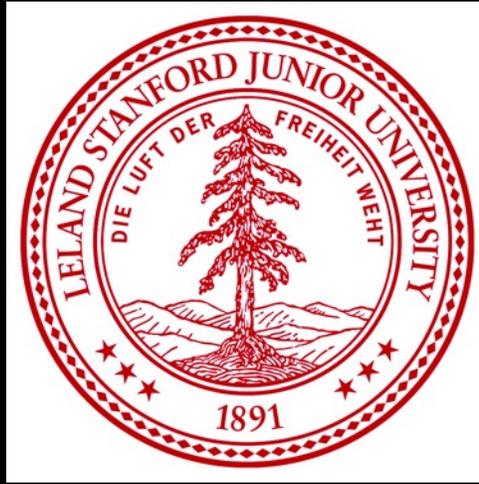
Cleanup

Record label

Traditional Deep Learning Doesn't Work

Label student code





Can we provide feedback
by dynamic analysis?



Starter code

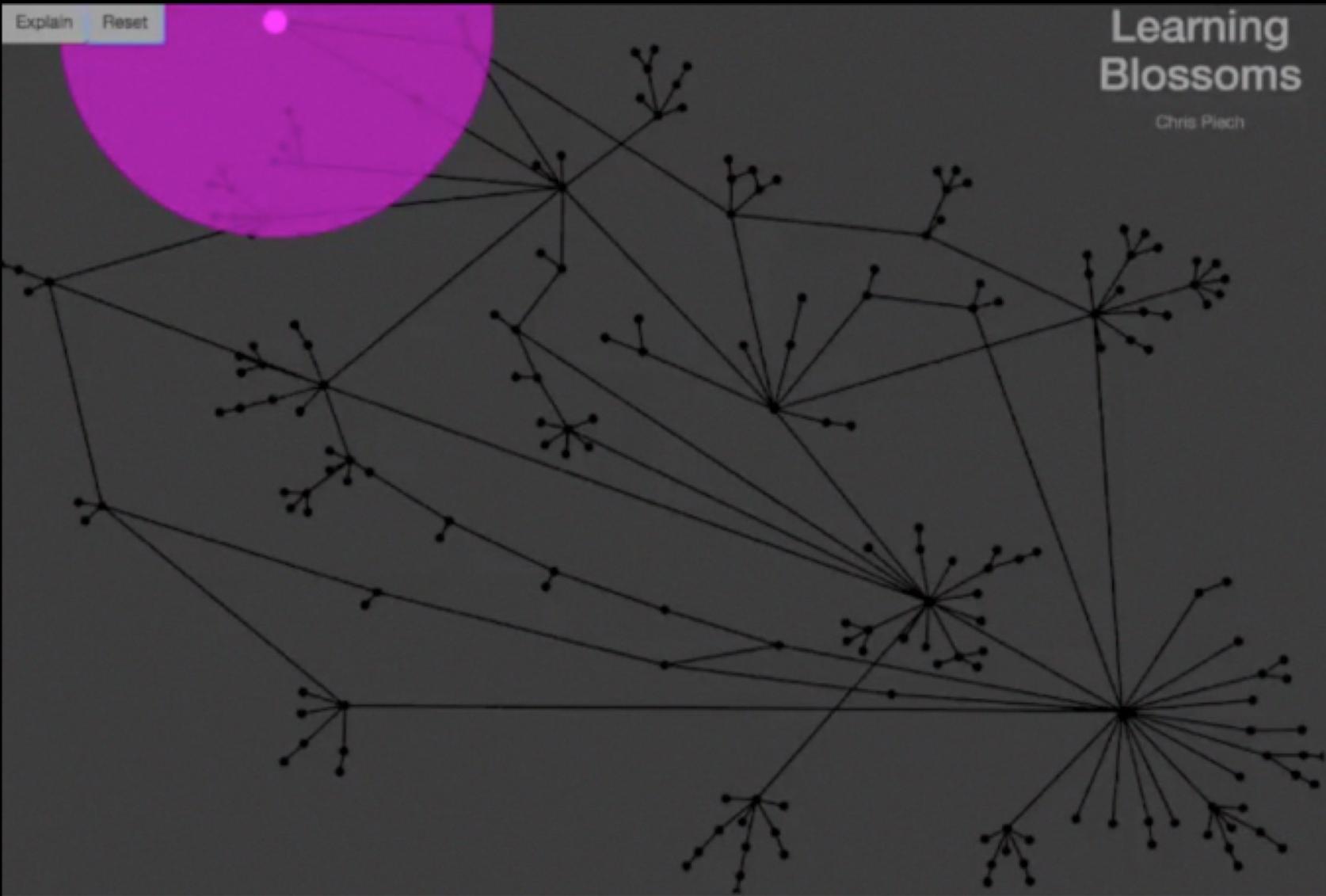
First attempt

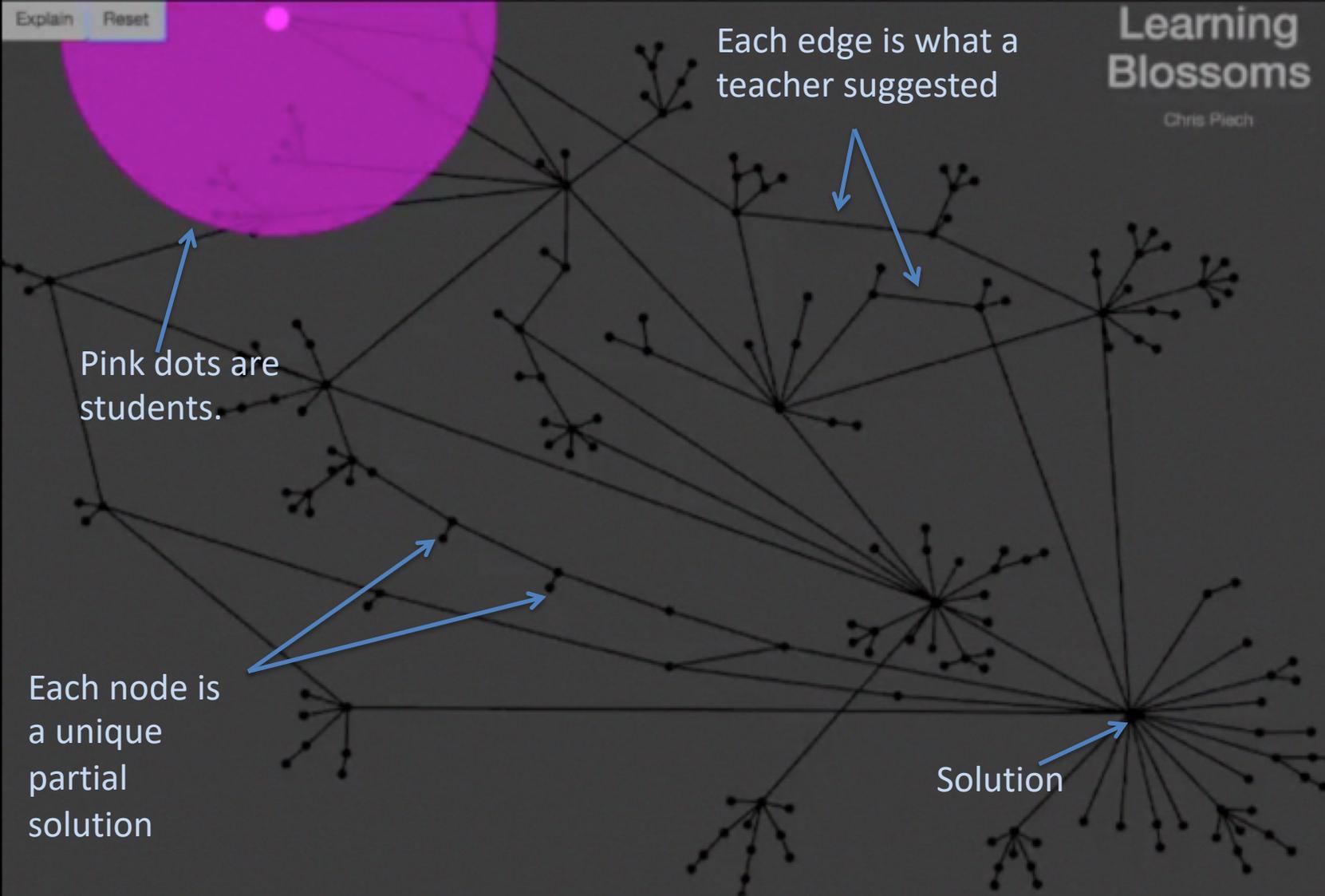
...

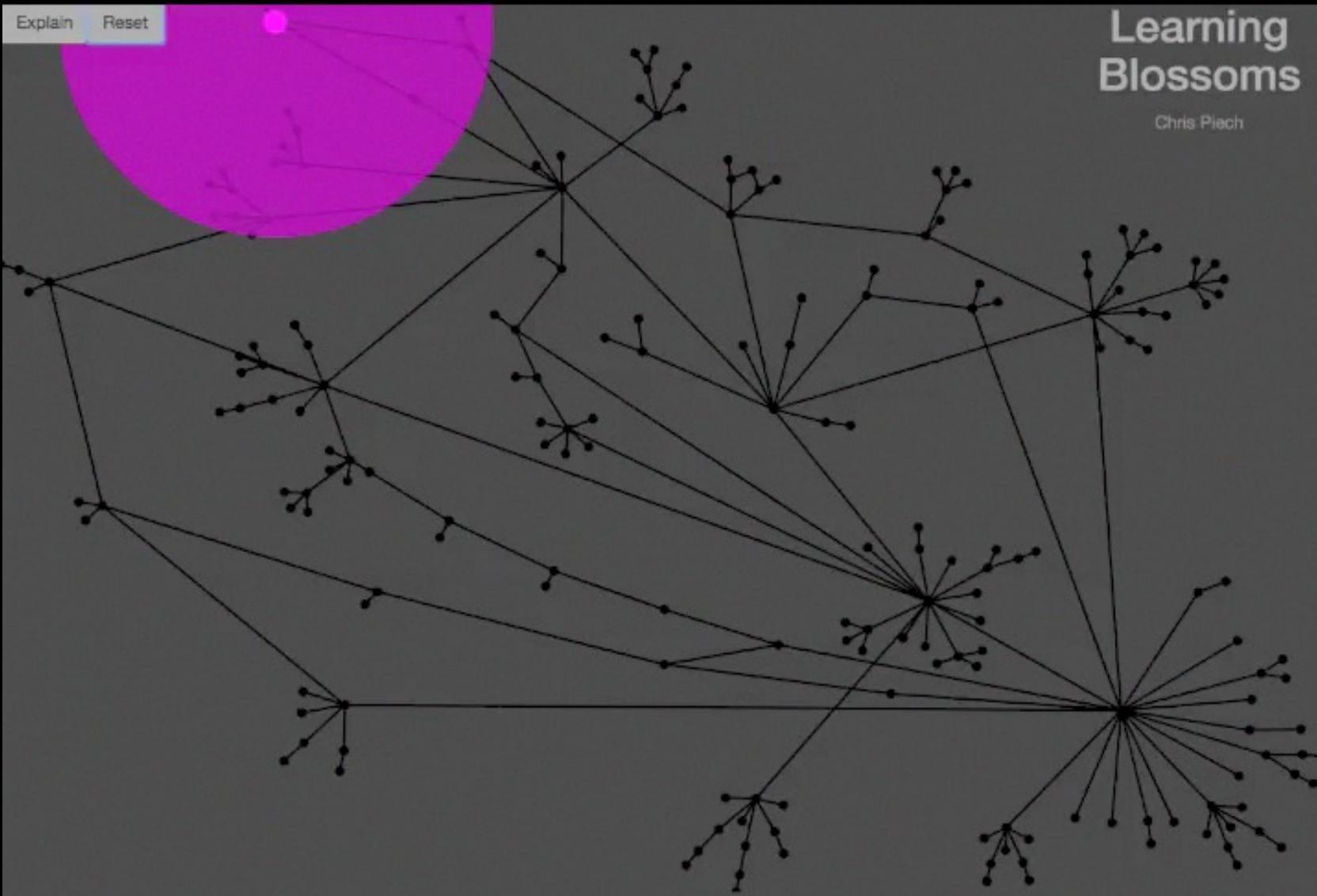
Final solution



Chapter 1: Better data source?







The Crowd is Un-wise

Temporal methods tried:
Shortest path
Min Time
Expected Success
Reinforcement learning
Most Common Next
Most Popular Path



18%



45%



12%



Desirable Path Algorithms

Poisson Common Path

$\gamma(s) =$ First step in the *most frequent* path to the solution from s , taken by *average* students. Assume poison process.

Predicted next
partial solution



Desirable Path Algorithms

Poisson Common Path

$$\gamma(s) = \arg \min_{p \in Z(s)} \sum_{x \in p} \frac{1}{\lambda_x}$$

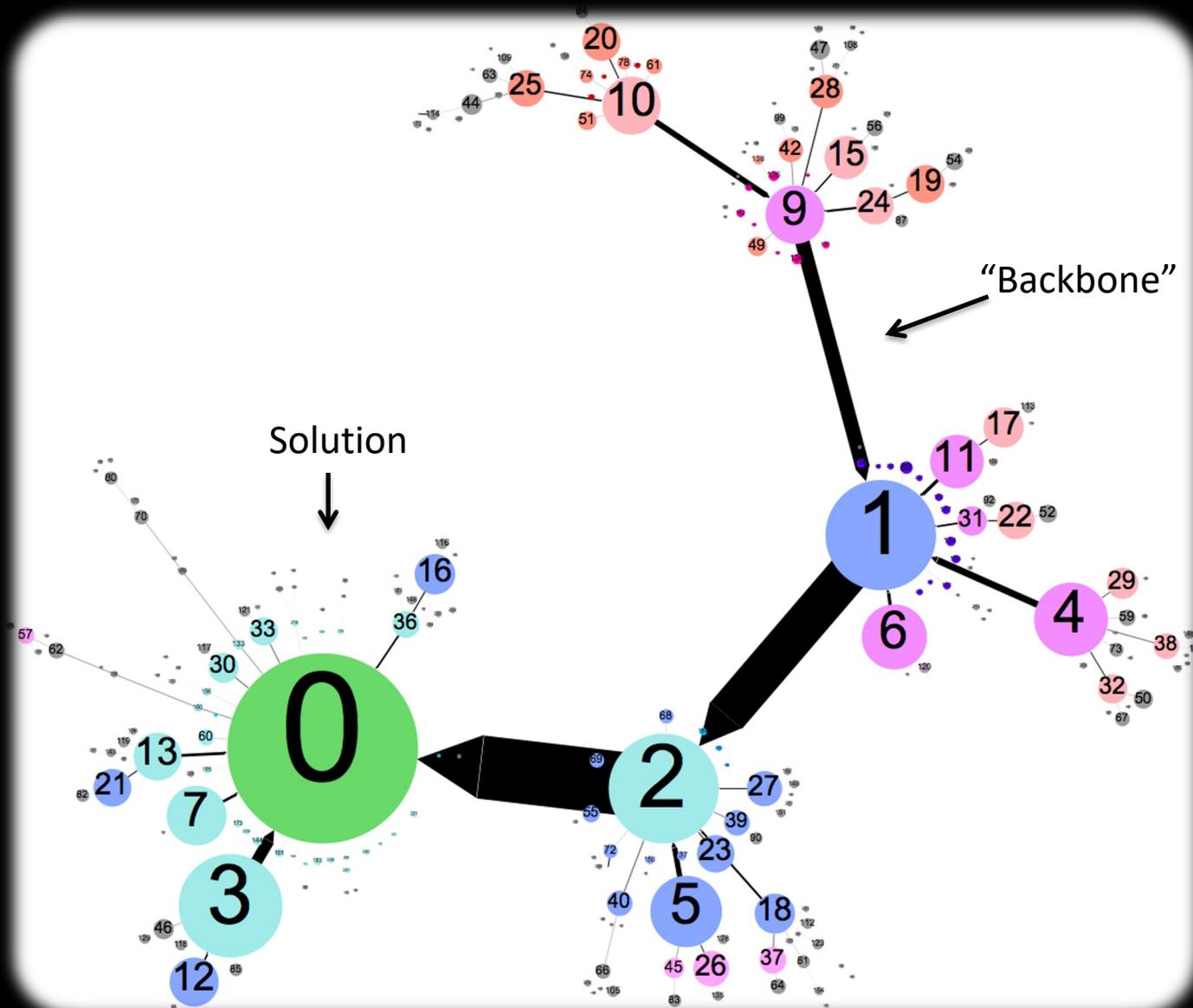
Diagram illustrating the Poisson Common Path algorithm. The equation is annotated with labels and arrows:

- $\gamma(s)$: Predicted next partial solution
- $p \in Z(s)$: Paths to solution
- $\sum_{x \in p}$: Partial solutions in the path
- $\frac{1}{\lambda_x}$: Submission count of partial solution

The term $\sum_{x \in p} \frac{1}{\lambda_x}$ is labeled as Path Cost.



Learned Problem Solving Policy



Only worked well for 6 line programs...

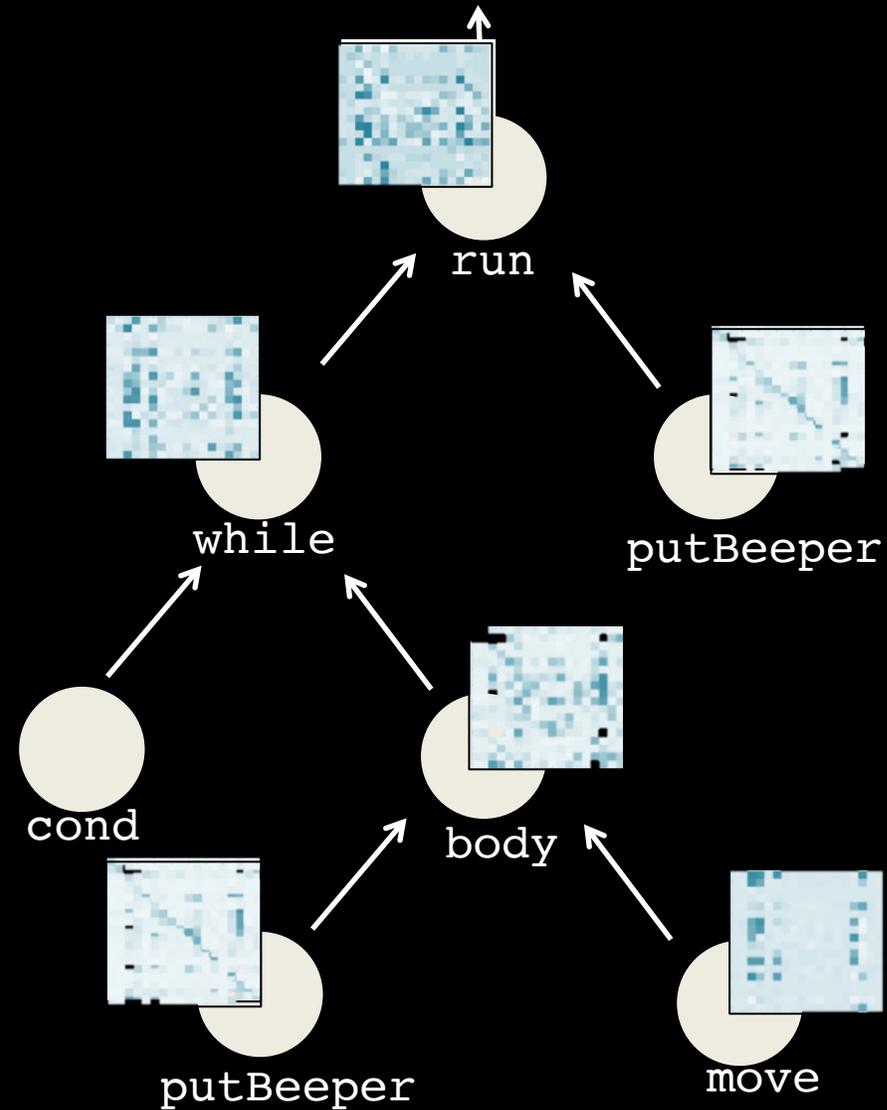
Chapter 2: Start to invent new algorithms...



Neural Network to Encode a Program

It looks like you have a fencepost error!

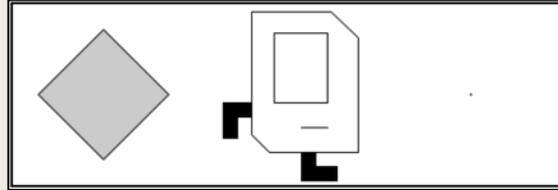
```
// User defined method
private void run() {
  while(isClear()){
    putBeeper();
    move();
  }
  putBeeper();
}
```



*Note: this was coded pre-tensor flow



Collect Triples



Precondition

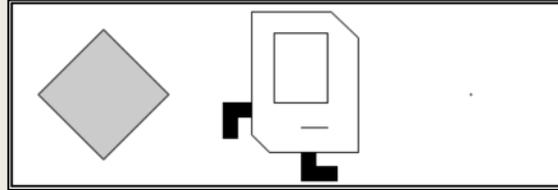
```
putBeeper ( ) ;  
move ( ) ;
```

Code

About 5 million triples per assignment



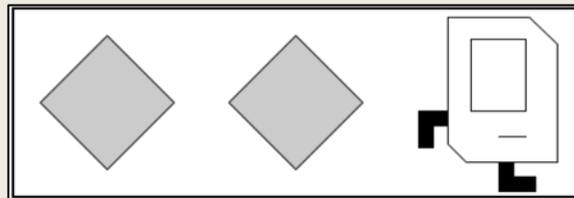
Collect Triples



Precondition

```
putBeeper();  
move();
```

Code



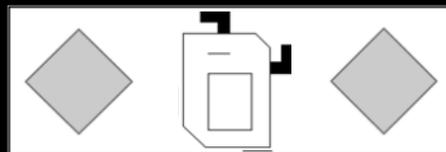
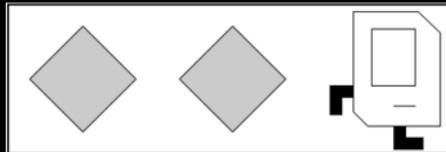
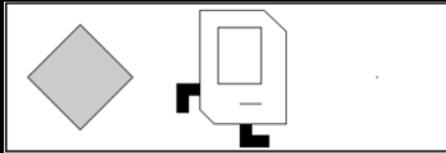
Postcondition

About 5 million triples per assignment



A Code Phrase is a Mapping

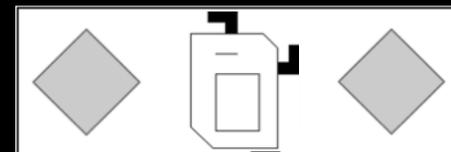
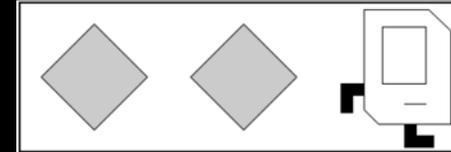
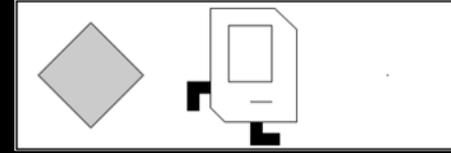
All possible preconditions



⋮



All possible postconditions



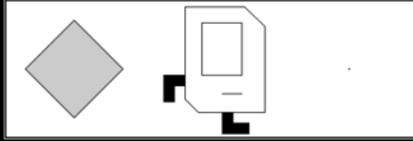
⋮



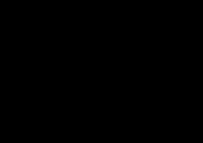
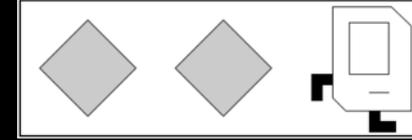
```
putBeeper();  
move();
```



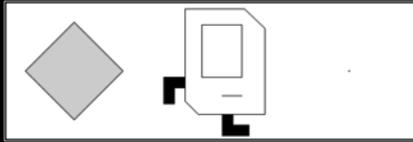
Neural Network for Programs



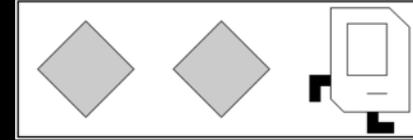
```
method step() {  
    putBeeper();  
    move();  
}
```



Neural Network for Programs



```
method step() {  
  putBeeper();  
  move();  
}
```



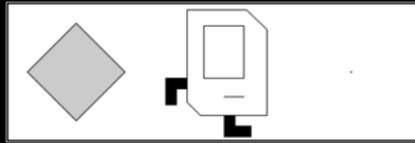
Raw Precondition

Program Matrix

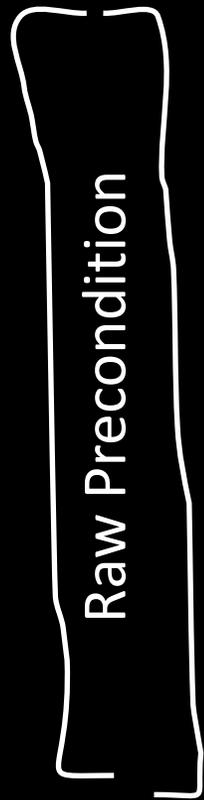
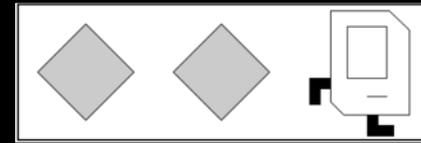
*coded pre-tensor flow



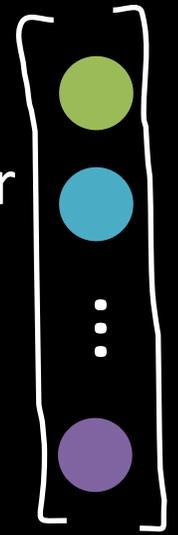
Neural Network for Programs



```
method step() {  
  putBeeper();  
  move();  
}
```

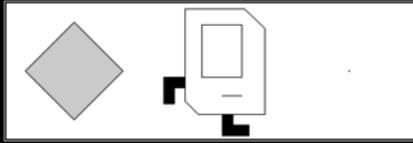


Encoder
→

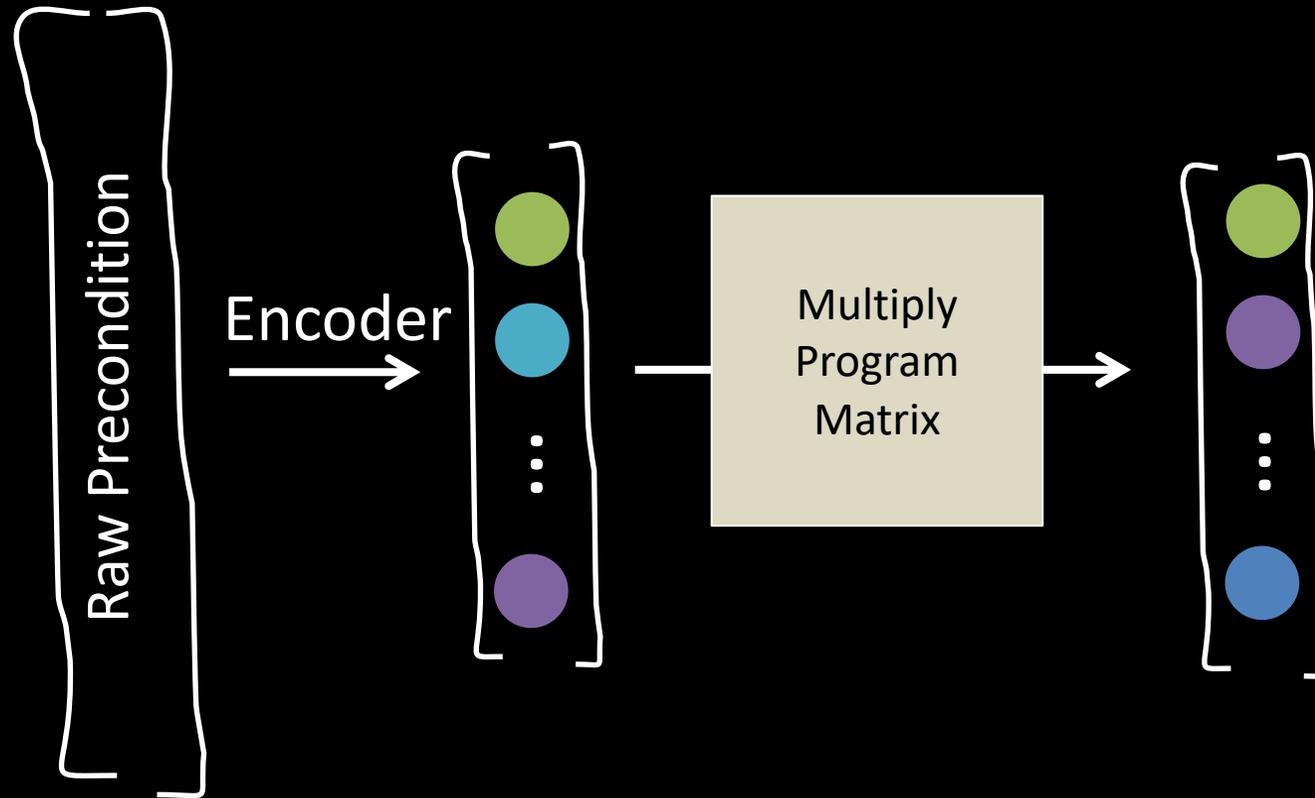
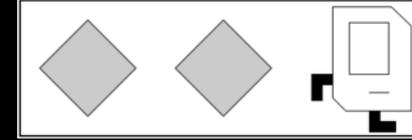


*coded pre-tensor flow

Neural Network for Programs



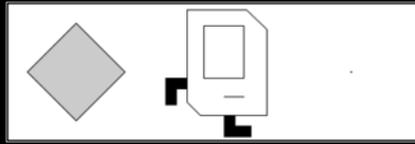
```
method step() {  
  putBeeper();  
  move();  
}
```



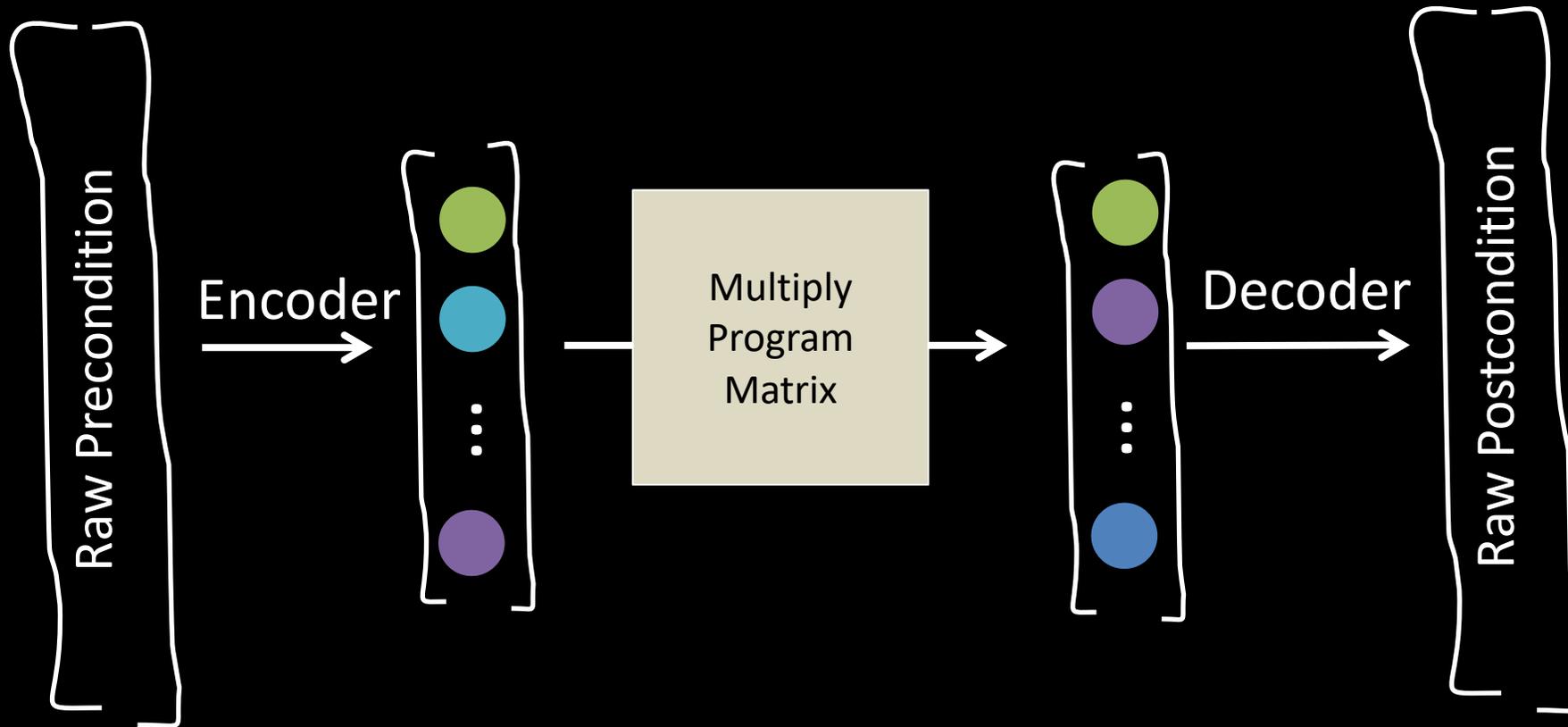
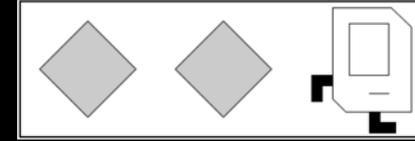
*coded pre-tensor flow



Neural Network for Programs



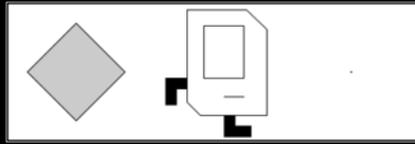
```
method step() {  
  putBeeper();  
  move();  
}
```



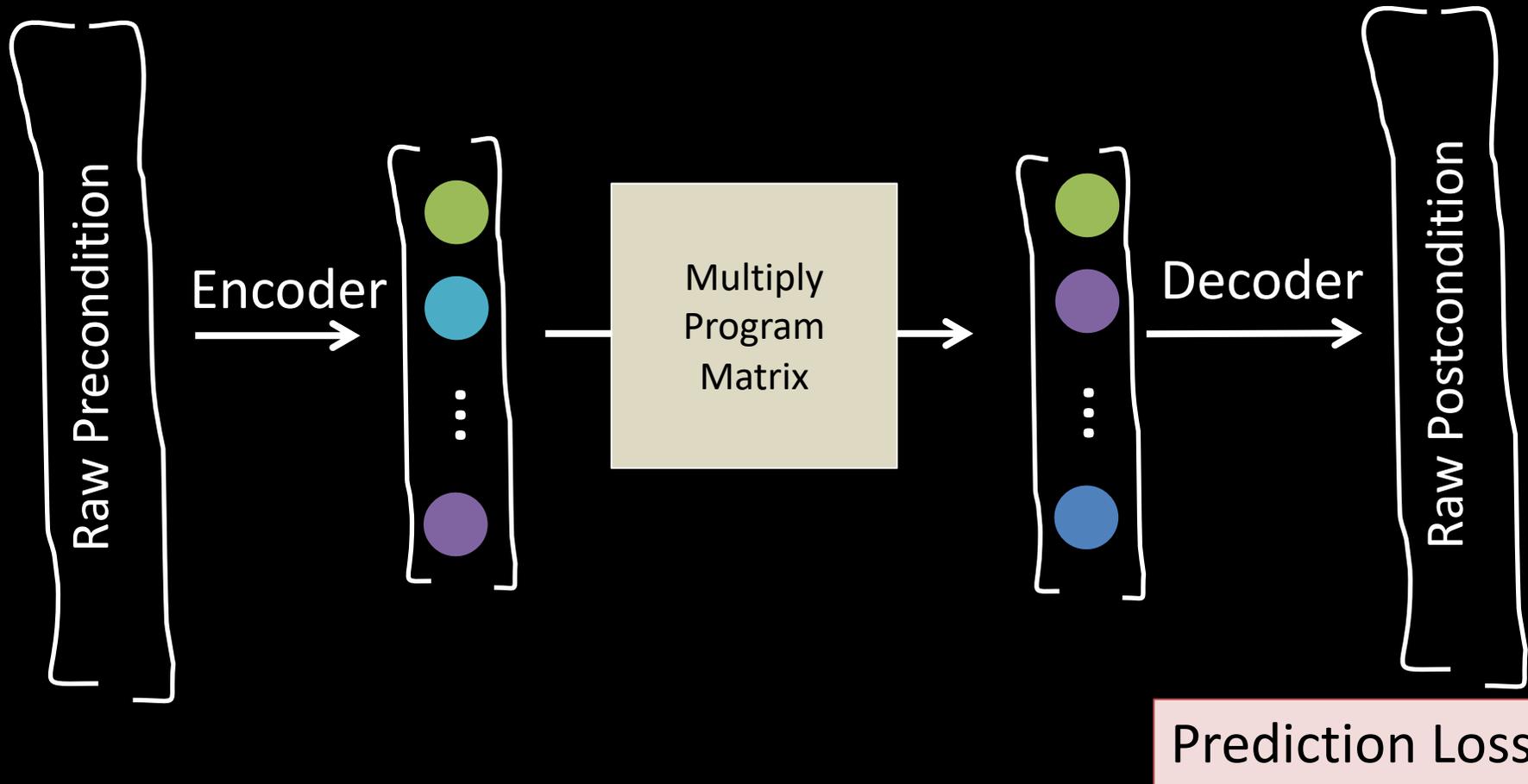
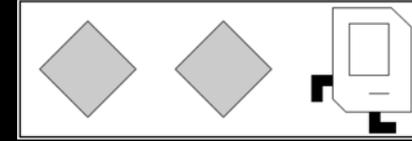
*coded pre-tensor flow



Neural Network for Programs



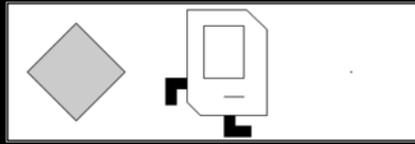
```
method step() {  
  putBeeper();  
  move();  
}
```



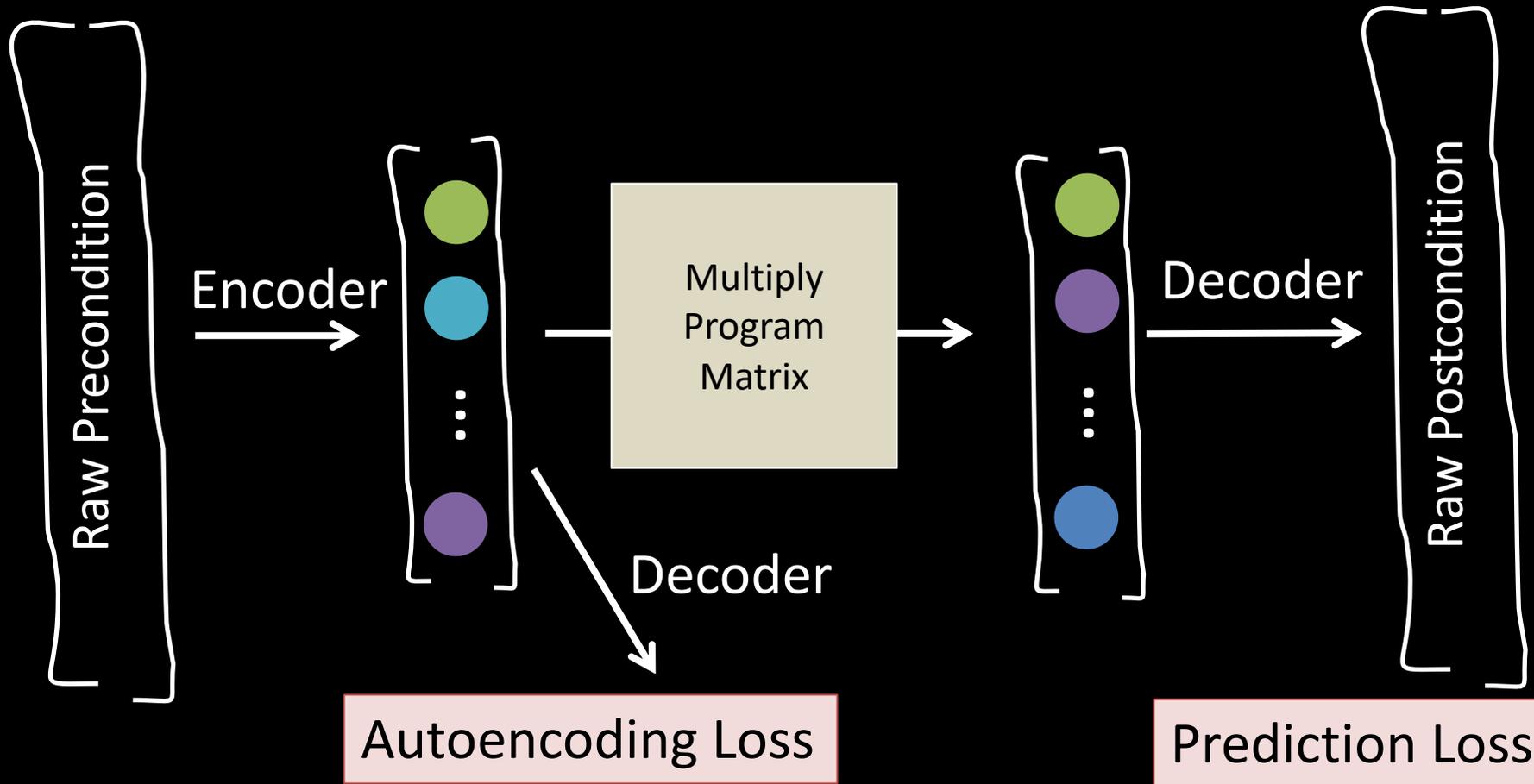
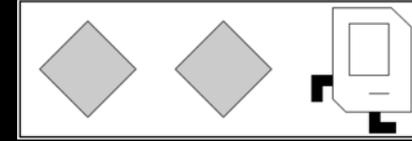
*coded pre-tensor flow



Neural Network for Programs



```
method step() {  
  putBeeper();  
  move();  
}
```



*coded pre-tensor flow

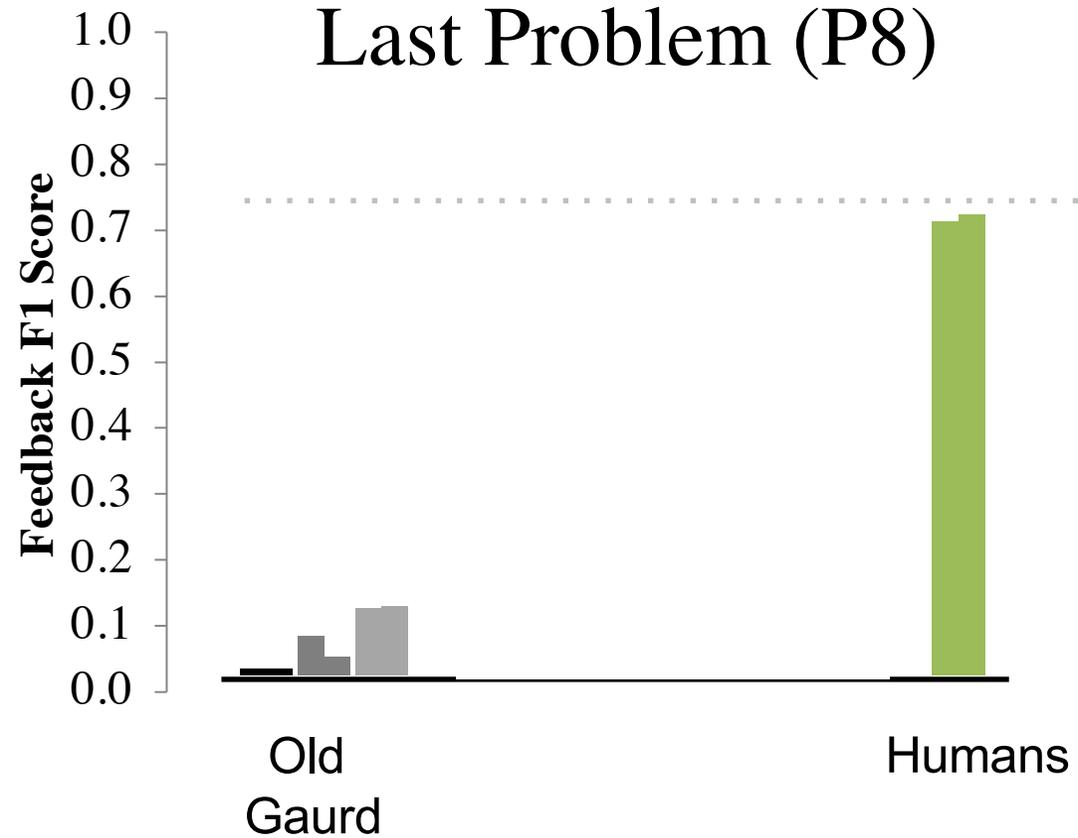


Does it work?



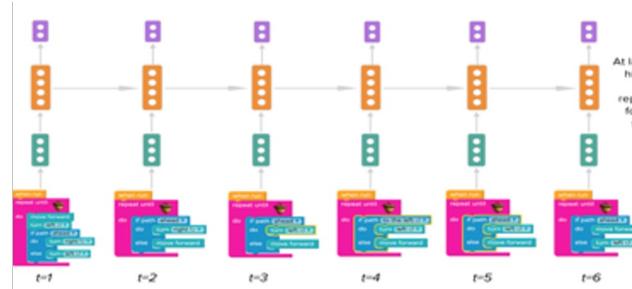
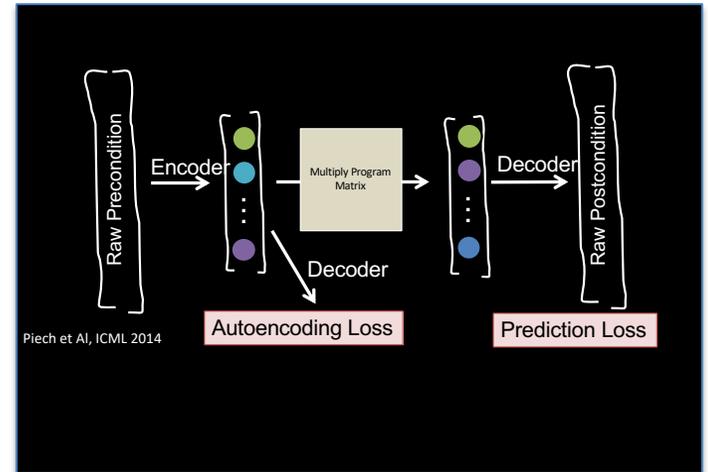
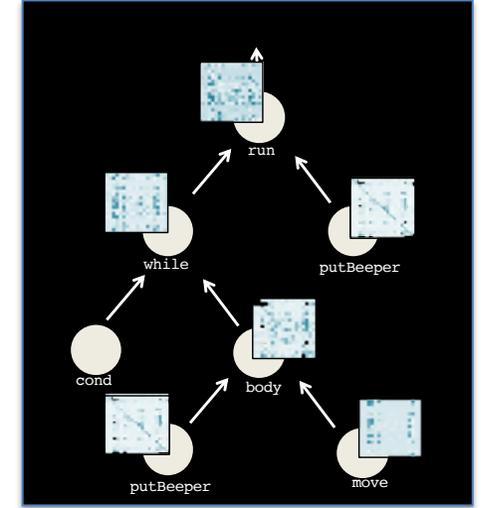
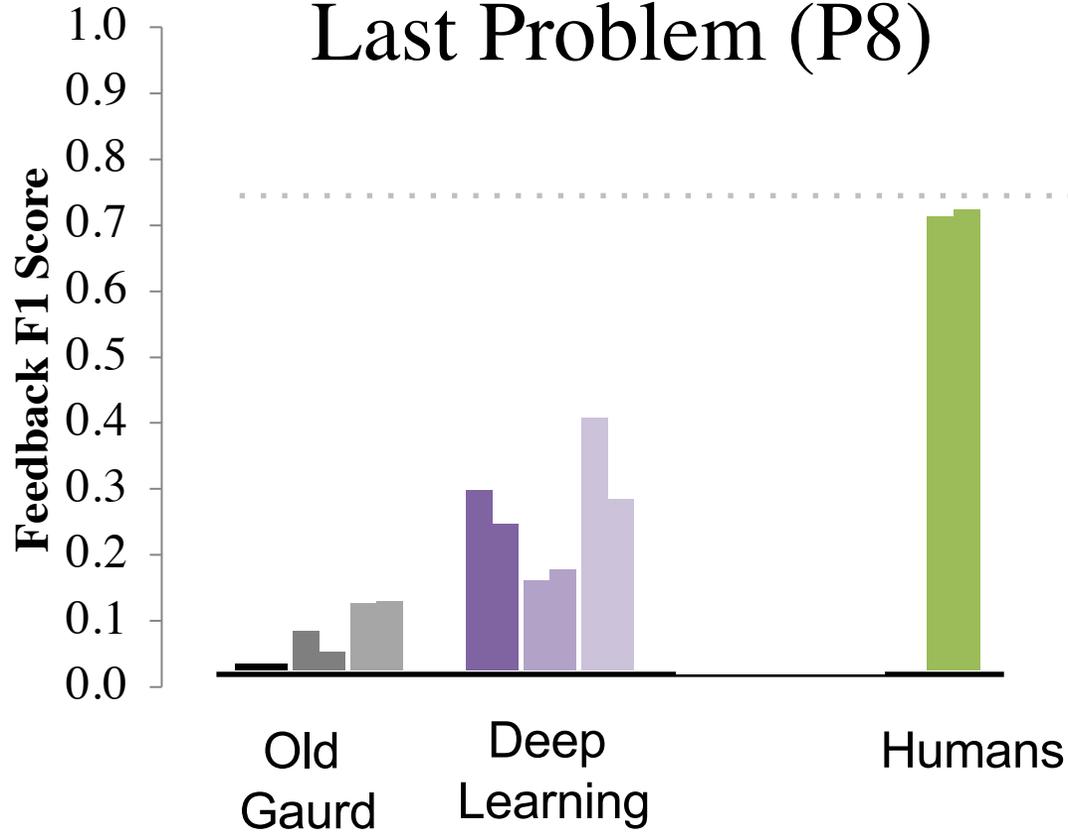
Traditional Deep Learning Doesn't Work

Label student code



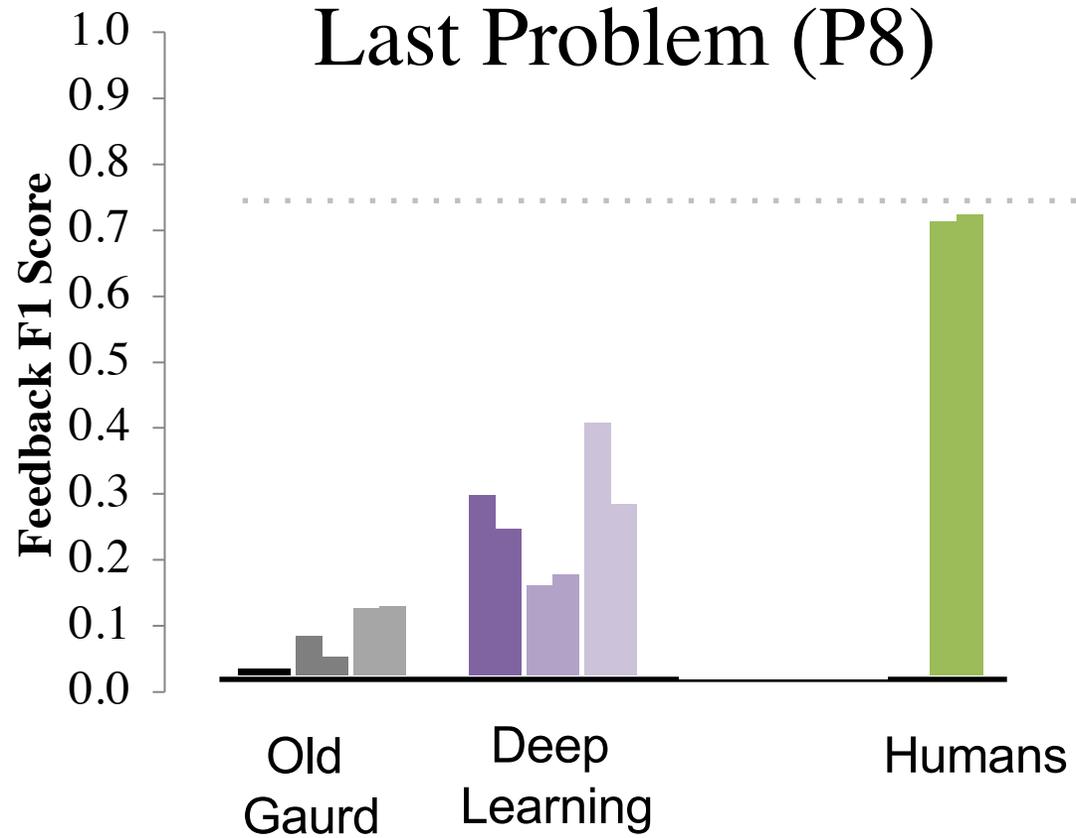
Inaccurate, Uninterpretable, and Data Hungry

Label student code



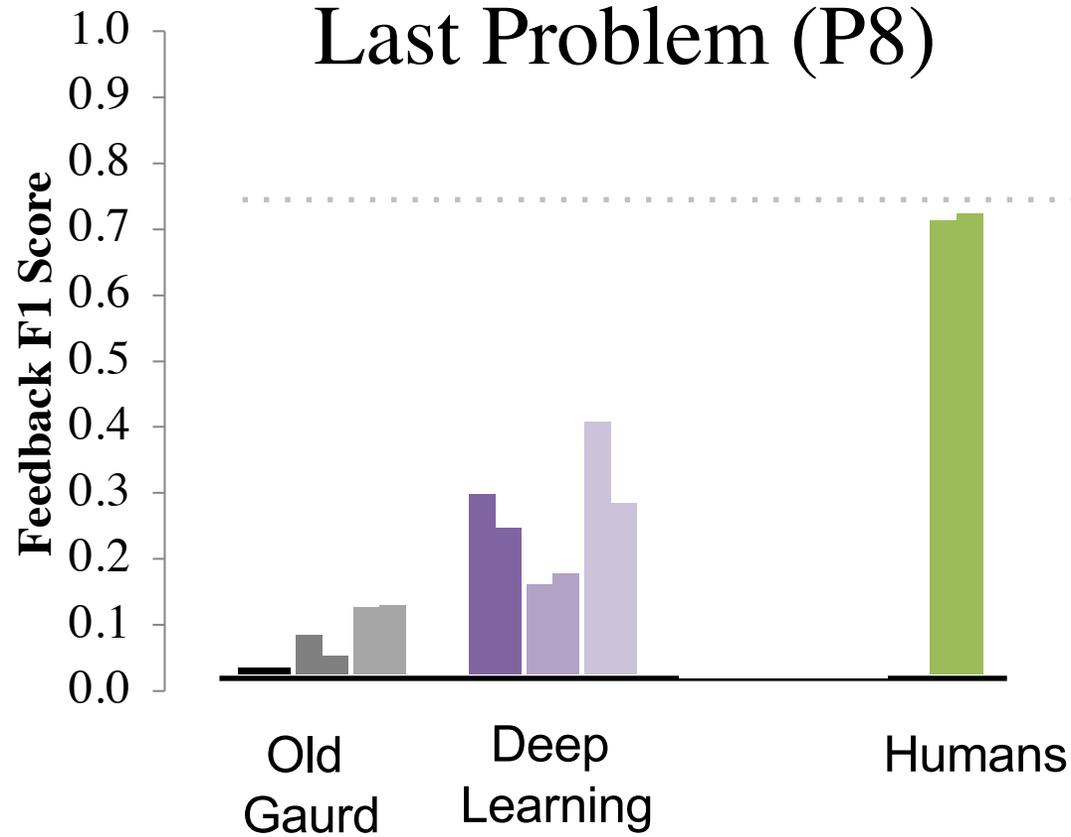
Inaccurate, Uninterpretable, and Data Hungry

Label student code



Inaccurate, Uninterpretable, and Data Hungry

Label student code



We need one shot learning

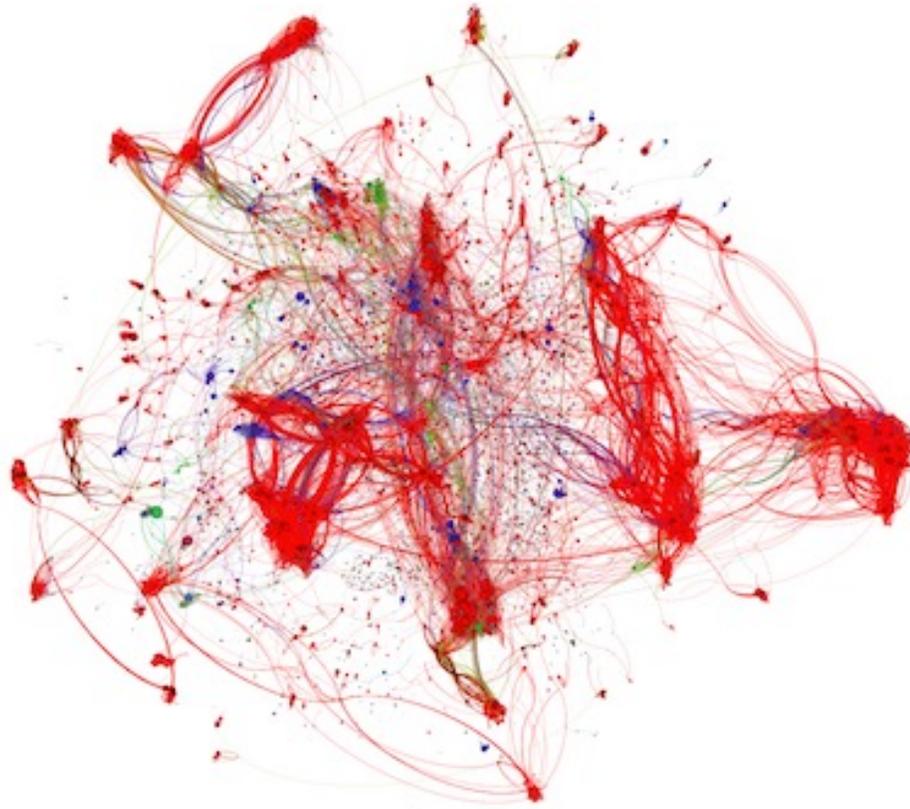
We need verifiability



Why is it so hard?

Hard Problem

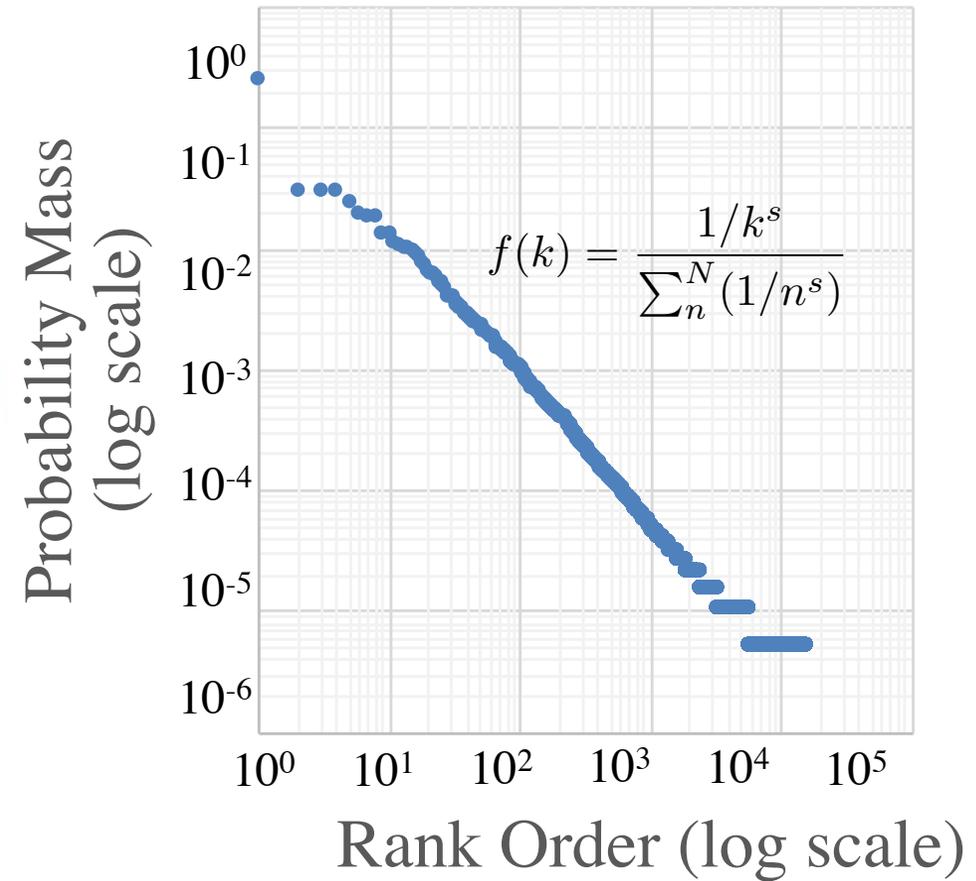
Brute force solution?



1 million unique solutions to
programming Linear Regression

WWW 2014

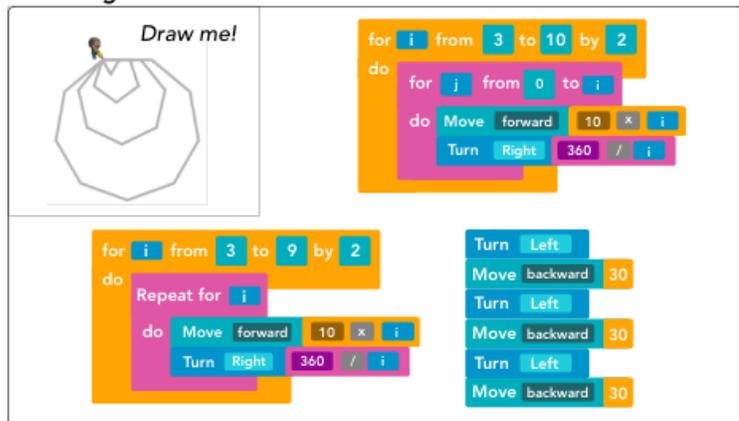
Code Zipf Plot



They are all Zipf!

(a) Datasets in Computational Education

Code.org Problem 8



Draw me!

```

for i from 3 to 10 by 2
do
  for j from 0 to i
  do
    Move forward 10 * i
    Turn Right 360 / i
  
```

```

for i from 3 to 9 by 2
do
  Repeat for i
  do
    Move forward 10 * i
    Turn Right 360 / i
  
```

```

Turn Left
Move backward 30
Turn Left
Move backward 30
Turn Left
Move backward 30
  
```

Powergrading P13

What is one reason the original colonists came to America?

- Religious freedom
- For religious freedom
- Freedom

- declared our independence from england
- religious freedom
- as a criminal punishment

- to create a new colony
- to find better economic prospects
- to break away from the church in great britain

CS1: Liftoff

Write a Java Program to print the numbers 10 down to 1 and then write liftoff. You must use a loop.

```

public void run() {
  for (int i=START; i>0; i -=1)
  {
    println(i);
  }
  println("Liftoff");
}
  
```

```

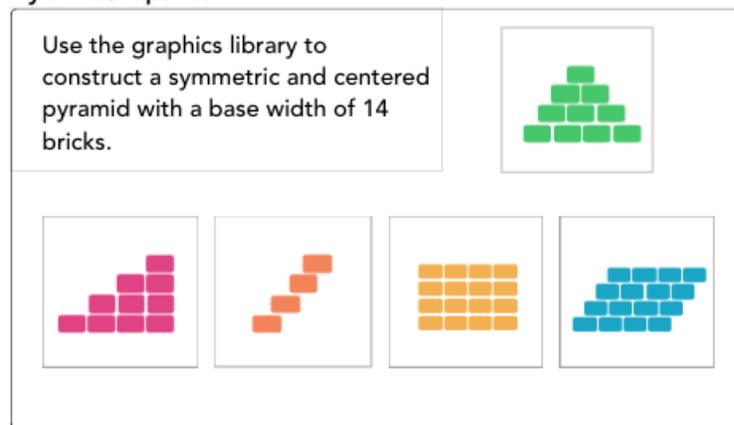
public void run() {
  int x = START;
  int y = 1;
  int z = 9;
  while (x>=1) {
    println(x);
    x=z;
    z=x-y;
  }
  println("Liftoff");
}
  
```

```

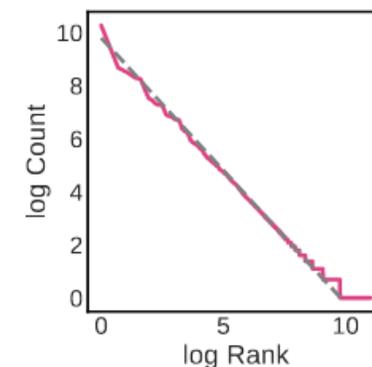
public void run() {
  for(int i = START; i>0; i--)
  {
    println(i);
    pause(1000);
  }
  println("Liftoff!");
}
  
```

PyramidSnapshot

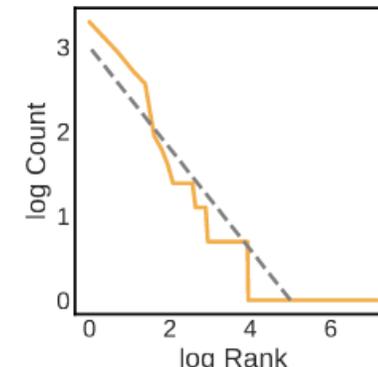
Use the graphics library to construct a symmetric and centered pyramid with a base width of 14 bricks.



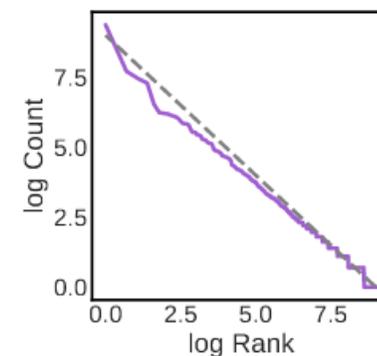
(b) Code.org P8



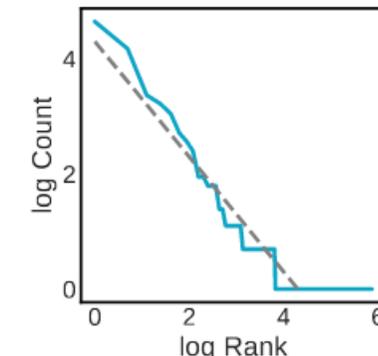
(c) CS1: Liftoff



(d) Pyramid



(e) Powergrading



[Suspense]

Chapter 3: Back to the drawing board

Humans Don't Need Much Data

Single training example:

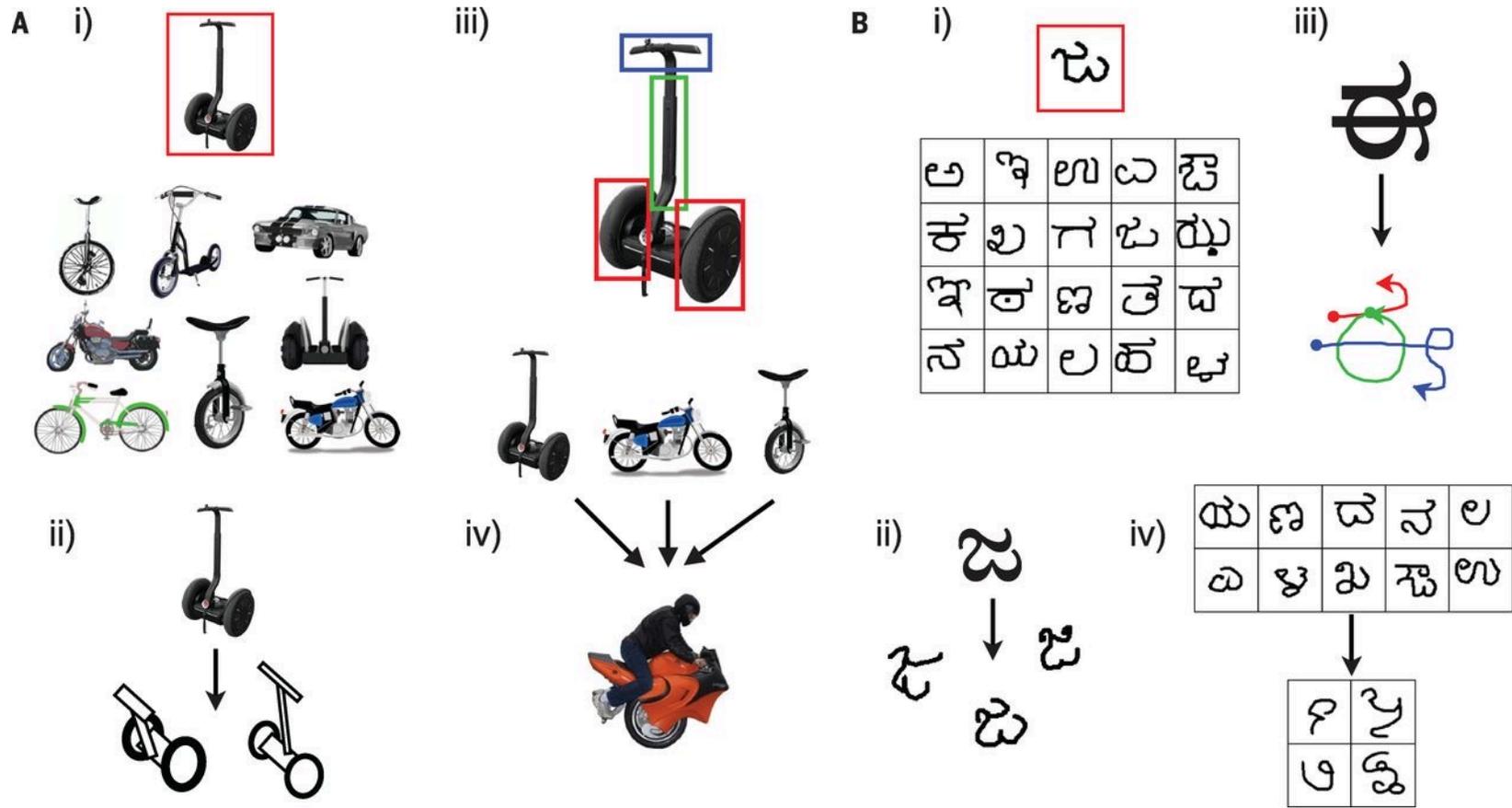
ॐ

Test set:

ॐ ॐ ॐ
ॐ ॐ ॐ
ॐ ॐ ॐ



Fig. 1 People can learn rich concepts from limited data.



Brenden M. Lake et al. Science 2015;350:1332-1338



Fig. 2 Simple visual concepts for comparing human and machine learning.



Brenden M. Lake et al. Science 2015;350:1332-1338



Bayesian Program Learning

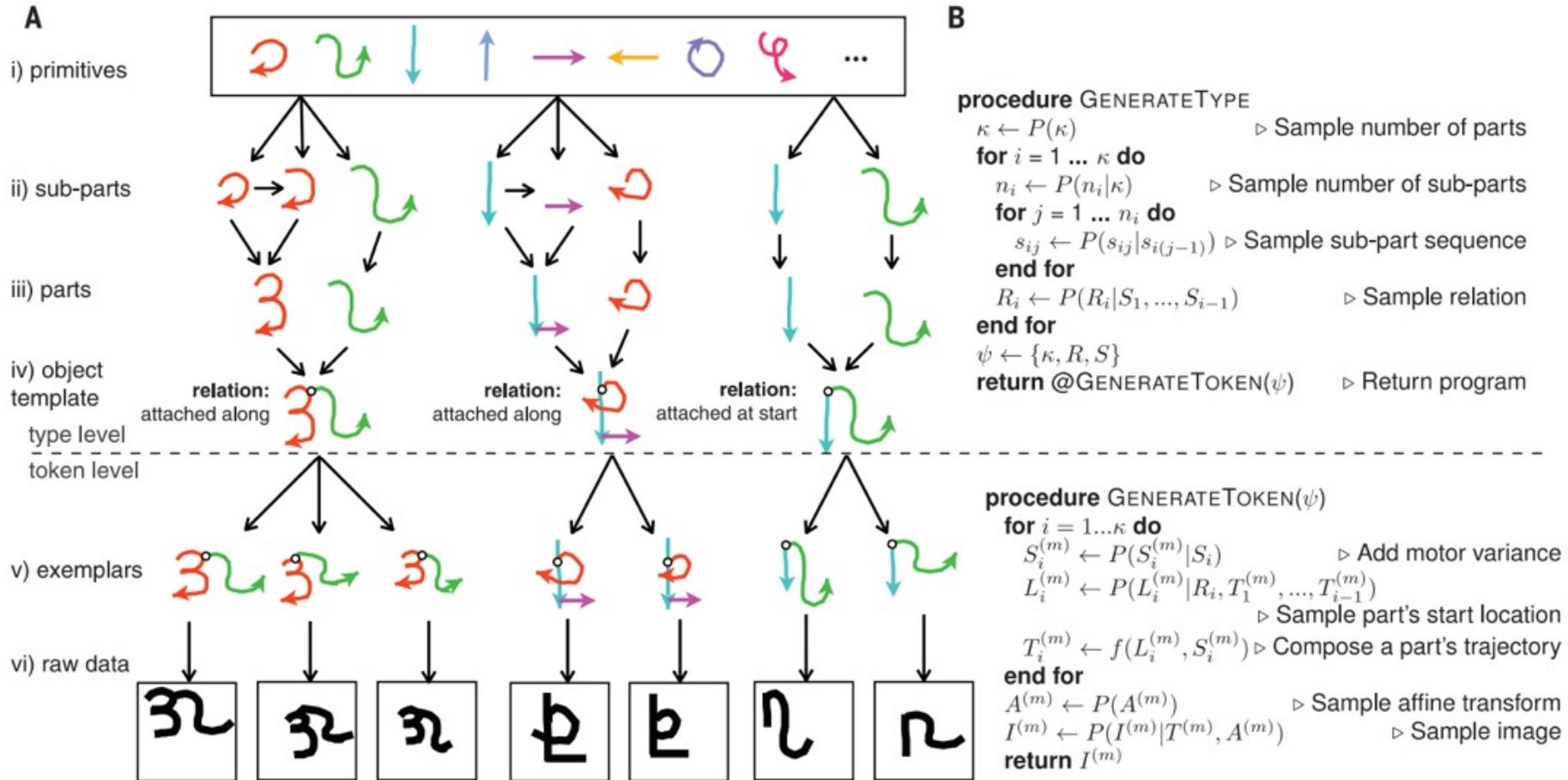
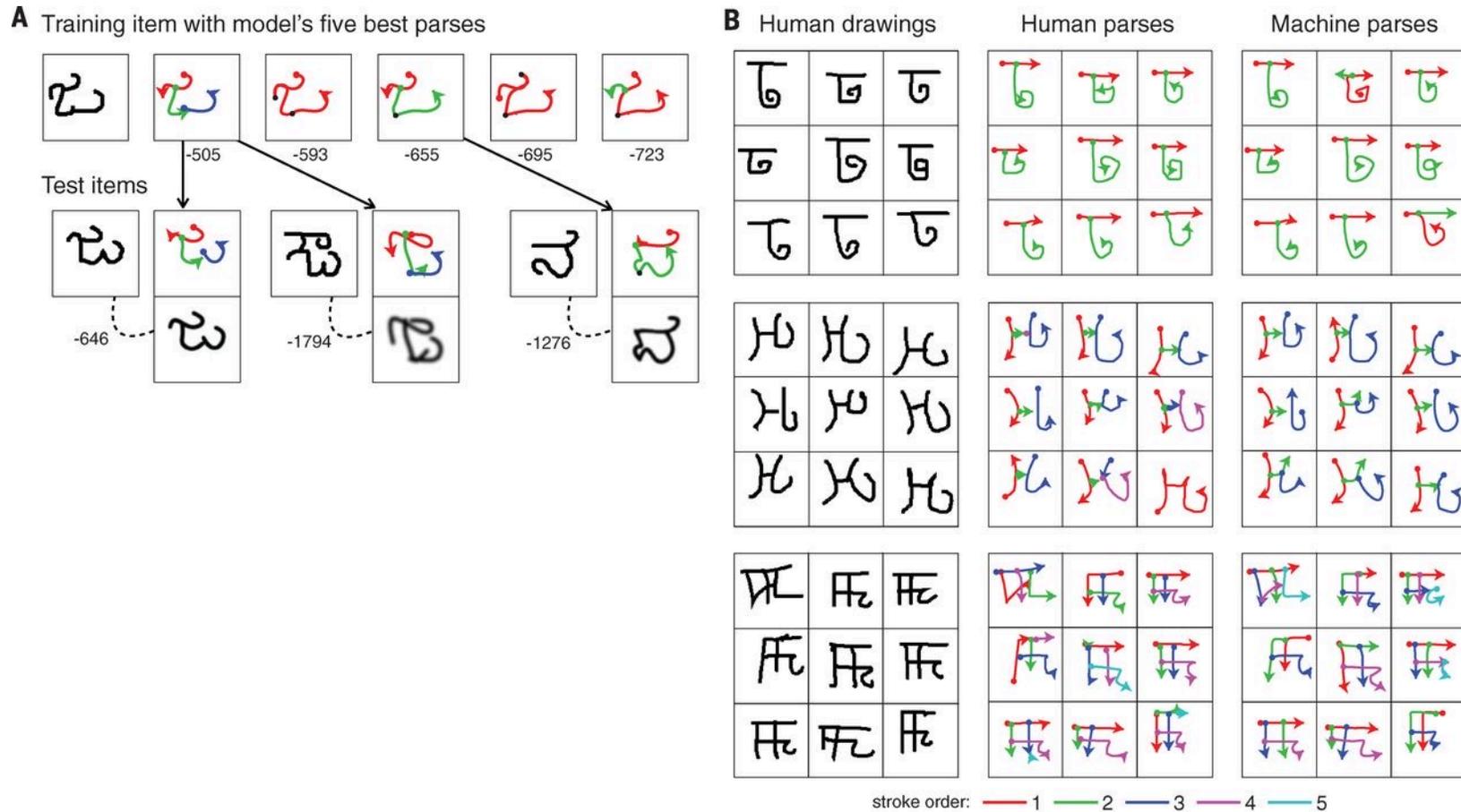


Fig. 4 Inferring motor programs from images.

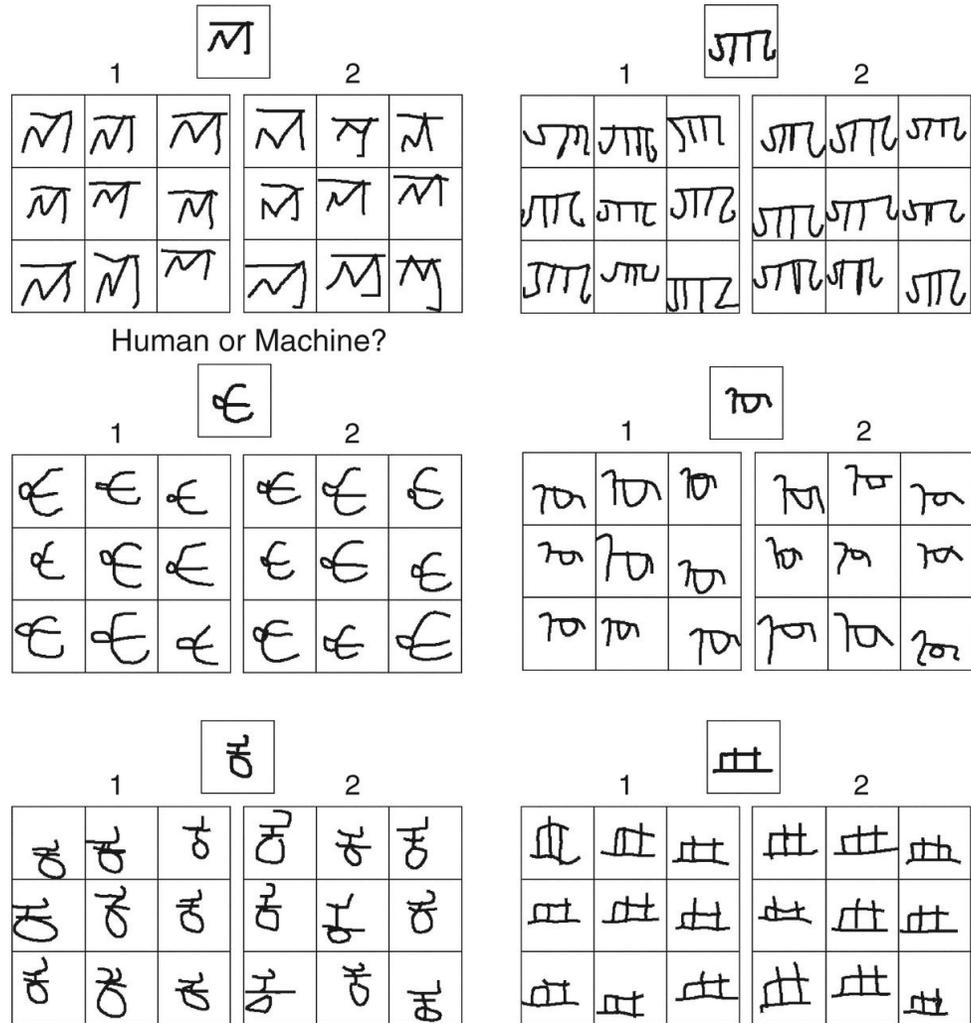


Brenden M. Lake et al. *Science* 2015;350:1332-1338

Copyright © 2015, American Association for the Advancement of Science



Fig. 5 Generating new exemplars.

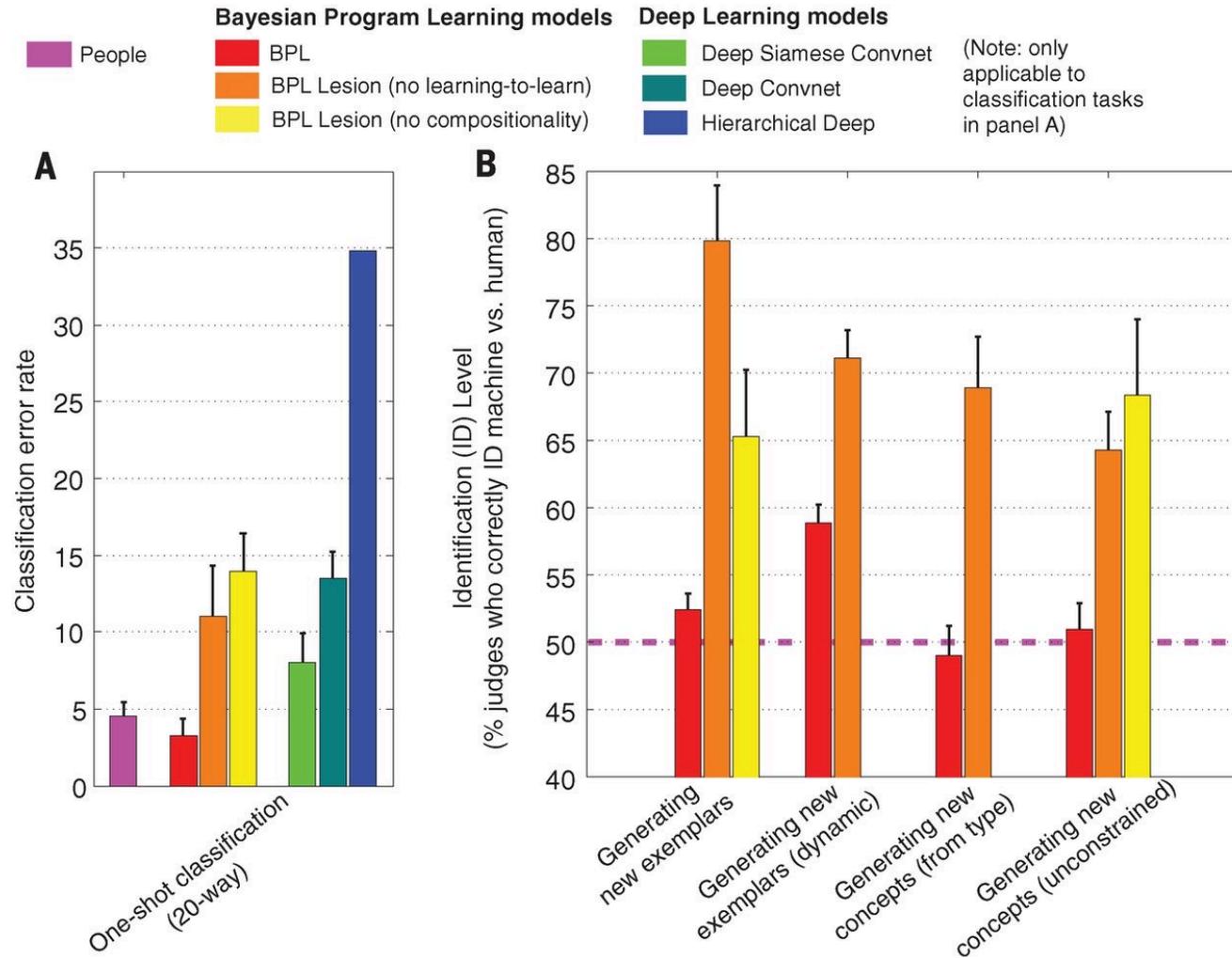


Human or Machine?

Brenden M. Lake et al. Science 2015;350:1332-1338



Fig. 6 Human and machine performance was compared on (A) one-shot classification and (B) four generative tasks.

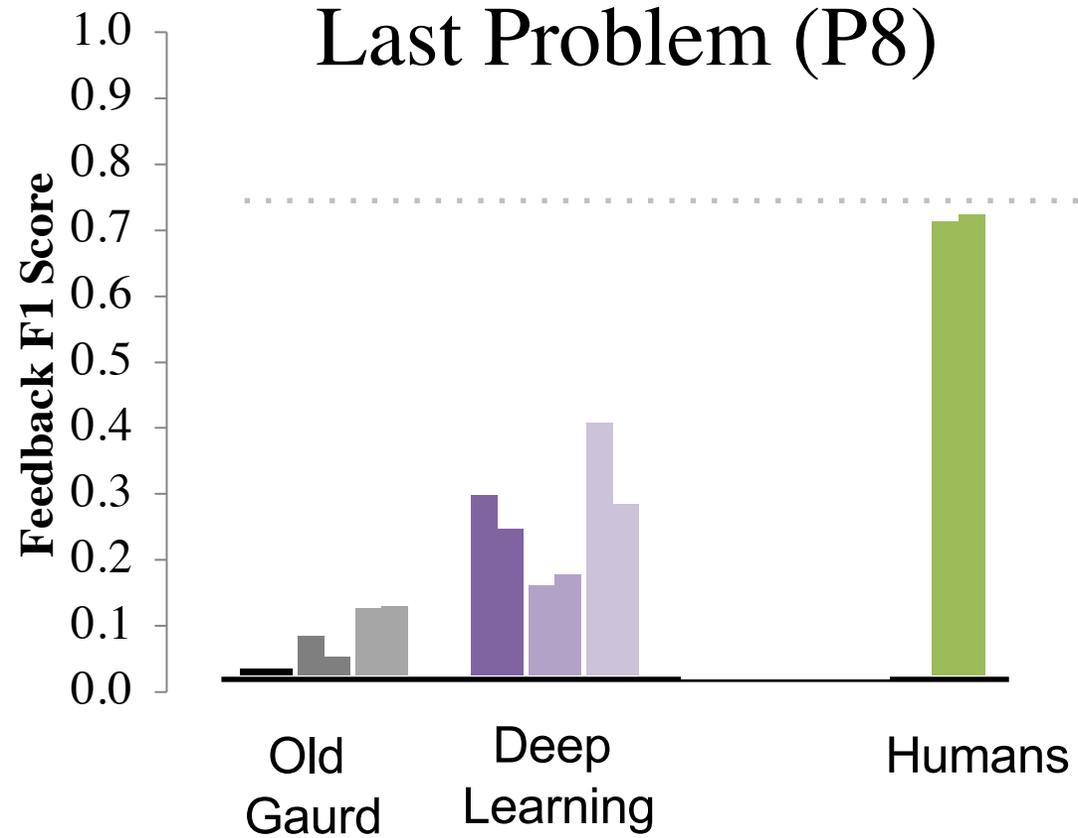


Brenden M. Lake et al. *Science* 2015;350:1332-1338



Generative Understanding

Label student code



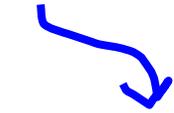
Grading is Hard



Grading:

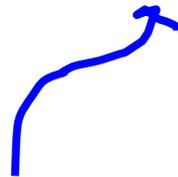
Infer ability and choices from code

Student ability



$$P(\Theta, C | \Pi)$$

Student choices

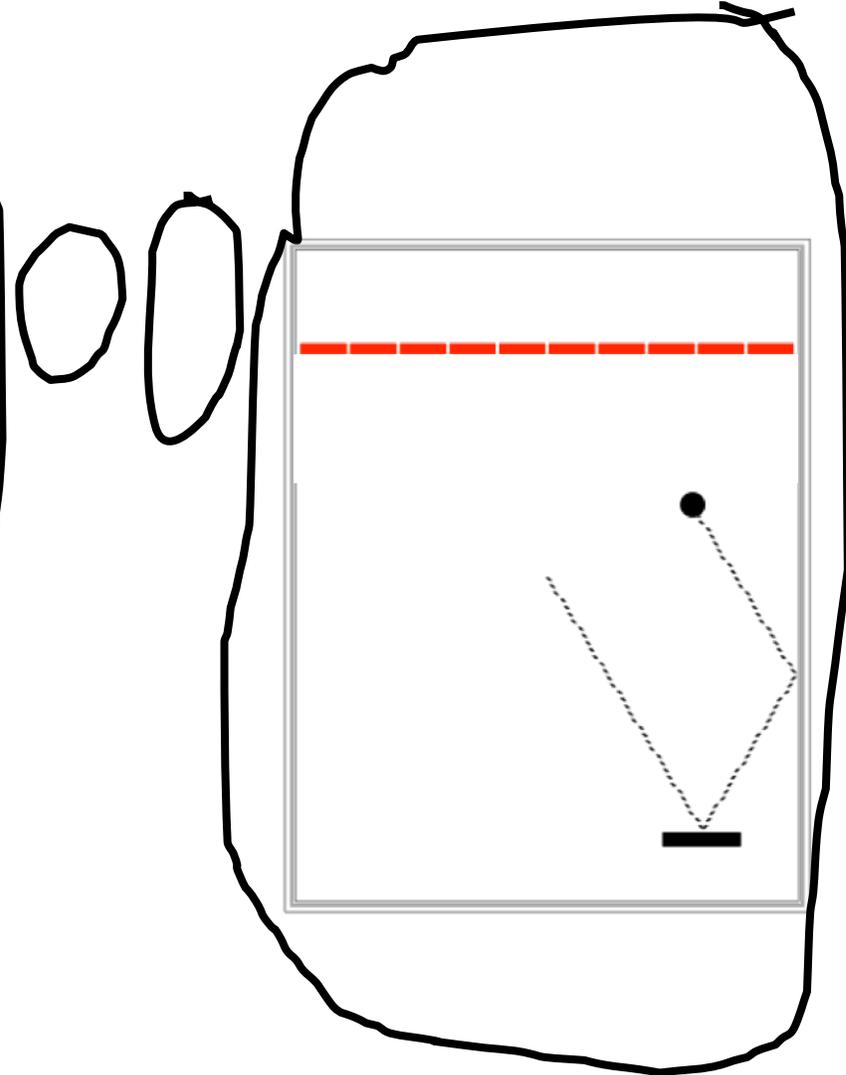


Student code

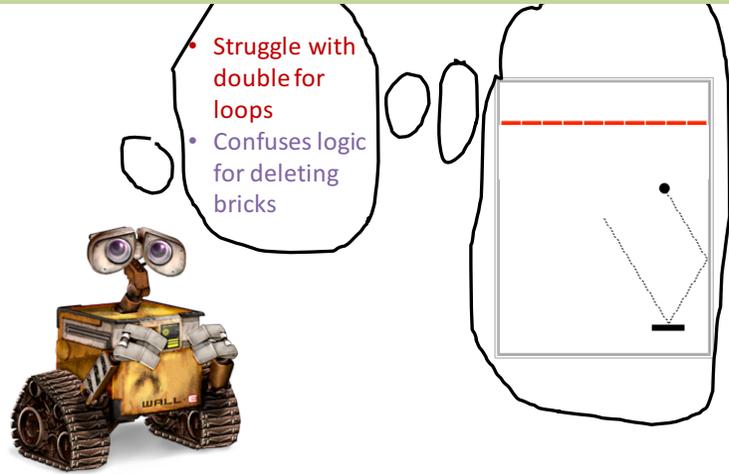


Generative Probabilistic Model?

- Struggle with double for loops
- Confuses logic for deleting bricks



Imagine Students



This is easy and exponential
human impact

A students
“ability”

$$\Theta \sim \text{pythonSample}$$

A students
“choices”

$$C \sim \text{pythonSample} | \Theta$$

The resulting
code

$$\Pi \sim \text{pythonSample} | C$$



This is hard and linear
human impact

$$P(\Theta, C | \Pi)$$

Infer ability and
choices from code



Bayesian Programming Language



ideaToText



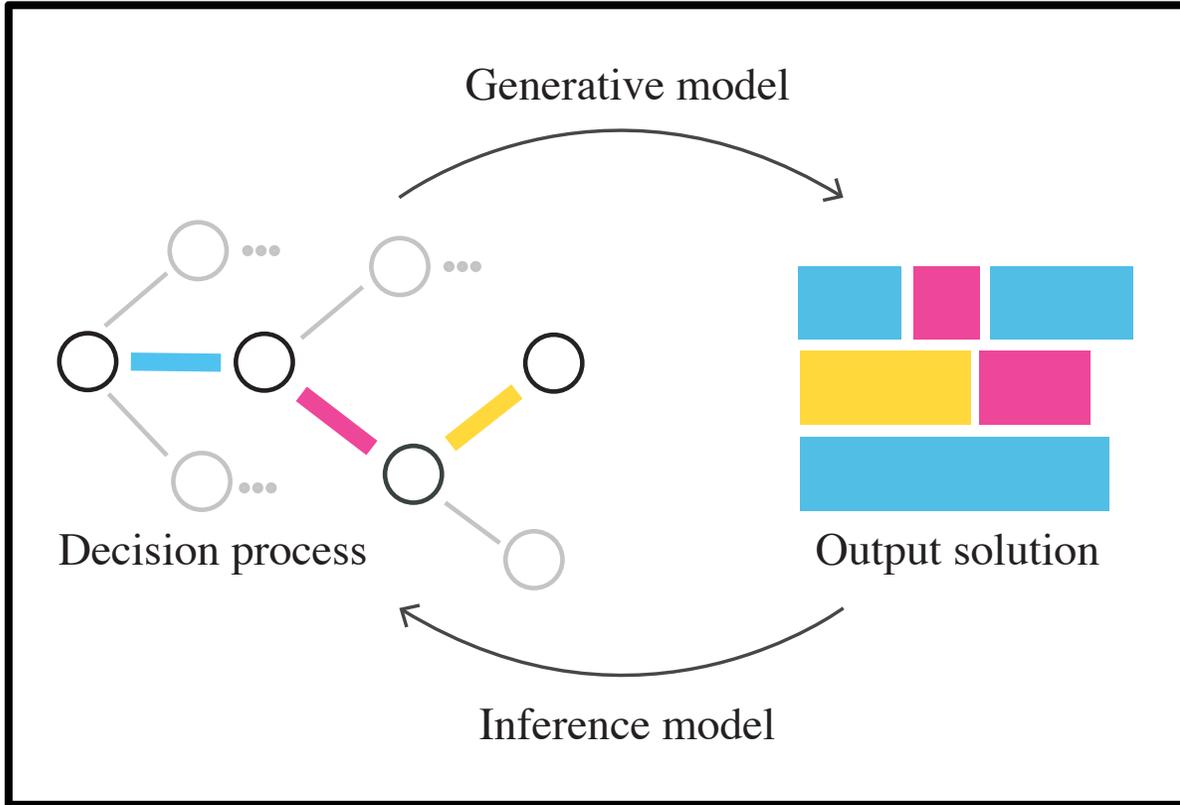
Teachers Articulate N misconceptions

```
7
8 # This python class is a RubricSampling Decision
9 # it generates programs that print the numbers 10 -> 1
10 class Countdown(Decision):
11
12     def registerChoices(self):
13         # these are the main strategies for printing out a
14         # countdown
15         self.addRubricChoices('loop-style', {
16             'for' :  $\theta_1$ ,
17             'while' :  $\theta_2$ ,
18             'none' :  $\theta_3$ ,
19             'empty' :  $\theta_4$ 
20         })
21
22     # we can make some grading choices based on which
23     # strategy they chose (did they actually use a loop?)
24     def processChoices(self):
25         style = self.getChoice('loop-style')
26         hasLoop = style != 'none' and style != 'empty'
27         self.addLabel('rubric-hasLoop', hasLoop)
28
29
30     # Based on their strategy render a different decision
31     def renderCode(self):
32         style = self.getChoice('loop-style')
33         if style == 'for': return '{ForSoln}'
34         if style == 'while': return '{WhileSoln}'
35         if style == 'none': return '{NoLoopSoln}'
36         if style == 'empty': return ''
37
```

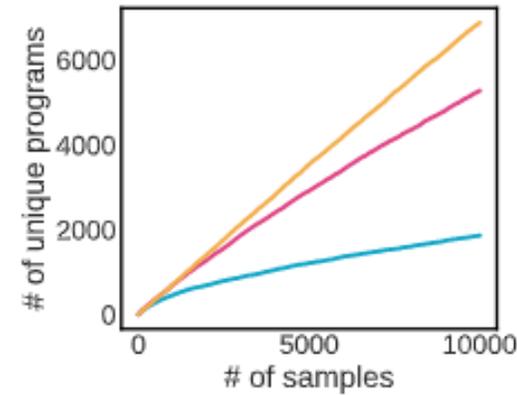
1. This is code for a single decision point
2. Give a name to the choice that the student is making (it is a random variable)
3. How do those choices translate into feedback?
4. What does the code look like? Often evokes other decision points



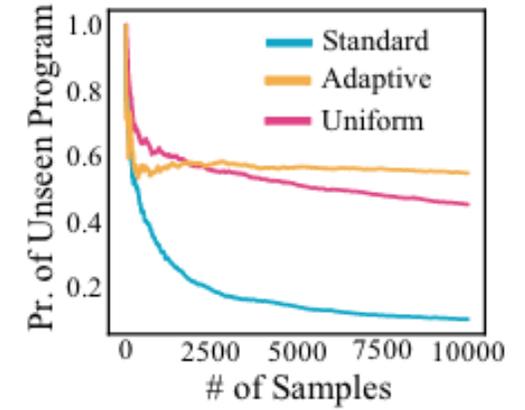
Generative Understanding



Idea: (1) sample a ton, then (2) build a neural network to learn to predict decisions



(a) Uniqueness



(b) Good-Turing

Decision process

Next choice

$$p_G(x_{a_1}, \dots, x_{a_T} | y) = \prod_{t=1}^T p_G(x_{a_t} | y, \mathbf{x}_{<a_t})$$

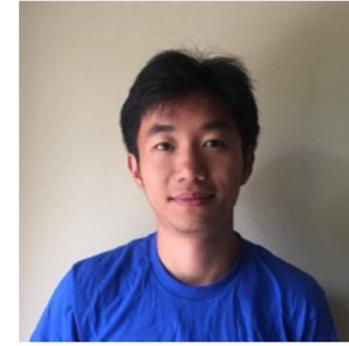
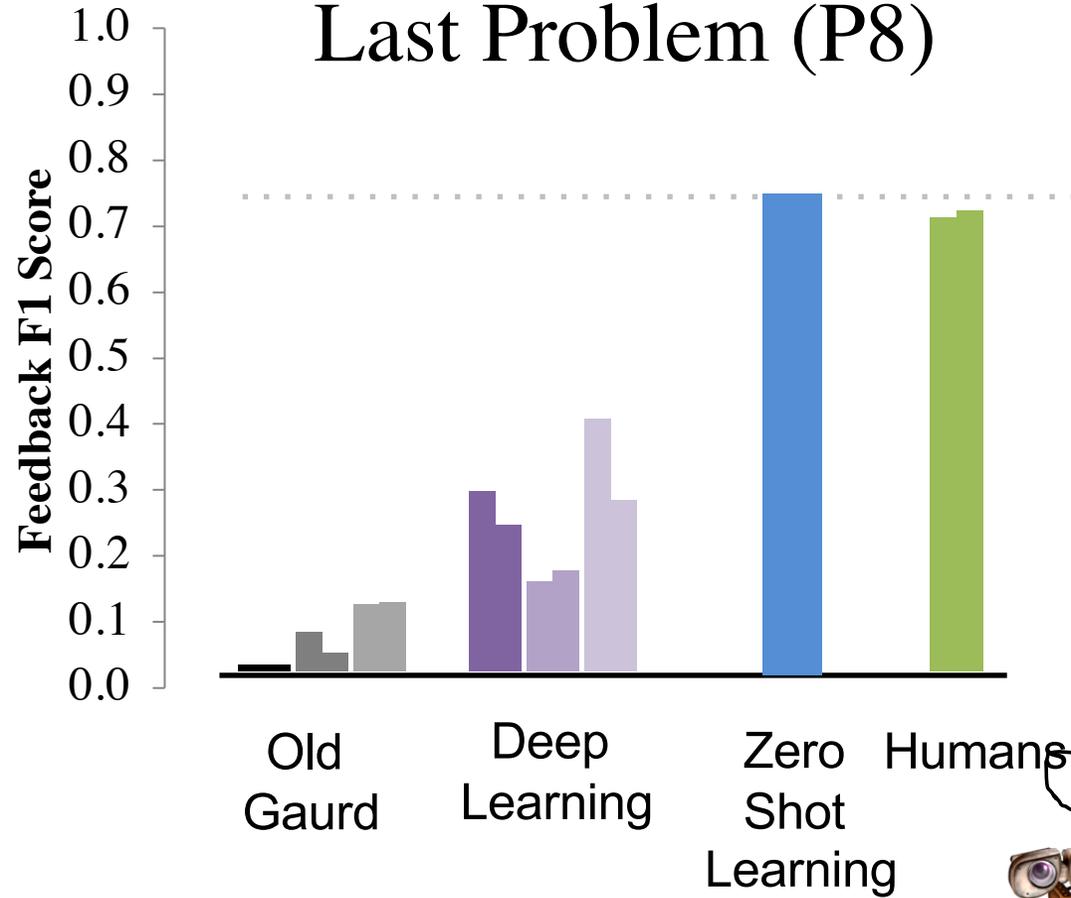
solution

Previous choices

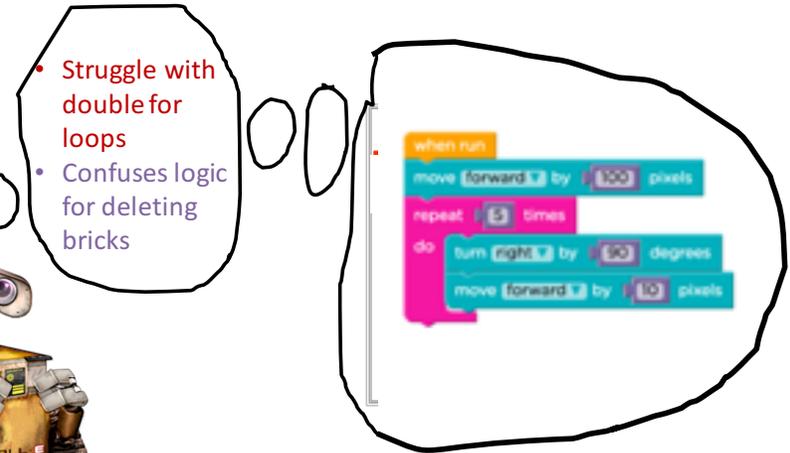


Generative Understanding

Label student code



*Outstanding Student
paper award, AAI 2019*



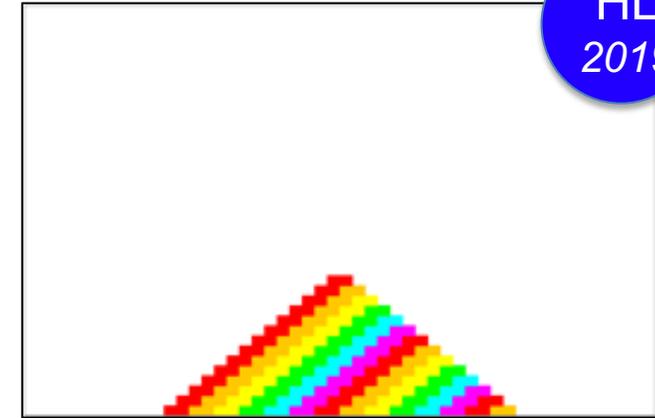
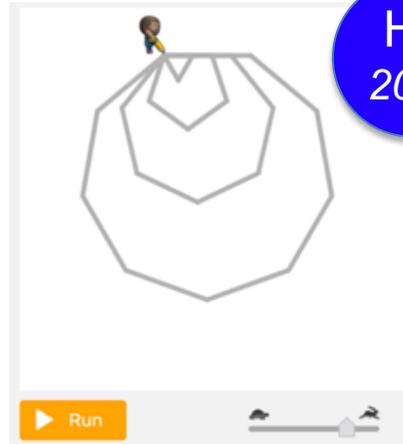
Not just for code

Results from early 2019

Many domains of student work

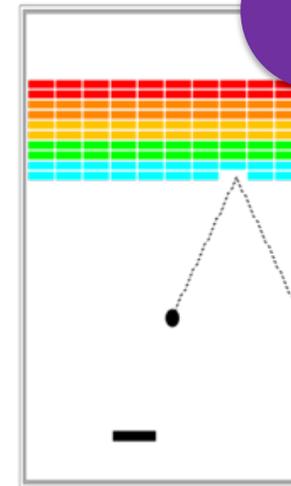
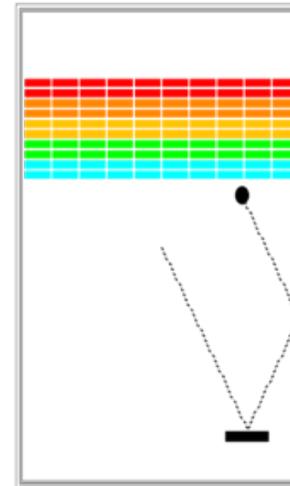
HL
2017

$$\begin{aligned} & 9 \times 6 \\ &= (10 - \boxed{}) \times 6 \\ &= 10 \times 6 - \boxed{} \times 6 \\ &= 60 - \boxed{} \\ &= \boxed{} \end{aligned}$$



NH
2019

Why did the original colonists come to America?



?



NH

Near Human

HL

Human Level

SH

Super Human Level



So what?

What does this mean for me?!?

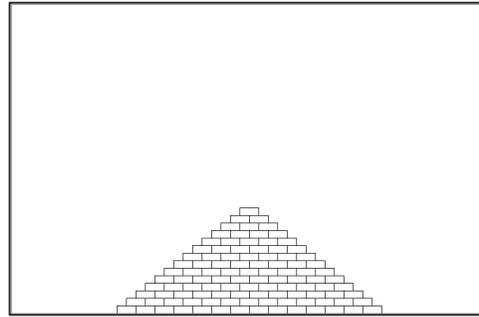
Understanding Process

2,600 students

130,000 partial solutions

μ snapshots per student = 50

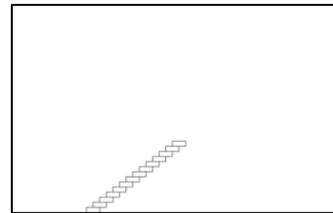
μ time per student = 2 hours



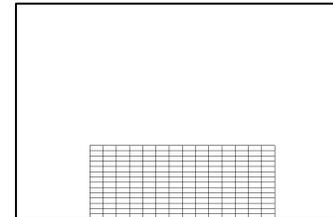
Step 1



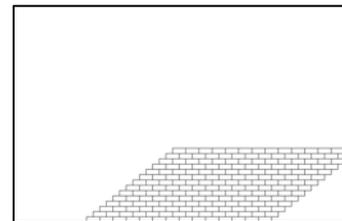
Step 2



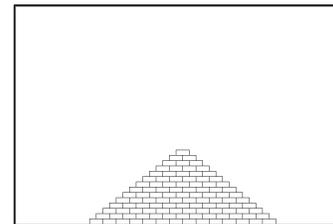
Step 3



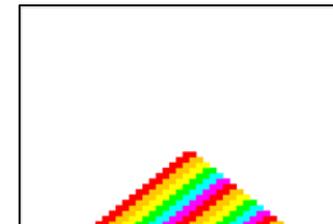
Step 4



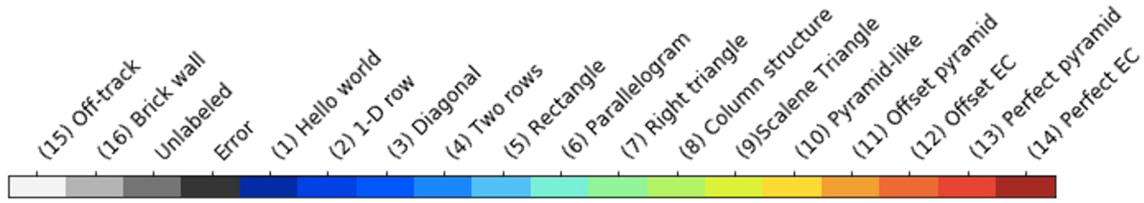
Step 5



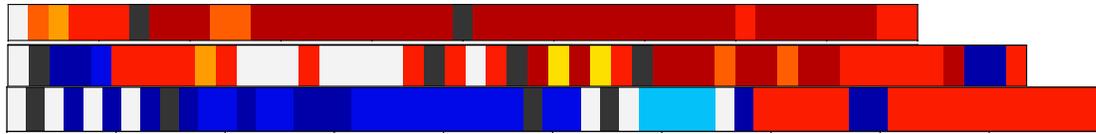
Step 6



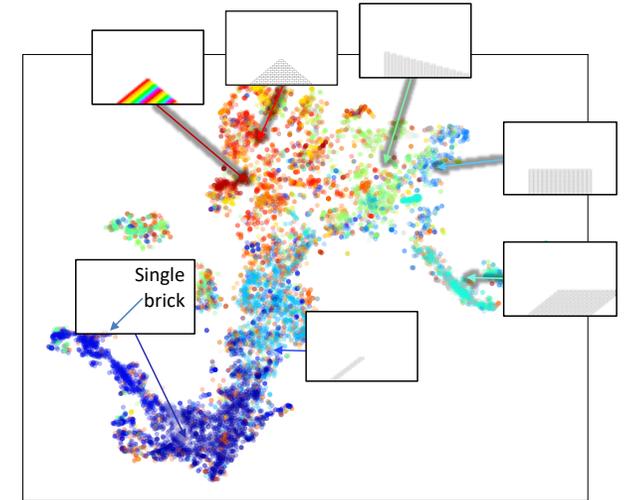
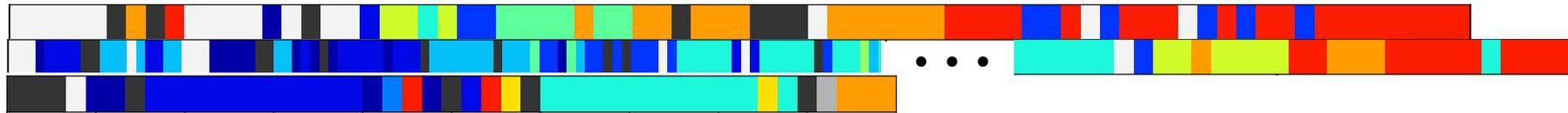
Understanding Process



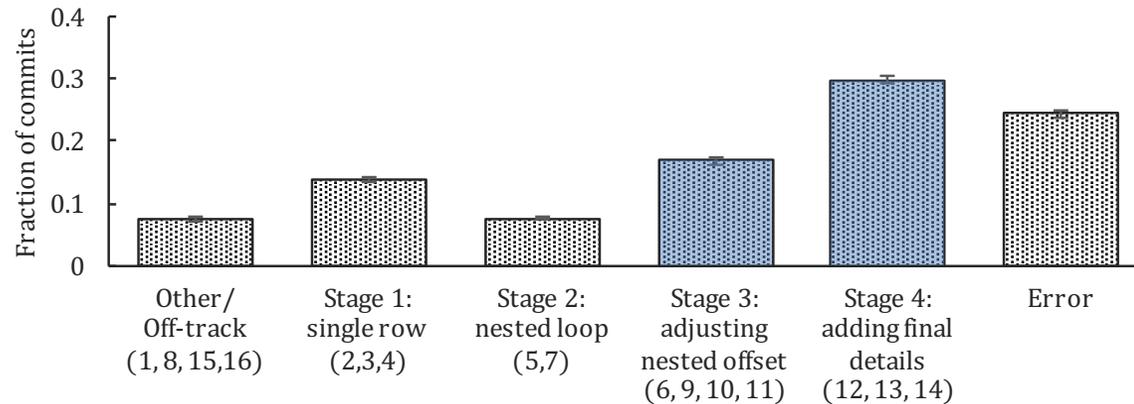
Students scoring in 99th percentile on midterm exam



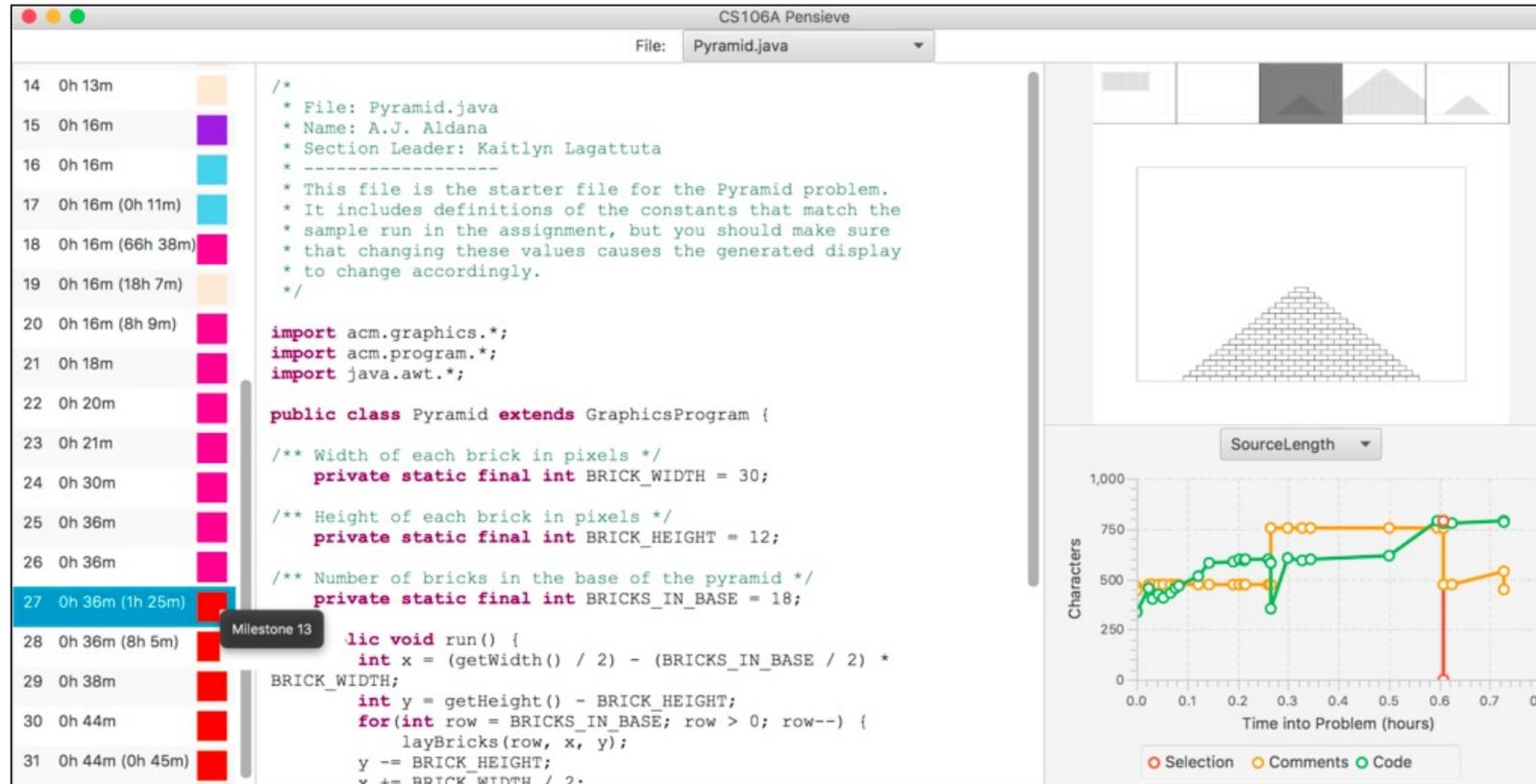
Students scoring in $\leq 3^{\text{rd}}$ percentile on midterm exam



t-SNE embedding of 130,000 partial solutions



Understanding Process



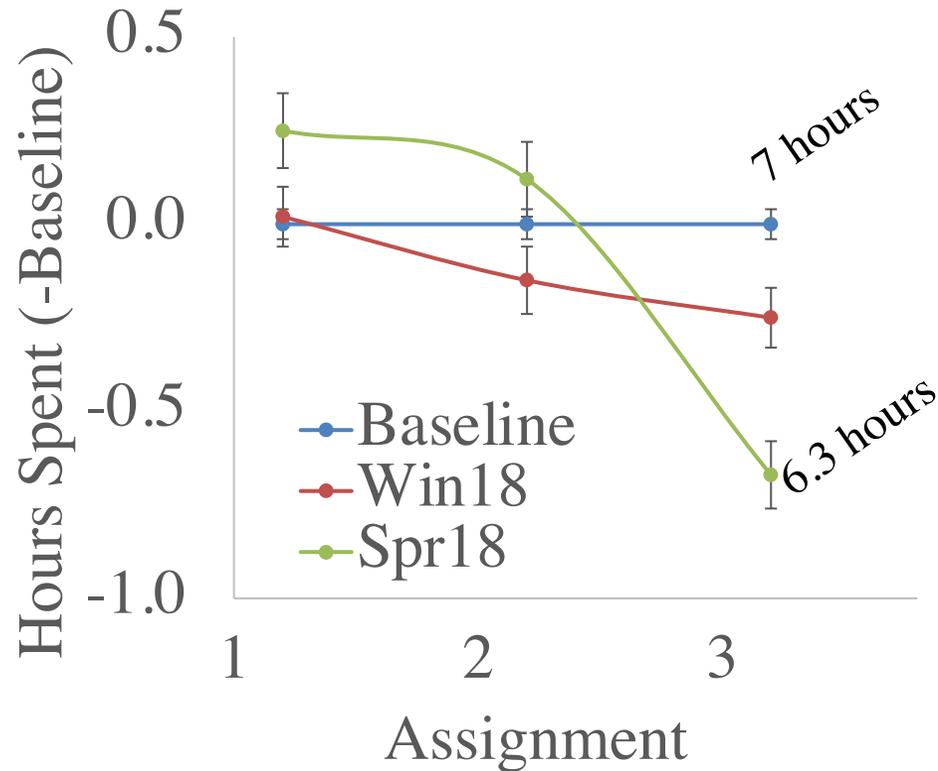
One quarter, all students were *shown* their progress:

- Early correction of bad habits.
- Chance to teach the art of programming.
- Academic dishonesty becomes much harder.



Understanding Process

Using assignment *timing* as pre-post

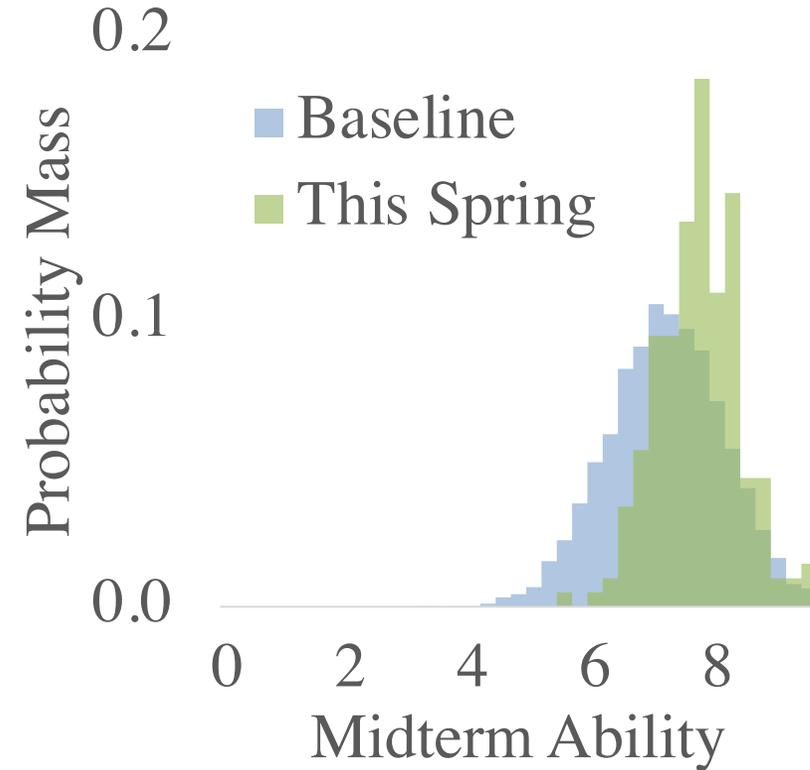


$$E[\hat{X}_3 | X_1] - E[X_3] = 42 \text{ mins}$$

Predicted time Actual time $p < 0.00001$

Assignments are taking less time

Item Response Theory based ability assessment



$$S_{i,j} = n \cdot \sigma(a_i - d_j)$$

Score points ability difficulty

Students perform better than expected



Feedback on *process* is an open problem

Our understanding of student work is not
perfect

Hit a ceiling with cs106a midterm

Midterm Grading Challenge

Browse Q3 | DynamoDB · AWS Console

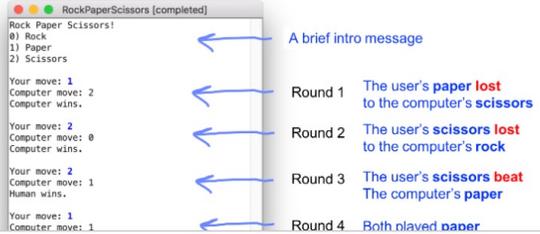
localhost:3000/#/view/cs106aMidWin19/3/abigail8/cpiech/8daa714e79a10cb

Question | Solution | Starter | Student | Answer

Rock Paper Scissors (30 points)



Write a `ConsoleProgram` that has a user play rock paper scissors against a computer until either the user or the computer has **three** "wins". To make the code simpler, use integers to represent the different plays (0 is rock, 1 is paper, 2 is scissors). Example run:



```
1 public class RockPaperScissors extends ConsoleProgram {
2
3     /* constants */
4     private static final int ROCK = 0;
5     private static final int PAPER = 1;
6     private static final int SCISSORS = 2;
7     private static final int N_WINS = 3;
8
9     private RandomGenerator rg = new RandomGenerator();
10
11    public void run() {
12        introMessage();
13        for (int i = 0; i < N_WINS; i++) {
14            inputNumber();
15            roundWinner();
16        }
17        gameWinner;
18    }
19
20    private void introMessage() {
21        println("Rock Paper Scissors!");
22        println("0) Rock");
23        println("1) Paper");
24        println("2) Scissors");
25        println(" ");
26    }
27
28    private void inputNumber() {
29        int a = readInt("Your move: ");
30        int computerMove = rg.nextInt(3);
31        int b = readInt("Computer move: " + computerMove);
32
33        private int computer = 0;
34        private int human = 0;
35
36        private void roundWinner {
37            if (a == 1 && b == 0) {
38                println("Human wins.");

```

Browse Answers

Grade: 26/30 points ✓ grade submitted!

Rubric: Standard

- For loop instead of a while loop (should go up to 3 wins by computer or user not 3 games)

User Inputting Move (3 points)

- Perfect (0 points)
- Minor Error (1 points)
- Major Error (2 points)
- Totally Wrong (3 points)

Computer Generating Move (4 points)

- Perfect (0 points)
- Minor Error (1 points)
- Major Error (2 points)
- Major Errors/No Attempt (4 points)

Determining the winner of a round (6 points)

- Perfect (0 points)
- Minor Error (1 points)
- Major Error (2 points)

Exam / Question 3 / Student

< Previous | Submit Grade | Next >

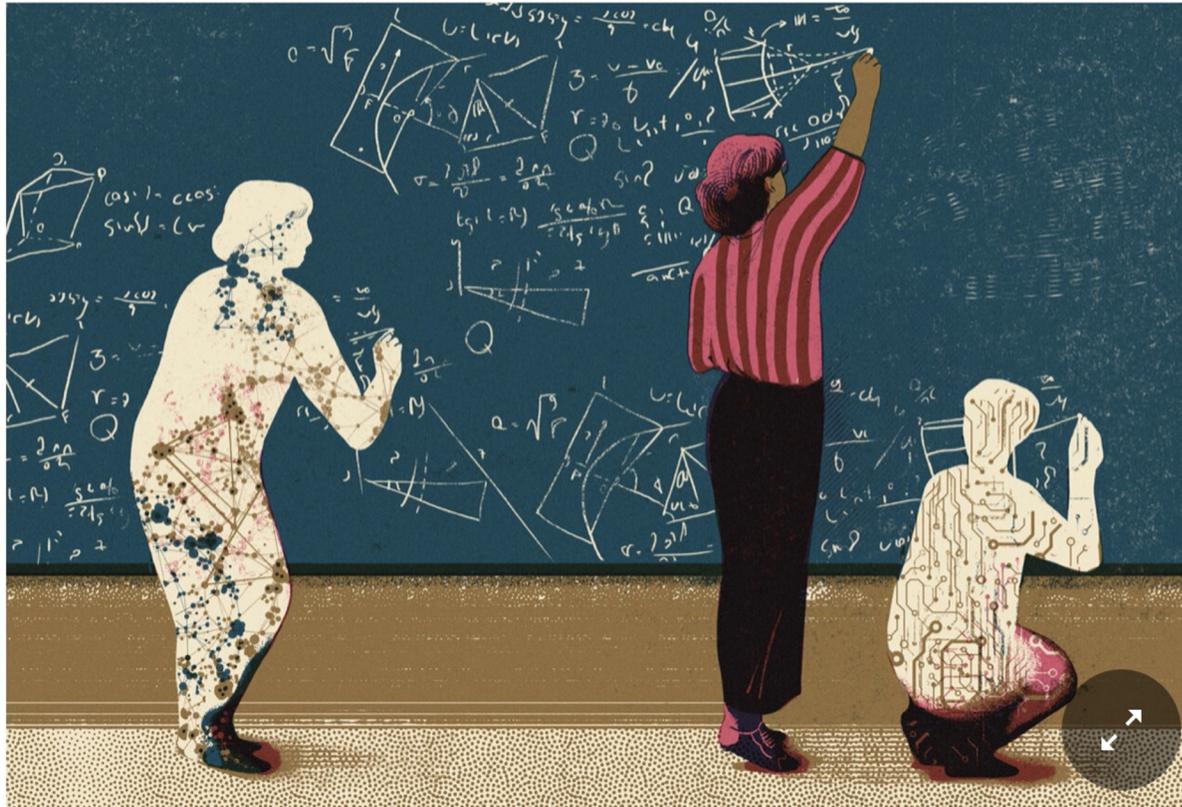
The
New York
Times

Can A.I. Grade Your Next Test?

Neural networks could give online education a boost by providing automated feedback to students.

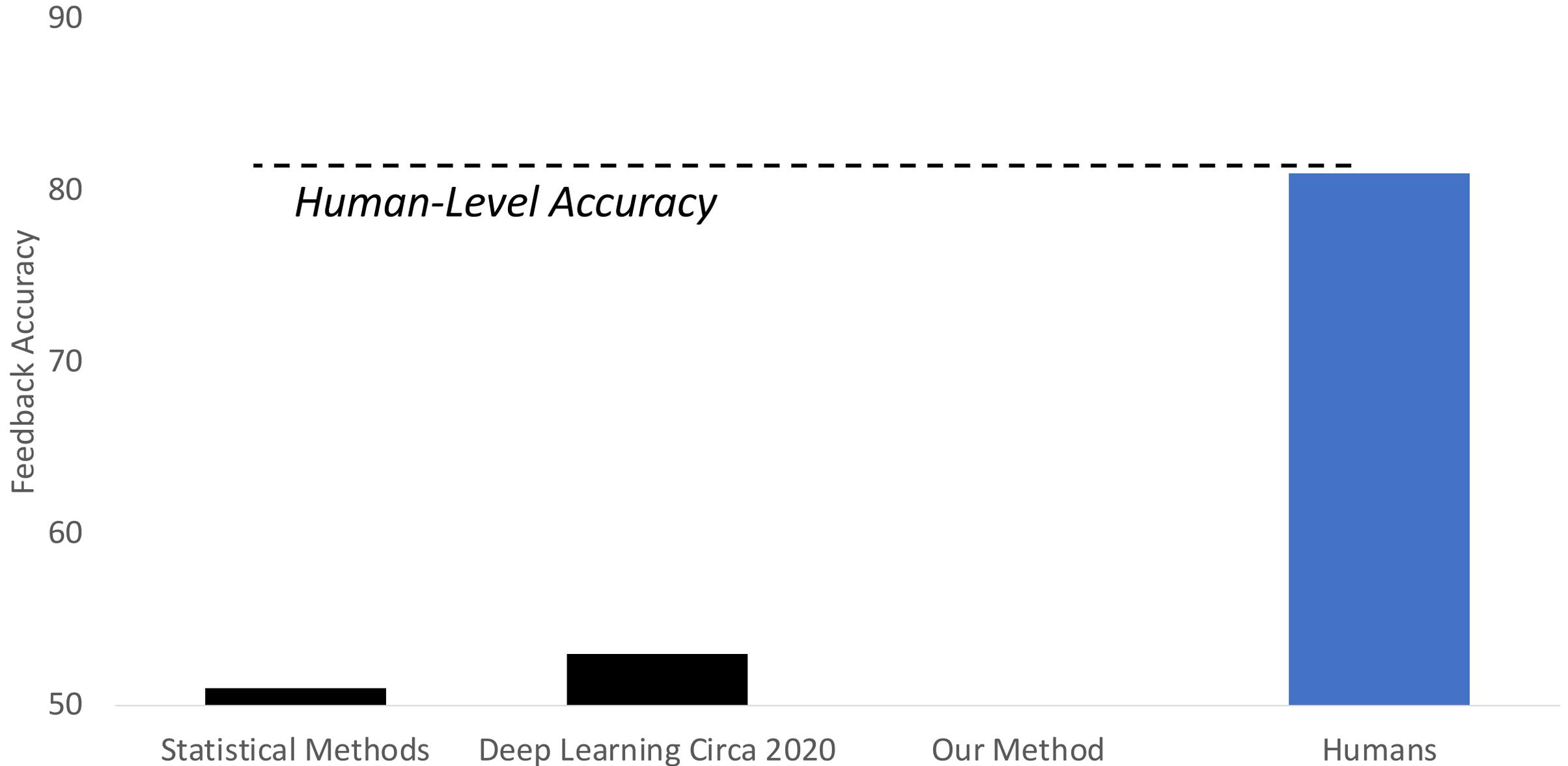


Stanford | News

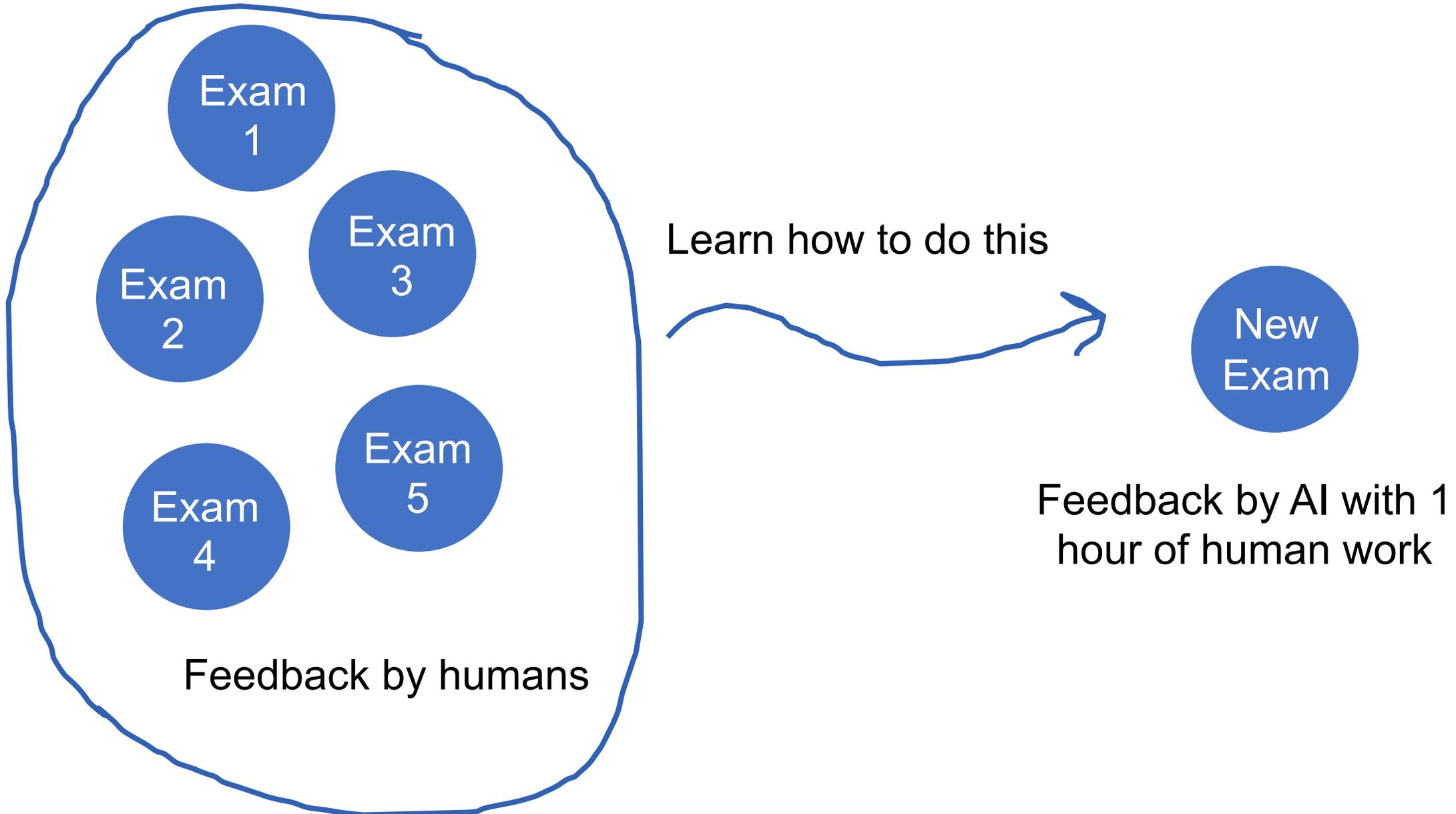


<https://www.nytimes.com/2021/07/20/technology/ai-education-neural-networks.html>

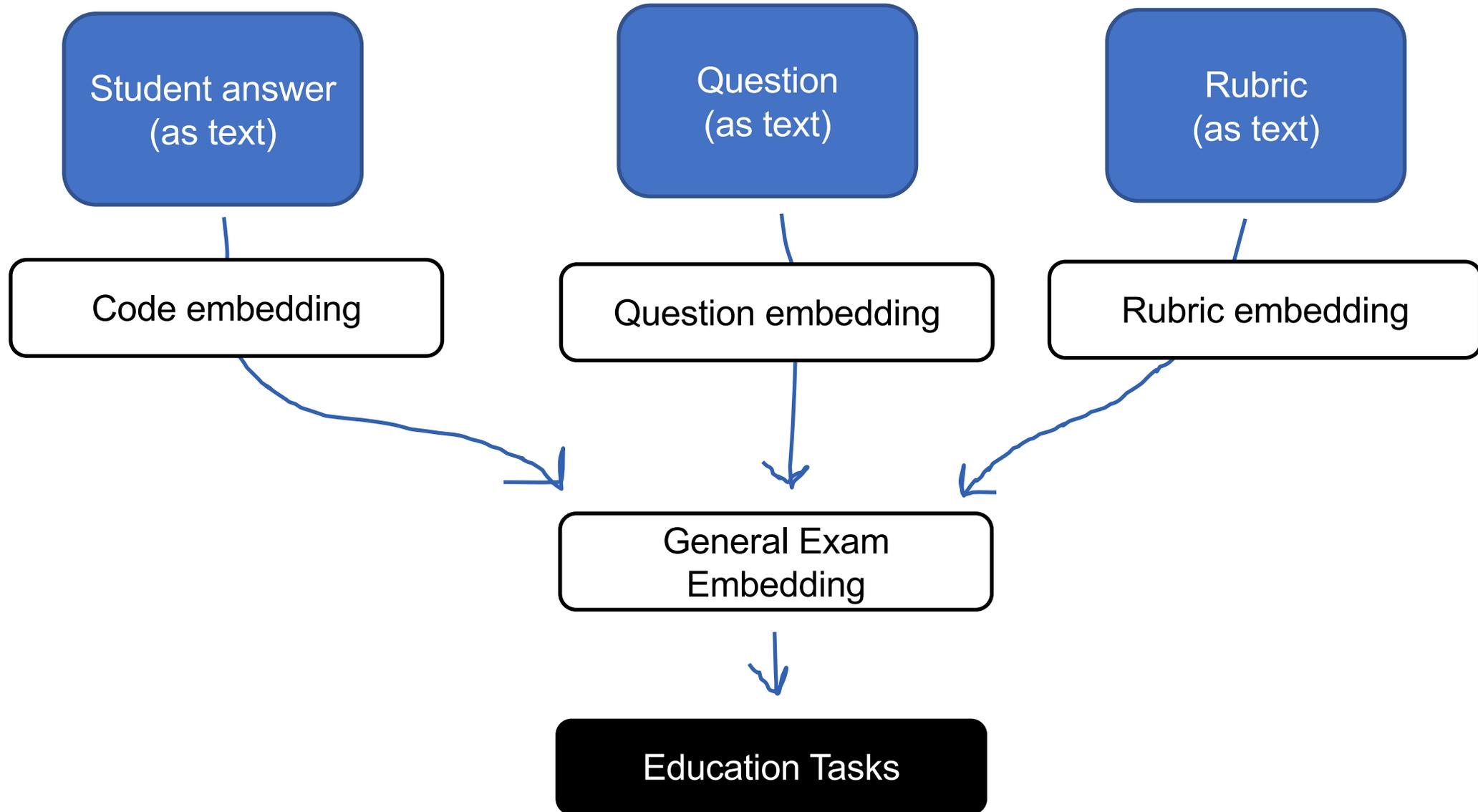
Rubric Level Accuracy on Few-Shot Grading a Novel Question



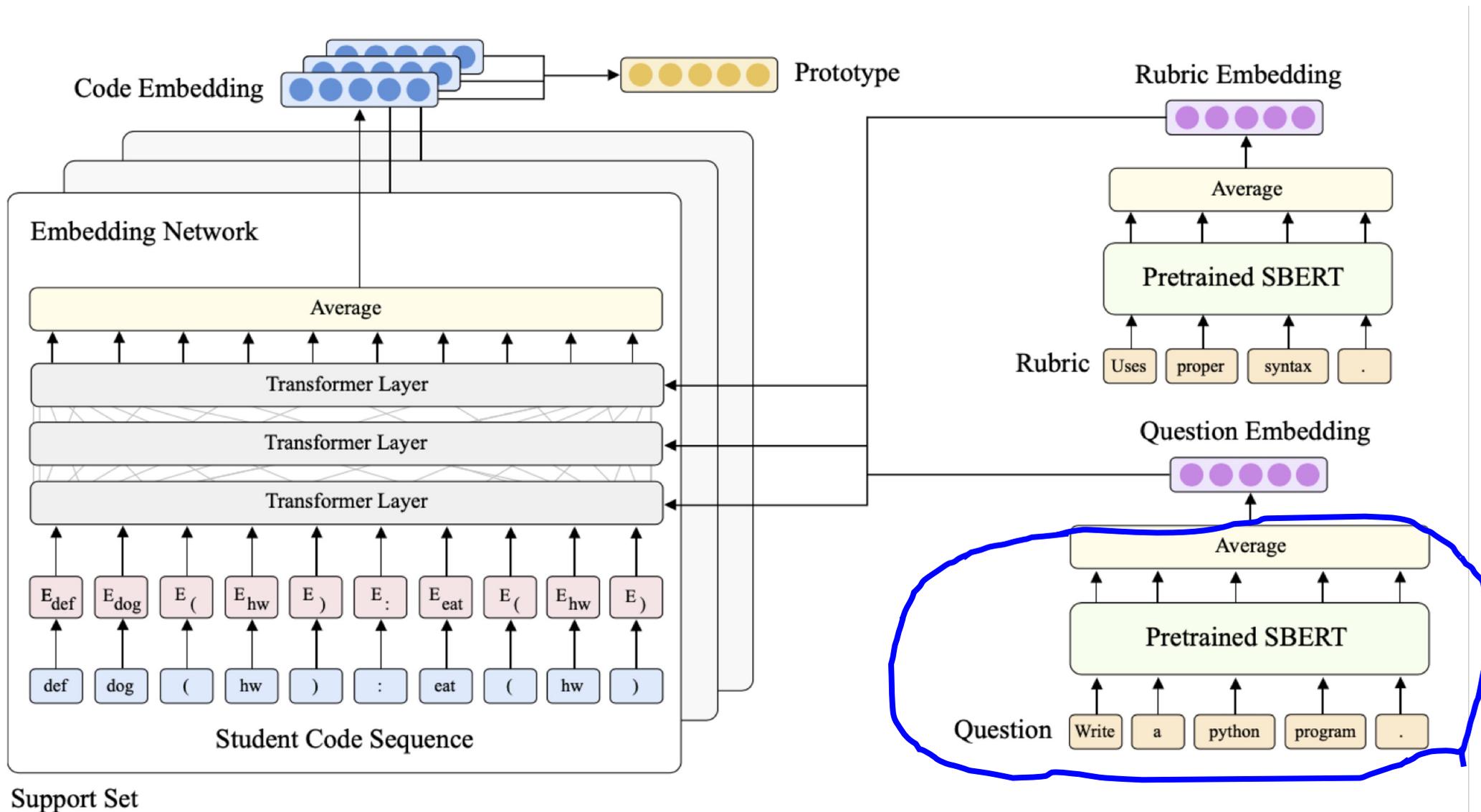
Give Feedback on Fresh Stanford Midterm



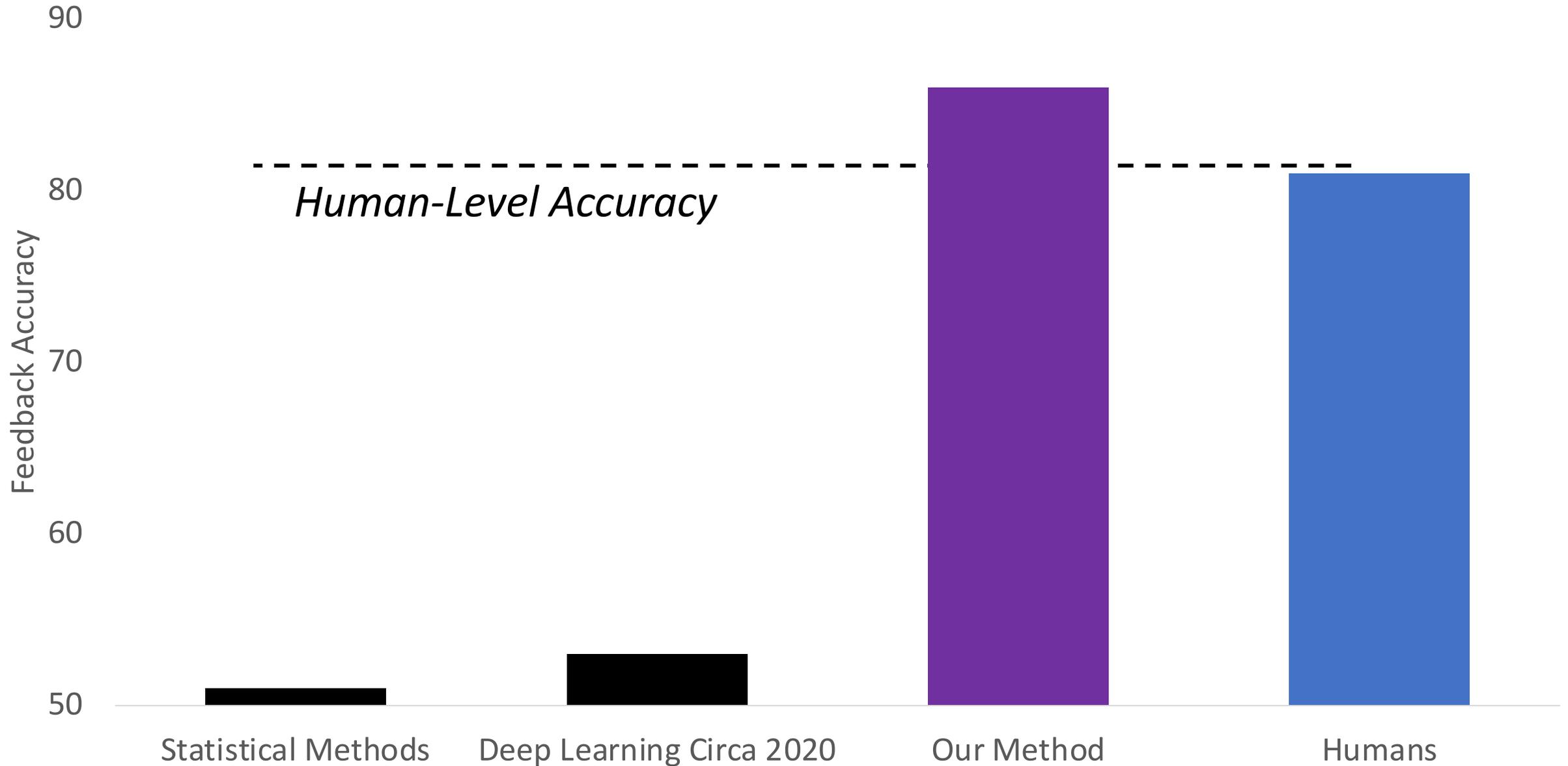
General Exam Grading Model



Invented the Proto-Transformer



Rubric Level Accuracy on Few-Shot Grading a Novel Question



Stanford Code in Place:



1100+ section leaders teach

12,000+ students

1/2 of CS106A

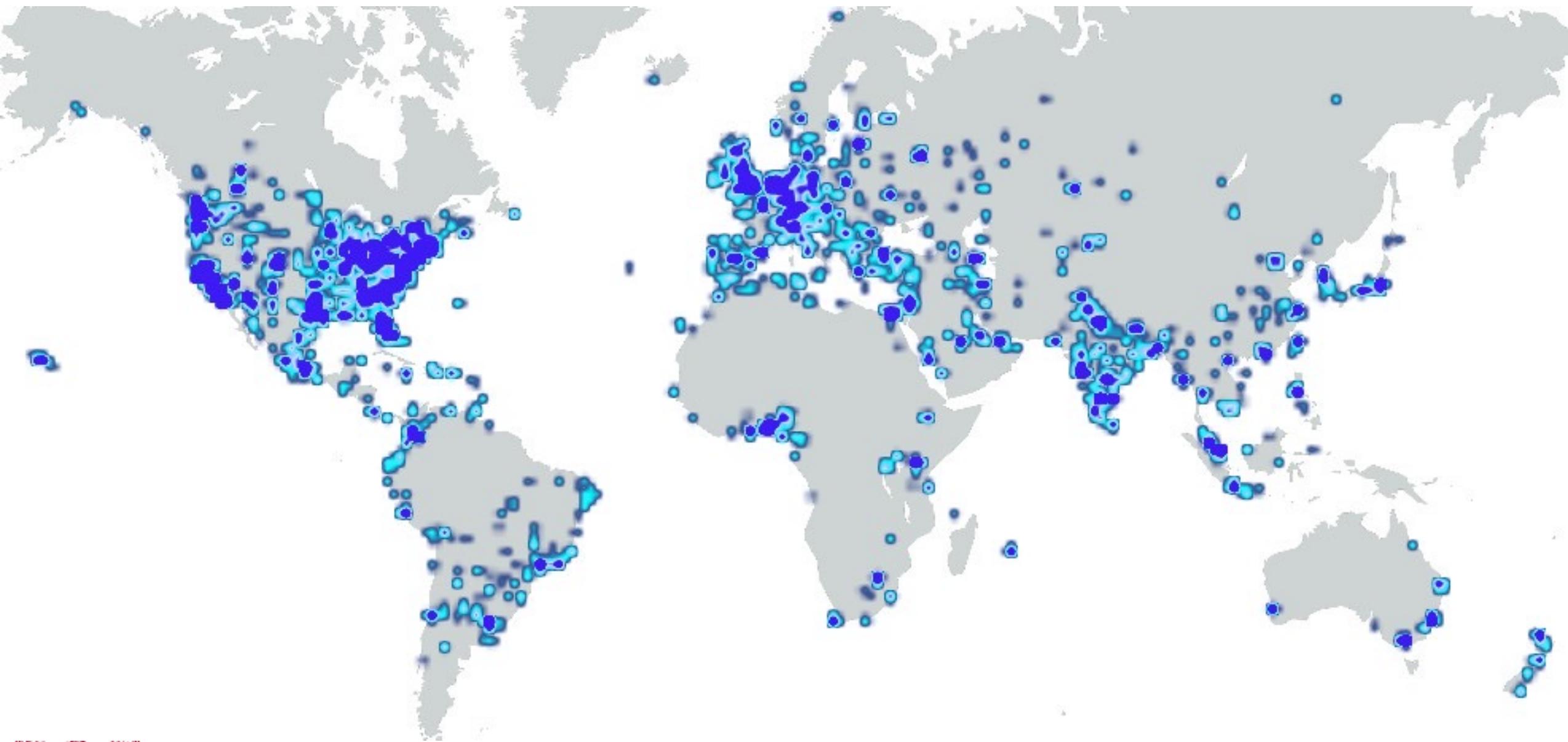
As Community Service

Featured in



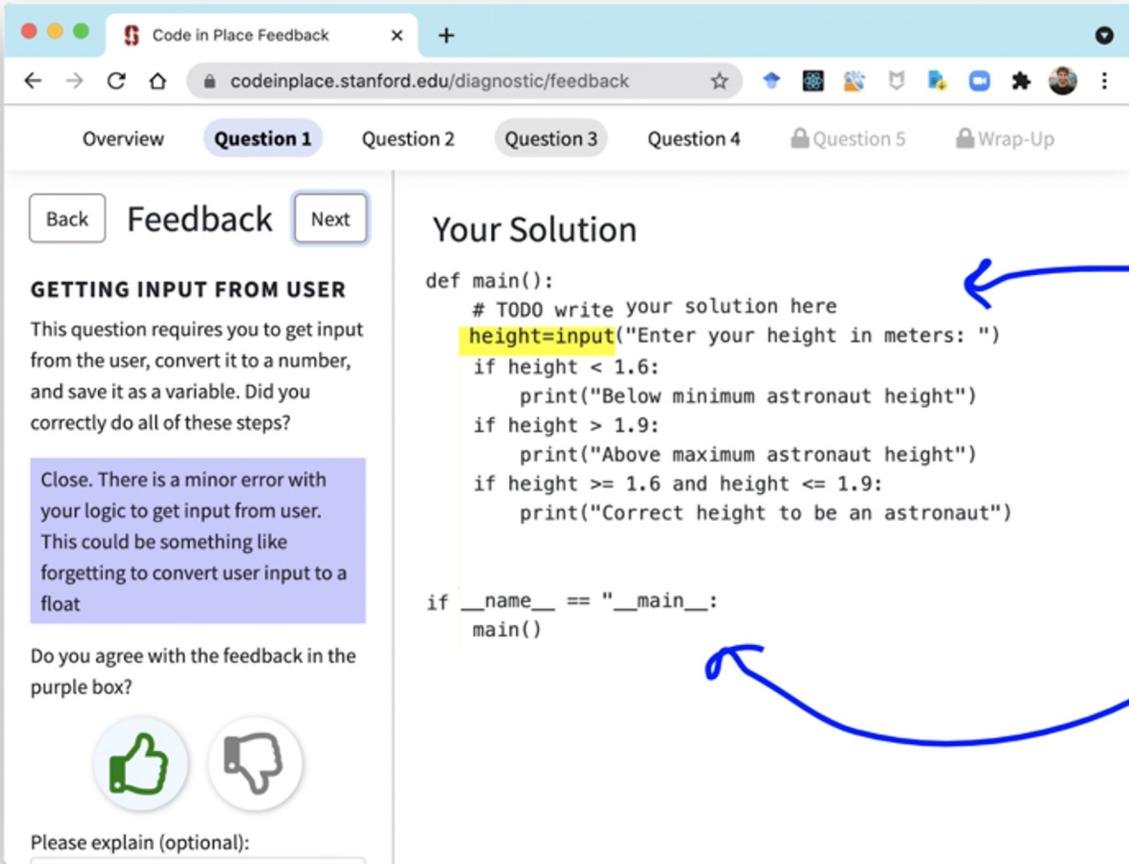
**SCIENTIFIC
AMERICAN**





Gave Feedback to 3,500 Real Students

Do you agree? AI feedback **97.9%**. Human feedback 96.7%



The screenshot shows a web browser window titled "Code in Place Feedback" with the URL "codeinplace.stanford.edu/diagnostic/feedback". The page is divided into two main sections: "GETTING INPUT FROM USER" and "Your Solution".

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

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

Do you agree with the feedback in the purple box?

Students evaluate the feedback

Your Solution

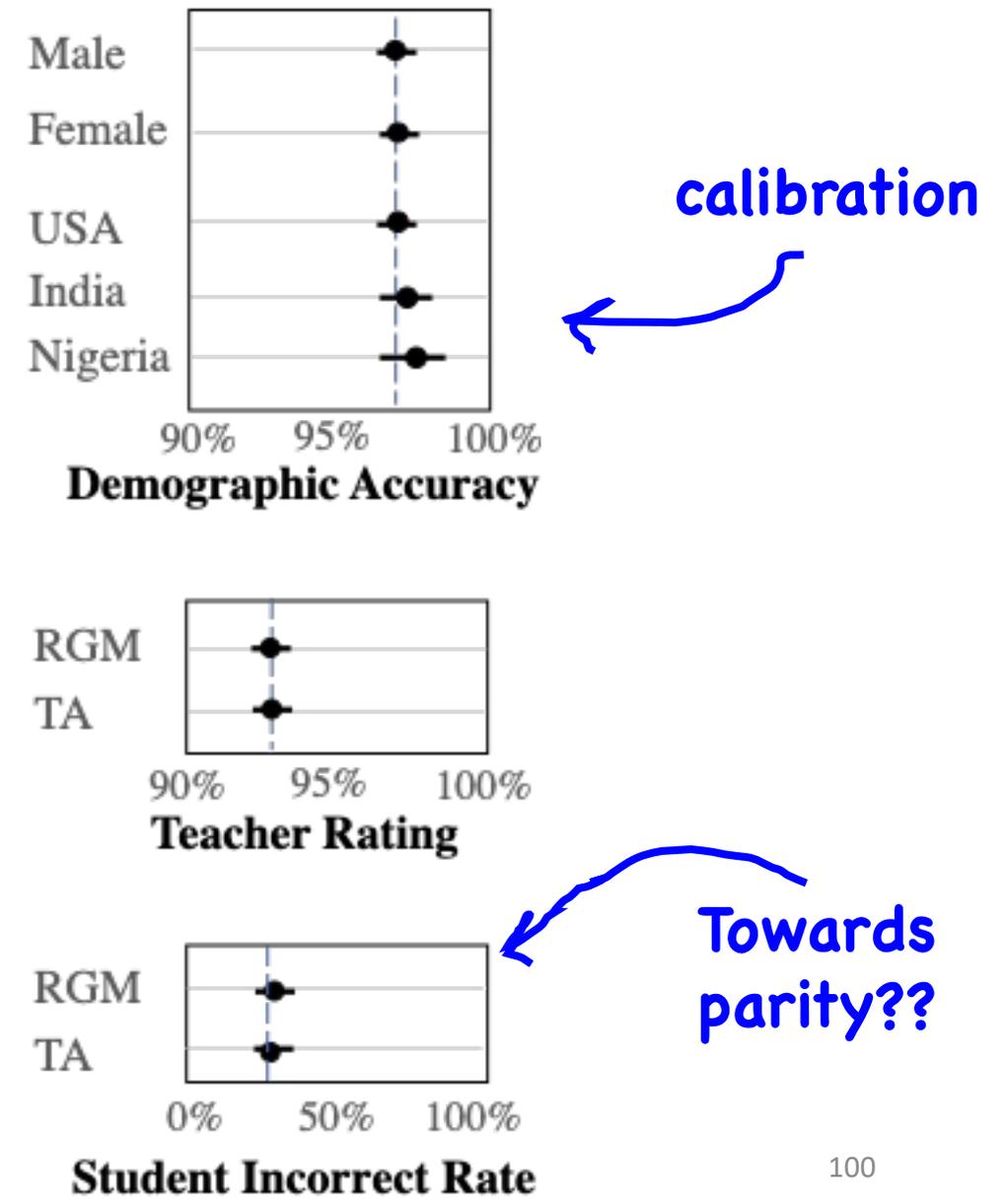
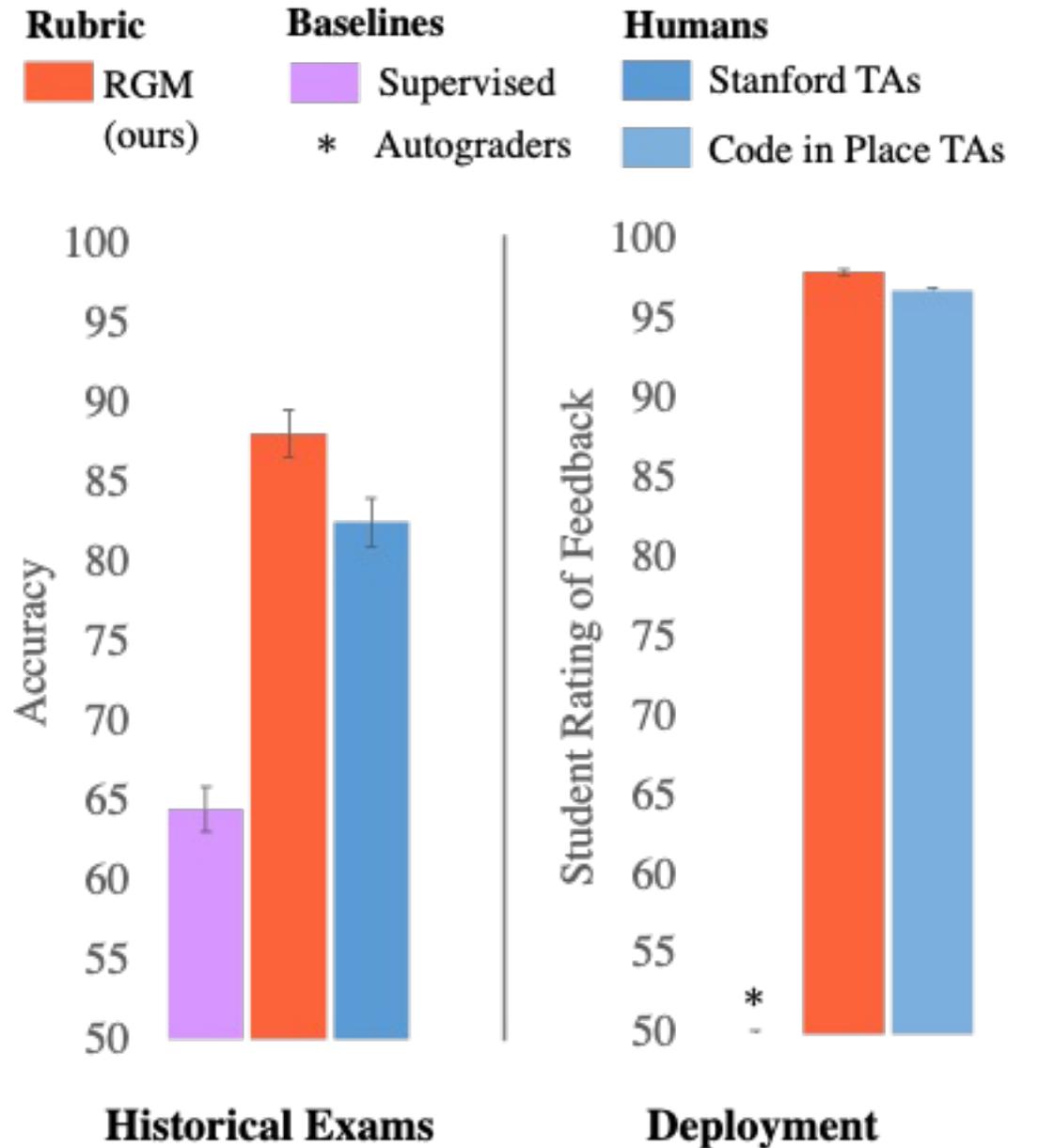
```
def main():  
    # TODO write your solution here  
    height=input("Enter your height in meters: ")  
    if height < 1.6:  
        print("Below minimum astronaut height")  
    if height > 1.9:  
        print("Above maximum astronaut height")  
    if height >= 1.6 and height <= 1.9:  
    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

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

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

Please explain (optional):

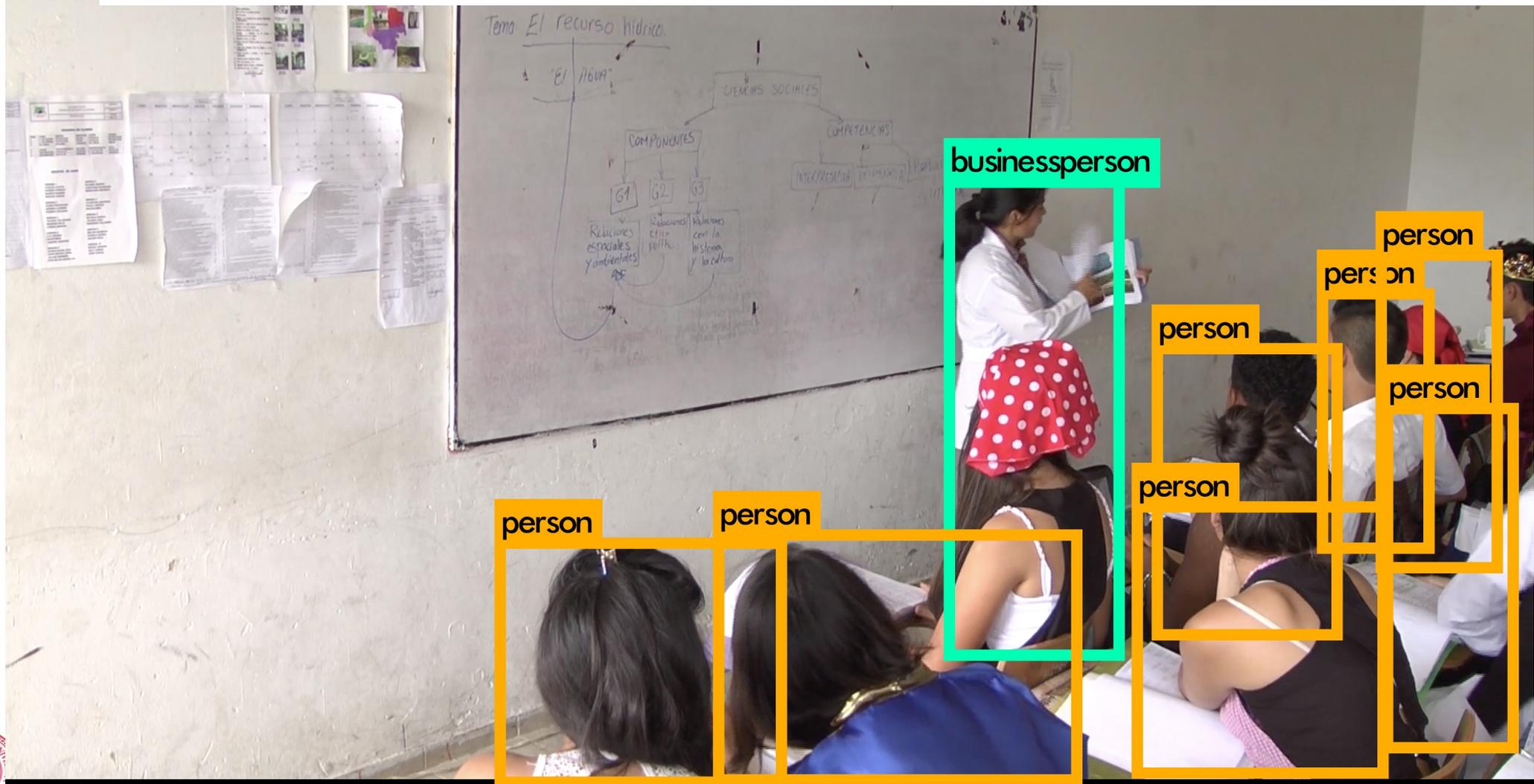
Impact and Fairness Analysis (work in progress)



That is just one example:
New problems

Feedback for Teachers

200,000 videos of teachers in Colombia, Chile and USA teaching



AI Teacher Training

1

Record Class



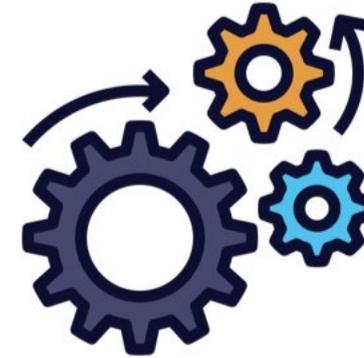
2

**Transcribe & Anonymize
Recording**



3

**Analyze
Transcripts**



4

**Generate
Insights**



Causal impact of AI teacher Training

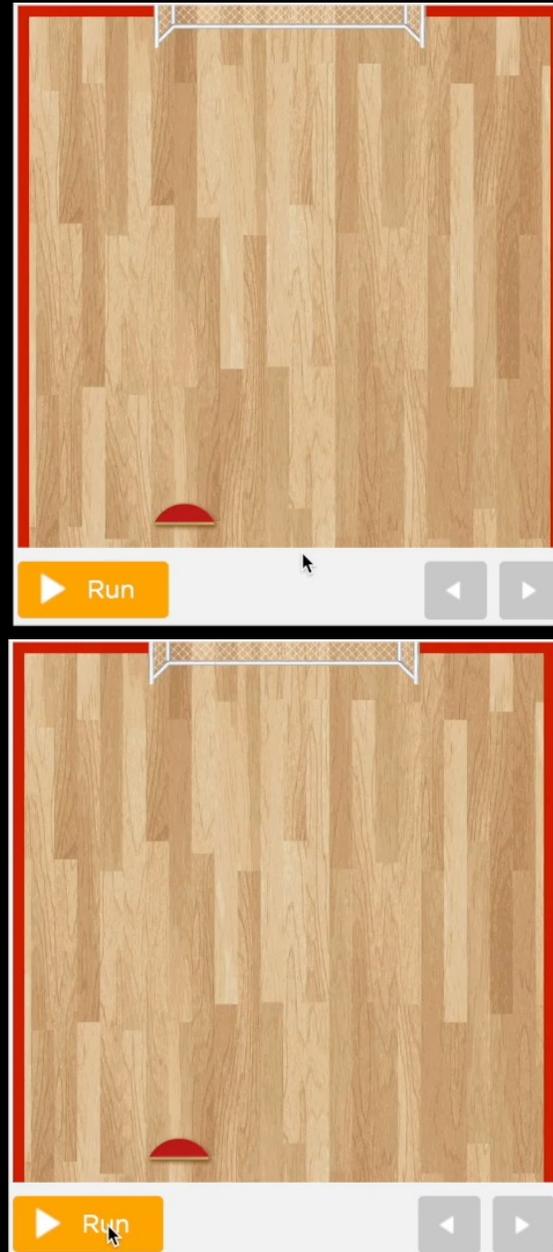
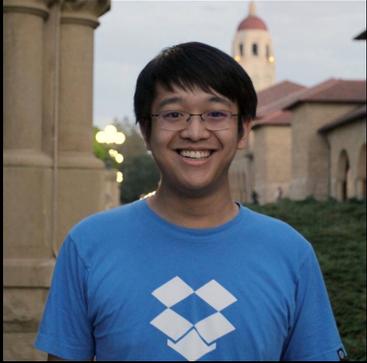
Within 4 lessons, as a result of training

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



Play 2 Grade

Allen Nie



Problem:

Grade the ~1M unique student implementations of this problem on code.org

Input:

Teacher gives you one example of each mistake on their rubric and one example of invariances

Theory contribution:

First model to build deep RL for a classification task . *Instead of learning an environment you are learning to test an environment*

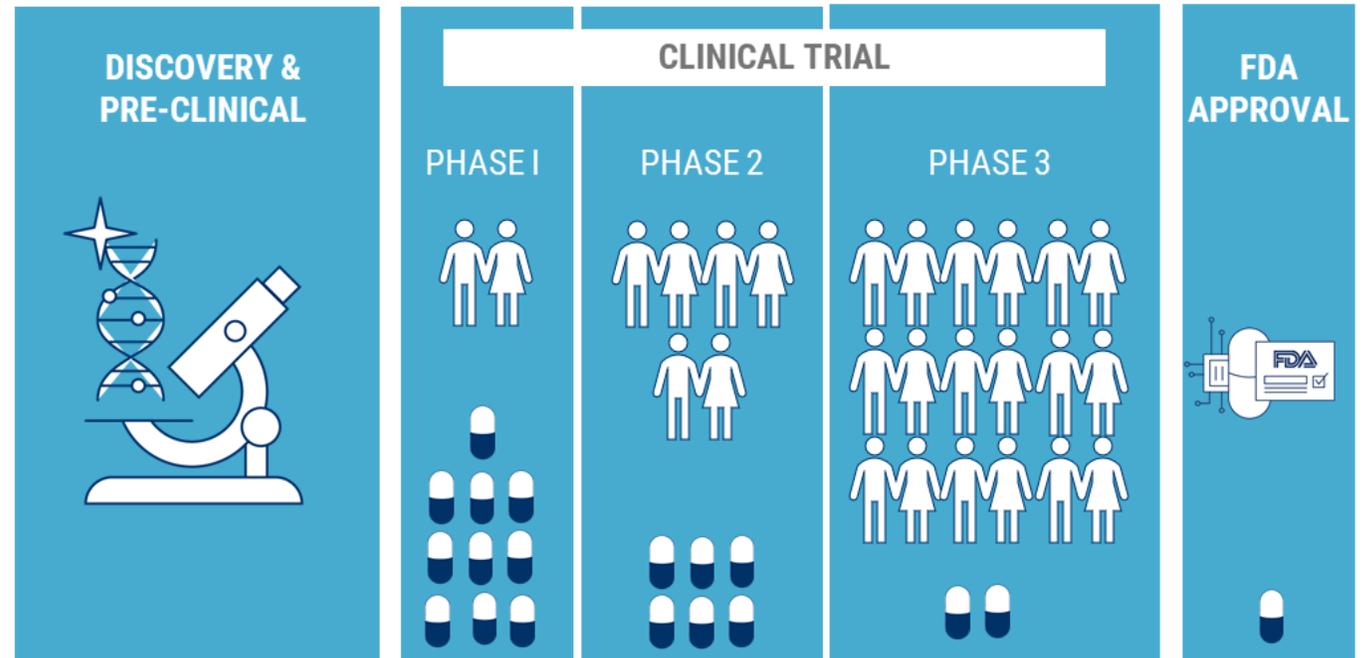
Impact:

Immediately change what sort of assignments are auto-gradable



More than education

 Bringing a drug to market is a drawn-out process



Source: cbinsights.com

 CBINSIGHTS

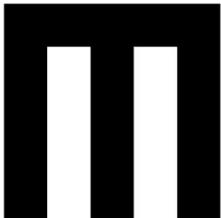


More than education

Vision Test

myeyes.ai/measure

Left Eye



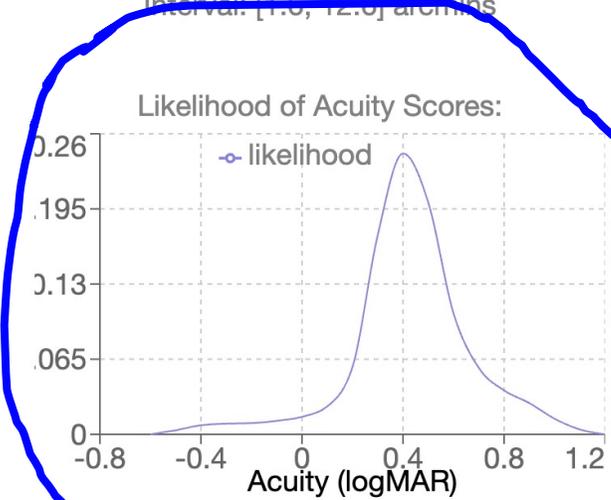
Featured in
THE LANCET

Progress: 10%

StAT Algorithm

N done: 2
Curr size: 3.3 arcmin
Curr size: 0.5 logMAR
MAP acuity: 2.5 arcmin
MAP acuity: 0.4 logMAR
Interval: [1.0, 12.0] arcmins

Likelihood of Acuity Scores:



Acuity (logMAR)	Likelihood
-0.8	0.00
-0.4	0.00
0.0	0.01
0.4	0.26
0.8	0.05
1.2	0.00



What else should be a **random variable**?

Grades??

Application -> Theory

Understand social science,
especially with small data

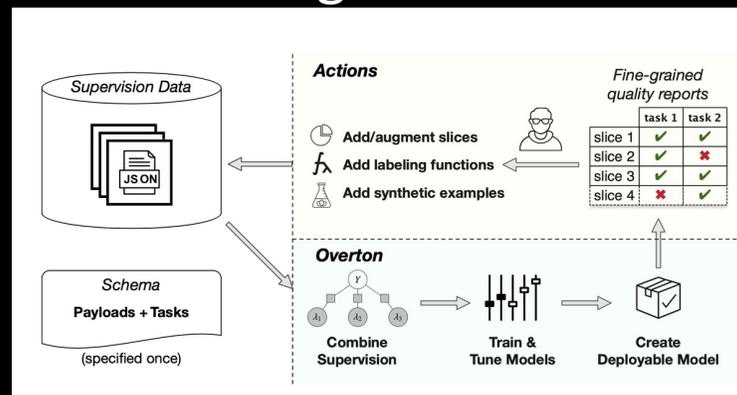
Explain why it made the
choices it did

What are things that AI
currently can't do?

Teach humans based on
what it has learned

Understand language

Design itself



What should you do
next?

Go solve amongst the abundance of important problems



Final Project | AA228/CS238

web.stanford.edu/clas...

Stanford University

AA228/CS238

Decision Making under Uncertainty

MENU

Final Project

The objective of the final project is to explore topics in decision making under uncertainty in greater depth than is permitted in class. The choice of topic is up to you, but it should be related to the general themes of the course. As part of the project you should:

- *describe* an approach (existing or newly developed),
- *apply* the approach to a problem of interest (which may or may not be related to aerospace), and
- *analyze* the performance of the approach according to a set of metrics.

CS221: Artificial Intelligence: P

stanford-cs221.github...



CS221: Artificial Intelligence: Principles and Techniques

Stanford / Autumn 2022-2023

[\[Calendar\]](#) [\[Modules\]](#) [\[Coursework\]](#) [\[Schedule\]](#)

- Lectures: Mon/Wed 1:30-2:50pm in NVIDIA Auditorium.
- Problem sessions: Fridays 1:30-2:20pm in Huang 018.
- Office hours, homework parties: see the [Calendar](#).
- To contact the teaching staff, please use Ed; for more personal/sensitive matters, email cs221-aut22-23-lead-staff@lists.stanford.edu.

Teaching Staff



Percy Liang
Instructor



Dorsa Sadigh
Instructor

CS229: Machine Learning

cs229.stanford.edu

CS229

CS229: Machine Learning

Instructors



Andrew Ng



Moses Charikar



Carlos Guestrin

Course Description This course provides a broad introduction to machine learning and statistical pattern recognition. Topics include: supervised learning (generative/discriminative learning, parametric/non-parametric learning, neural networks, support vector machines); unsupervised learning (clustering, dimensionality reduction, kernel methods); learning theory

CS 228 - Probabilistic Graphical

ermongroup.github.io/...

CS 228 - Probabilistic Graphical Models

Winter 2021-22

[Ed](#)
[Calendar](#)
[Course Notes](#)

[Logistics](#) | [Course Info](#) | [Syllabus](#) | [Other Resources](#)

Logistics

- **Lectures:** Tue, Thu, 9:45am-11:15am, Nvidia Auditorium
- **Office Hours and Sections:** [Google Calendar](#)

Statistics 200: Introduction to

web.stanford.edu/clas...

Statistics 200: Introduction to Statistical Inference

Zhou Fan, Stanford University, Autumn 2016

Home

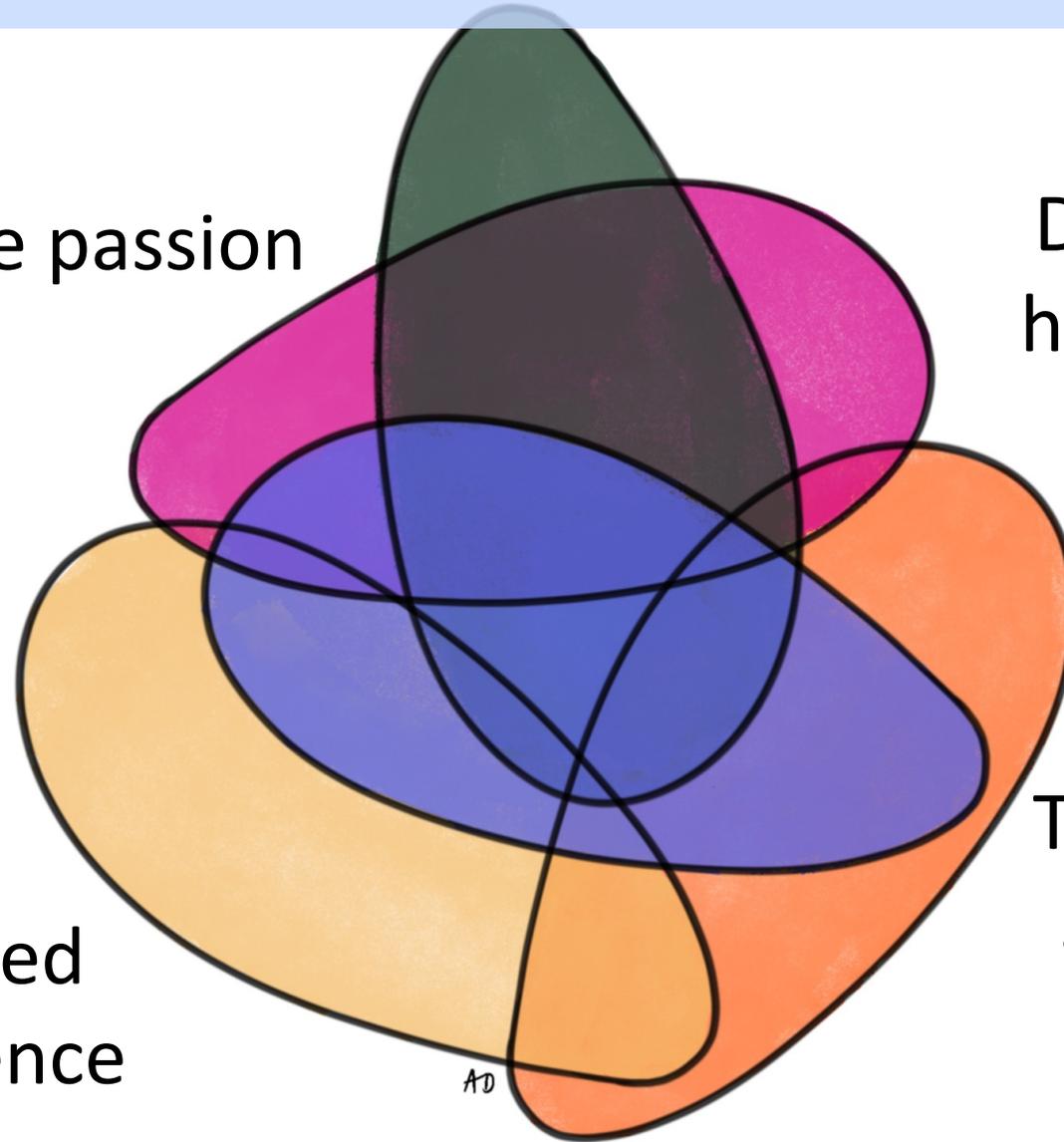
- [Lectures](#)
- [Homework](#)
- [Grades](#)
- [Piazza](#)



Think about intersectionality

Your side passion

Data that you
have access to



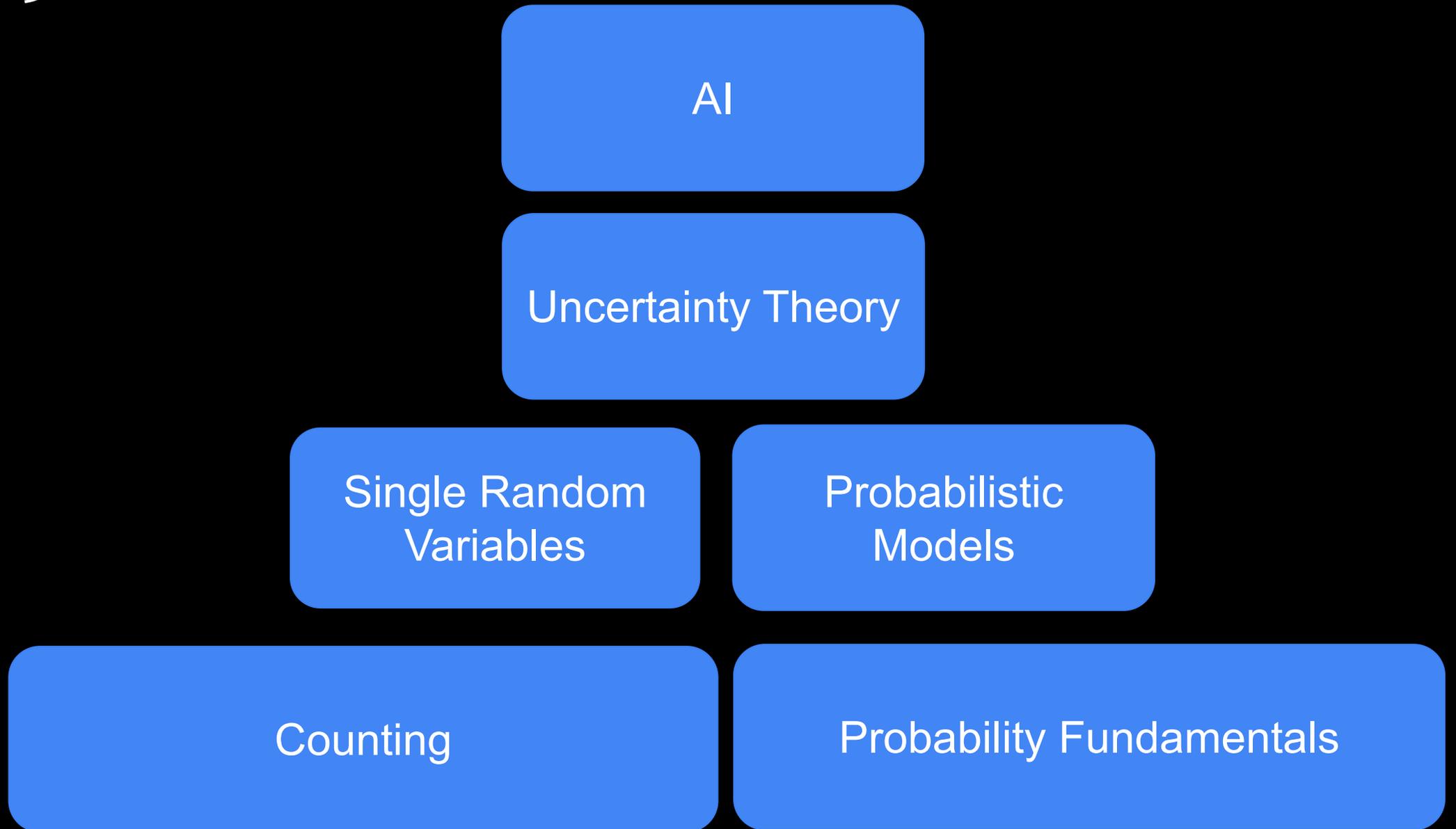
Your lived
experience

Thompson
sampling

AD

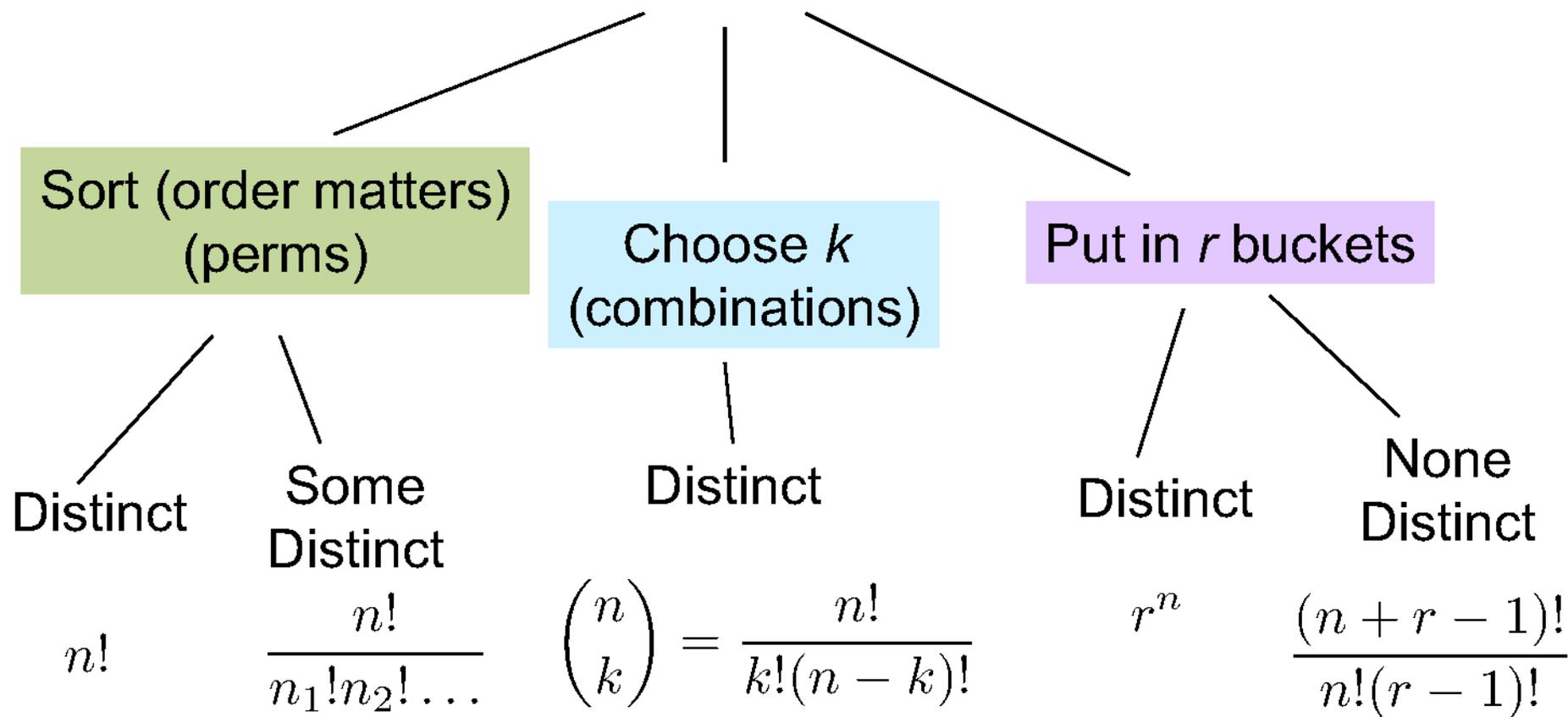


Last Class...



Counting Rules

Counting operations on n objects



Counting



Ayesha



Tim



Irina



Joey



Waddie



COURSE VALUES

Everyone is welcome.
Intellectual joy. Be kind. Be humane. Social connection.
Learn by doing. Thrill of building. Adapt to new contexts. Especially in a hard time 🦠.



WHEN YOU MEET YOUR BEST FRIEND

Somewhere you didn't expect to.



Trailing the dovetail shuffle to it's lair – Persi Diaconosis

What is a Probability?

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



Netflix and Learn

$$P(E|F) = \frac{P(EF)}{P(F)}$$

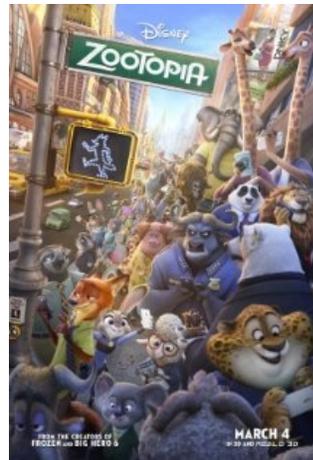
Definition of
Cond. Probability

- Let E be the event that a user watches the given movie.
- Let F be the event that the same user watches CODA (2021).



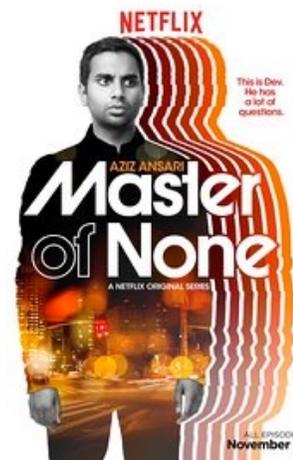
$$P(E) = 0.19$$

$$P(E|F) = 0.14$$



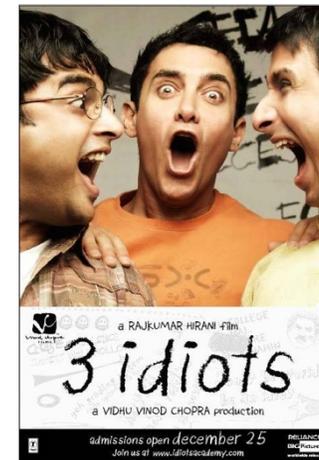
$$P(E) = 0.32$$

$$P(E|F) = 0.35$$



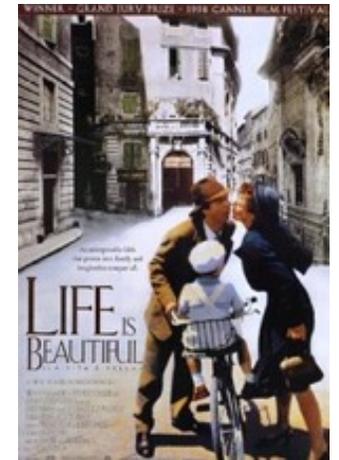
$$P(E) = 0.20$$

$$P(E|F) = 0.20$$



$$P(E) = 0.09$$

$$P(E|F) = 0.72$$



$$P(E) = 0.20$$

$$P(E|F) = 0.42$$

Monty Hall Problem

and Wayne Brady



Marilyn discovers the
Probability Bug

Zika Test



Positive Zika.

What is the probability of zika?

-
- *0.1% of people have zika*
 - *90% positive rate for people with zika*
 - *7% positive rate for people without zika*

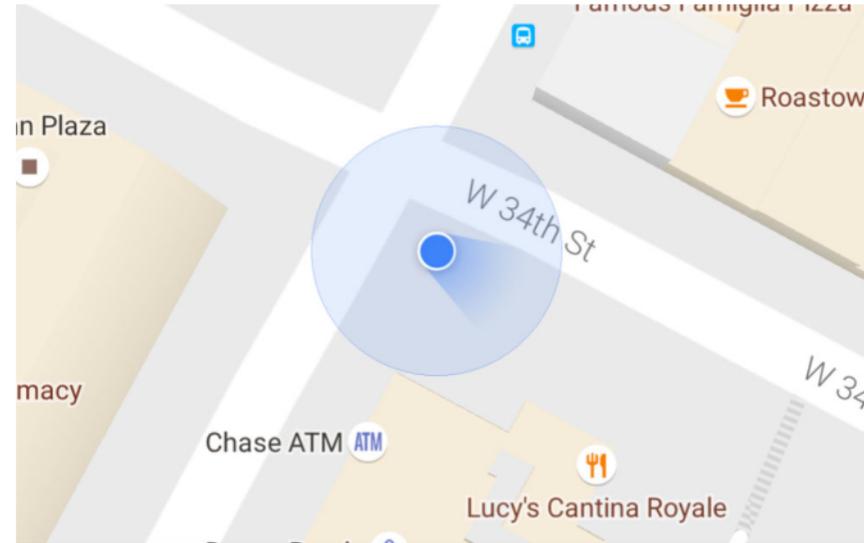
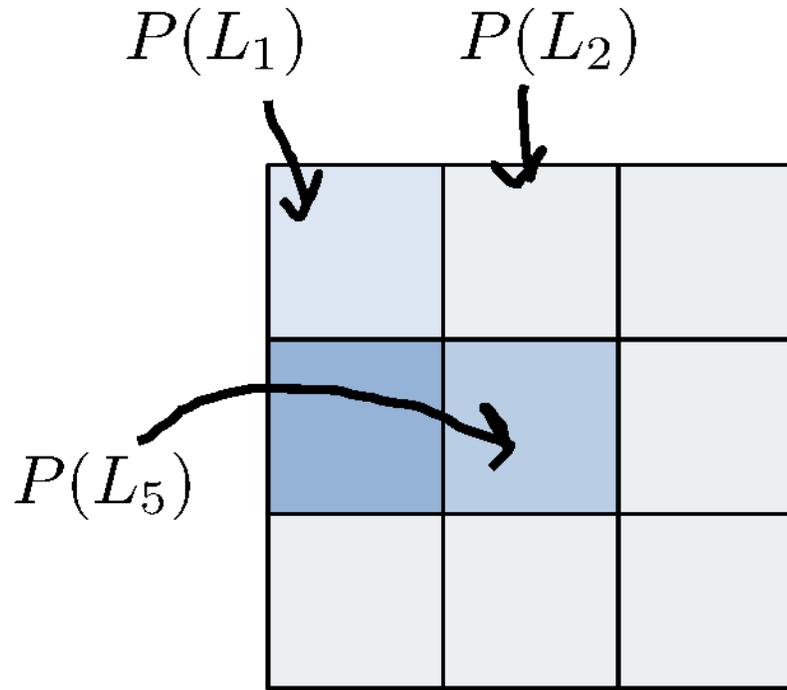
The right answer is 1%



Bayes Theorem Intuition



Update Belief

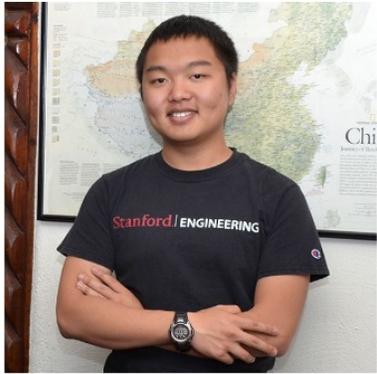


Before Observation



Recall our Ebola Bats





Third Year of Sections



I'm not a robot


reCAPTCHA
[Privacy - Terms](#)

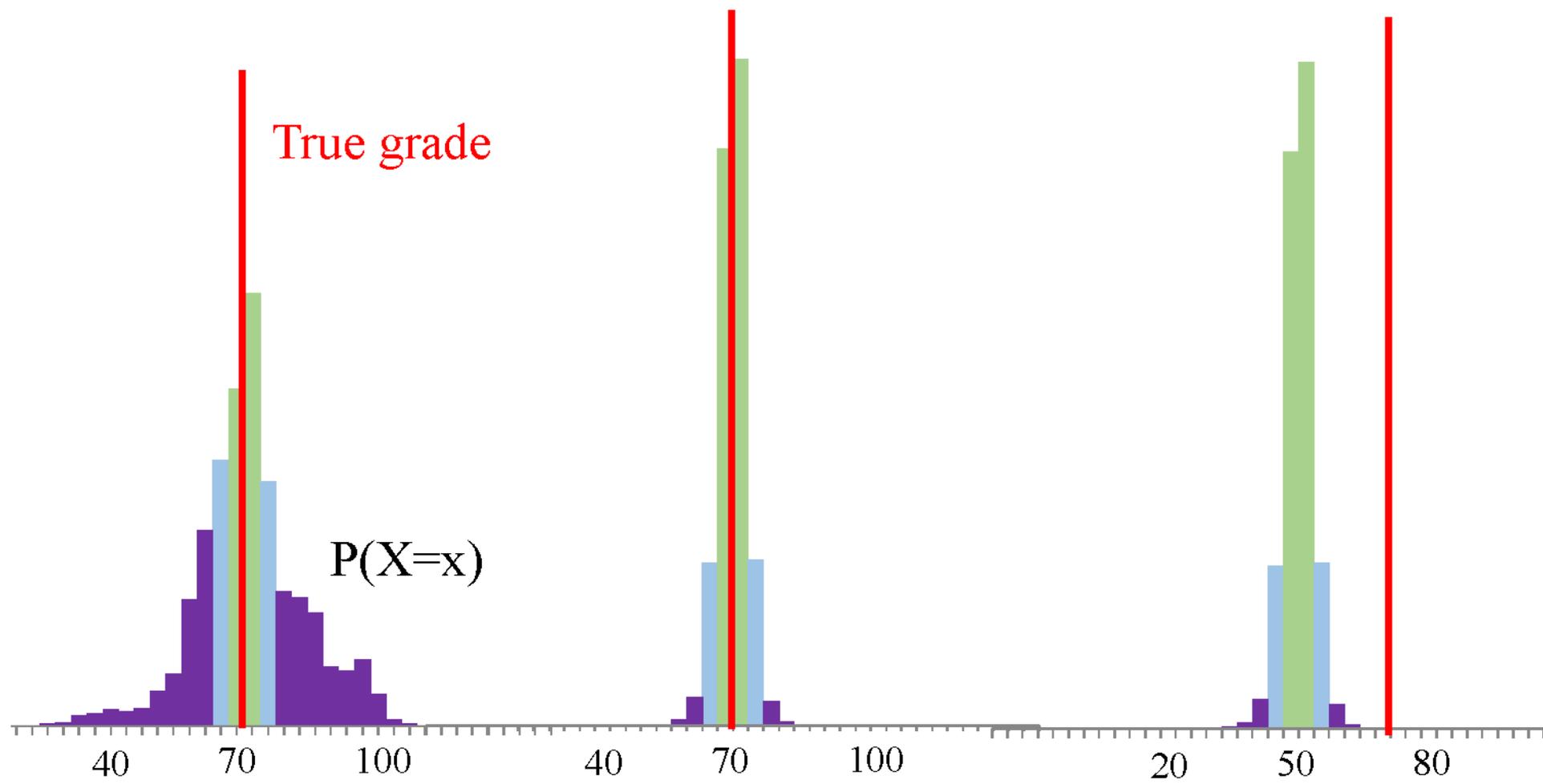


X		O
O	X	
		X



Random Variables

X is the score a peer grader gives to an assignment submission

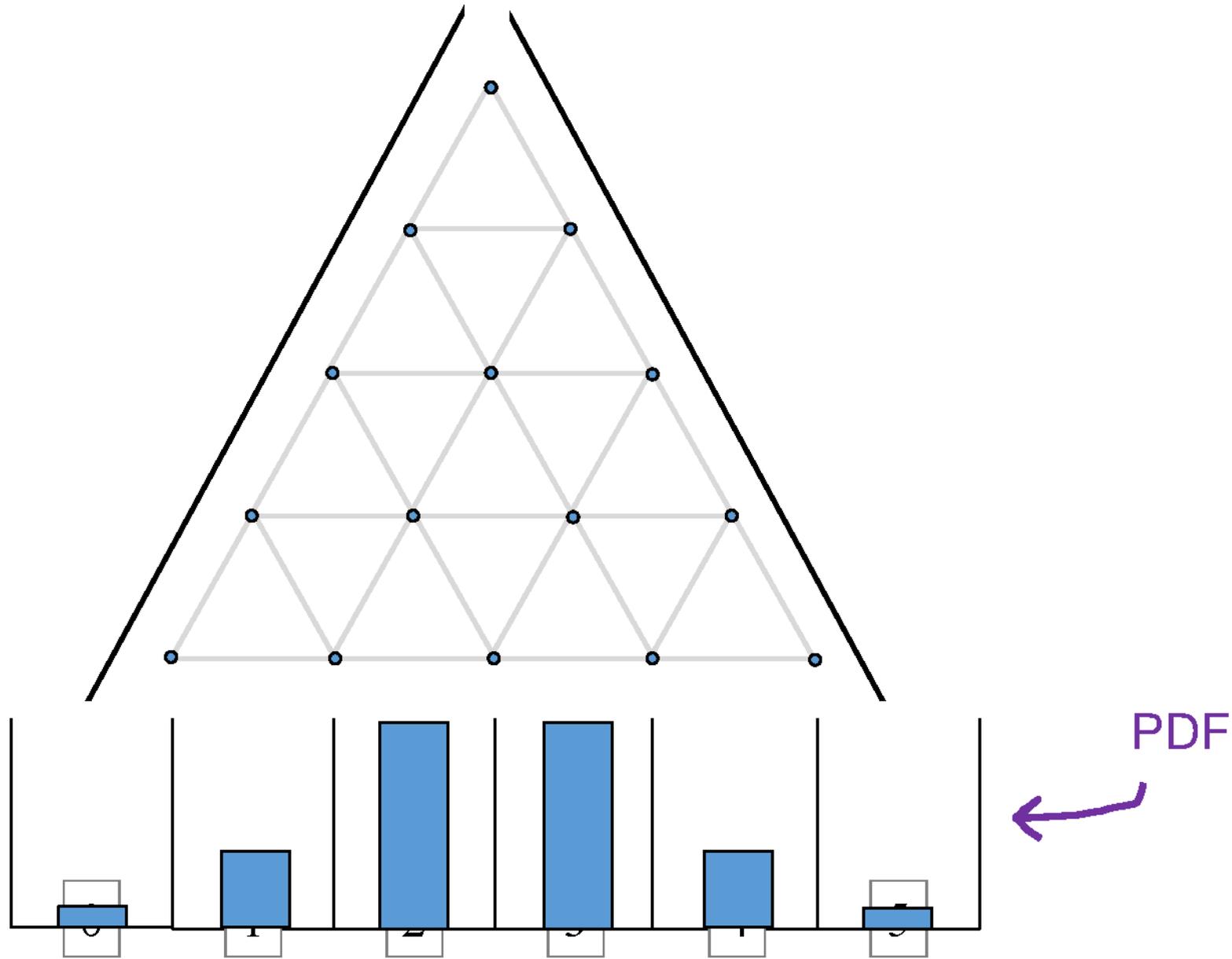


A

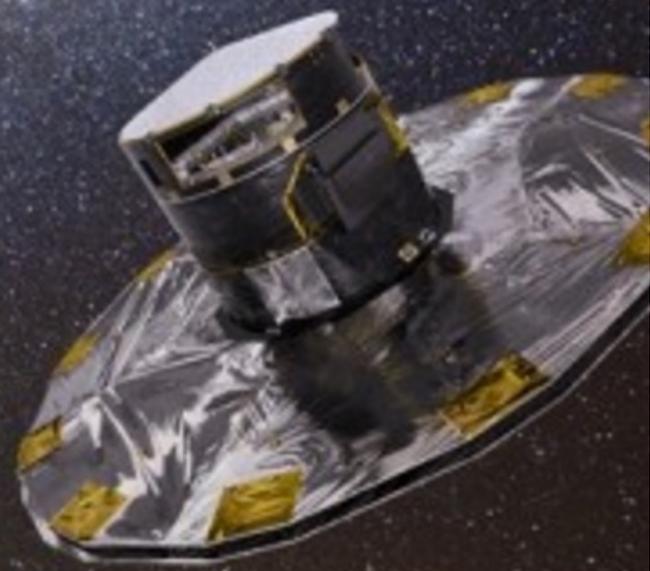
B

C

Binomial



1001

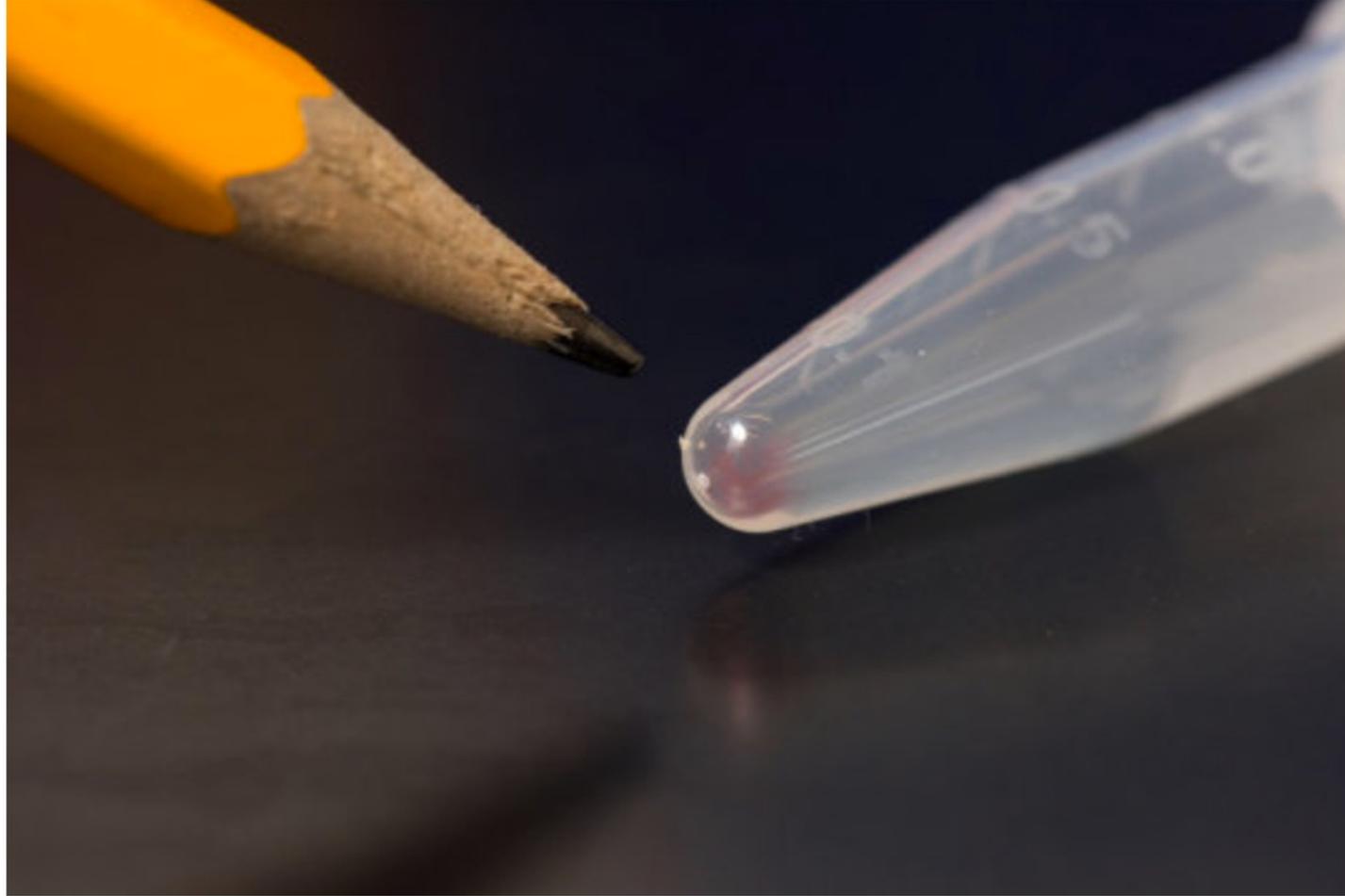


Geometric

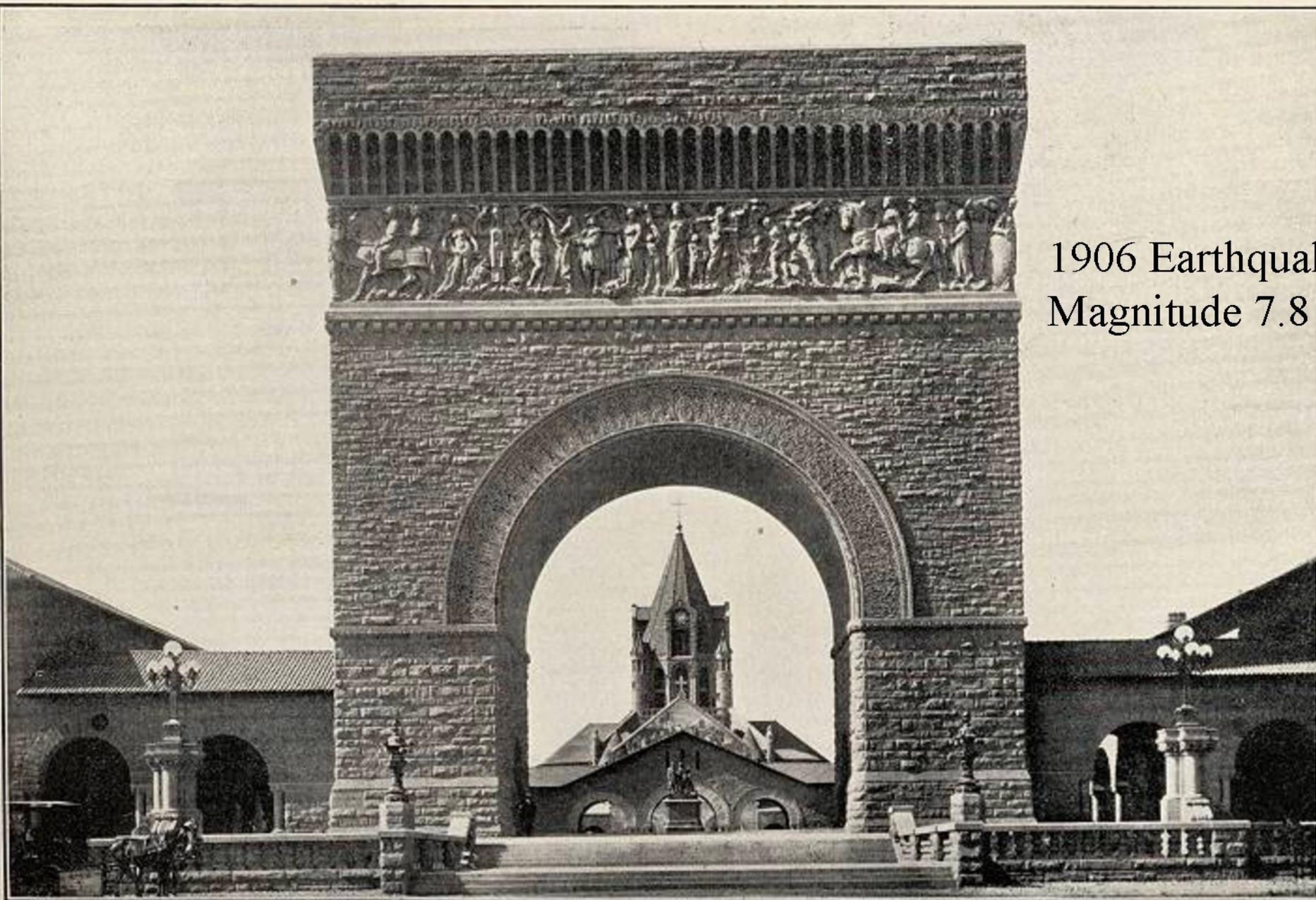
Sequence 1:

TTHTHTTHTTTHTTTHTTTHTTHTHTHT
HTTHTTTHHTHTHTTHTTHTTHTTHTT
HTHTHTHTHTTHTTHTTHTHTHTTHTTHT
TTHTHTTHTHTHTHTHTHTHTHTHTHTHT
TTHTHTHTHTHTHTTHTTHTTHTTHTHTHT

Storing Data on DNA



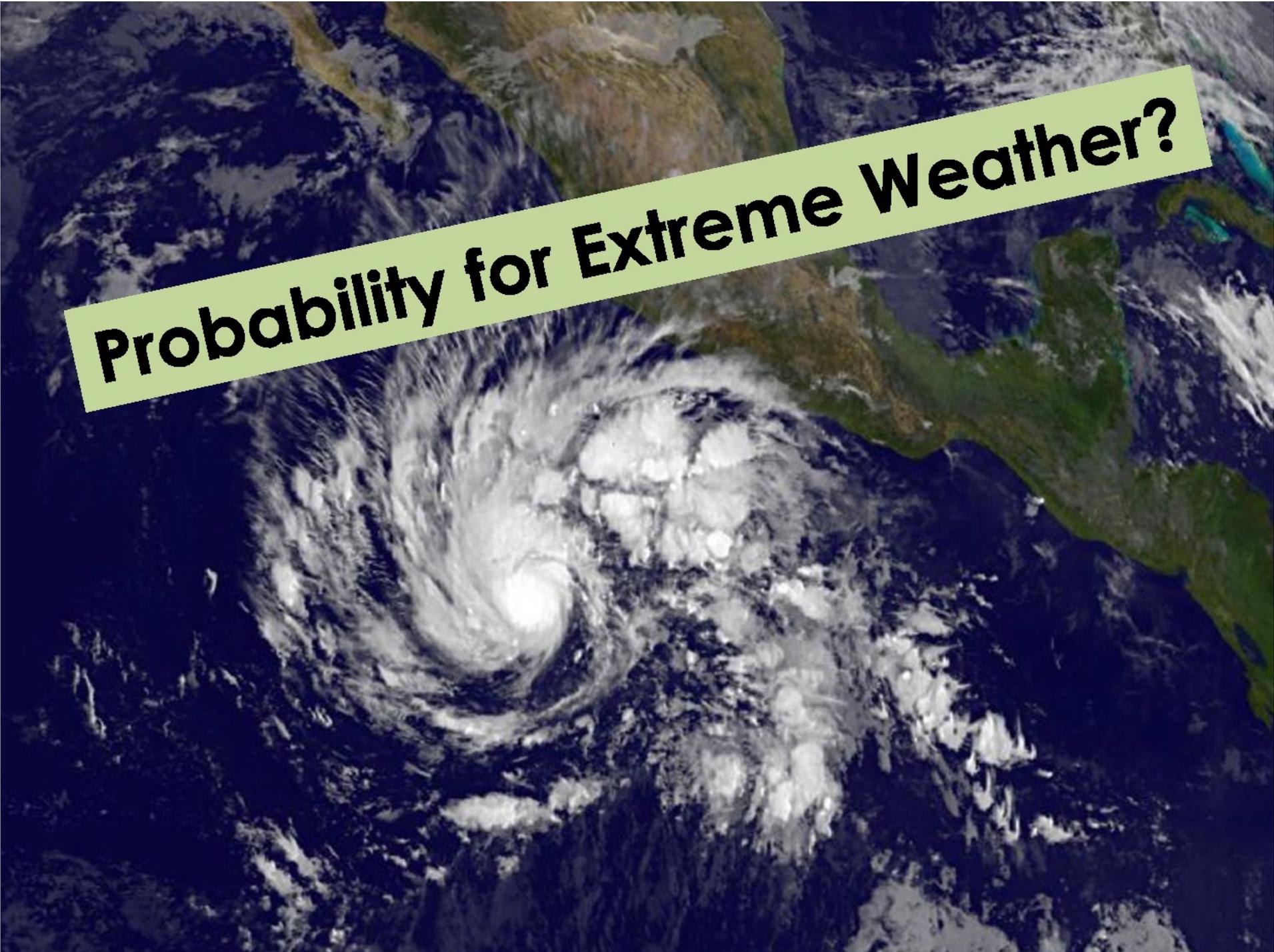
All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.



1906 Earthquake
Magnitude 7.8

ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WORK WITH THE SANDSTONE.

Probability for Extreme Weather?

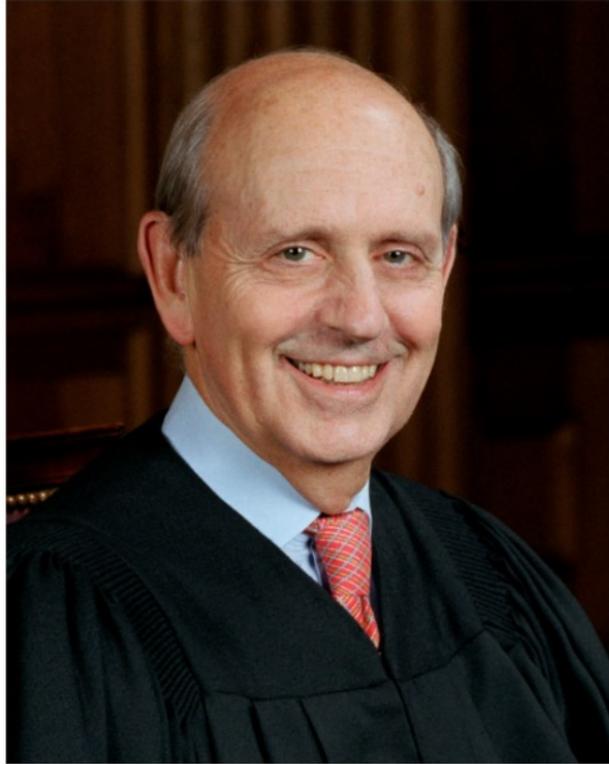


Bit Coin Mining

You “mine a bitcoin” if, for given data D , you find a number N such that $\text{Hash}(D, N)$ produces a string that starts with g zeroes.

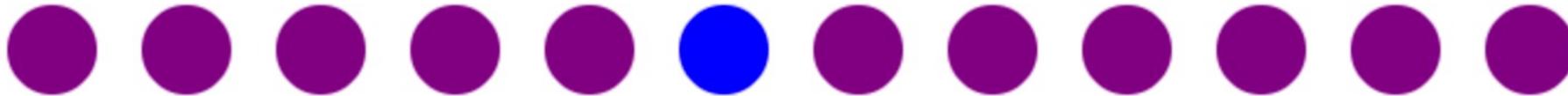


Representative Juries



Simulate

Simulation:

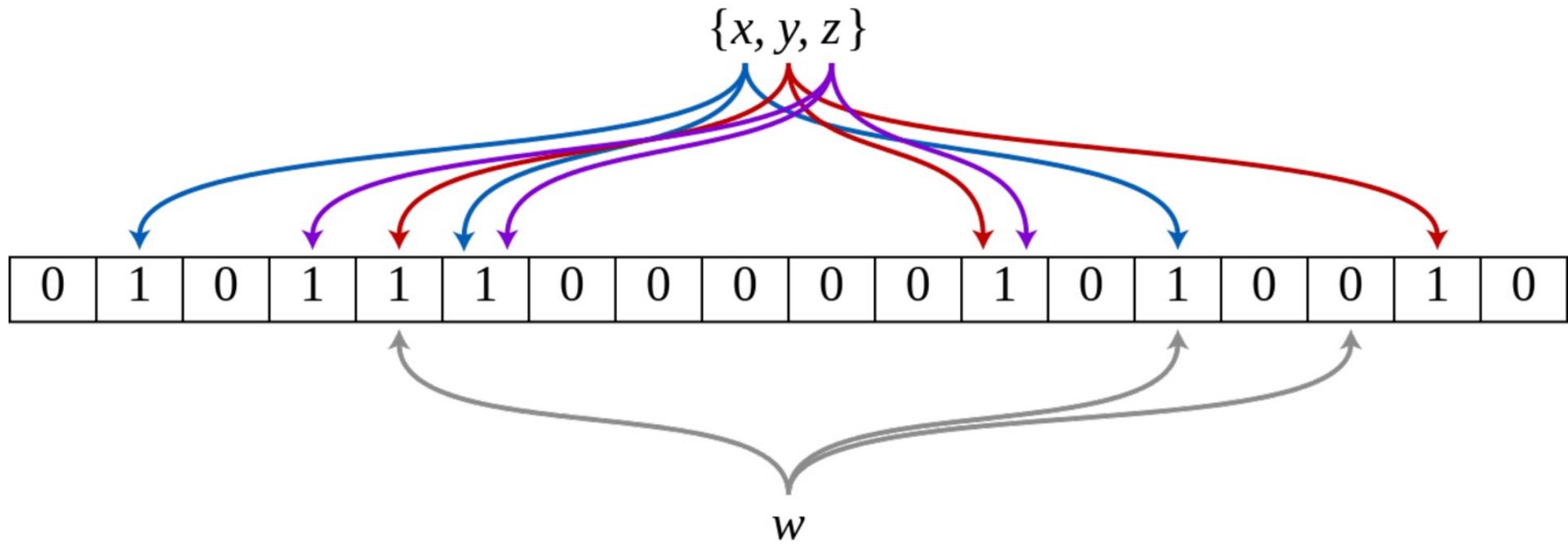


Dating at Stanford

Each person you date has a 0.2 probability of being someone you spend your life with. What is the average number of people one will date? What is the standard deviation?



Bloom Filter



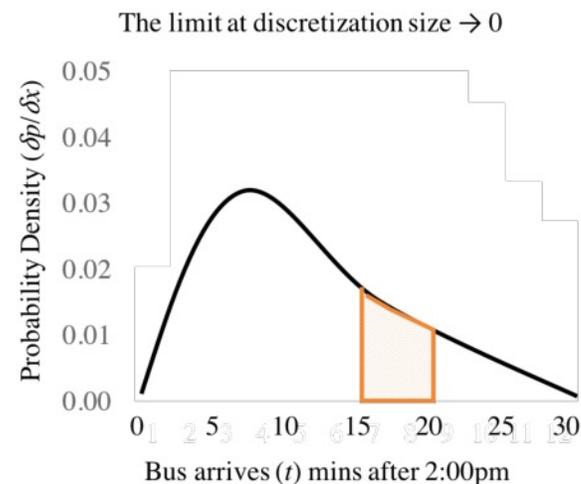
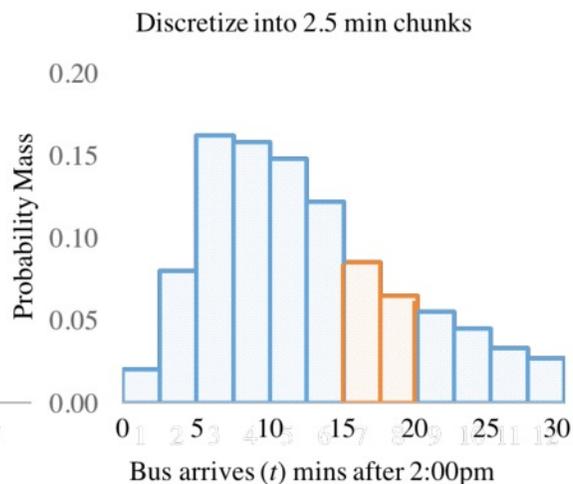
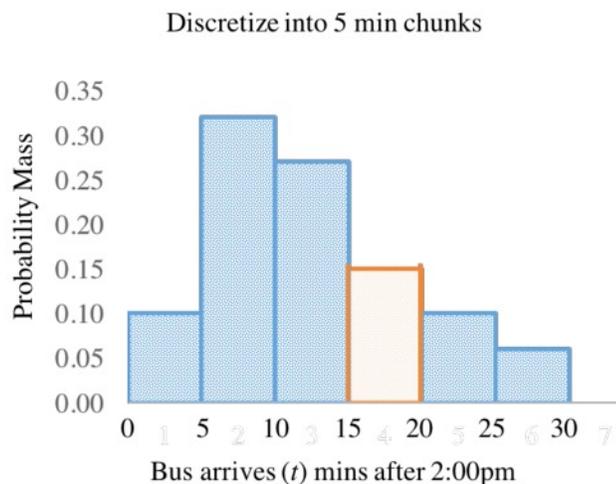
random() ?

Riding the Marguerite

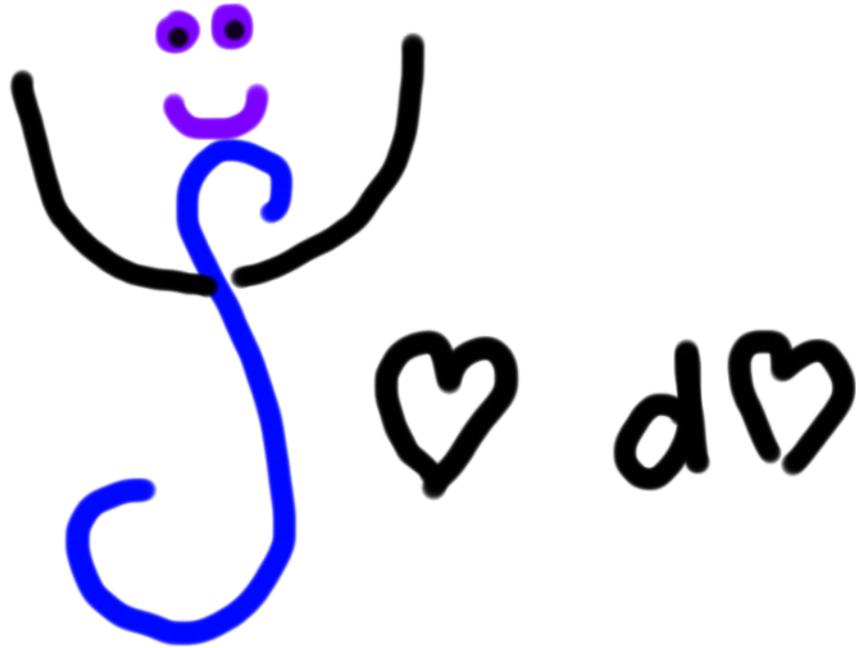


You are running to the bus stop.
You don't know exactly when
the bus arrives. You arrive at
2:20pm.

What is $P(\text{wait} < 5 \text{ min})$?



Integrals

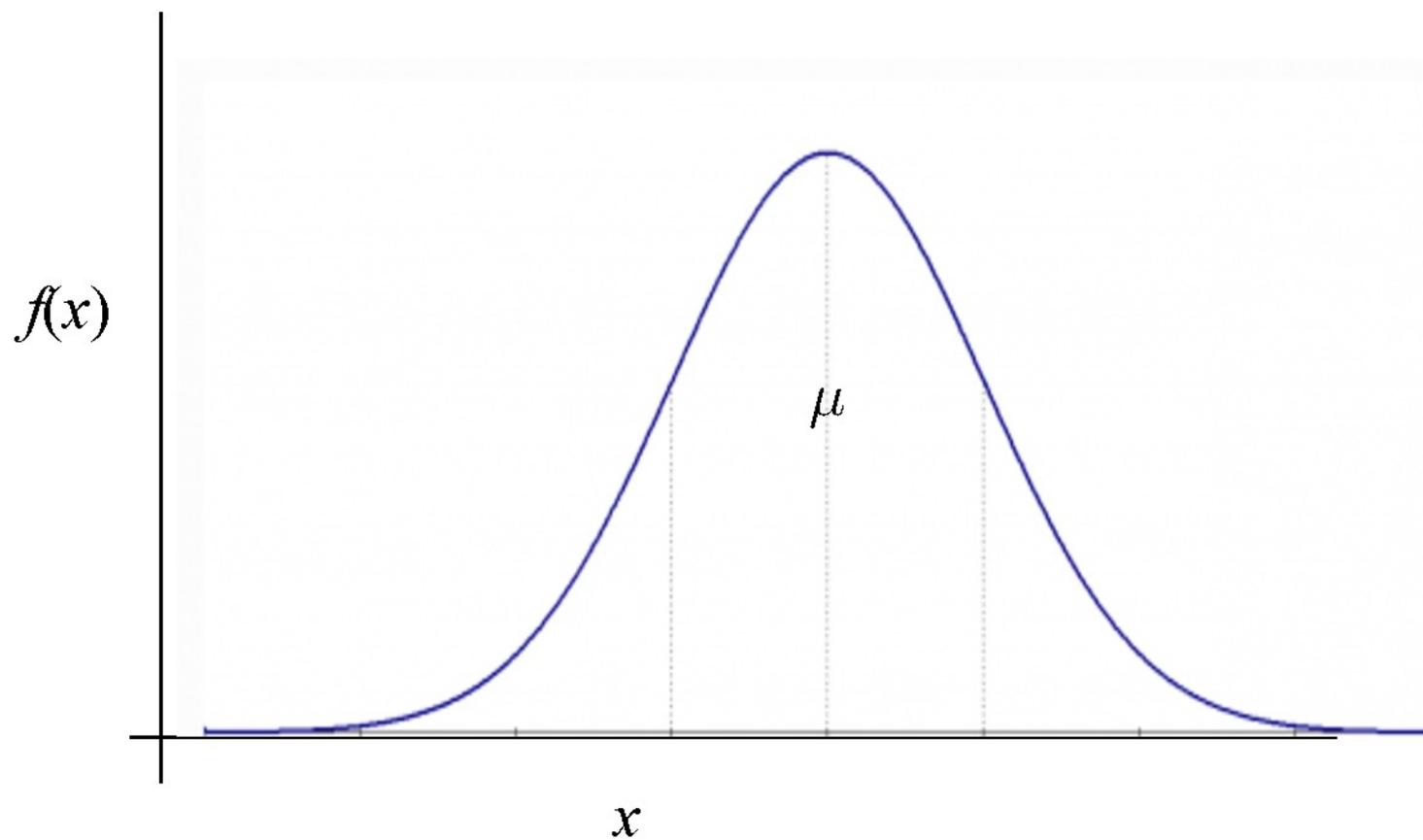


*loving, not scary

Probability Density Function

$$\mathcal{N}(\mu, \sigma^2)$$

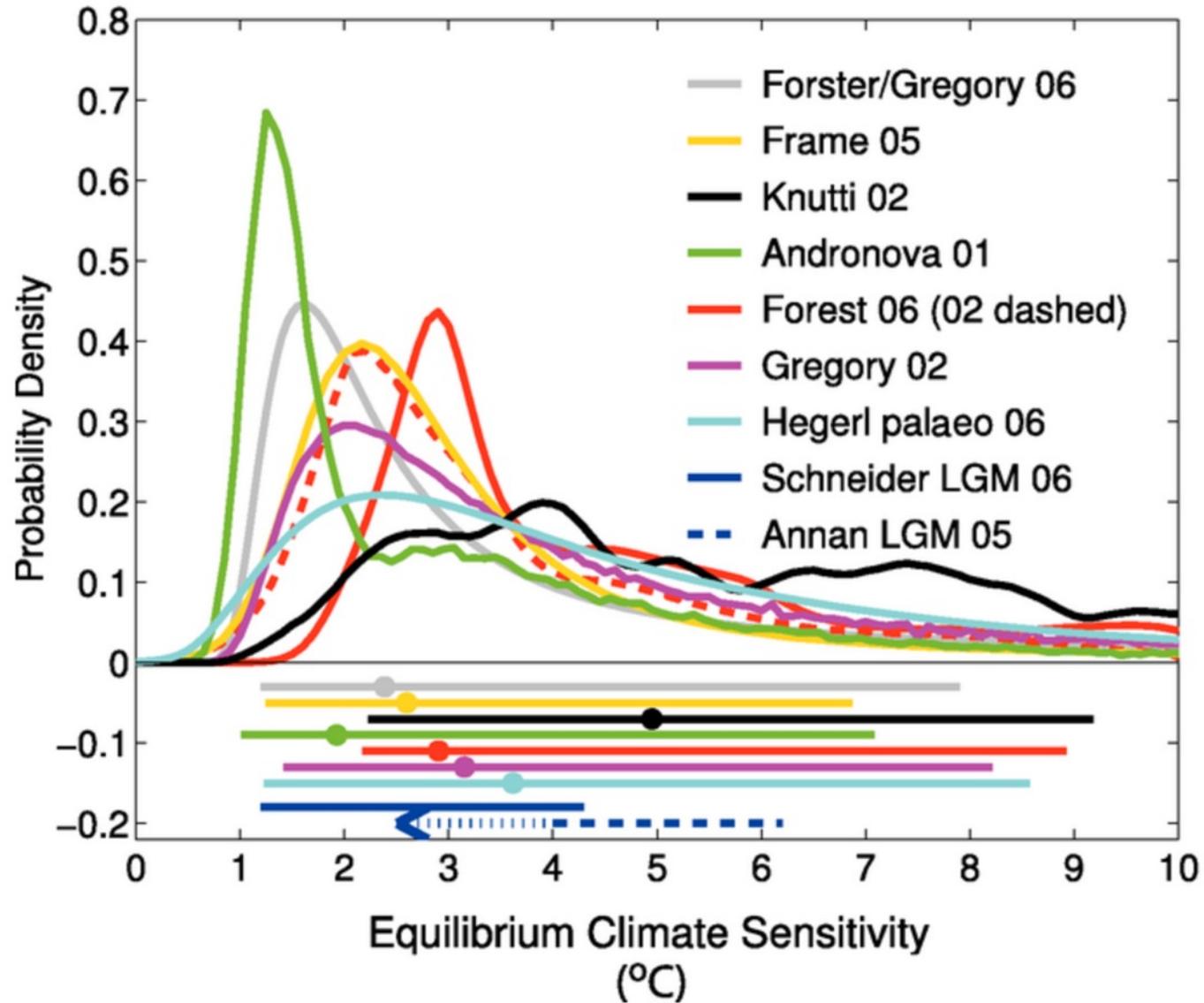
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



What do you get if you
integrate over a
probability *density* function?

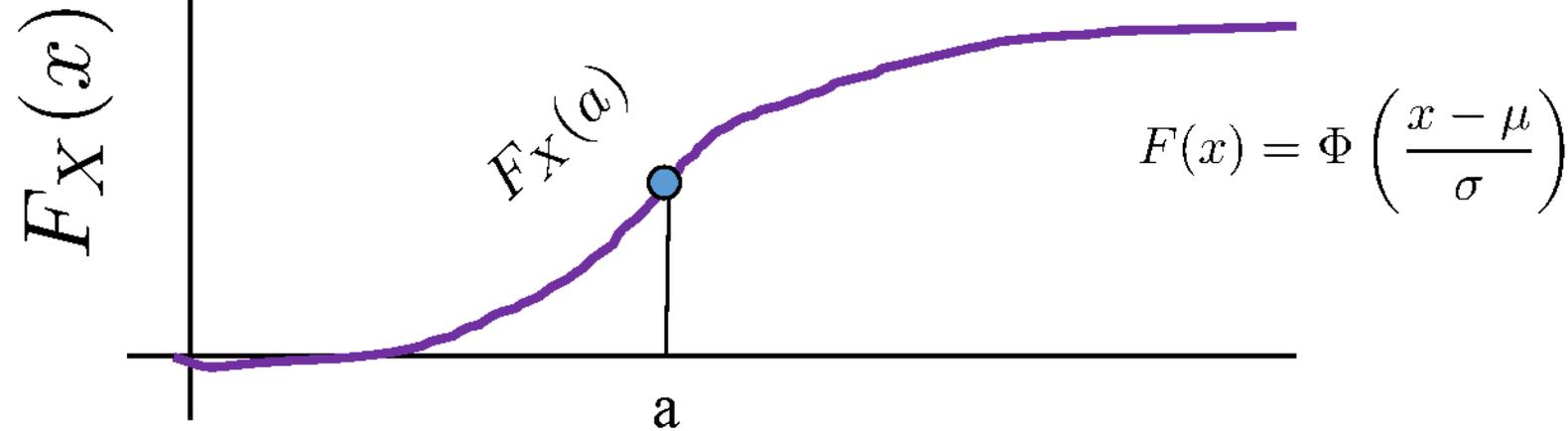
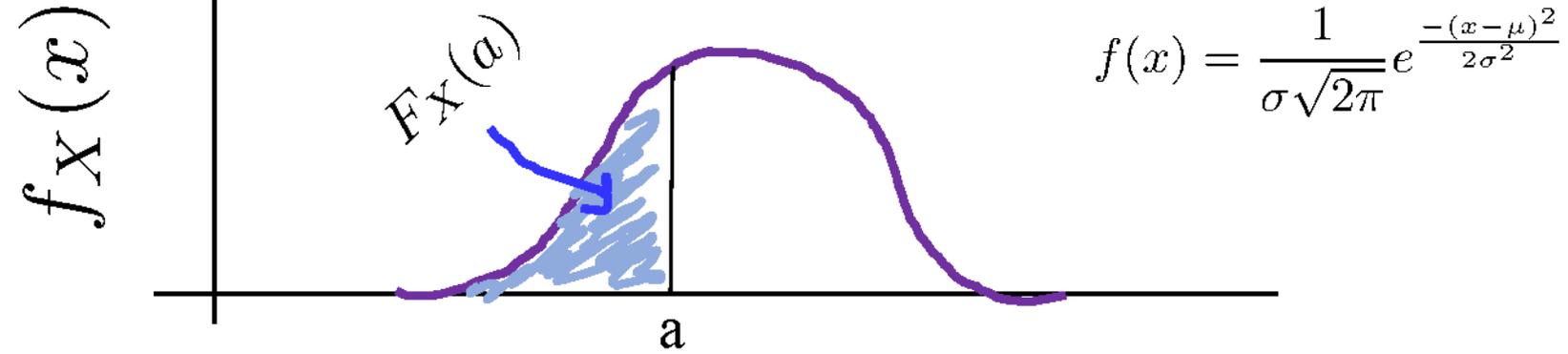
A probability!

Climate Sensitivity



PDF and CDF of a Normal

$$X \sim N(\mu, \sigma^2)$$



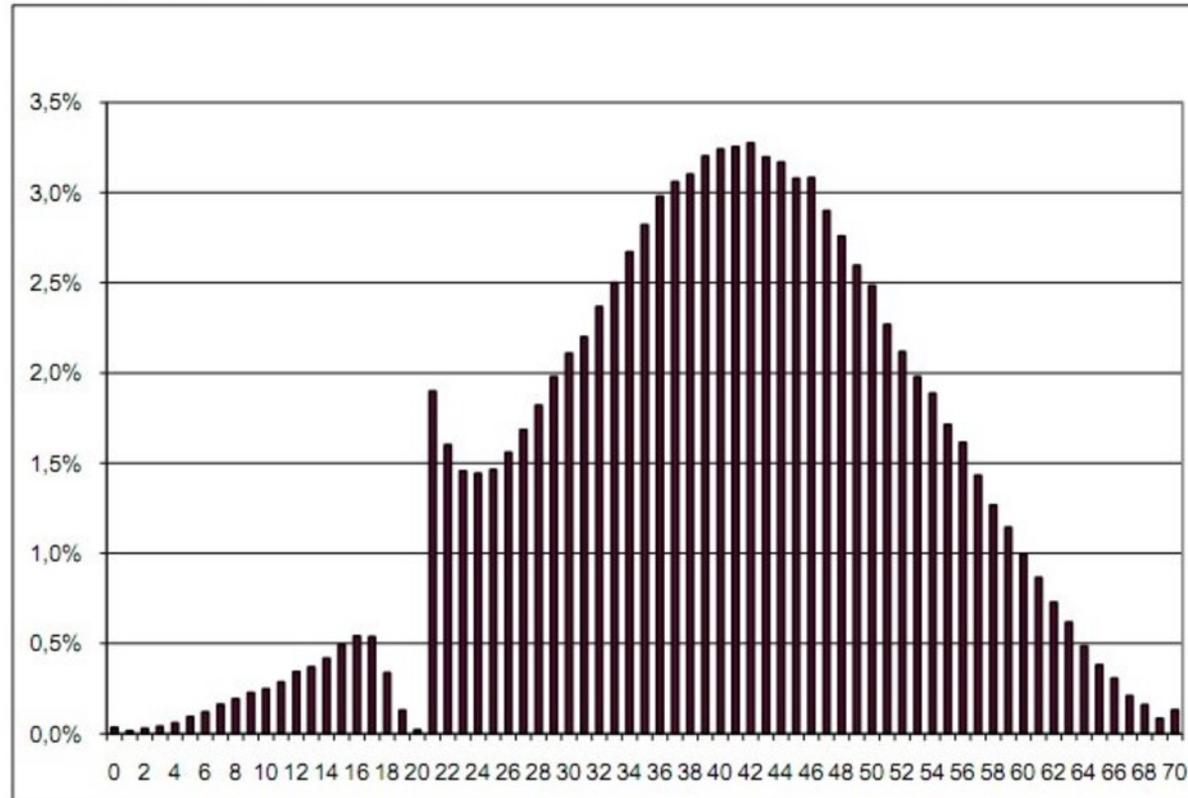
A CDF is the integral from $-\infty$ to x of the PDF

Altruism?

Scores for a standardized test that students in Poland are required to pass before moving on in school

See if you can guess the minimum score to pass the test.

2.1. Poziom podstawowy



Wykres 1. Rozkład wyników na poziomie podstawowym

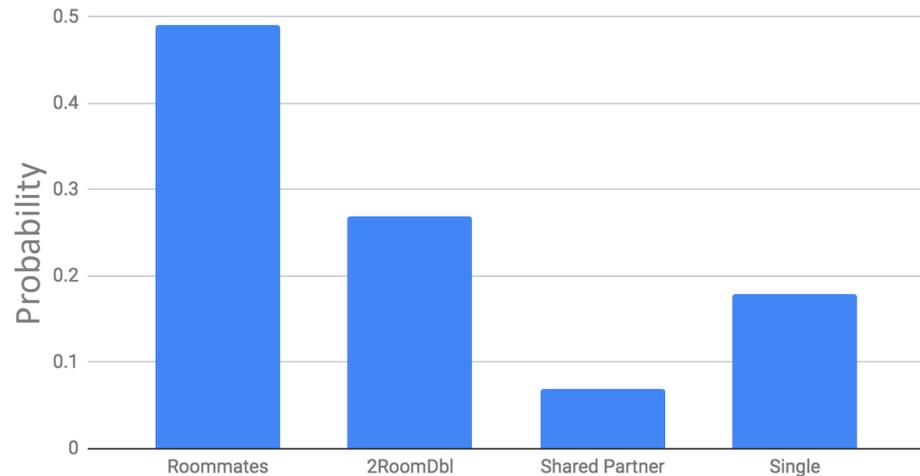
Probabilistic Models



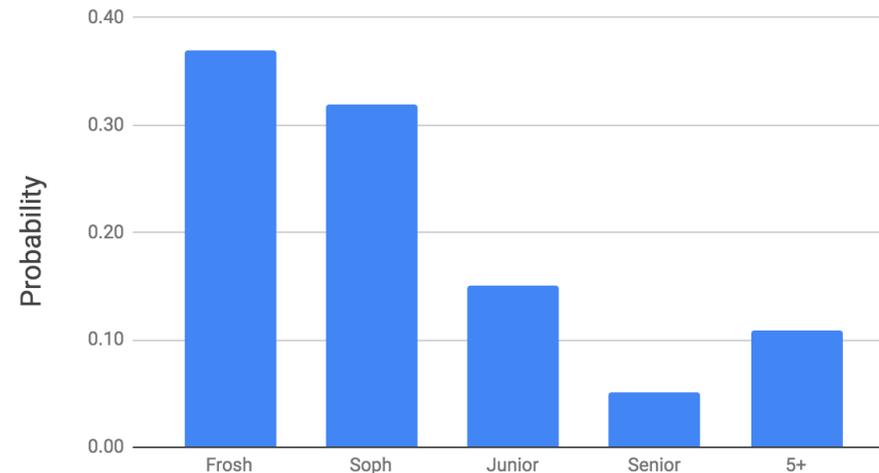
Joint Probability Table

	Roommates	2RoomDbI	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00

Marginal Room type



Marginal Year



Multinomial

Example document:

“Pay for Viagra with a credit-card. Viagra is great.
So are credit-cards. Risk free Viagra. Click for free.”

$n = 18$

$$P \left(\begin{array}{l} \text{Viagra} = 2 \\ \text{Free} = 2 \\ \text{Risk} = 1 \\ \text{Credit-card: } 2 \\ \dots \\ \text{For} = 2 \end{array} \middle| \text{spam} \right) = \frac{n!}{2!2! \dots 2!} p_{\text{viagra}}^2 p_{\text{free}}^2 \dots p_{\text{for}}^2$$

It's a Multinomial!

Probability of seeing
this document | spam

The probability of a word in
spam email being viagra



General “Inference”



General “Inference”

WebMD Symptom Checker BETA

INFO SYMPTOMS QUESTIONS CONDITIONS DETAILS TREATMENT

Add more symptoms

Type your main symptom here

or Choose common symptoms

bloating cough diarrhea dizziness fatigue

fever headache muscle cramp nausea

throat irritation

AGE 30 GENDER Male

MY SYMPTOMS

cough × throat irritation ×

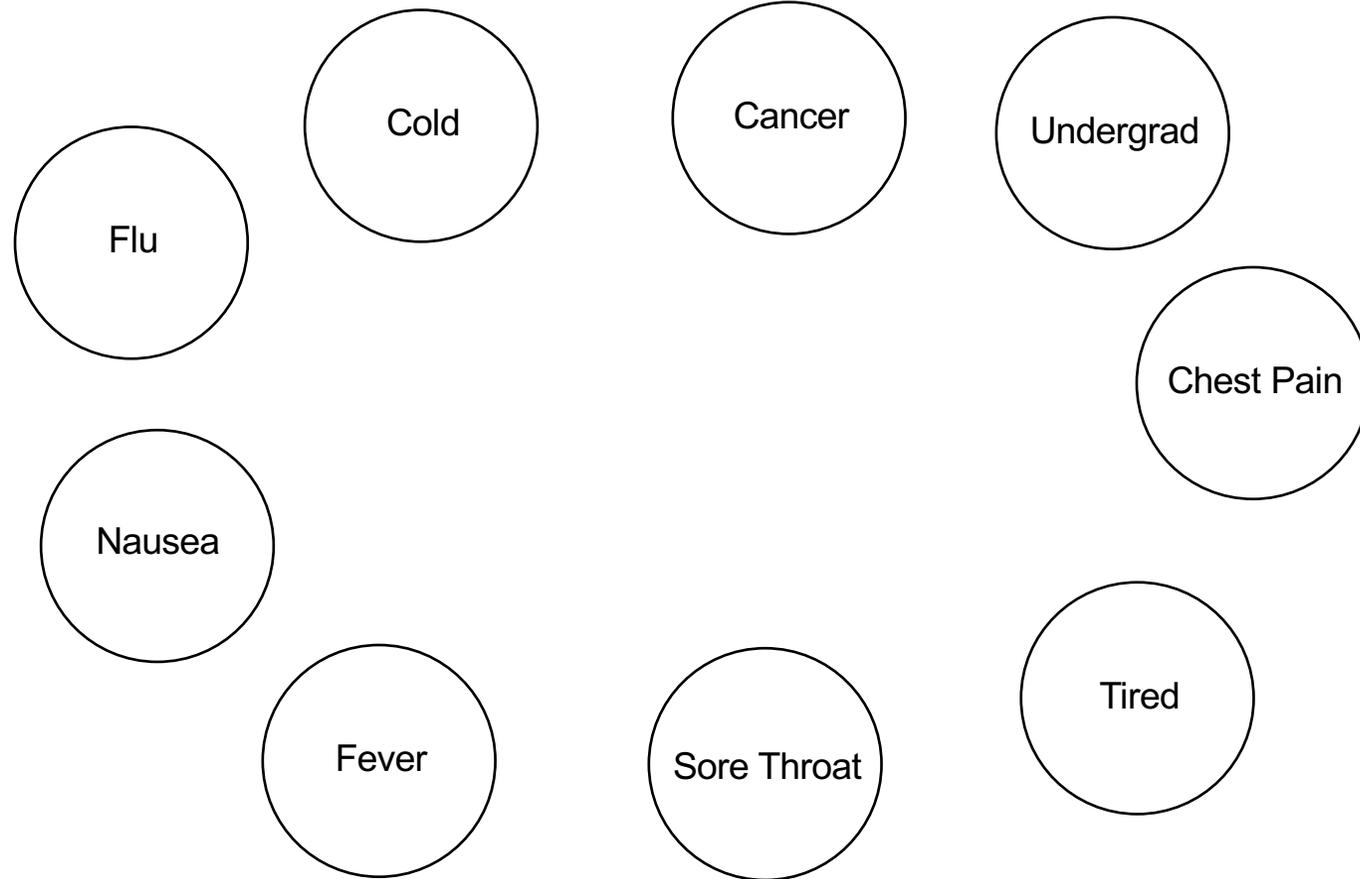
sneezing ×

Results Strength: **MODERATE**

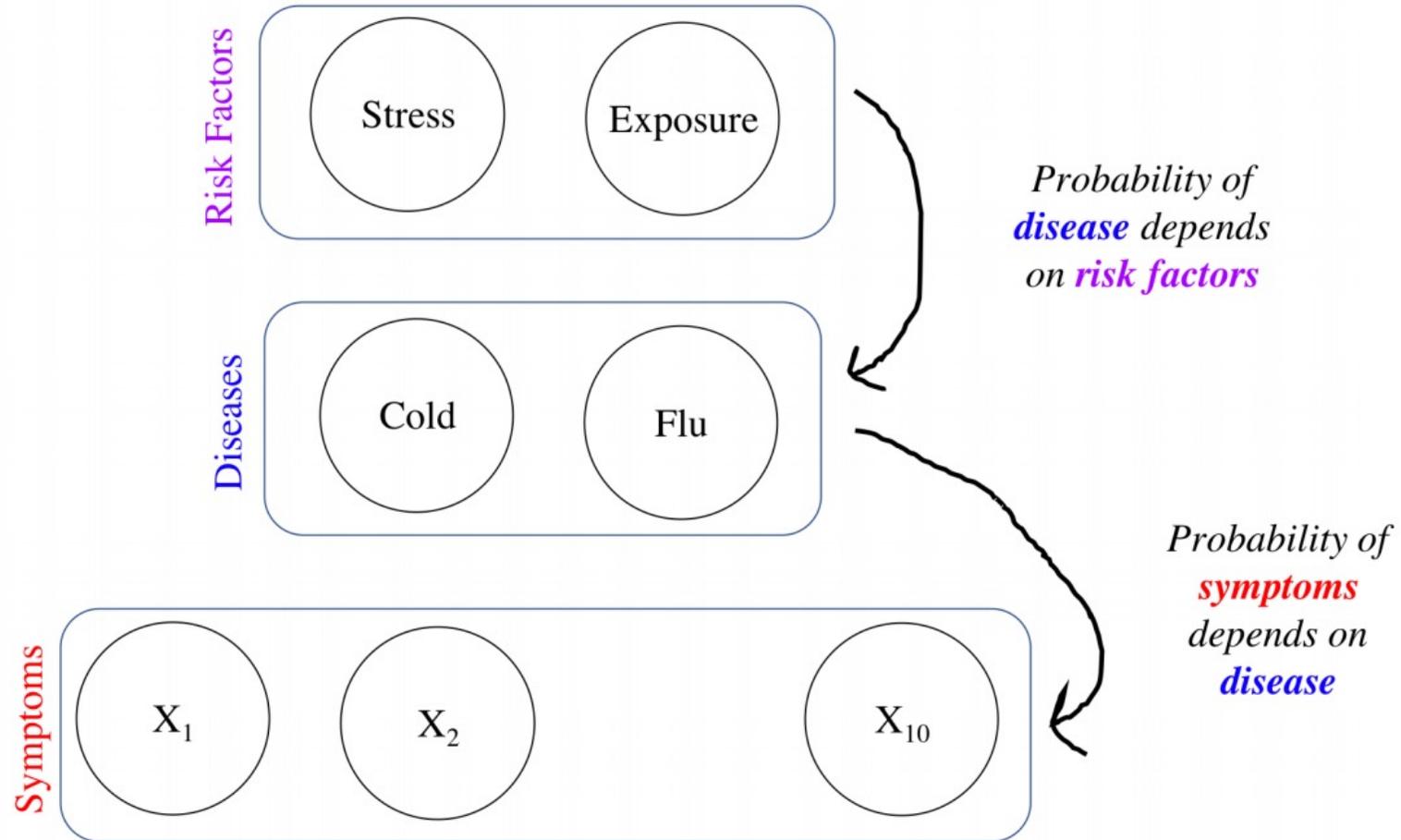
< Previous

Info
Continue >

Lots of Random Vars?



Bayes Nets!

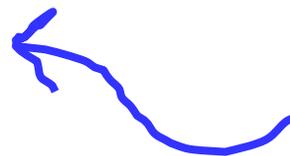


Alg #1: Rejection Sampling

```
3 N_SAMPLES = 100000
4
5 # Program: Joint Sa
6 # -----
7 # we can answer any
8 # with multivariate
9 # where conditioned
10 def main():
11     obs = getObserv
12     print 'Observat
13
14     samples = sampl
15     prob = probFluG
16     print 'Pr(Flu)
```

```
[0, 0, 0, 0]
[0, 1, 0, 1]
[1, 0, 1, 0]
[1, 1, 1, 1]
[0, 1, 0, 1]
[0, 1, 0, 0]
[0, 0, 0, 0]
[0, 1, 1, 1]
[0, 1, 0, 0]
[0, 1, 0, 1]
[0, 1, 0, 0]
[0, 1, 0, 1]
[0, 1, 0, 1]
[0, 0, 0, 0]
[1, 1, 1, 1]
[0, 0, 0, 0]
[0, 0, 0, 0]
[1, 1, 1, 1]
[0, 1, 0, 0]
Observation = [None, None, None, 1]
Pr(Flu | Obs) = 0.140635888502
>
```

Each one of these is
one posterior sample:



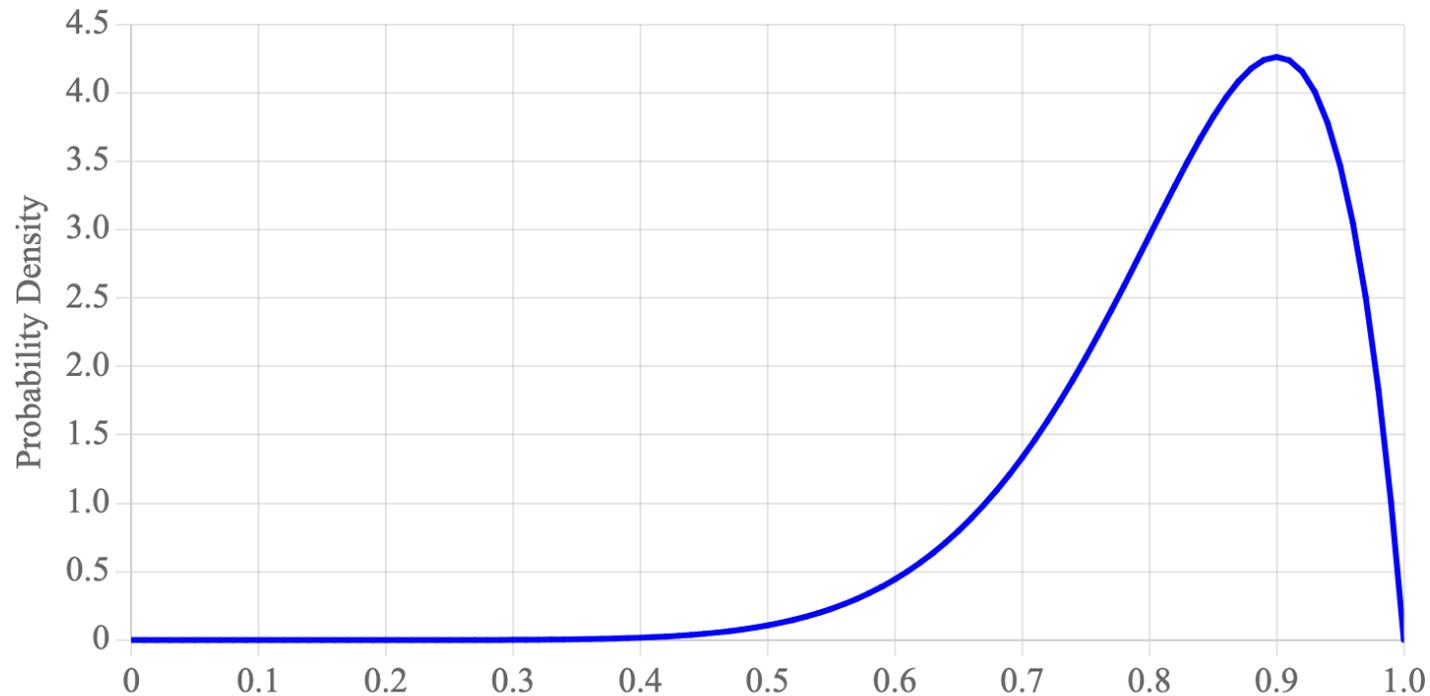
[Flu, Ugrad, Fever, Tired]

Uncertainty Theory



$$\begin{aligned}
& f(X = x|H = 9, T = 1) \\
&= \frac{P(H = 9, T = 1|X = x) \cdot f(X = x)}{P(H = 9, T = 1)} && \text{Bayes Theorem} \\
&= \frac{\binom{10}{9} x^9 (1-x)^1 \cdot f(X = x)}{P(H = 9, T = 1)} && \text{Binomial PMF} \\
&= \frac{\binom{10}{9} x^9 (1-x)^1 \cdot 1}{P(H = 9, T = 1)} && \text{Uniform PDF} \\
&= \frac{\binom{10}{9}}{P(H = 9, T = 1)} x^9 (1-x)^1 && \text{Constants to front} \\
&= K \cdot x^9 (1-x)^1 && \text{Rename constant}
\end{aligned}$$

Lets take a look at that function. For now we can let $K = \frac{1}{110}$. Regardless of K we will get the same shape, just scaled:



Lets Play!

Drug A

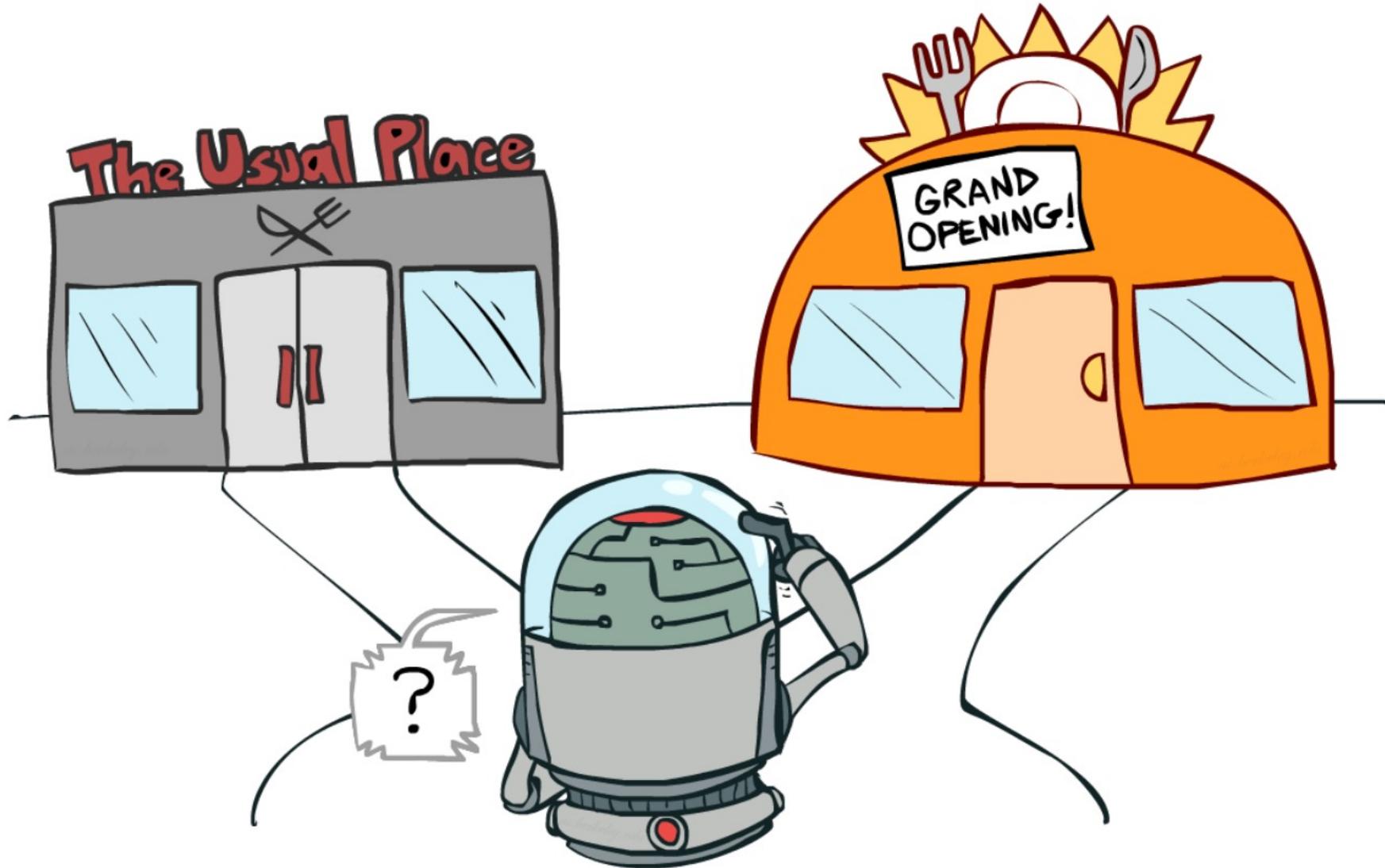


Drug B

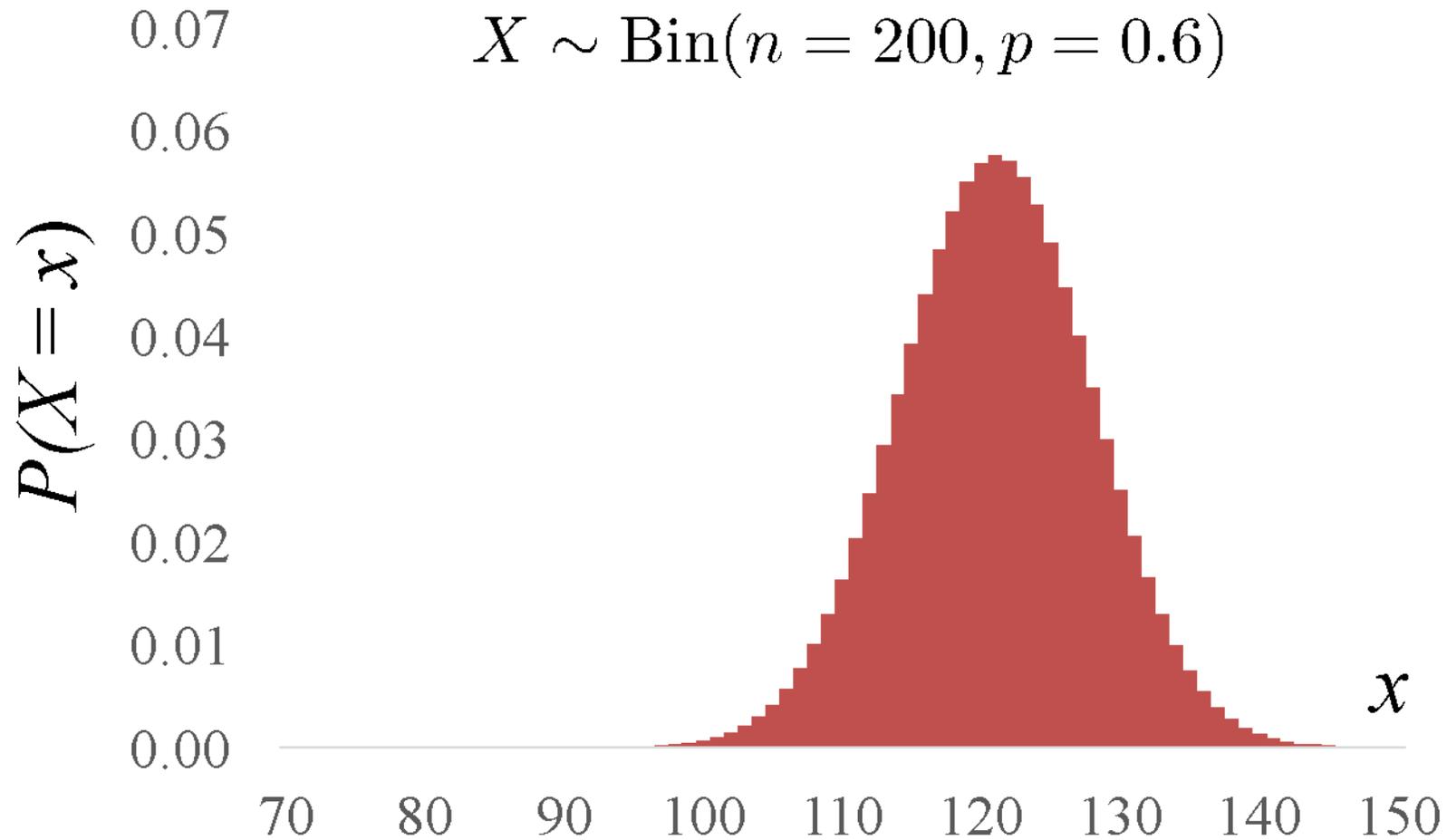


Which one do you give to a patient?

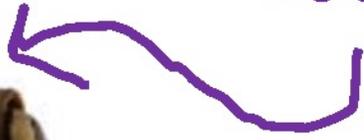
Thompson Sampling



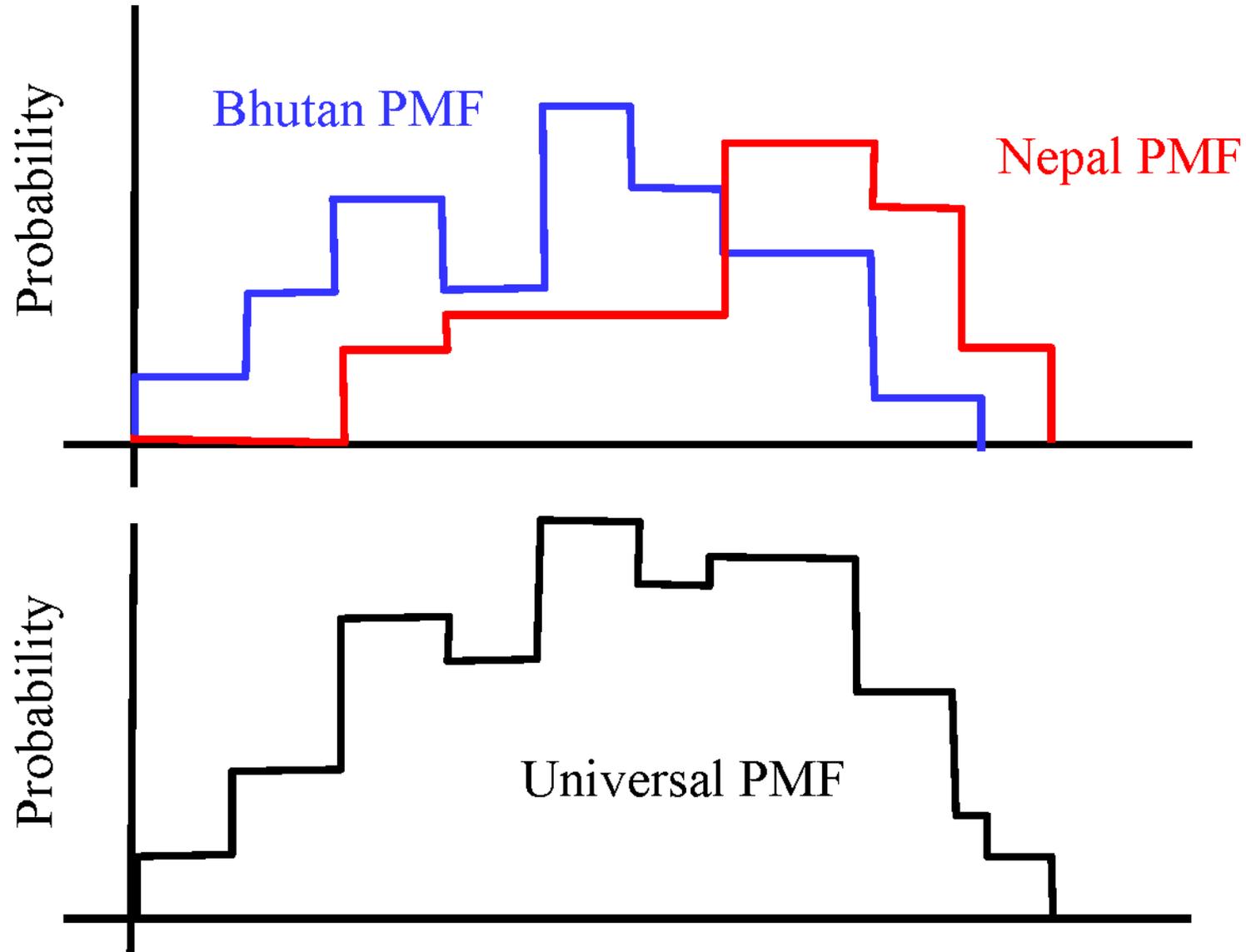
C.L.T. Explains This



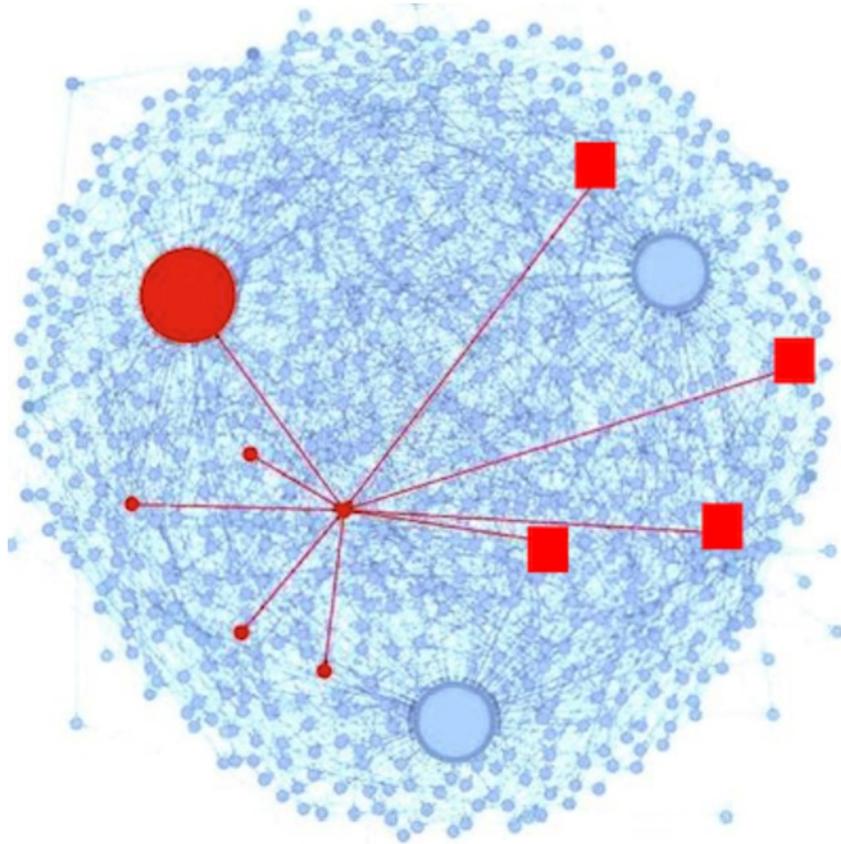
Bootstrap



Universal Sample



Peer Grading

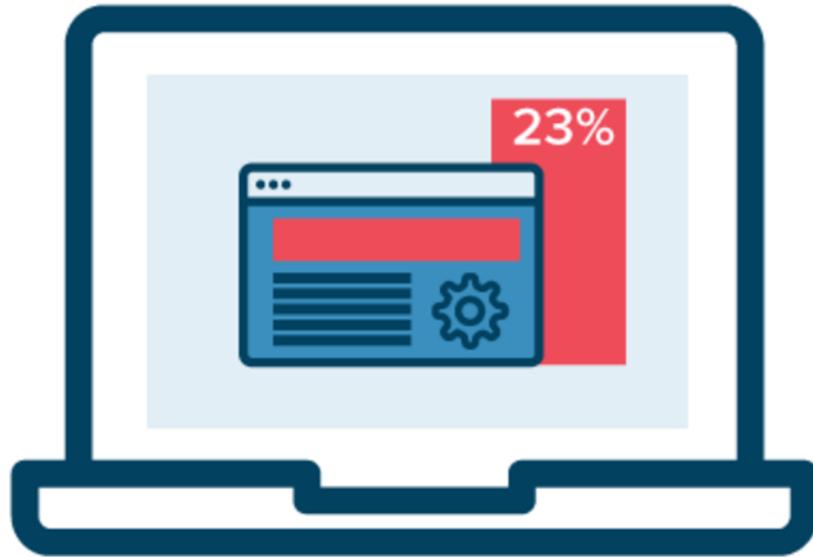


Peer Grading on Coursera
HCI.

31,067 peer grades for
3,607 students.

A/B Testing

A



CONTROL

B



VARIATION

Machine Learning



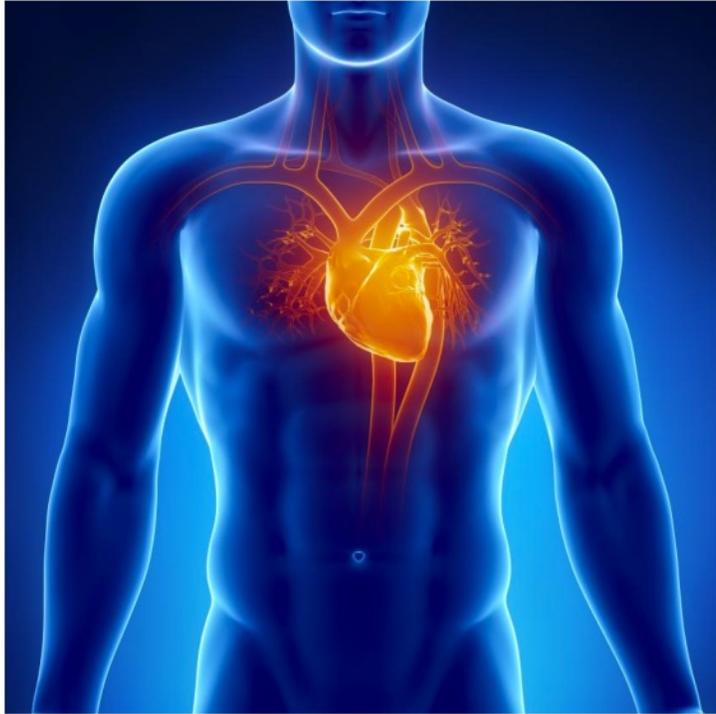
MAP: Most Probable Parameter

So good to see
you again!



Machine Learning

Heart



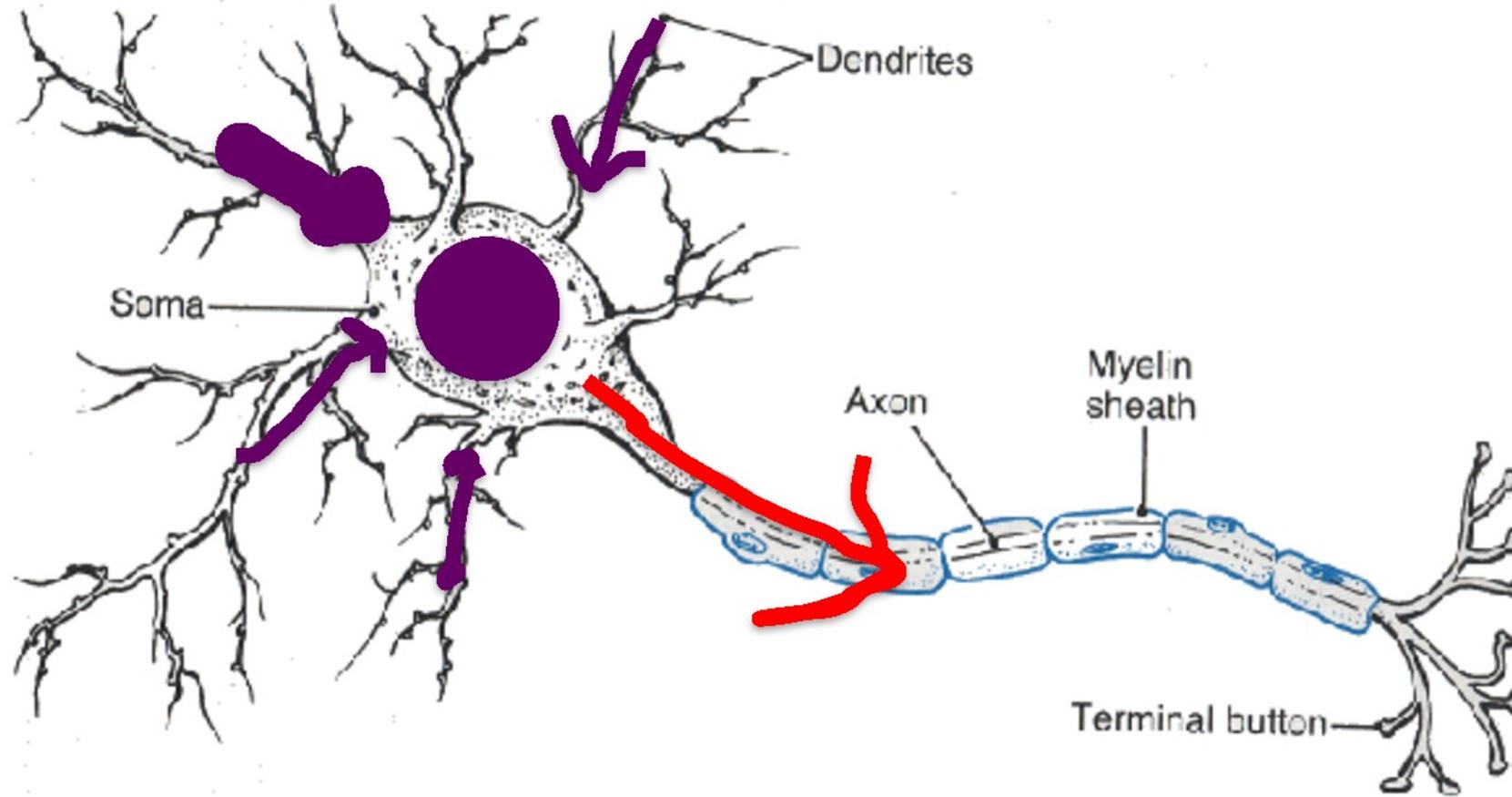
Ancestry



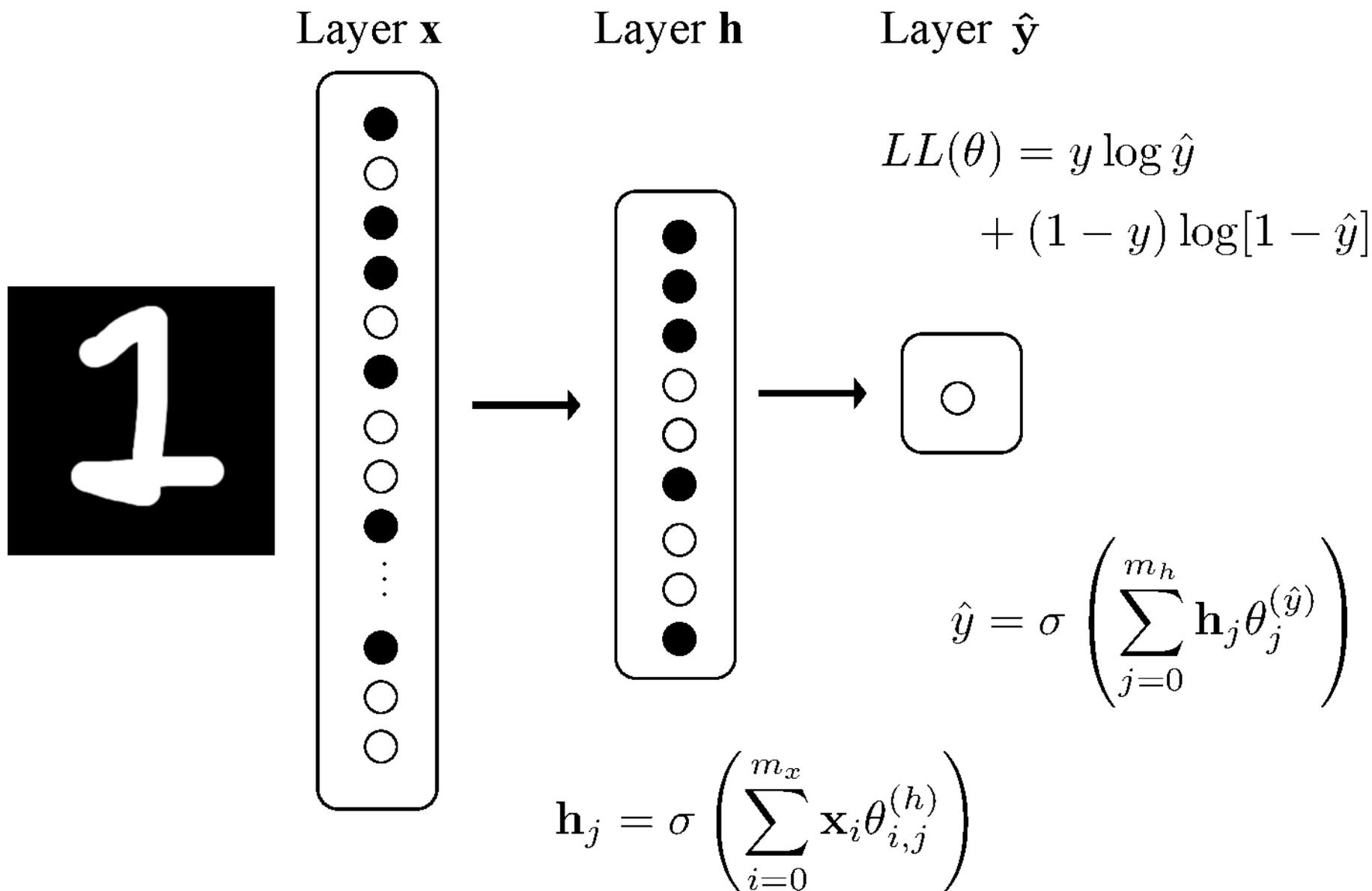
Netflix

NETFLIX

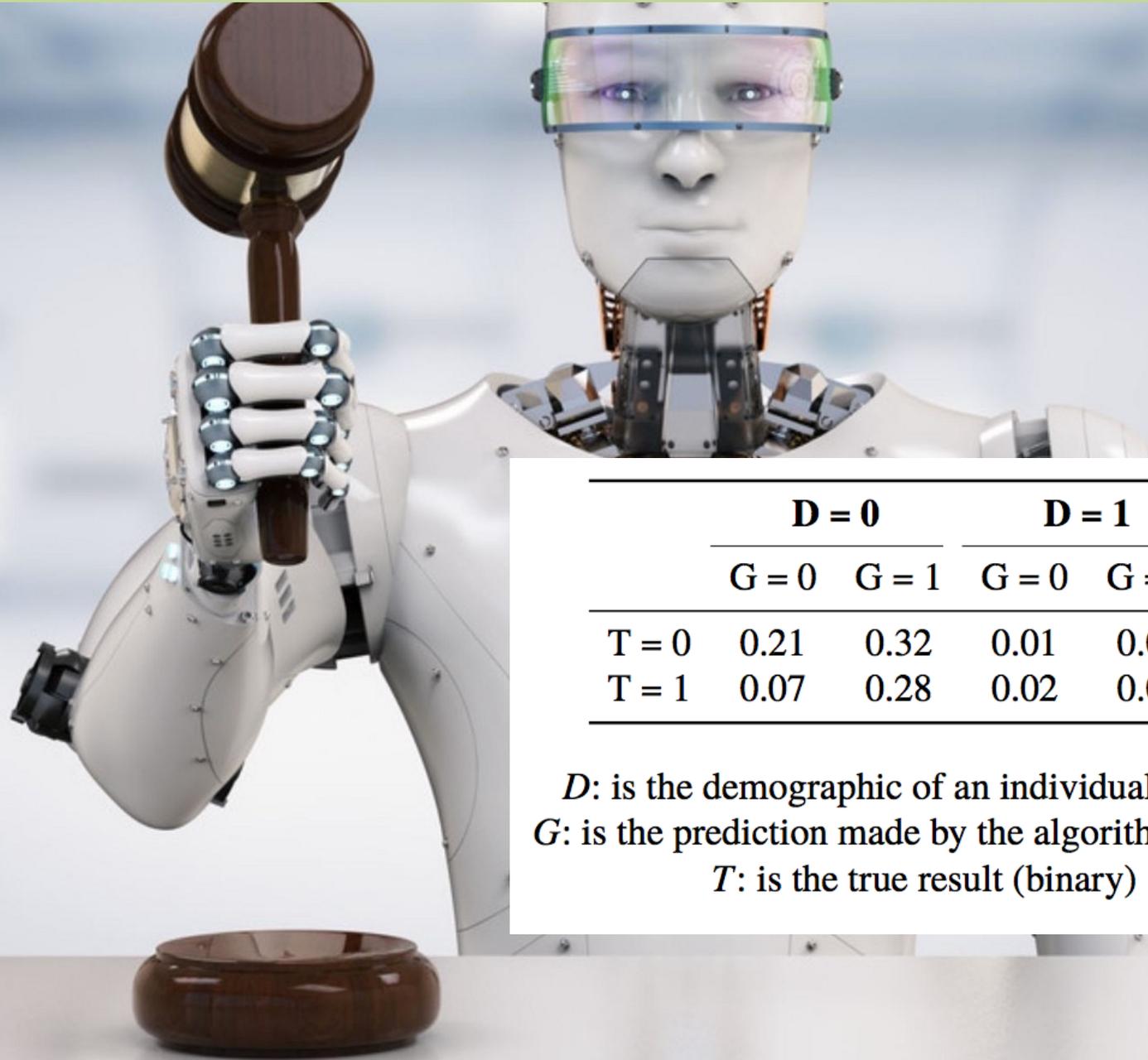
Logistic Regression



Deep Learning



Algorithmic Fairness

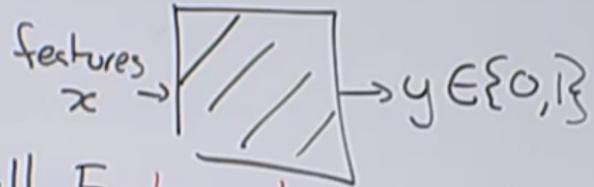


	D = 0		D = 1	
	G = 0	G = 1	G = 0	G = 1
T = 0	0.21	0.32	0.01	0.01
T = 1	0.07	0.28	0.02	0.08

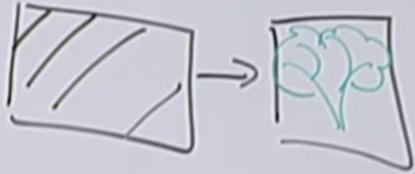
D: is the demographic of an individual (binary)
G: is the prediction made by the algorithm (binary)
T: is the true result (binary)

DallE and GPT

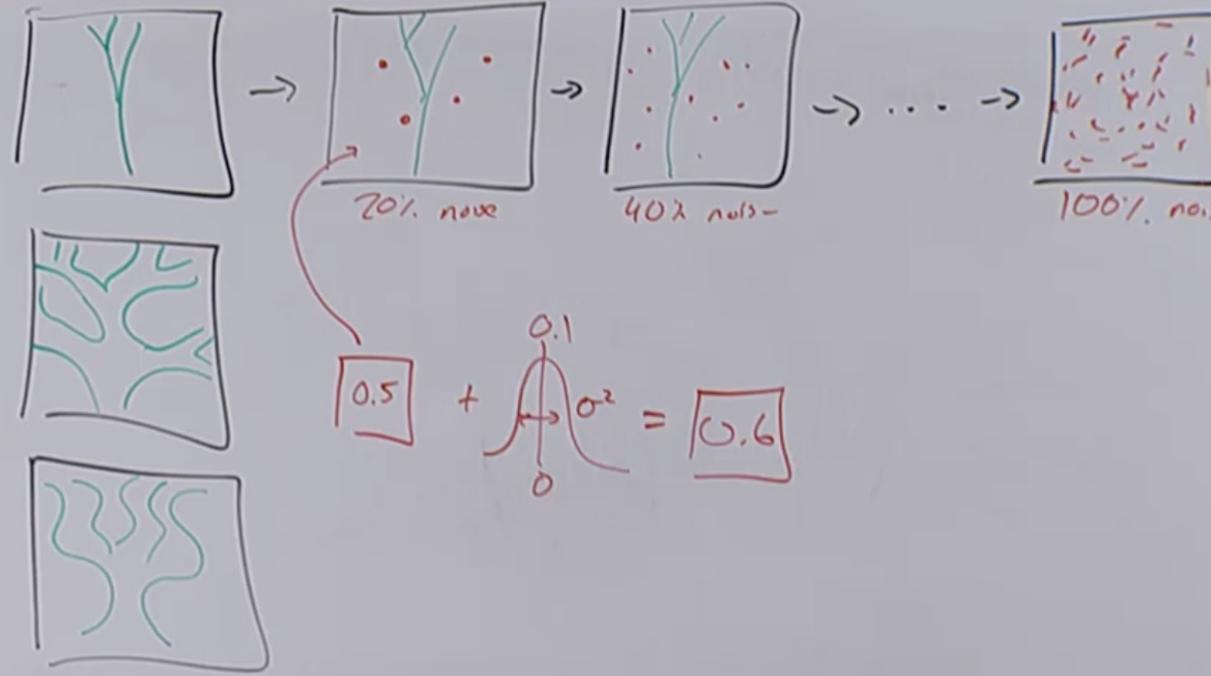
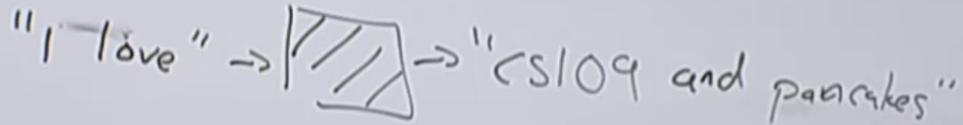
Pset 6:



Dall E tree black & white:



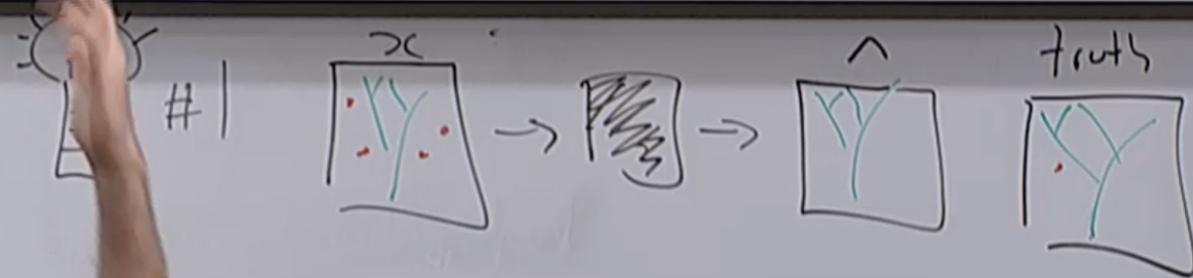
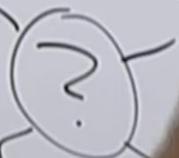
GPT:



reverse a classification predictor

scrape google

how to predict



https://chat.openai.com/chat

chat.openai.com/chat

hey GPT can you help me with my homework?

Of course! I'd be happy to help you with your homework. Can you provide some details about what you need help with?

A fair 6-sided die is repeatedly rolled until the total sum of all the rolls exceeds 300. Approximate the probability that at least 80 rolls are necessary to reach a sum that exceeds 300.

To solve this problem, we need to use the geometric distribution. This distribution is used to model the number of Bernoulli trials (i.e. experiments with only two possible outcomes, such as rolling a die) that must be conducted in order to obtain a success. In this case, rolling a 6 on the die is considered a success because it contributes the most to the total sum.

The probability of success on each roll is $1/6$, since there is only one favorable outcome (rolling a 6) out of a total of 6 possible outcomes. This means that the probability of failure on each roll is $5/6$.

The probability of obtaining at least 80 successes before the first failure is given by:

$$P(X \geq 80) = 1 - P(X < 80)$$

where X is the random variable representing the number of successes before the first

Try again

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

https://chat.openai.com/chat

chat.openai.com/chat

This is the probability that at least 80 rolls are necessary to reach a sum that exceeds 300.

I hope this helps! Let me know if you have any other questions.

hmm that doesn't seem correct. I am pretty sure you are supposed to use the central limit theorem

You are correct that the central limit theorem can be used to approximate the probability in this problem. The central limit theorem states that the sum of a large number of independent and identically distributed (i.i.d.) random variables will be approximately normally distributed, even if the individual random variables are not normally distributed.

In this problem, we are summing the outcomes of a large number of independent rolls of a fair 6-sided die. This means that the central limit theorem can be applied to approximate the distribution of the sum.

To use the central limit theorem, we first need to compute the mean and standard deviation of the sum. The mean of the sum is simply the mean of a single roll multiplied by the number of rolls. Since each roll has a mean of 3.5 (obtained by adding the possible outcomes and dividing by 6), the mean of the sum is $3.5 * 80 = 280$.

The standard deviation of the sum is equal to the standard deviation of a single roll multiplied by the square root of the number of rolls. Since the standard deviation of a single roll is 1.7 (obtained by using the formula $\sqrt{\frac{6^2 - 1}{12}}$), the standard deviation of the sum is $1.7 * \sqrt{80} = 15.4$.

Try again

Free Research Preview: ChatGPT is optimized for dialogue. Our goal is to make AI systems more natural to interact with, and your feedback will help us improve our systems and make them safer.

By the numbers

~ 30 Major Keys



Naïve Bayes Assumption:

$$P(\mathbf{x}|y) = \prod_i P(x_i|y)$$

64 Course Reader Chapters

The screenshot shows a web browser window with the following content:

- Browser Tab:** Name to Age
- URL:** chrispiech.github.io/probabilityForComputerScientists/en/examples/name2age/
- Table of Contents (Left Sidebar):**
 - 100 Binomial Problems
 - Jury Selection
 - Grading Eye Inflammation
 - Gaussian CDF Calculator
 - Grades are Not Normal
 - Curse of Dimensionality
 - Probability of Baby Delivery
 - Part 3: Probabilistic Models**
 - Joint Probability
 - Multinomial
 - Continuous Joint
 - Inference
 - Bayesian Networks
 - Independence in Variables
 - Correlation
 - General Inference
 - Worked Examples
 - CS109 Logo
 - Fairness in AI
 - Federalist Paper Authorship
 - Name to Age
 - Bridge Distribution
 - Tracking in 2D
 - Part 4: Uncertainty Theory**
 - Beta Distribution
 - Adding Random Variables
 - Central Limit Theorem
 - Sampling
 - Bootstrapping
 - Algorithmic Analysis
 - Worked Examples
- Main Content:**
 - Assumptions:
 - This data only is accurate for names of people in the US. The probability of age given names could be very different in other countries.
 - The US census is not perfect. It does not capture all people who are resident in the US, and there are demographics which are underrepresented. This will also skew our results.
 - Demo**
 - Query Name: ✓
 - Figure:** A line graph showing the probability of age for the name Katherine. The x-axis is labeled "Age" (0 to 120) and the y-axis is labeled "Probability" (0.002 to 0.020). The curve shows a primary peak around age 30 (probability ~0.019) and a secondary peak around age 65 (probability ~0.012).
 - Records with name: 589753
 - This demo is based on real data from US Social Security applications between 1914 and 2014. Thank you to <https://www.kaggle.com/kaggle/us-baby-names> for compiling the data. [Download Data](#)
 - Names that Give Away Your Age**

Some names have certain years where they were exceptionally popular. These names provide quite a lot of information about birth year. Let's look at some of the names with the highest max probability.

1 Personal Challenge



1 New Counter



t h a n k y o u

A row of ten light-colored wooden blocks, each with a single lowercase letter, spelling out the words 'thank you'. The blocks are arranged on a dark wooden surface. The background is a soft-focus bokeh of numerous warm, golden-yellow lights, creating a warm and appreciative atmosphere.