Future of Probability

Chris Piech

CS109

AI

Uncertainty Theory

Single Random Variables

Probabilistic Models

Counting

Probability Fundamentals

Learning Goal: Abundance of important problems

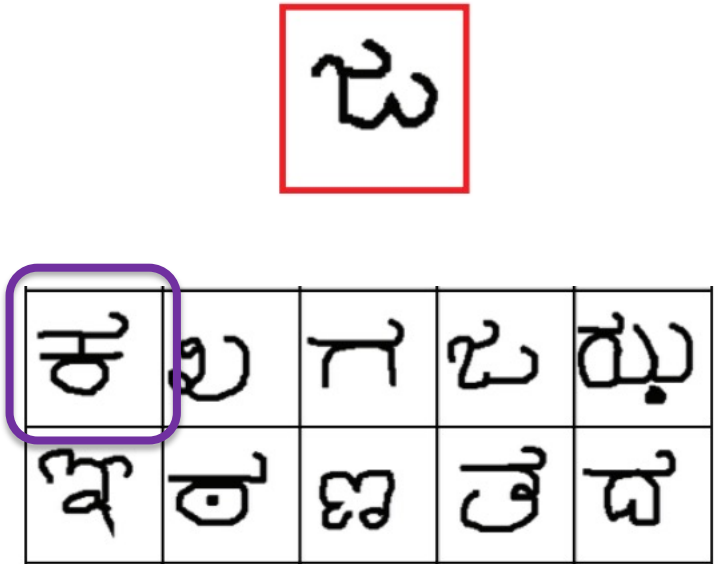
Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.

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

Open Problem: One Shot Learning

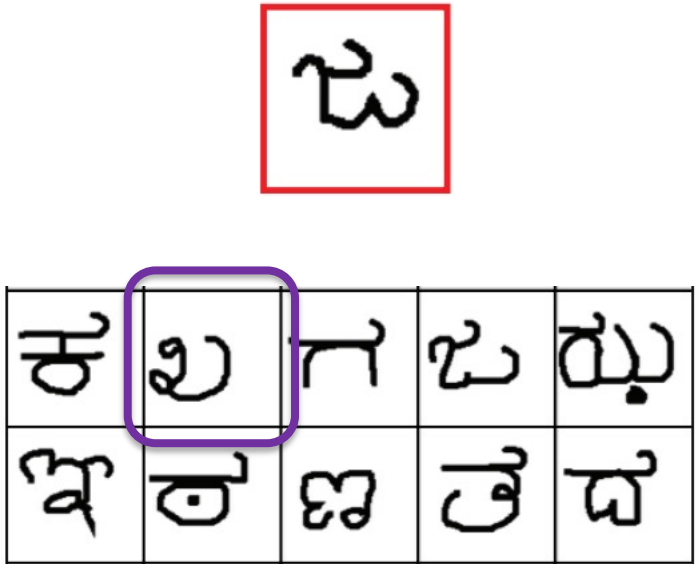
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

Open Problem: One Shot Learning

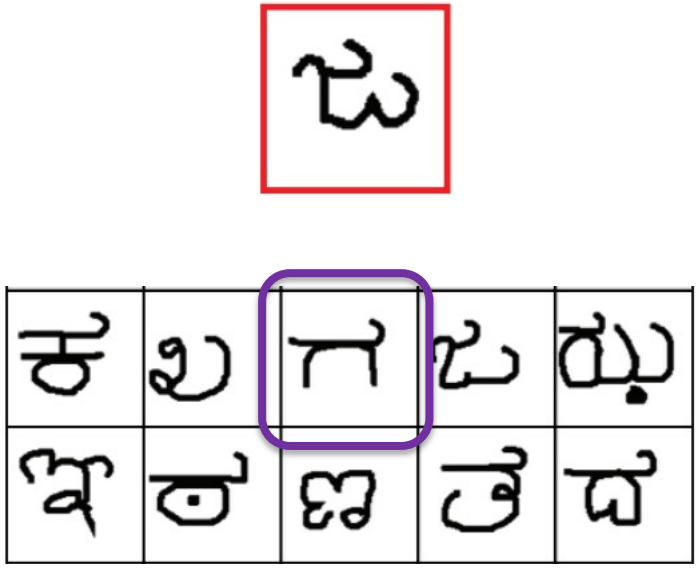
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

Open Problem: One Shot Learning

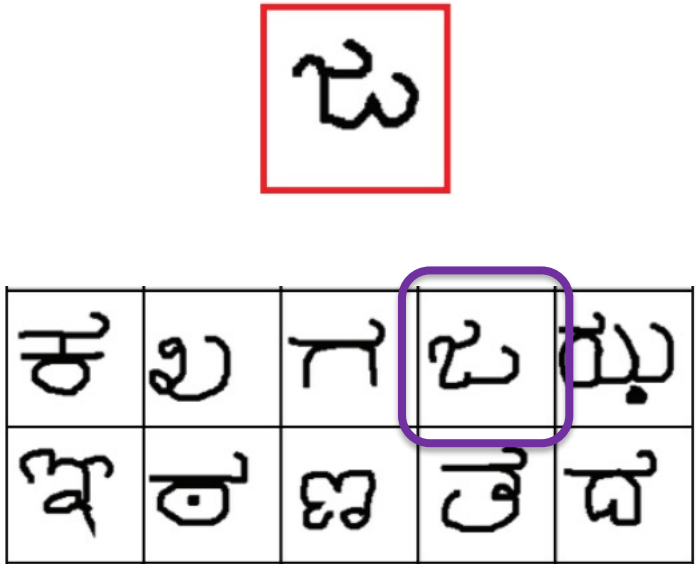
B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

Open Problem: One Shot Learning

B Lake, R Salakhutdinov, J Tenenbaum. Science 2015.
Human-level concept learning through probabilistic program induction.



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

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**

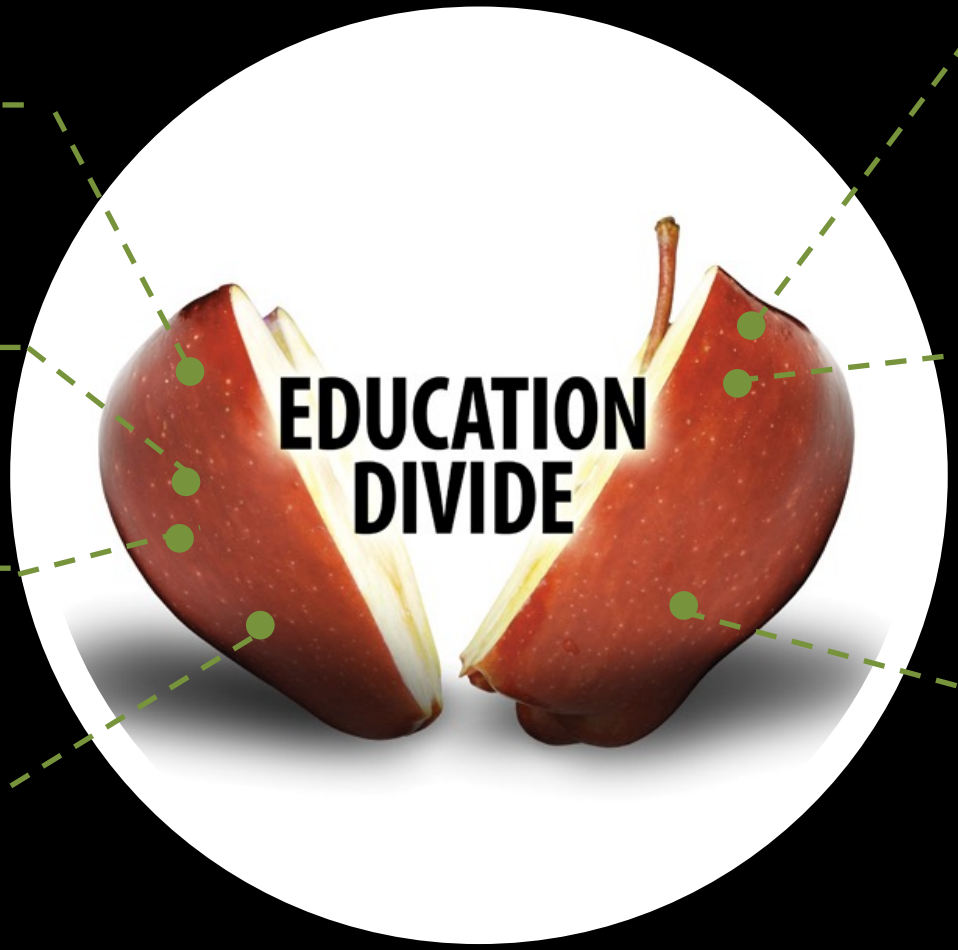
Quality Education Gap

90% of children enroll in primary education [1]

40% in secondary education [1]

20% in tertiary education [1]

Dramatic quality differences



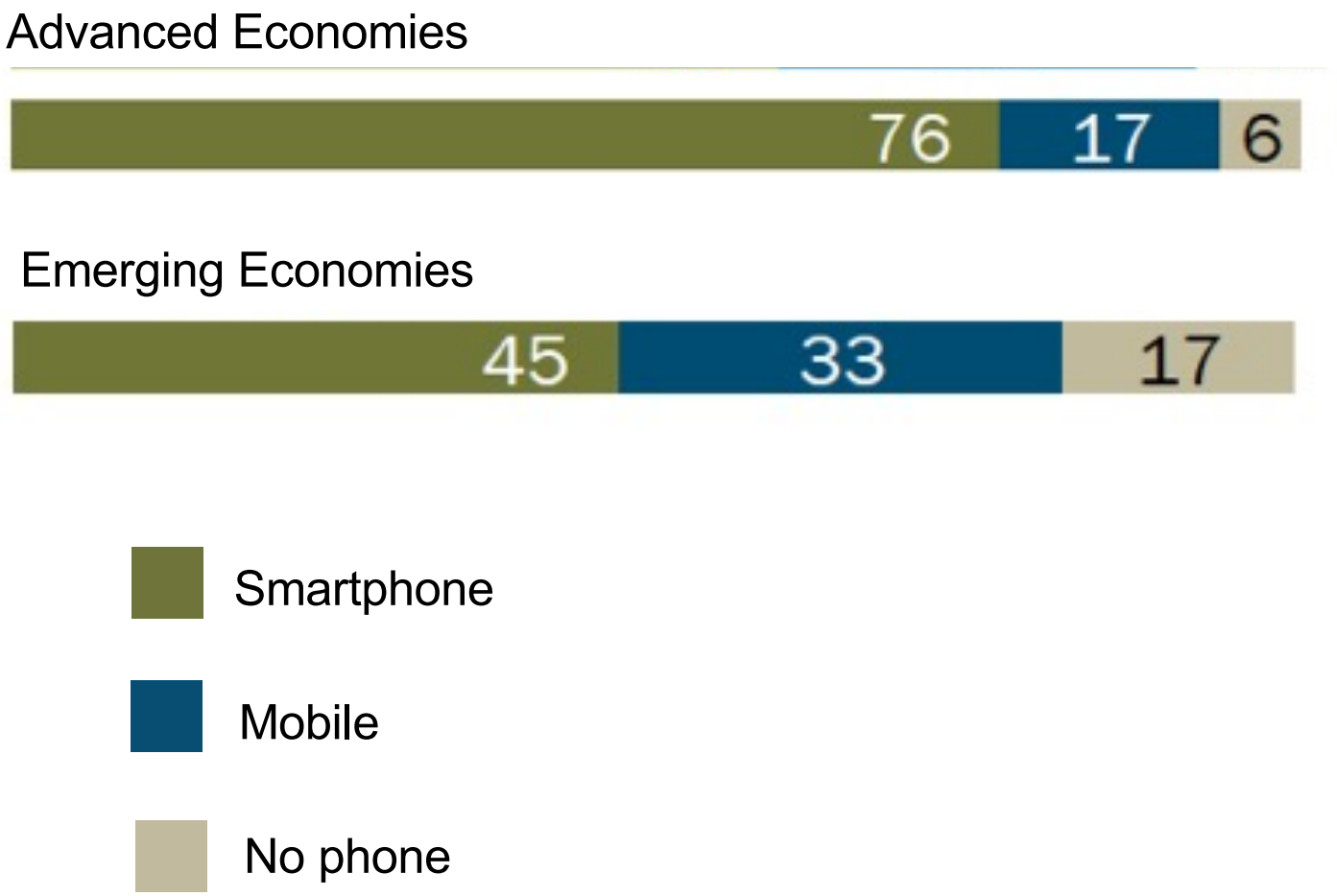
375 million workers need to be retrained by 2030 [3]

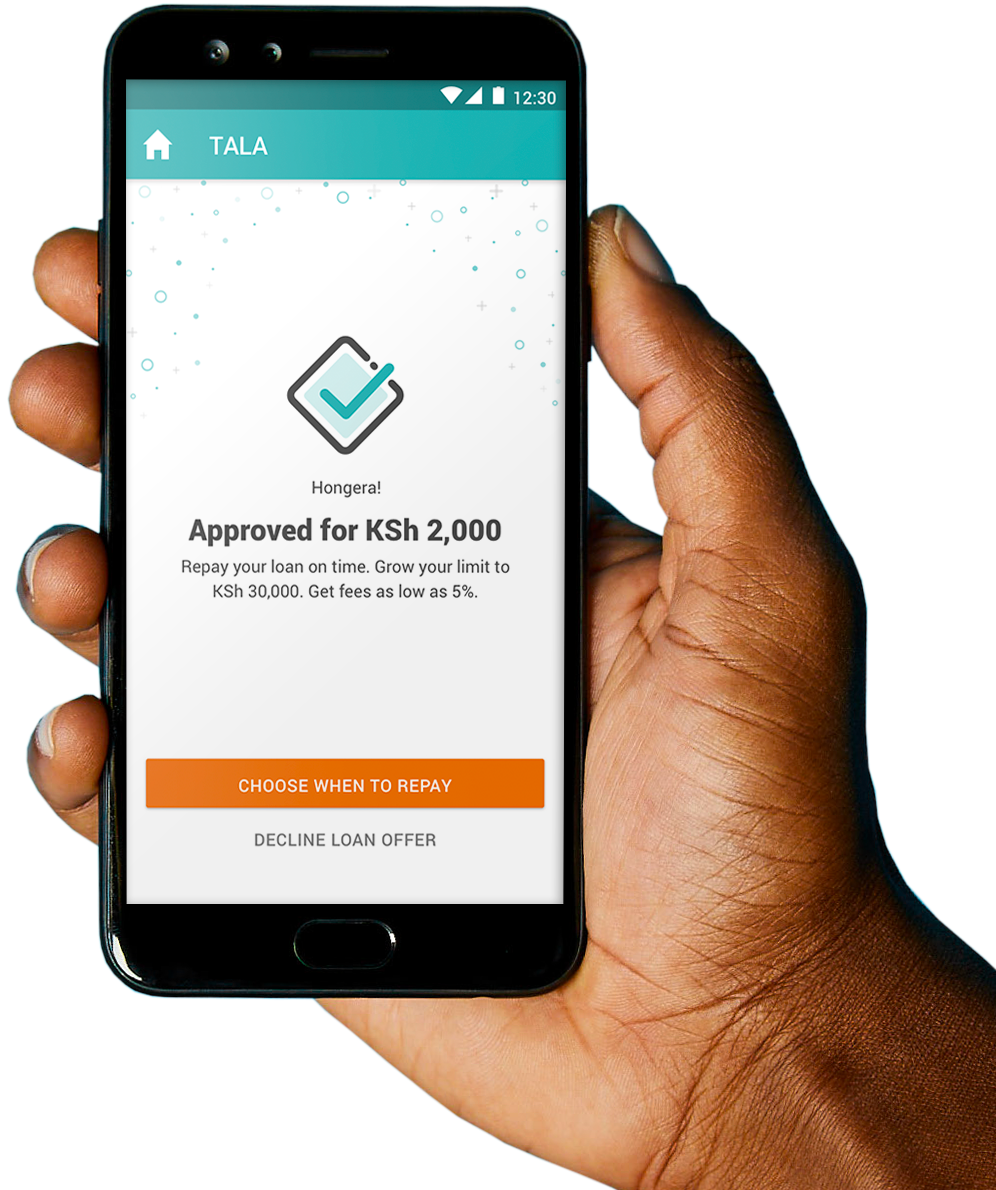
Half a million unfilled computer science jobs (60% of STEM jobs) [2]

For all learners we want *quality*.

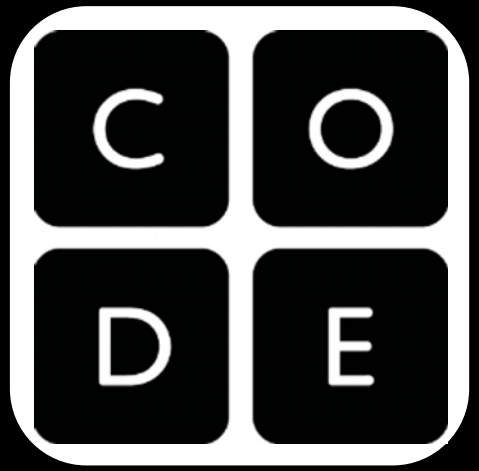
[1] *World development indicators 2015*. World Bank Publications, 2015.
[2] USA Bureau of Labor Statistics Employment Projections , 2016.
[3] Jobs lost, jobs gained. McKinsy Global Institute, 2017.

Smart Phone Access



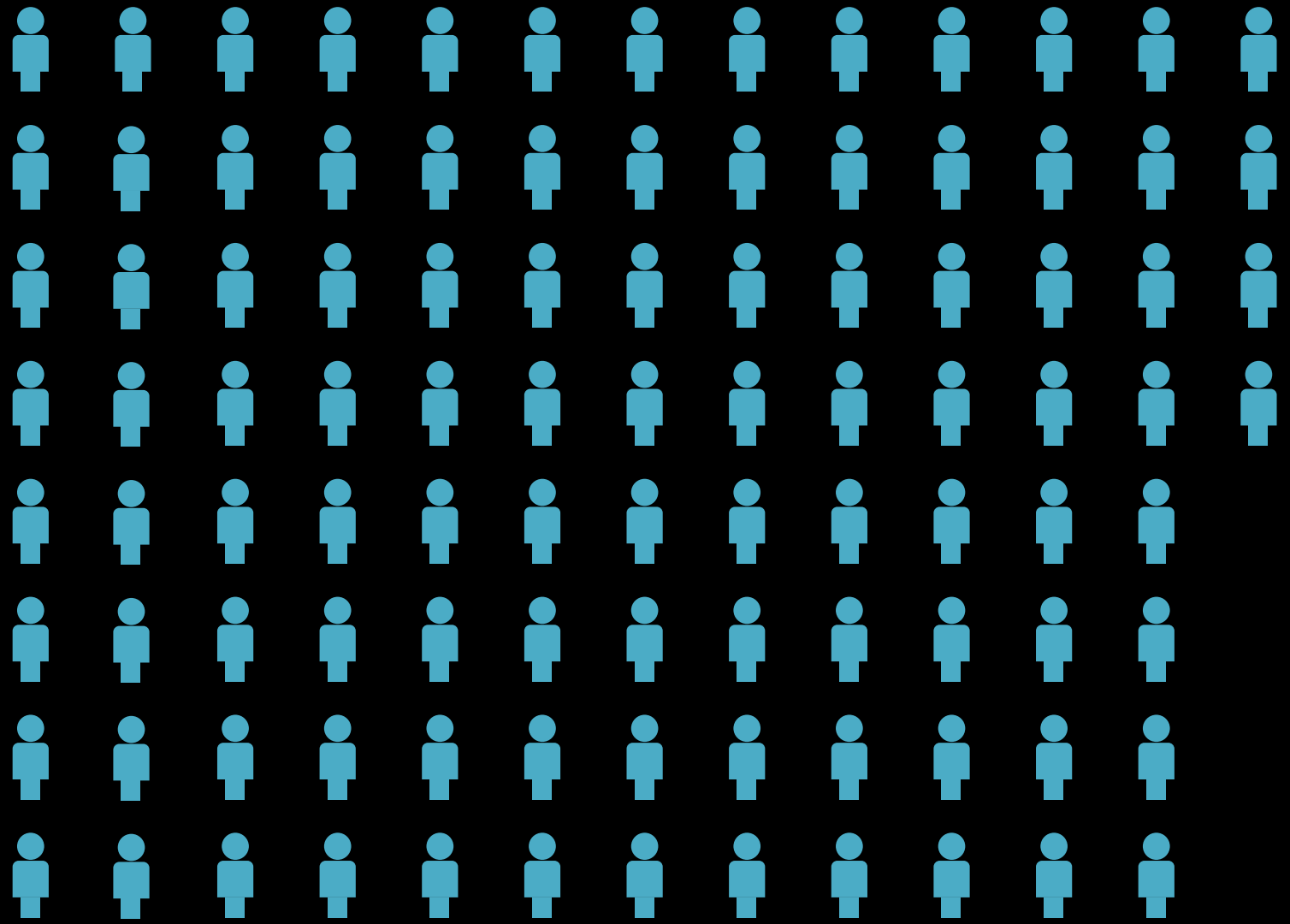


Unprecedented Data



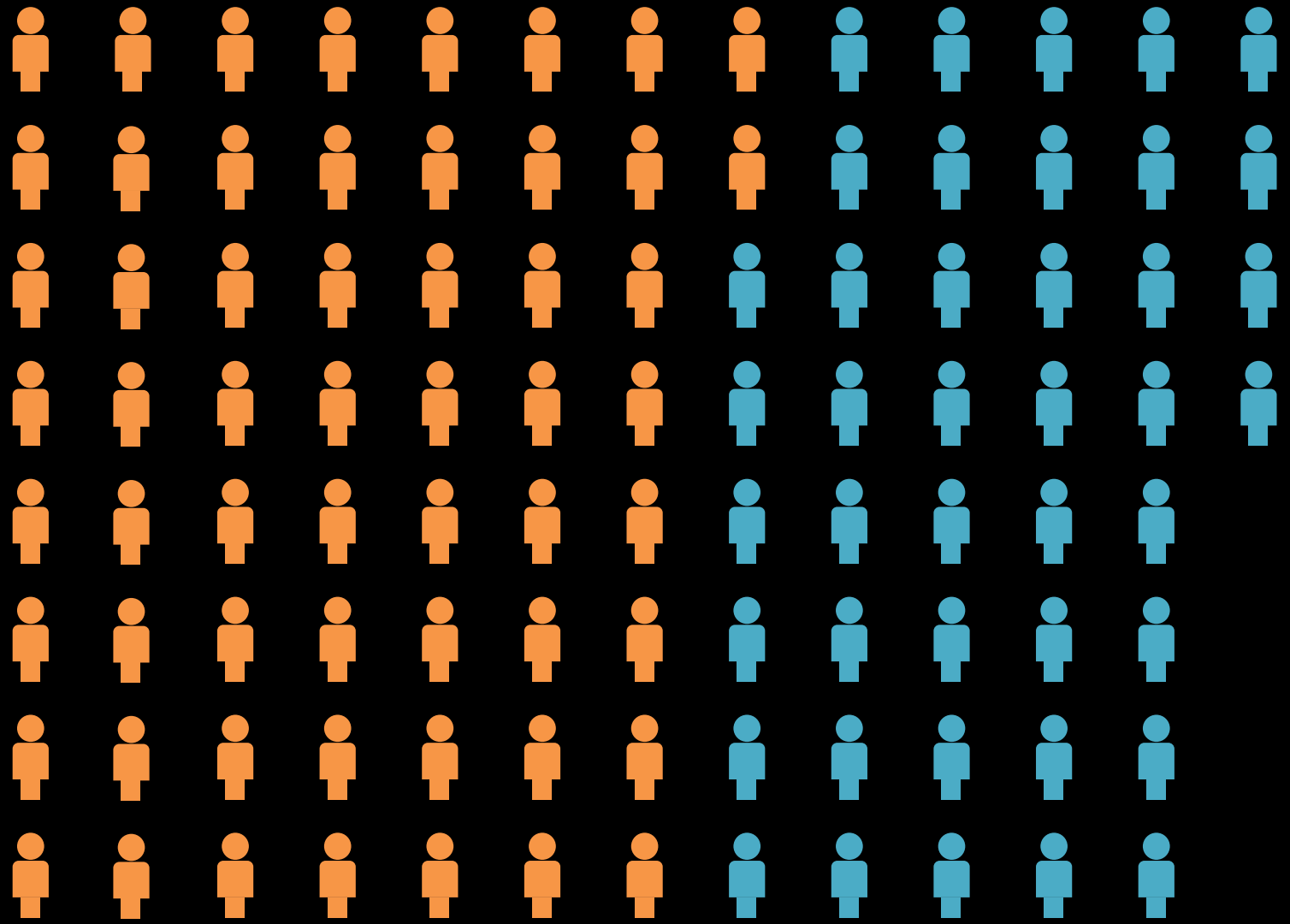
Over 50 million learners

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

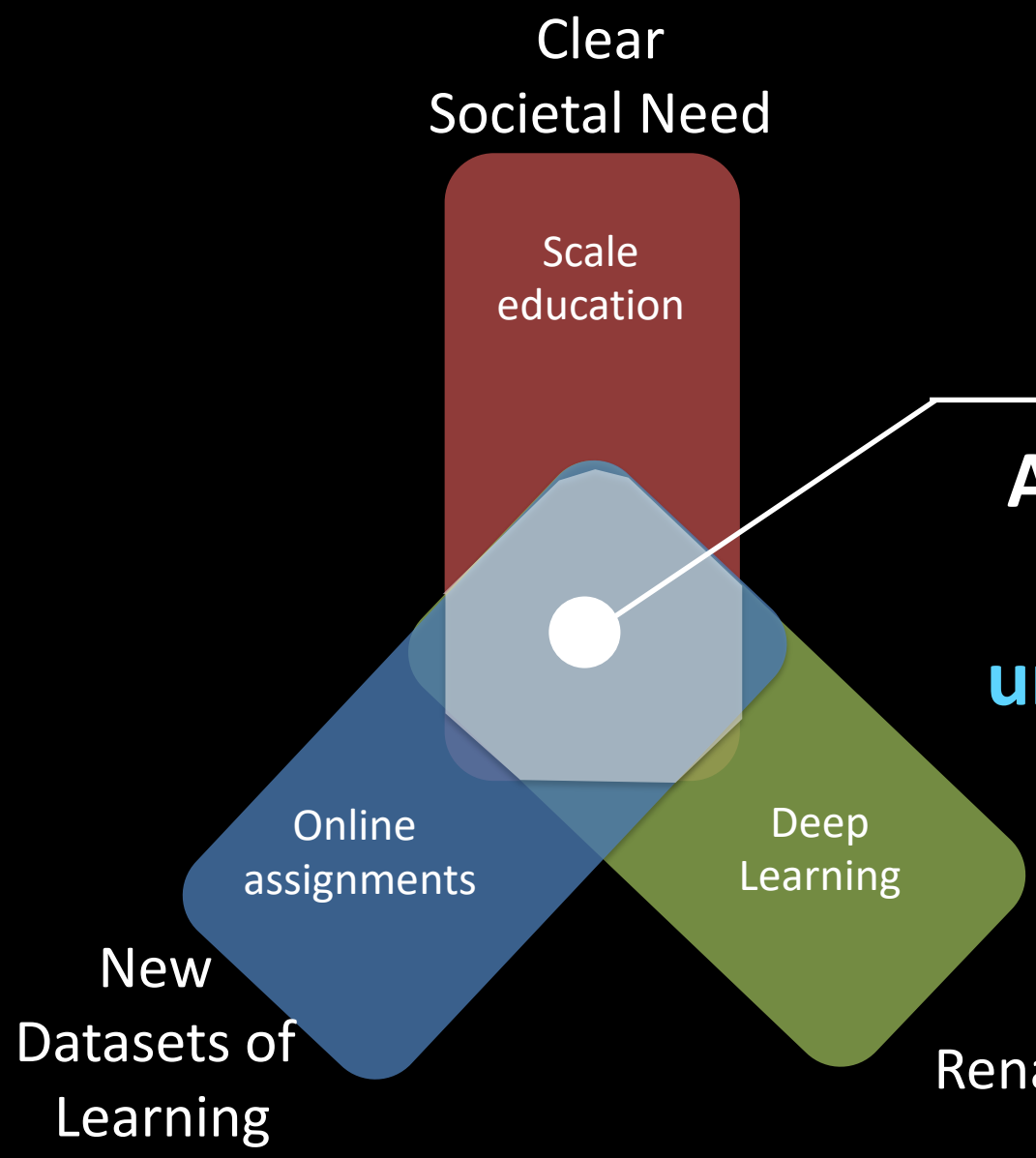
Used in **180+** countries

832M hour of code sessions

4 papers publish with our lab

50M K12 students in the US

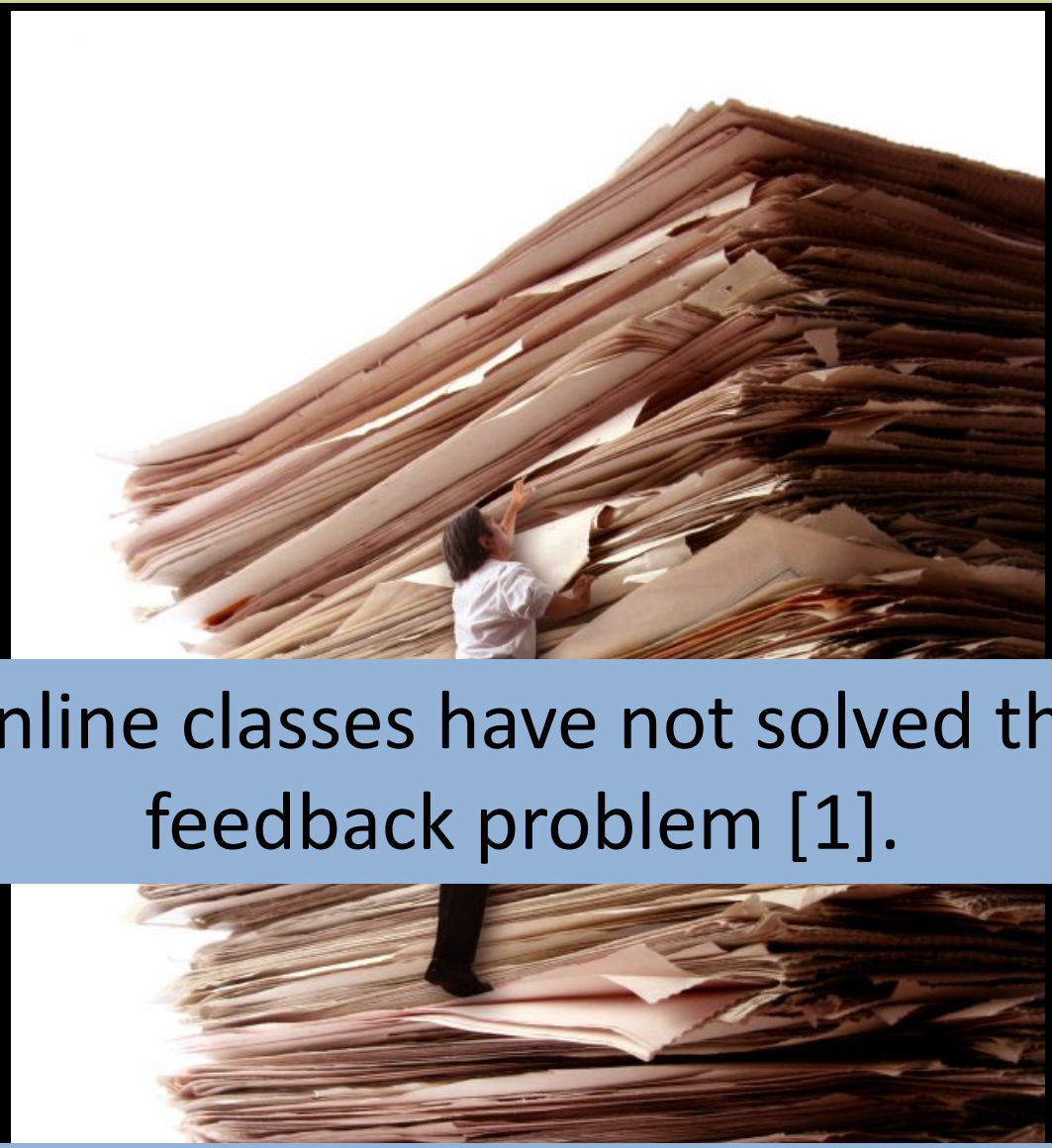
Speech Recognition



Grand Challenge in Education

Autonomously support education by better understanding students.

Feedback is Labor Intensive

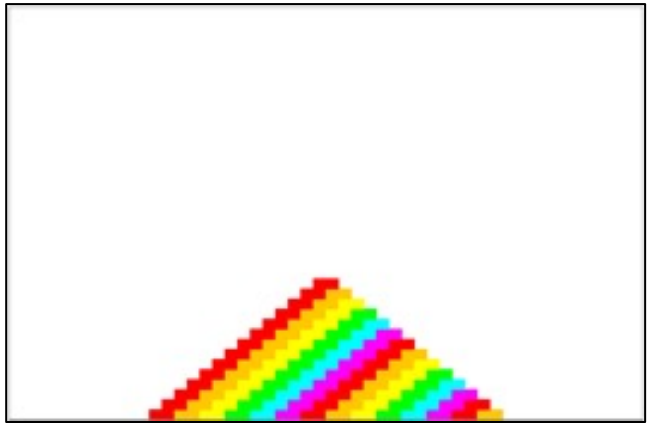


Online classes have not solved the feedback problem [1].

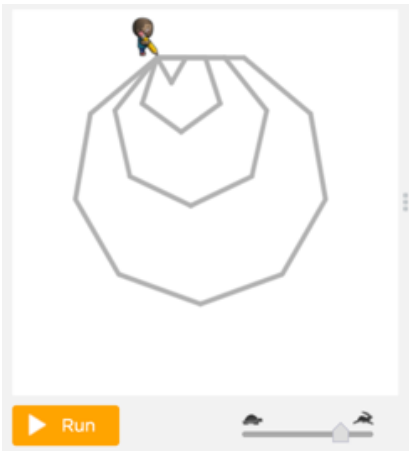
[1] Deconstructing Disengagement. Analysing learner subpopulations in MOOCs. Kizilcec, Piech, Schneider. Over 600 citations since 2013

Many domains of student work

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



Why did the original pilgrims come to America?



Chapter 0: Always start simple

First deep learning for education

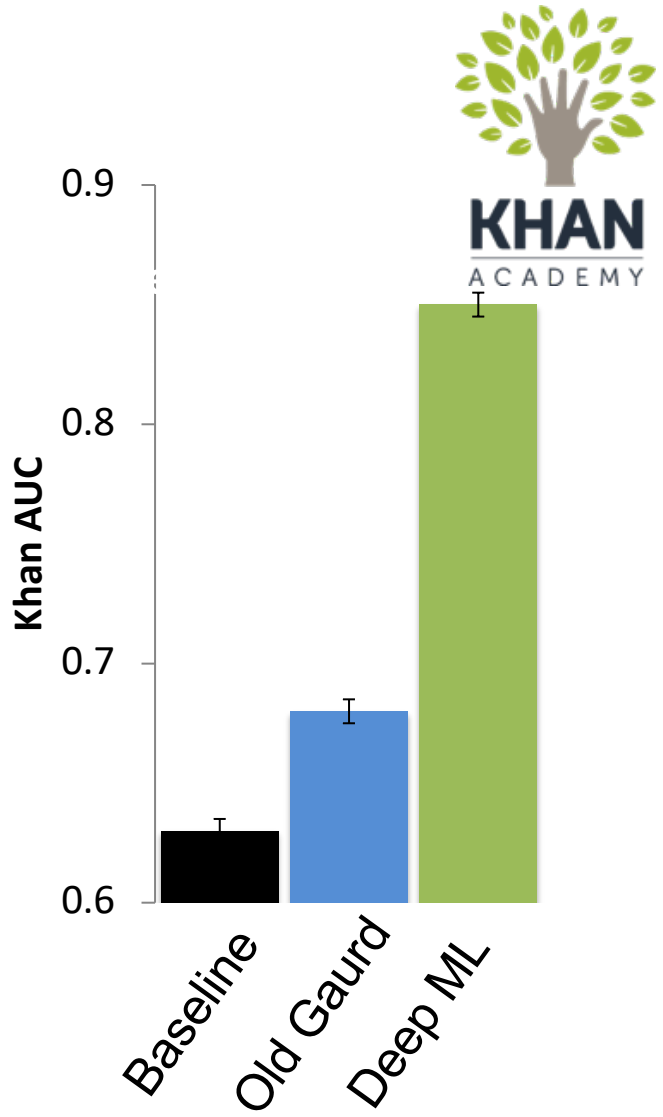
KHAN
Student

Exercise Type:

- Solving for x-intercept
- Solving for y-intercept
- Graphing linear equations
- Square roots
- Slope of a line

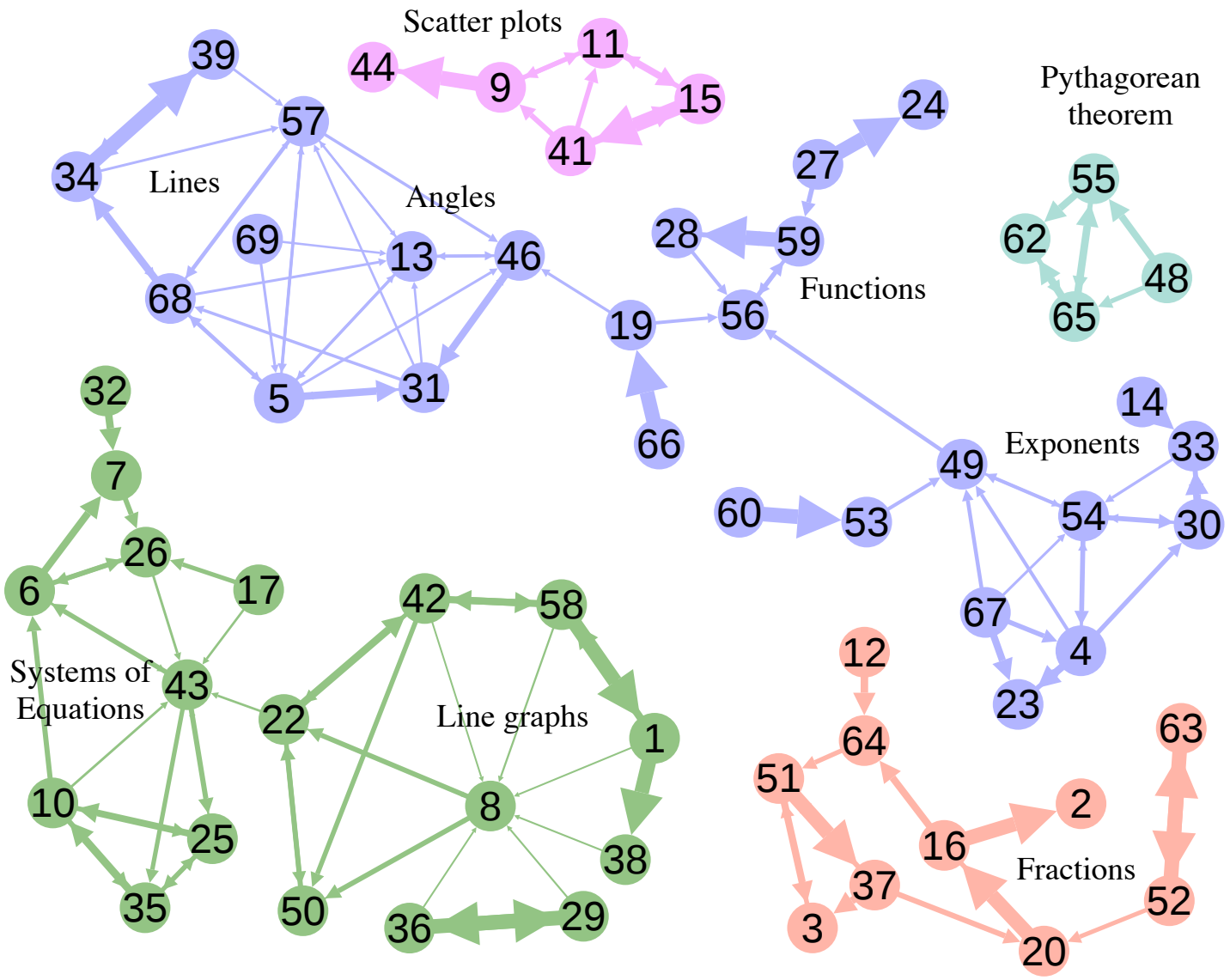
Answer:

- Correct
- Incorrect

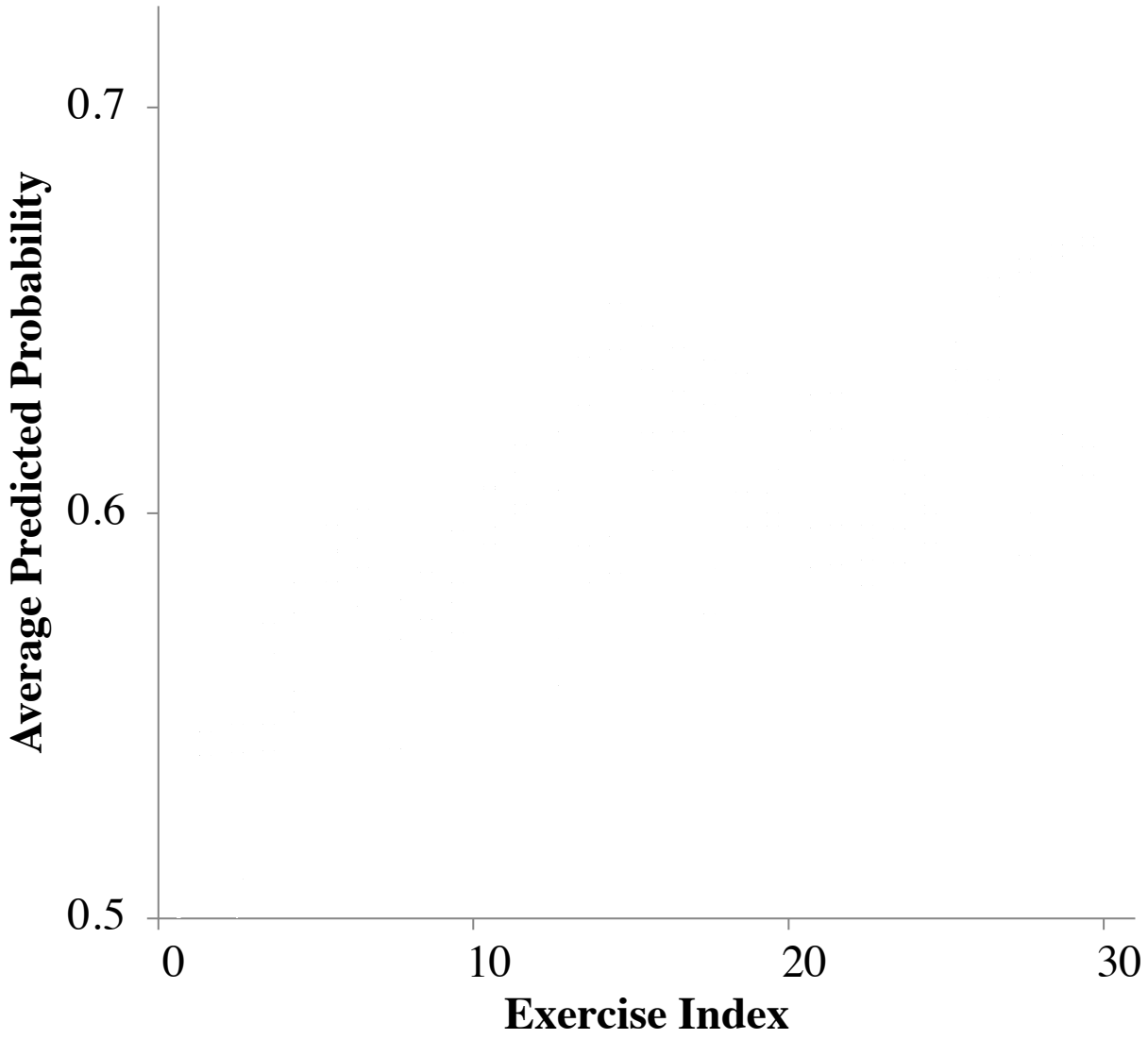


Old Problem

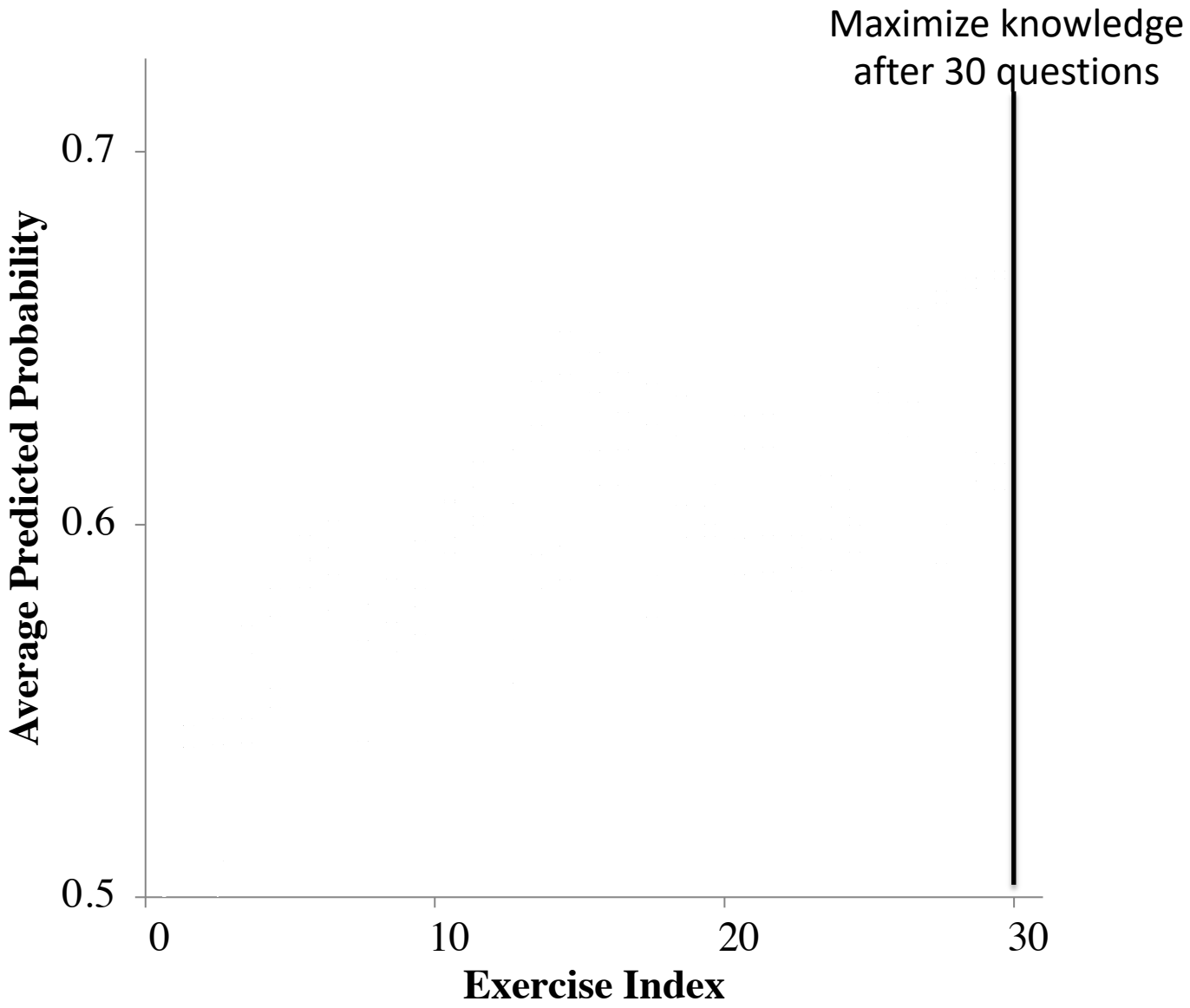
Learns Concept Relationships



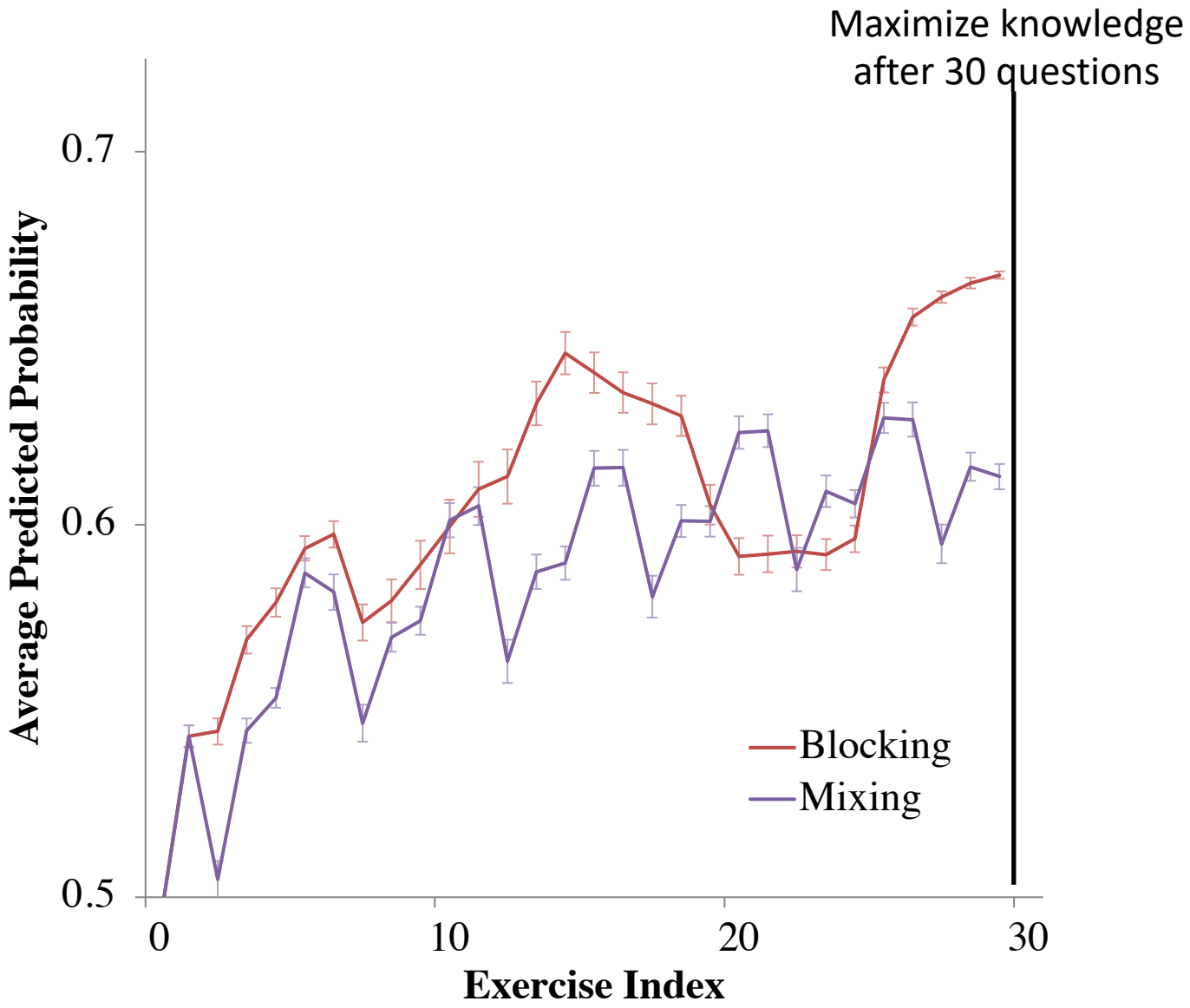
Optimal Teaching



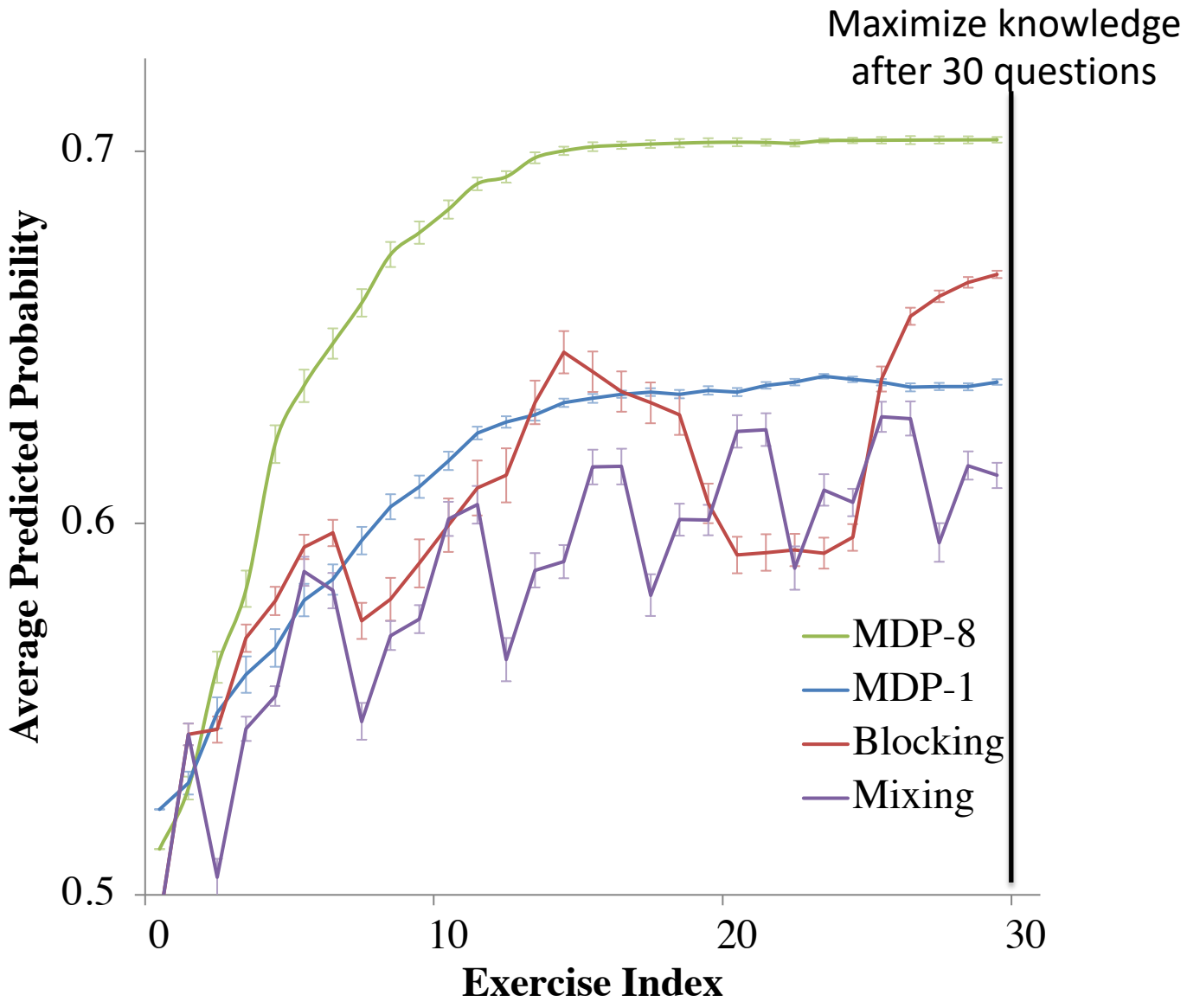
Optimal Teaching



Optimal Teaching

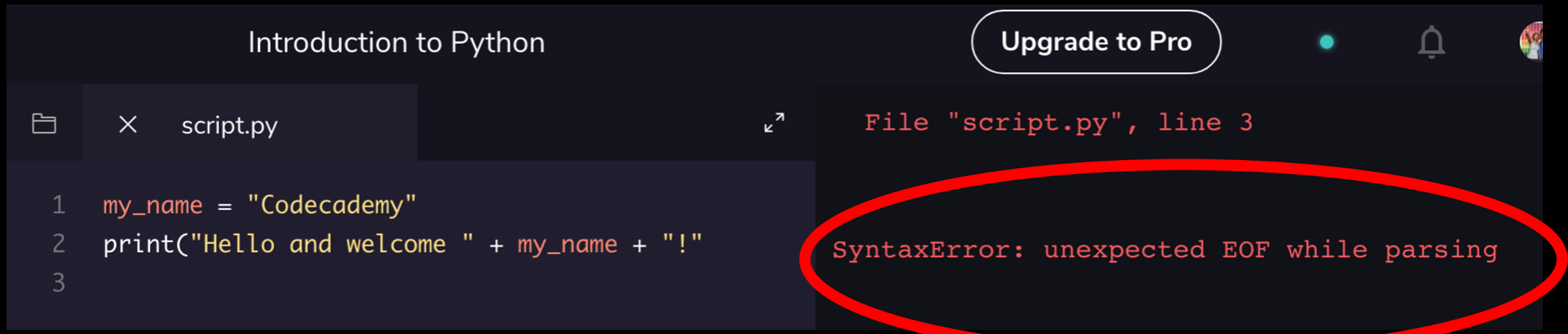


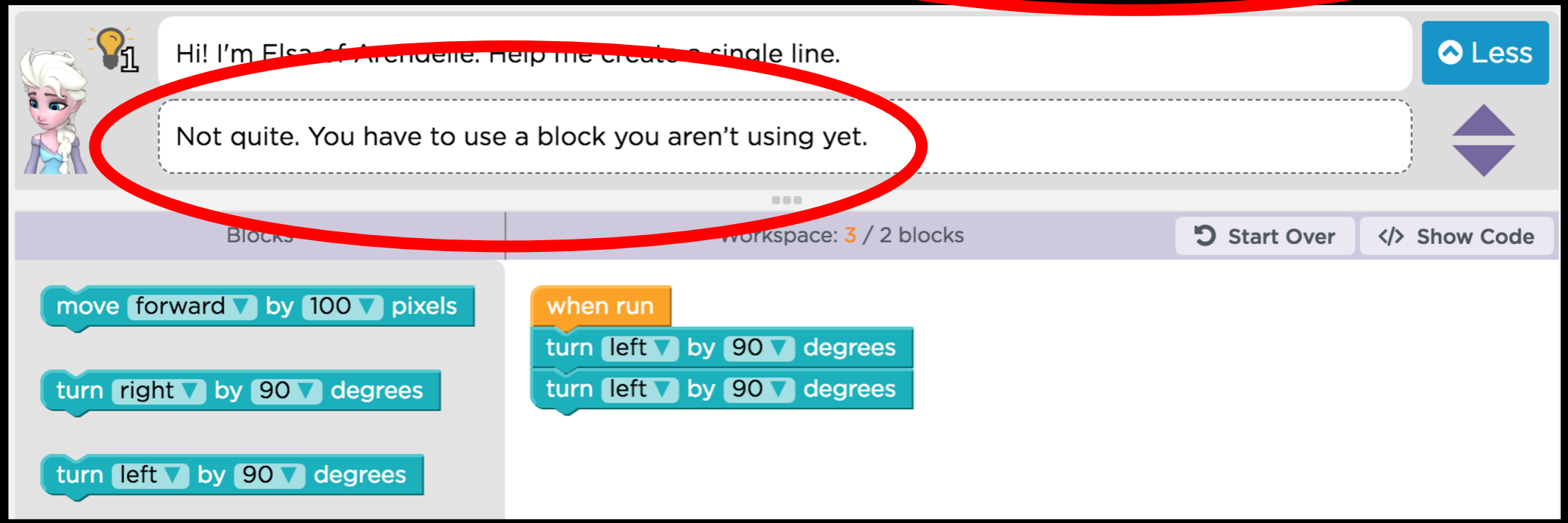
Optimal Teaching



We truly would rather move beyond correct / incorrect

Some domains are very hard





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

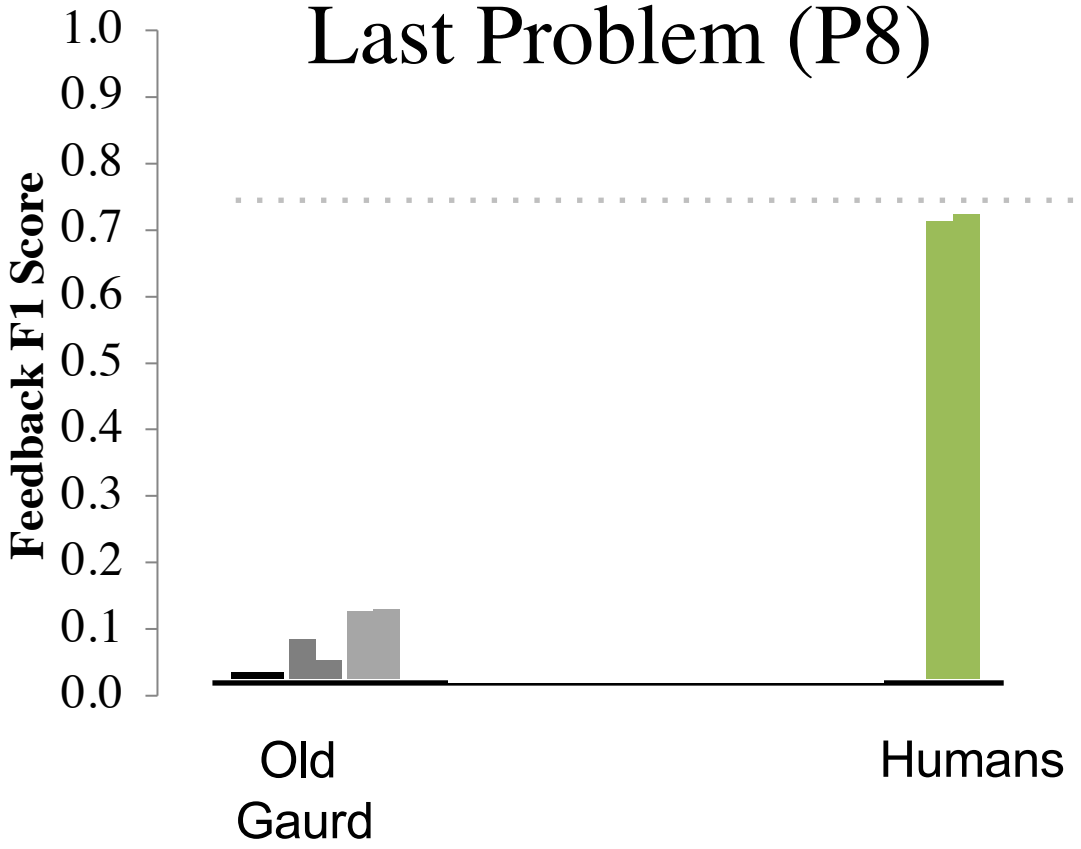
- ☑ Loop post condition doesn't match preconditi
- ☐ Repetition of bodies

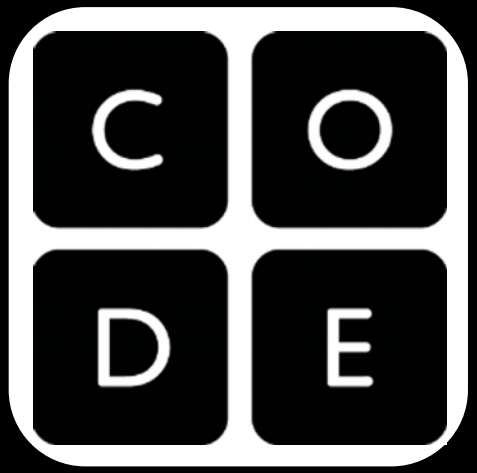
Cleanup

Record label

Traditional Deep Learning Doesn't Work

Label student code





Can we provide feedback by dynamic analysis?

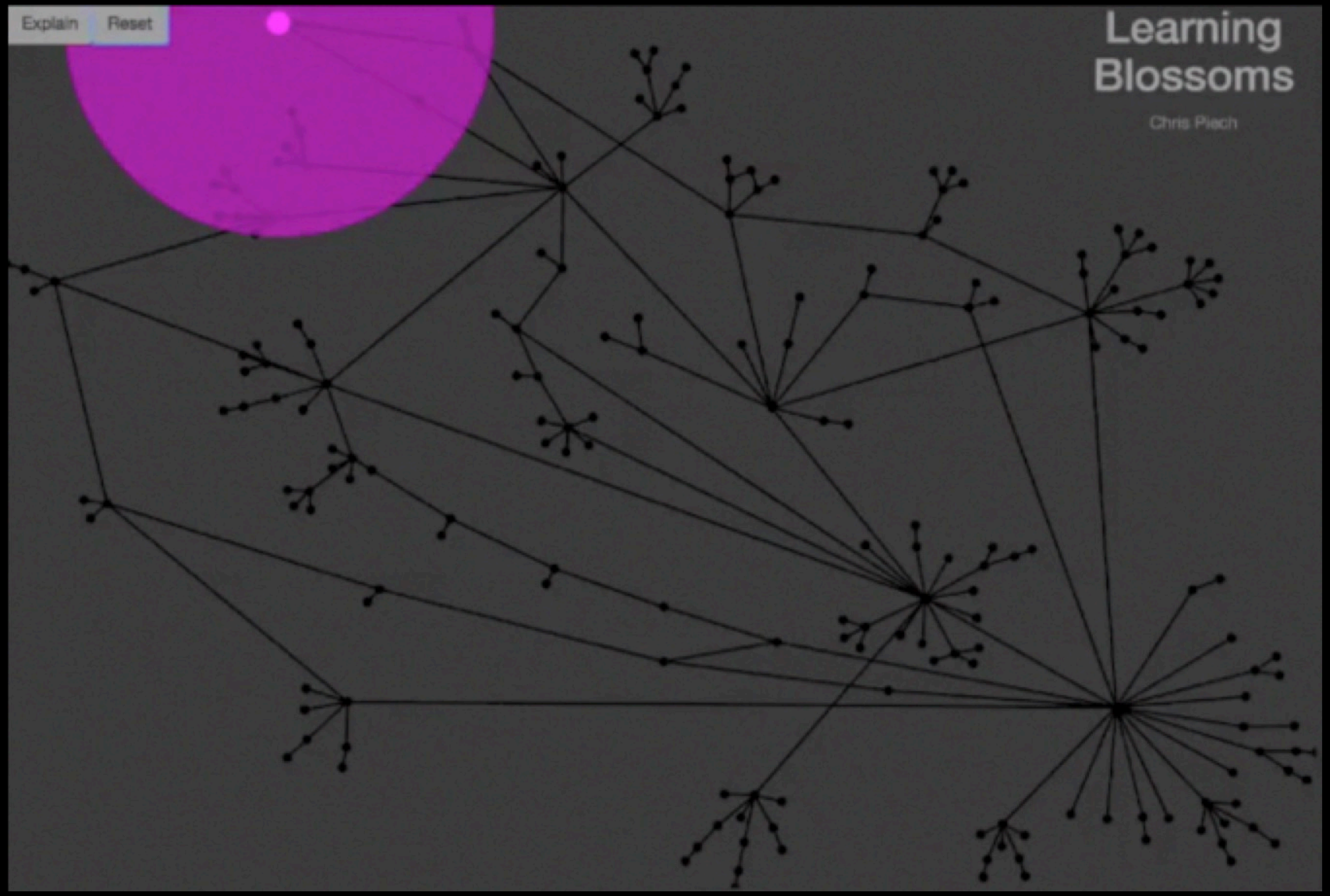
Starter code

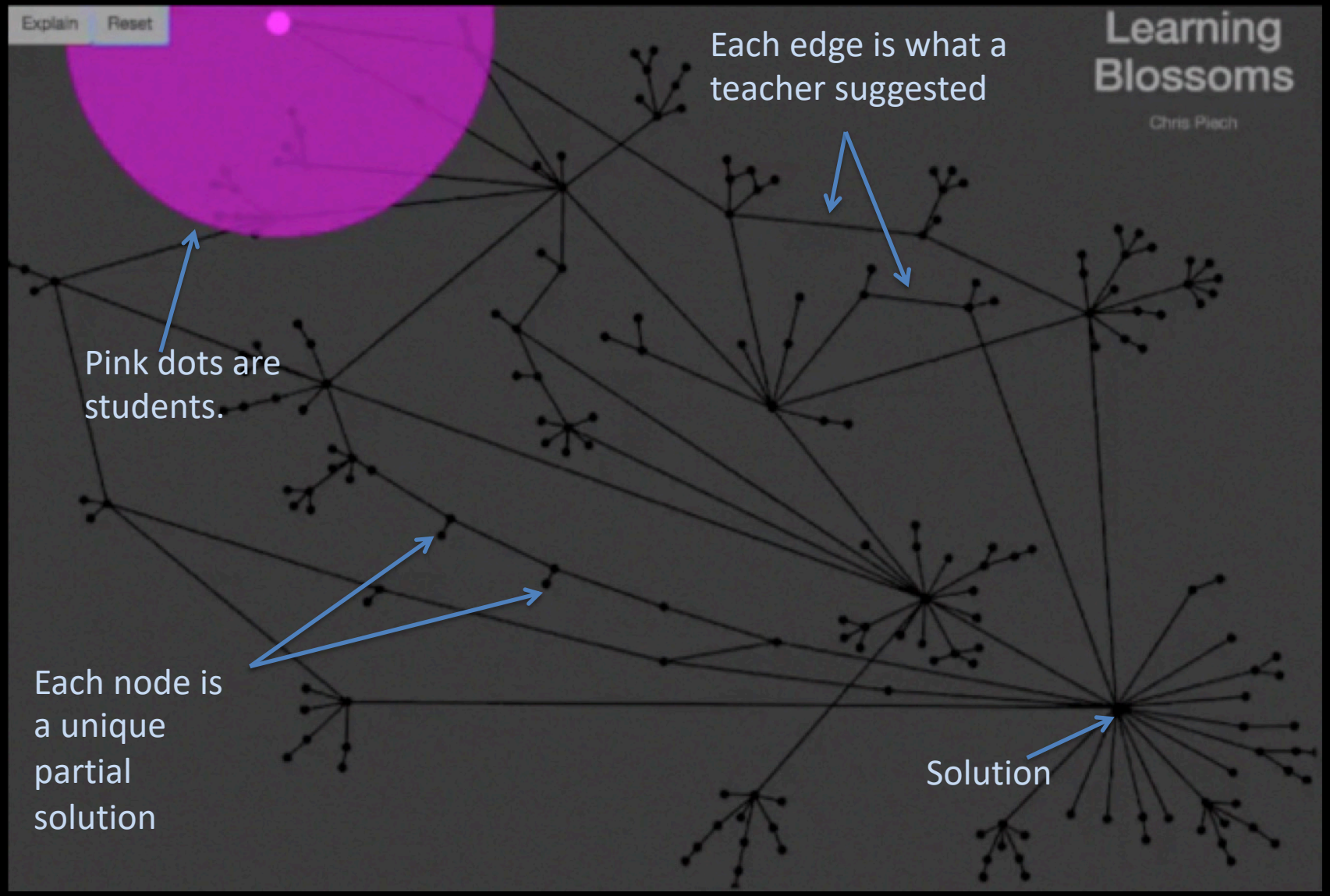
First attempt

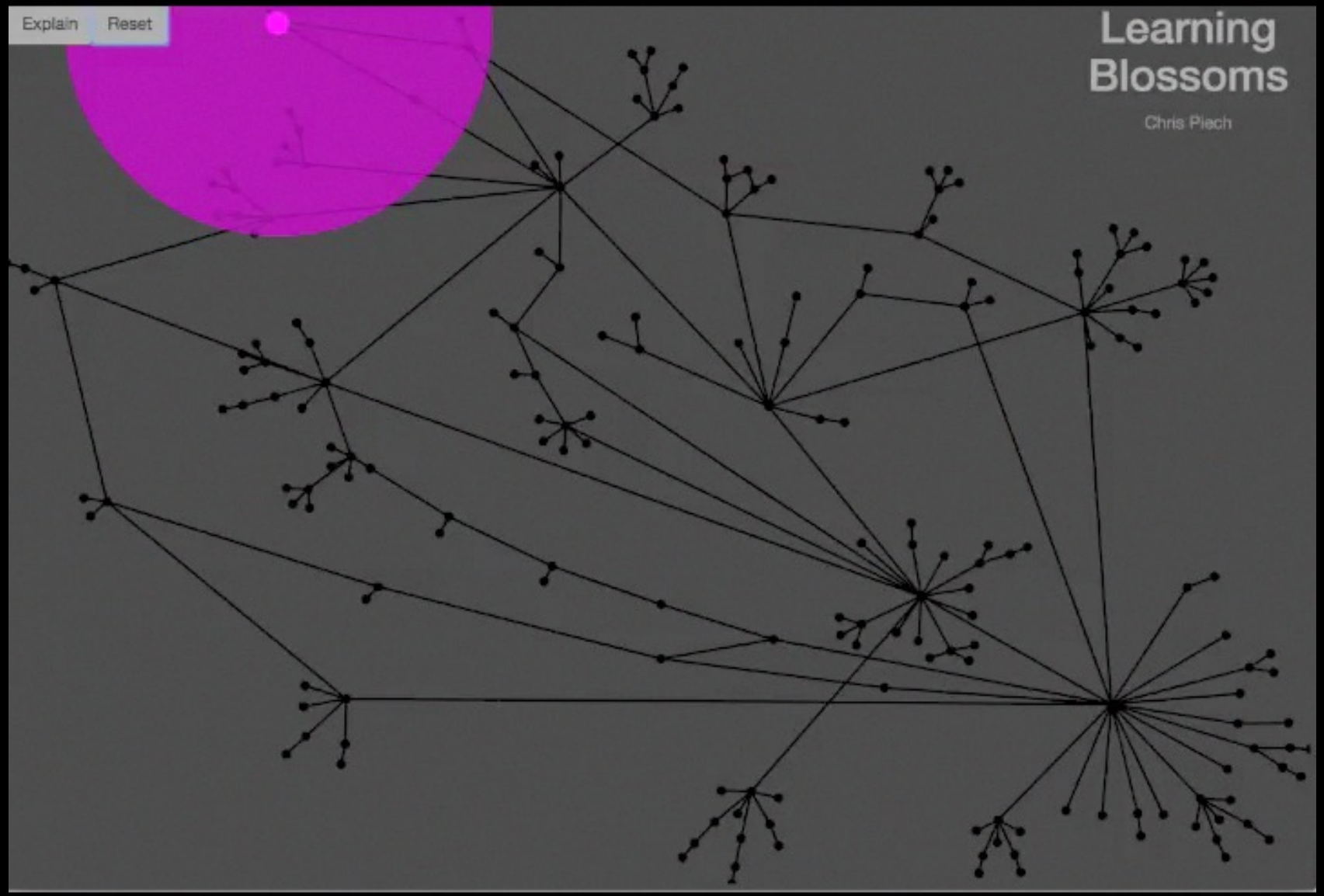
. . .

Final solution

Chapter 1: Better data source?


Explain
Reset
Learning Blossoms
Chris Piech


Explain
Reset
Learning Blossoms
Chris Piech
Each edge is what a teacher suggested
Pink dots are students.
Each node is a unique partial solution
Solution


Explain
Reset
Learning Blossoms
Chris Piech

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

$$\gamma(s) =$$

Predicted next partial solution

First step in the ***most frequent*** path to the solution from *s*, taken by ***average*** students. Assume poison process.

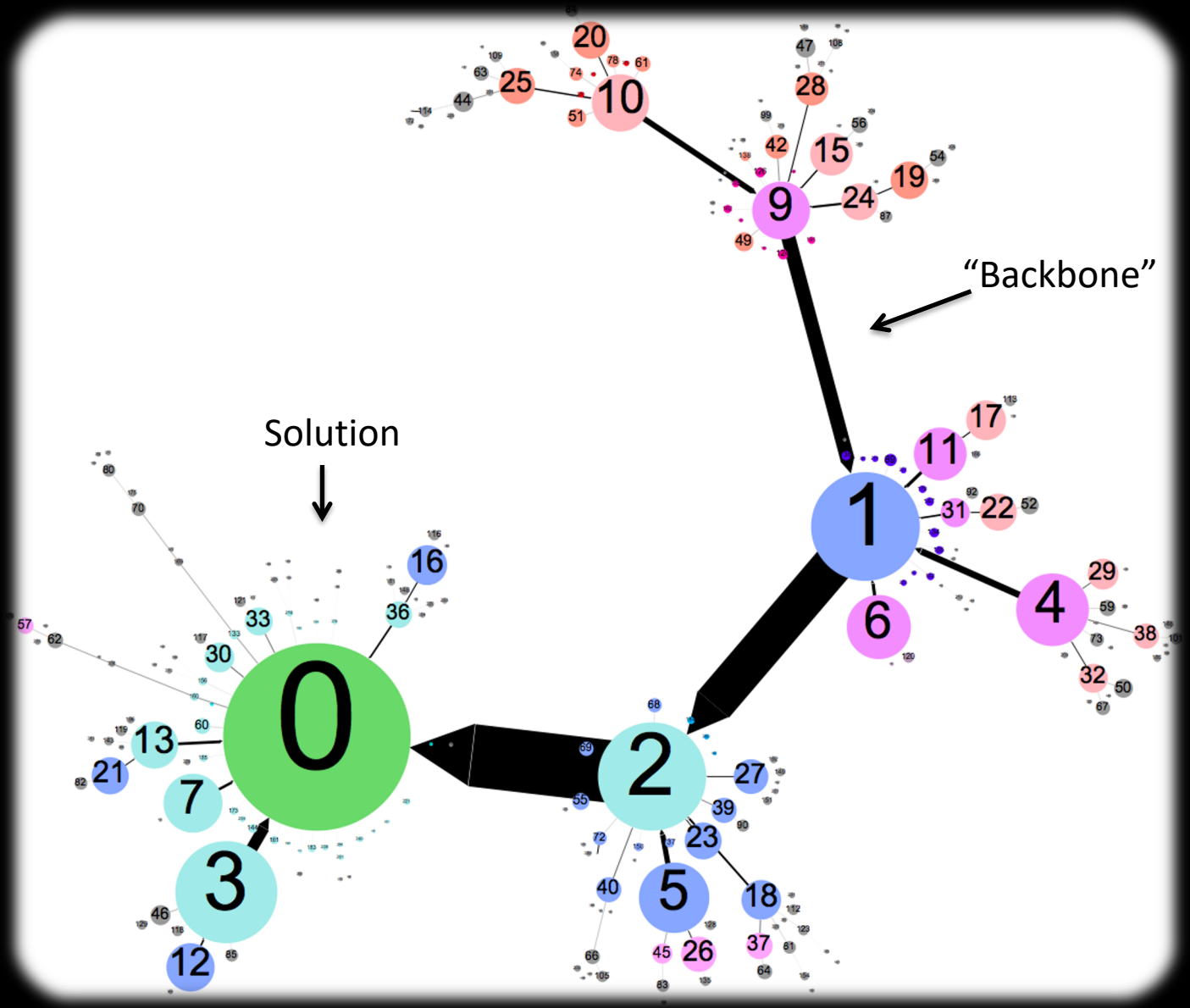
Desirable Path Algorithms

Poisson Common Path

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

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

Learned Problem Solving Policy



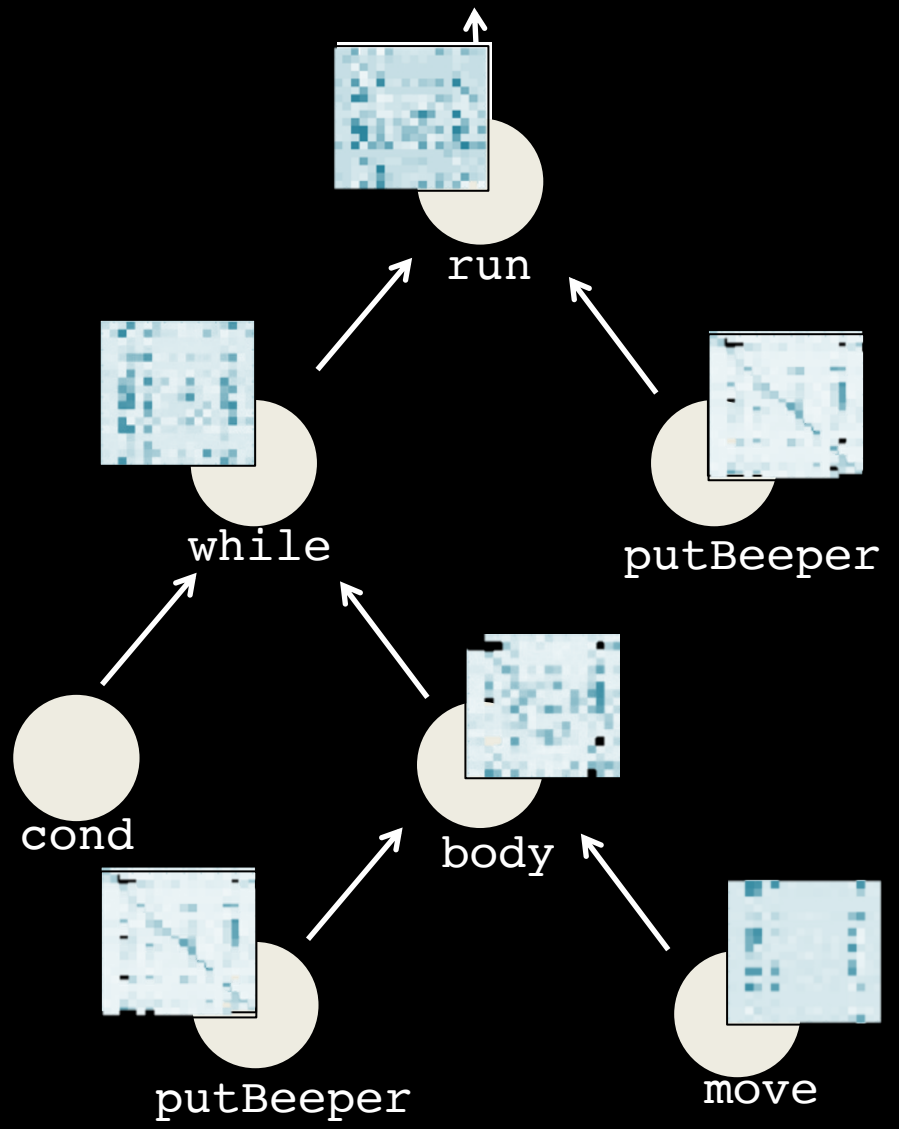
Only worked well for 6 line programs…

Chapter 2: Start to invent new algorithms…

Neural Network to Encode a Program

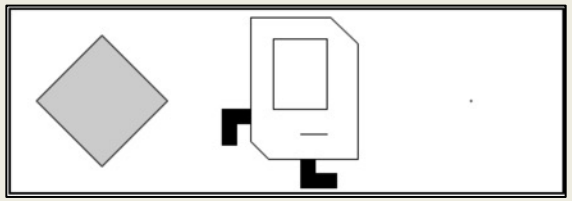
```
// User defined method
private void run() {
   while(isClear()){
      putBeeper();
      move();
   }
   putBeeper();
}
```

It looks like you have a fencepost error!



*Note: this was coded pre-tensor flow

Collect Triples



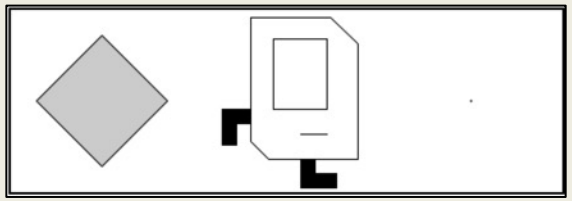
Precondition

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

Code

About 5 million triples per assignment

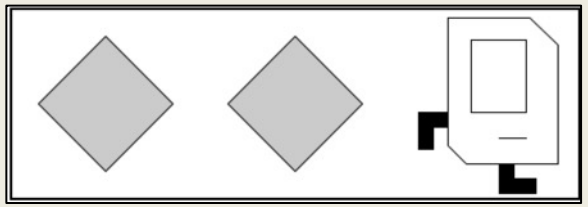
Collect Triples



Precondition

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

Code

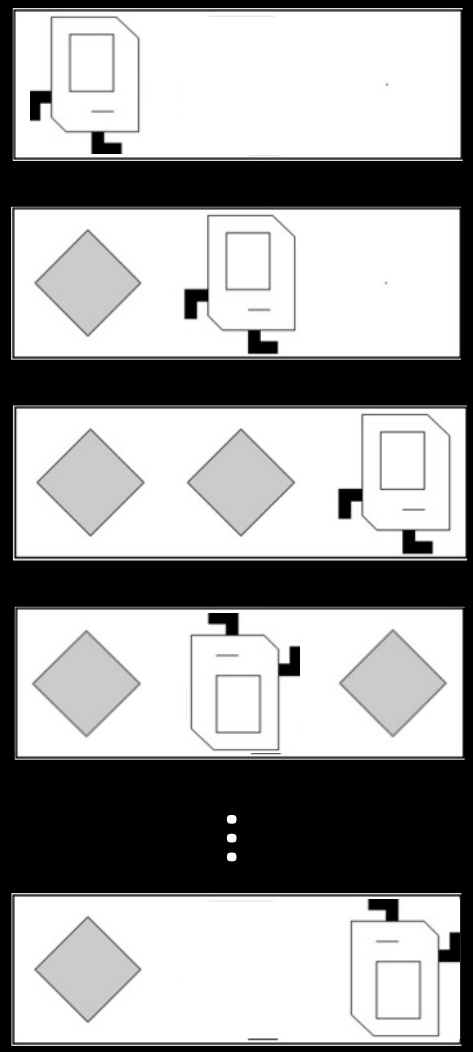


Postcondition

About 5 million triples per assignment

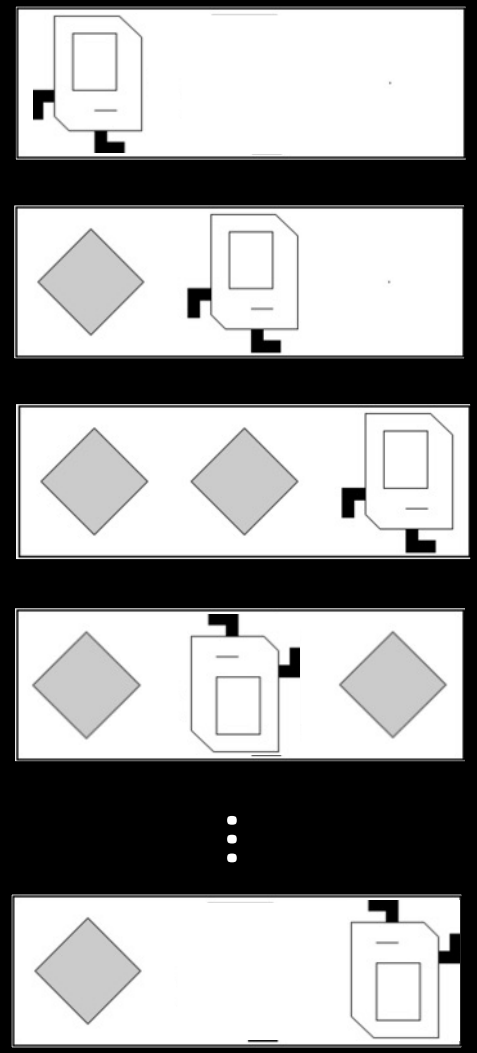
A Code Phrase is a Mapping

All possible preconditions

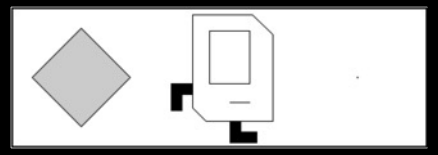


```
putBeeper();
move();
```

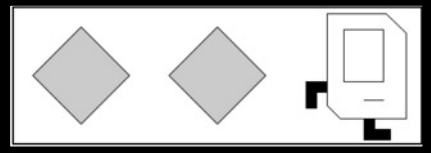
All possible postconditions



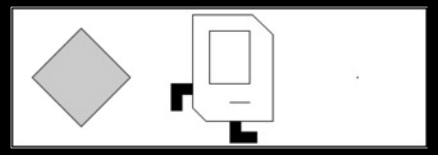
Neural Network for Programs



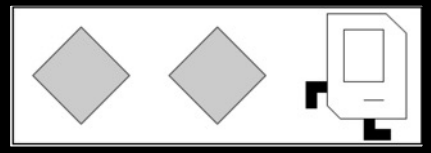
```
method step() {
    putBeeper();
    move();
}
```



Neural Network for Programs



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

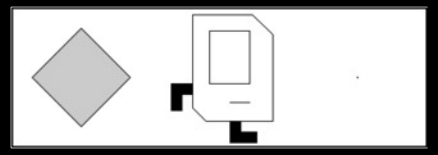


Raw Precondition

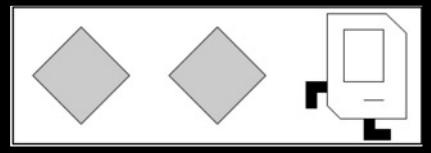
Program Matrix

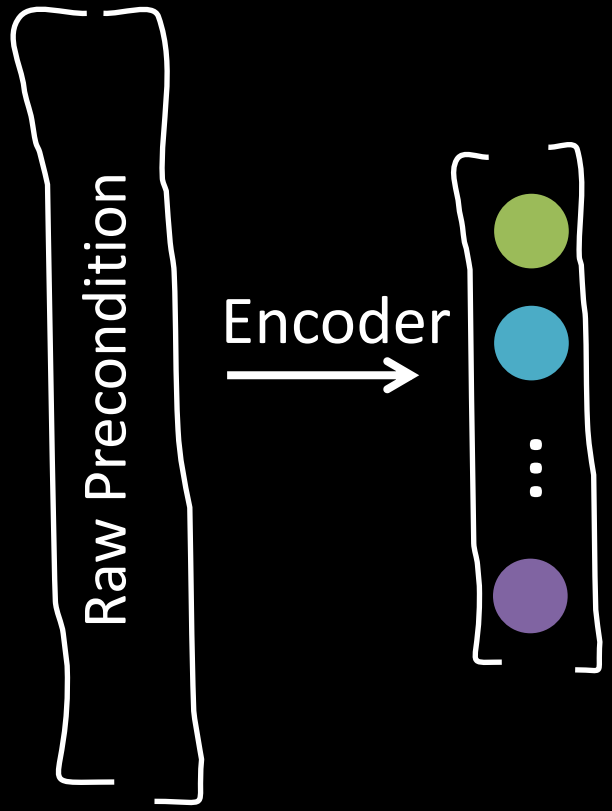
*coded pre-tensor flow

Neural Network for Programs



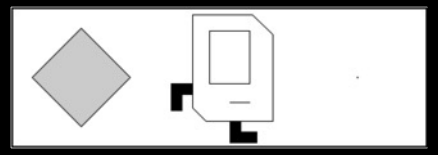
```
method step() {
    putBeeper();
    move();
}
```



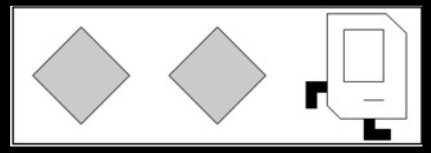


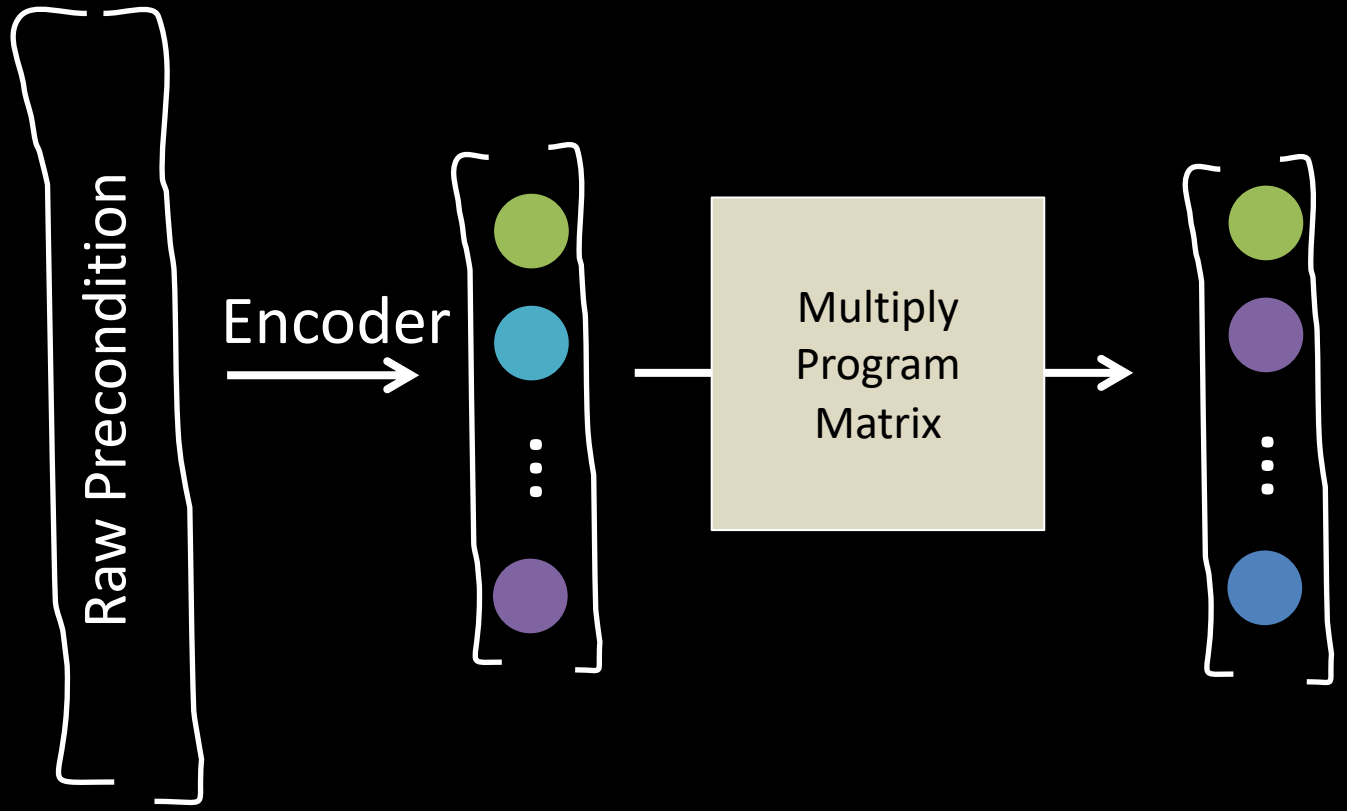
*coded pre-tensor flow

Neural Network for Programs



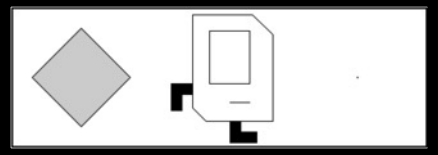
```
method step() {
    putBeeper();
    move();
}
```



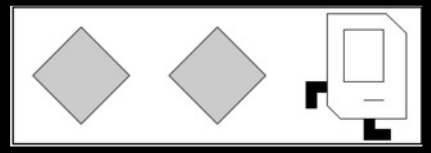


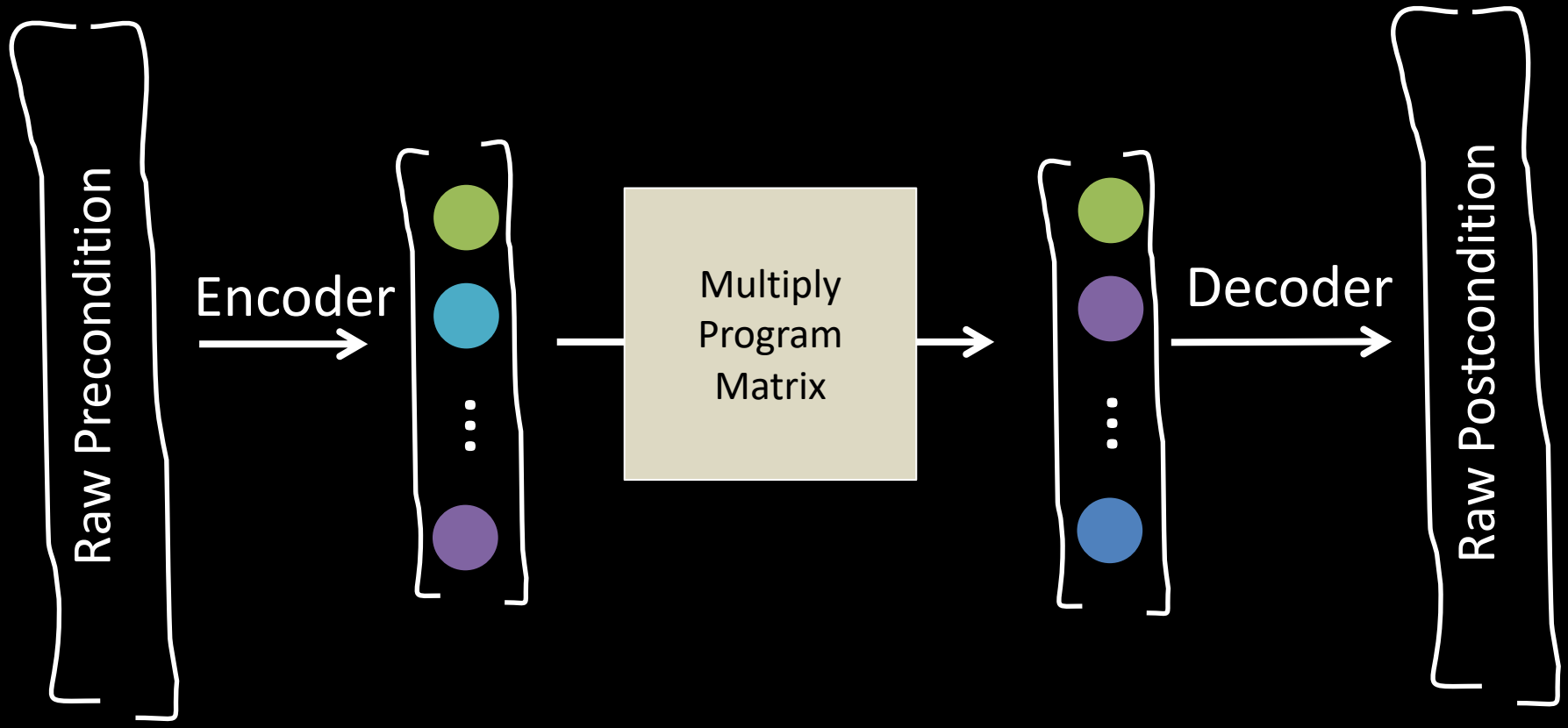
*coded pre-tensor flow

Neural Network for Programs



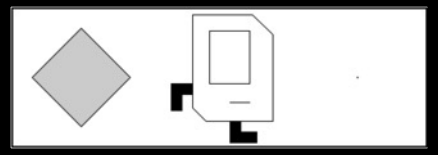
```
method step() {
    putBeeper();
    move();
}
```



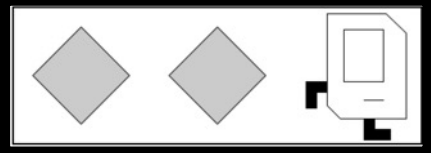


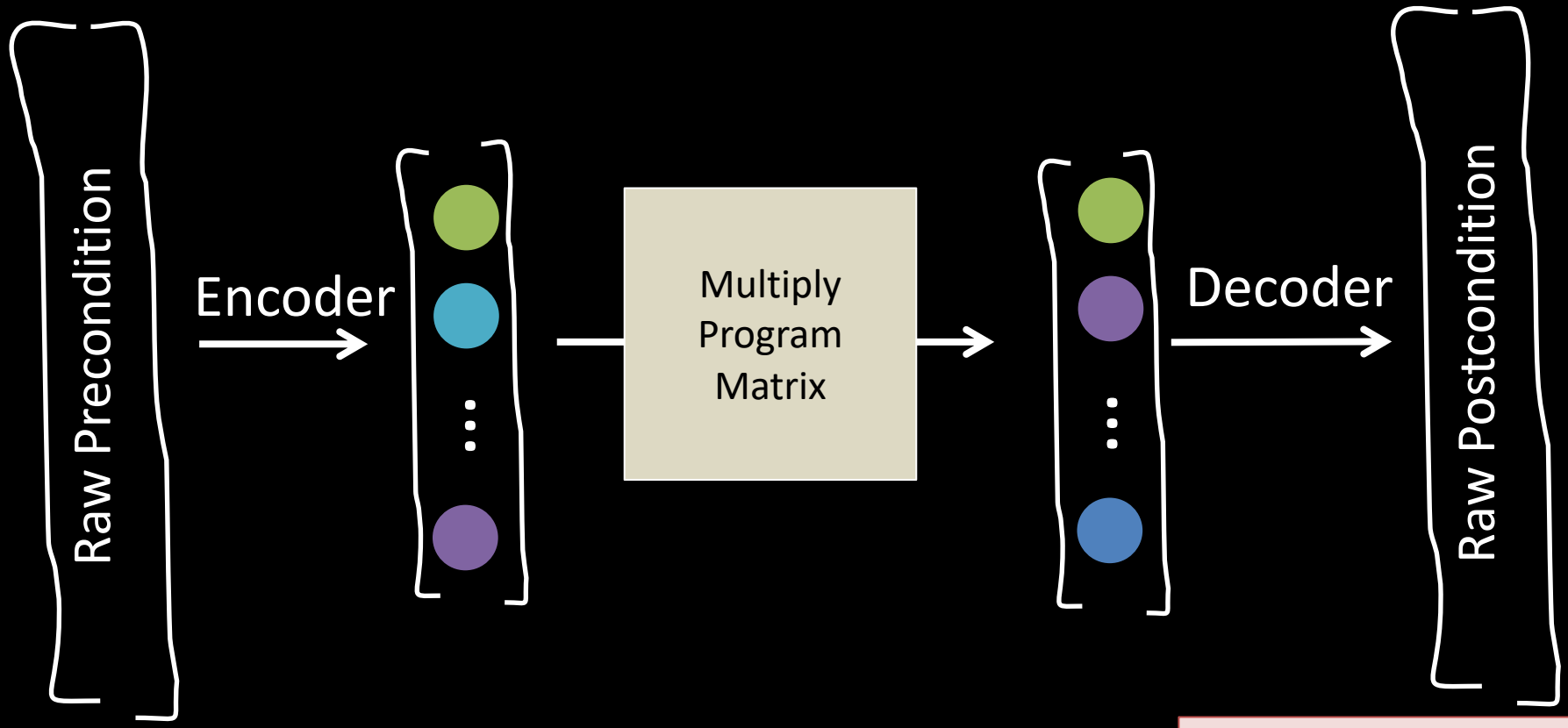
*coded pre-tensor flow

Neural Network for Programs



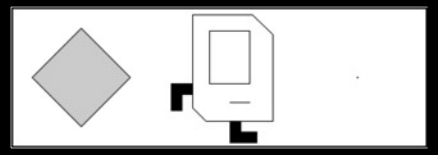
```
method step() {
    putBeeper();
    move();
}
```



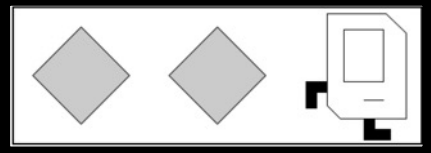


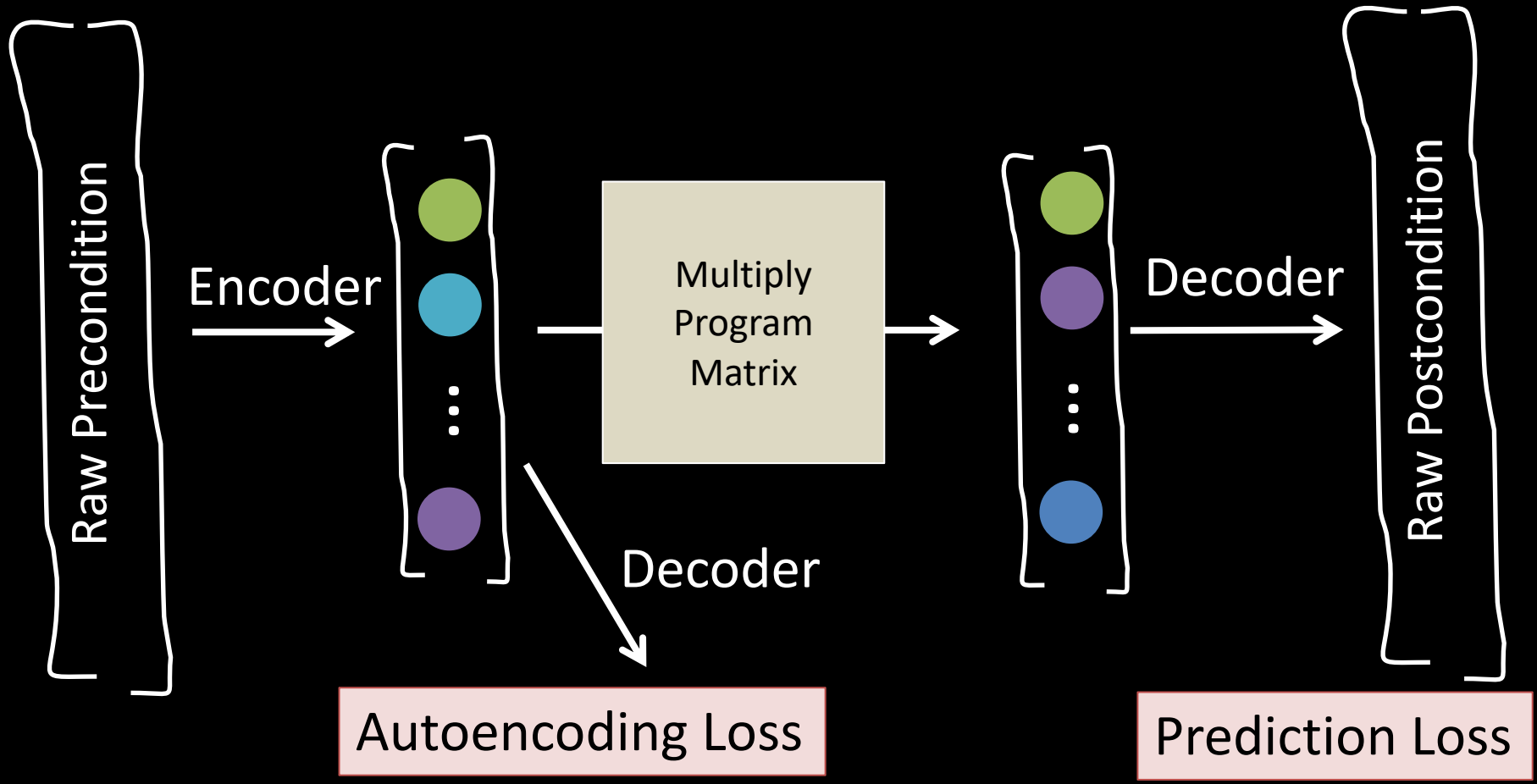
*coded pre-tensor flow

Neural Network for Programs



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





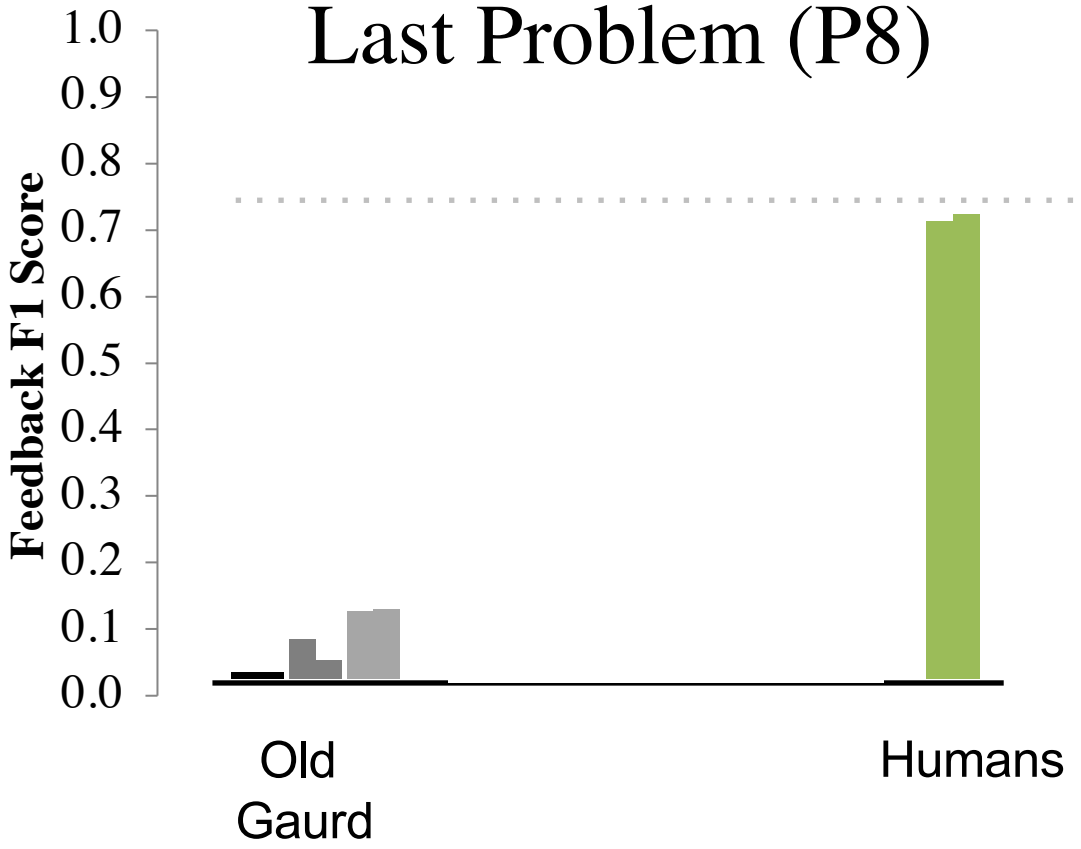
*coded pre-tensor flow

Does it work?


LELAND STANFORD JUNIOR UNIVERSITY
DIE LUFT DER FREIHEIT WEHT
1891

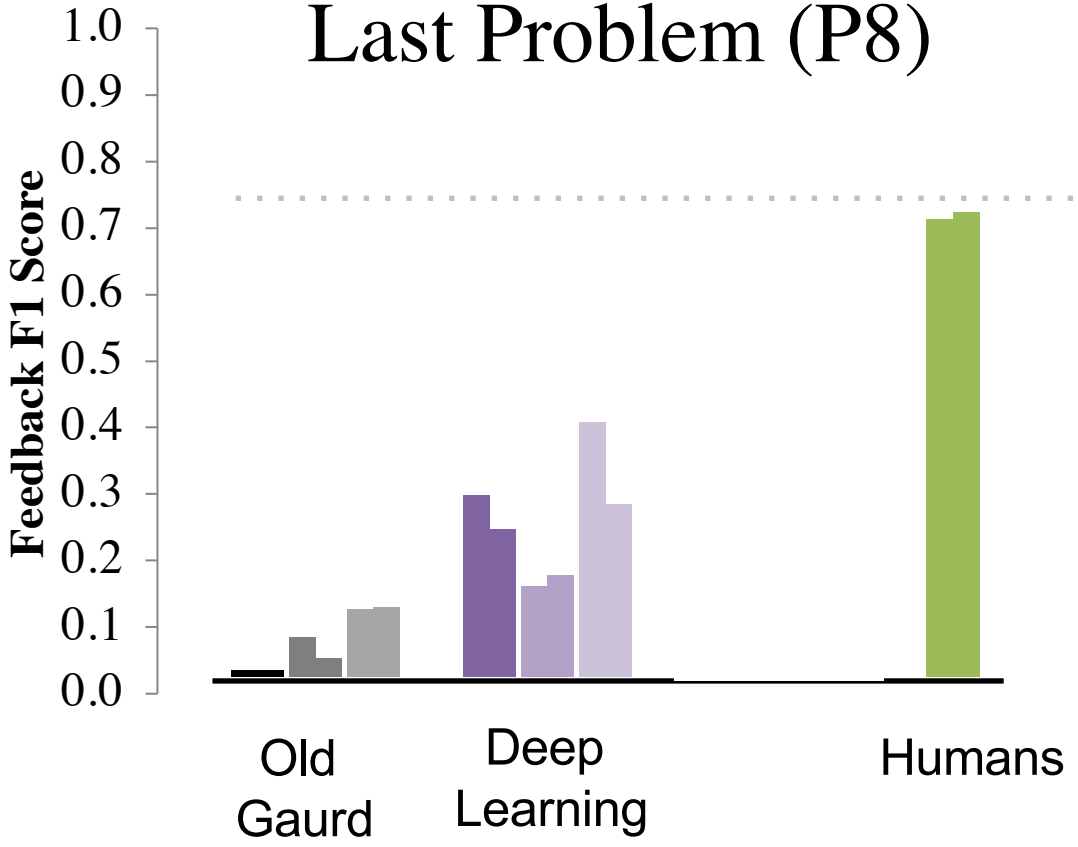
Traditional Deep Learning Doesn't Work

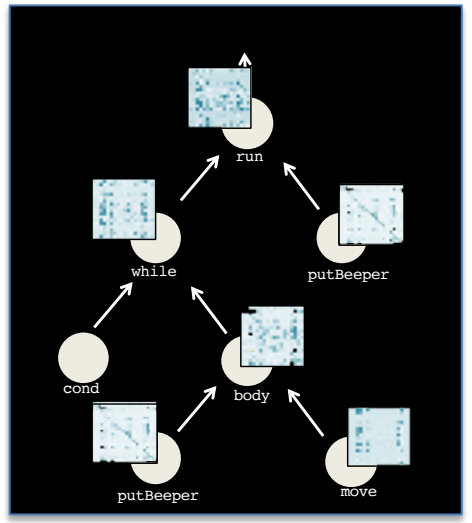
Label student code

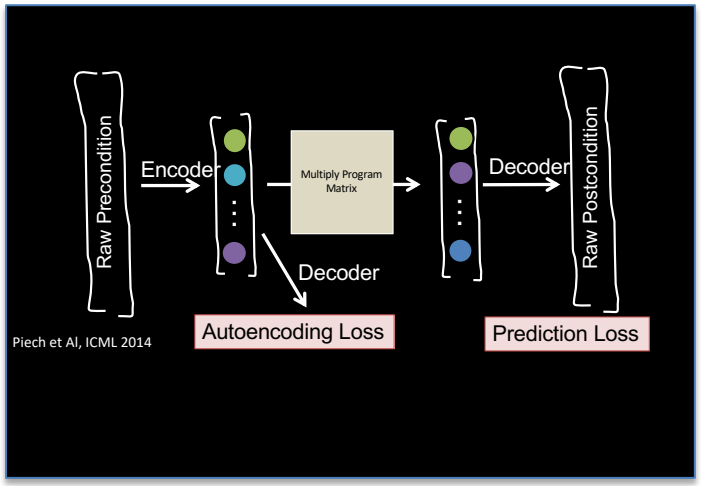


Inaccurate, Uninterpretable, and Data Hungry

Label student code

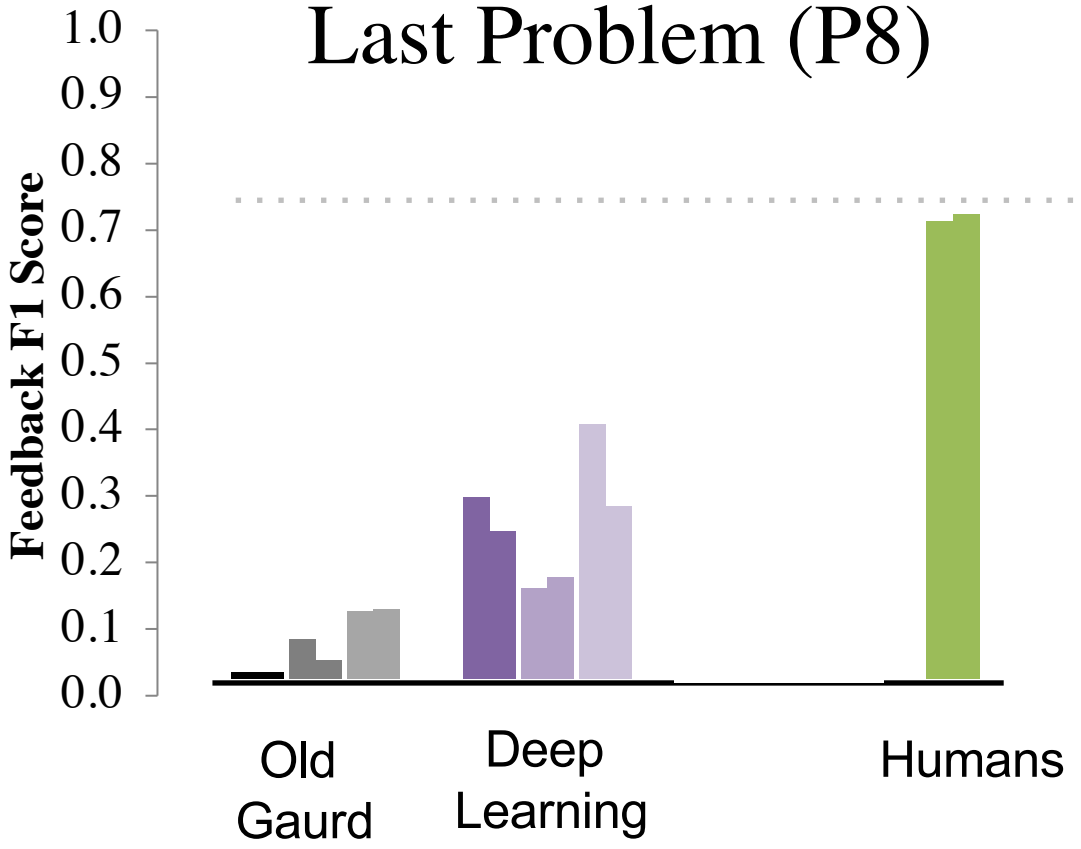






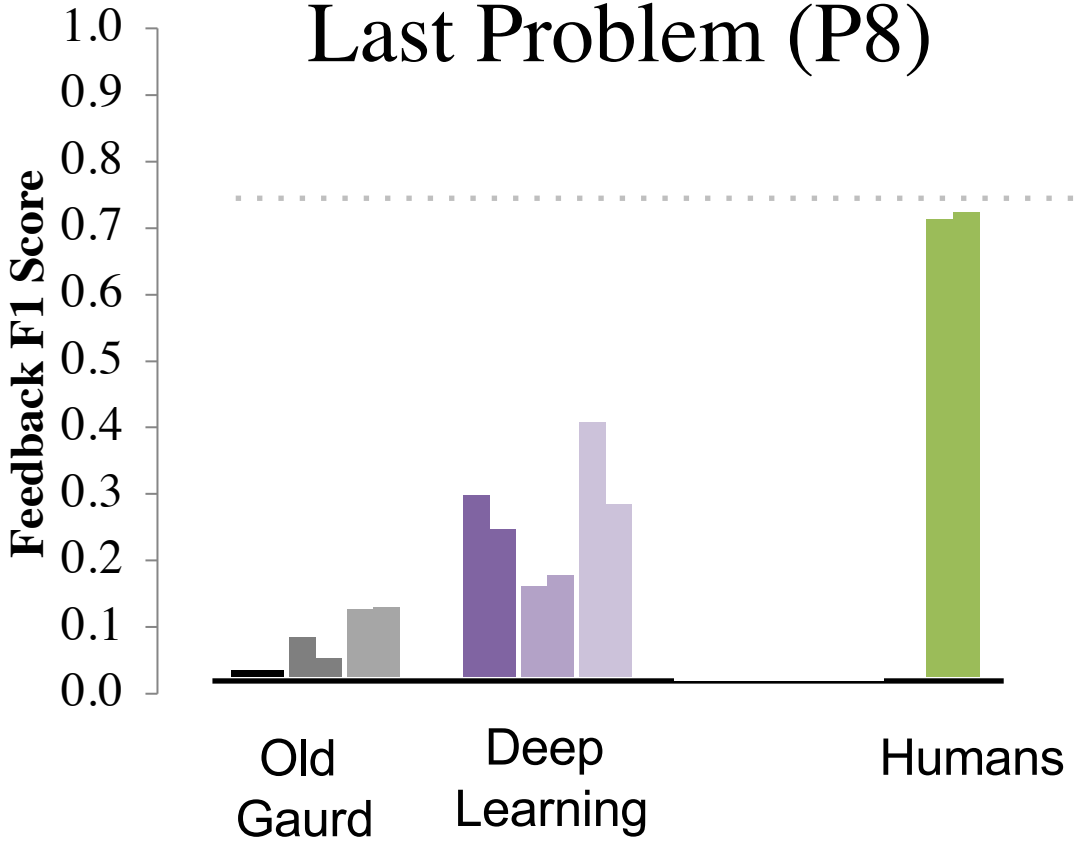
Inaccurate, Uninterpretable, and Data Hungry

Label student code



Inaccurate, Uninterpretable, and Data Hungry

Label student code



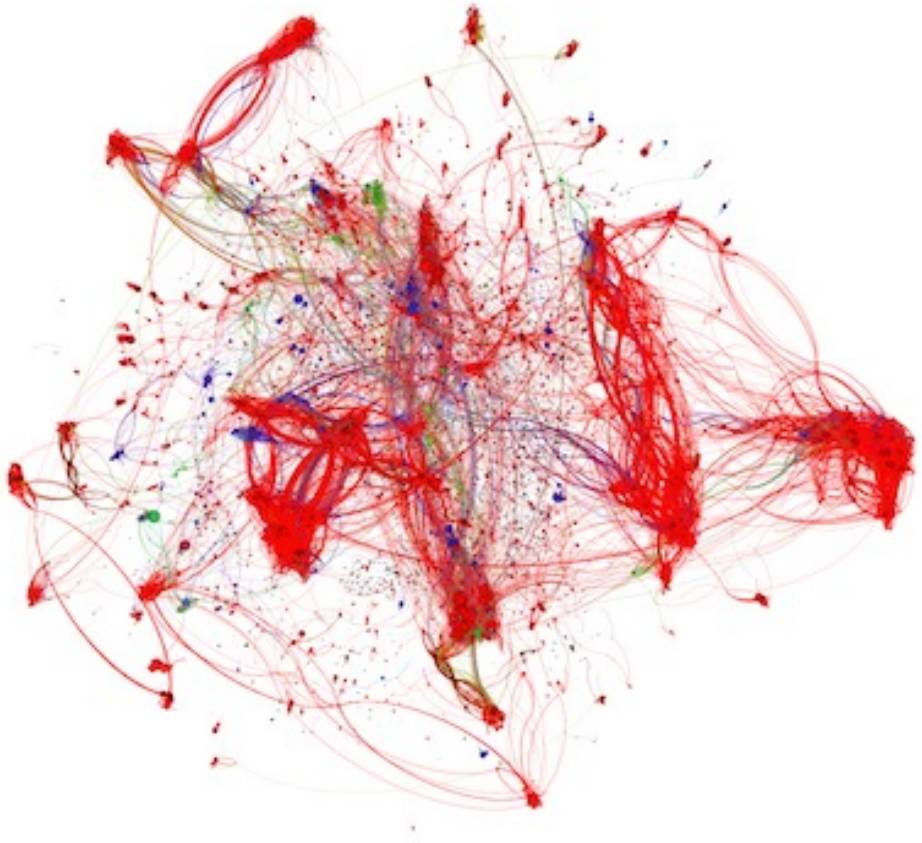
We need one shot learning

We need verifiability

Why is it so hard?

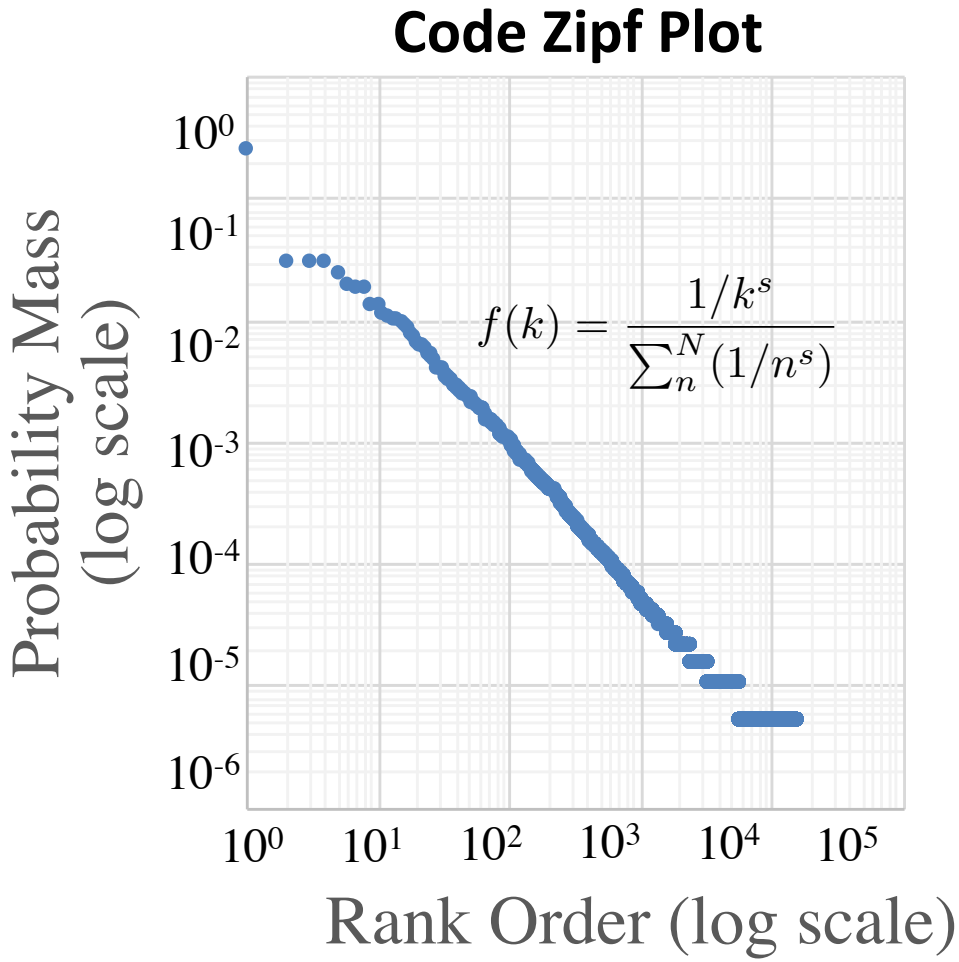
Hard Problem

Brute force solution?



1 million unique solutions to programming Linear Regression

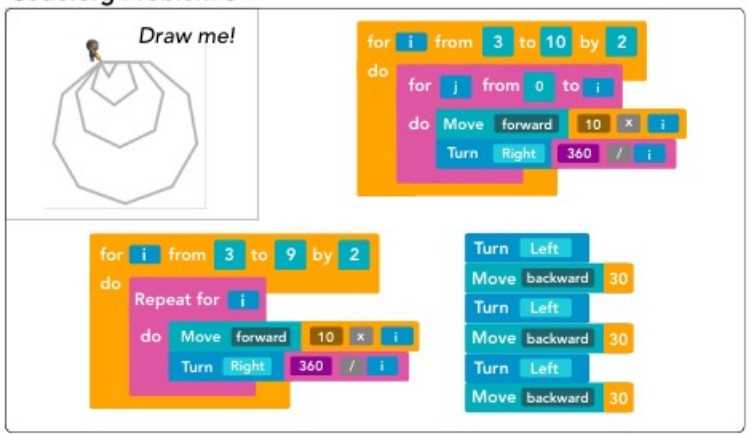
WWW 2014



They are all Zipf!

(a) Datasets in Computational Education

Code.org Problem 8



Powergrading P13

What is one reason the original colonists came to America?

- Religuous freedom
- For religious freedom
- Freedom

- declared our independence from england
- religeous 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.

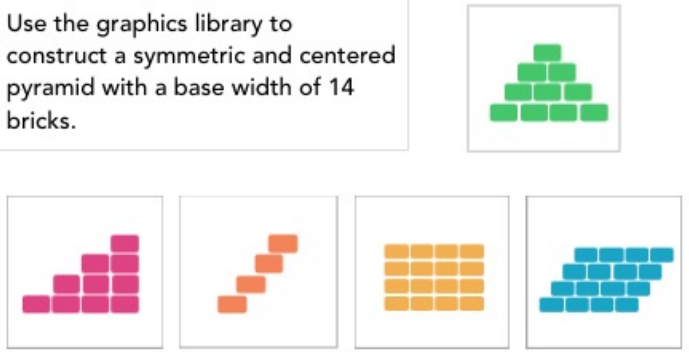
```
public void run() {
  for (int i=START; i>0; i -=1)
  {
    println(i);
  }
  println("Liftoff");
}
```

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

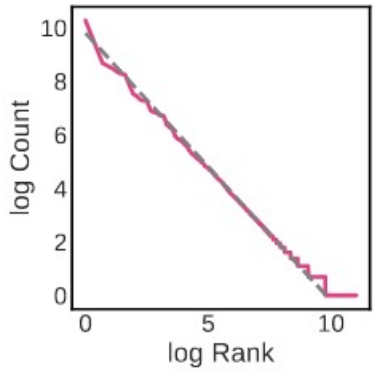
```
public void run() {
  int x = START;
  int y = 1;
  int z = 9;
  while (x>=1) {
    println(x);
    x=z;
    z=x-y;
  }
  println("Liftoff");
}
```

PyramidSnapshot

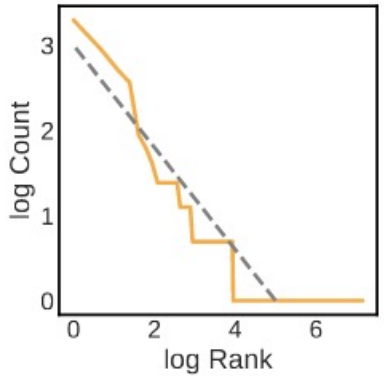
Use the graphics library to construct a symmetric and centered pyramid with a base width of 14 bricks.



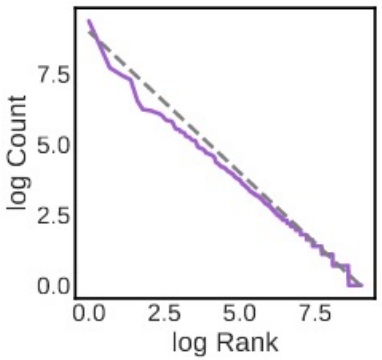
(b) Code.org P8



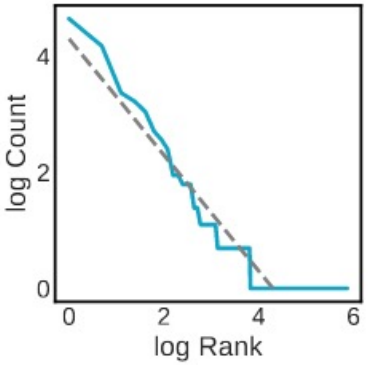
(c) CS1: Liftoff



(d) Pyramid



(e) Powergrading



[Suspense]

Chapter 3: Back to the drawing board

Humans Don't Need Much Data

Single training example:

Test set:

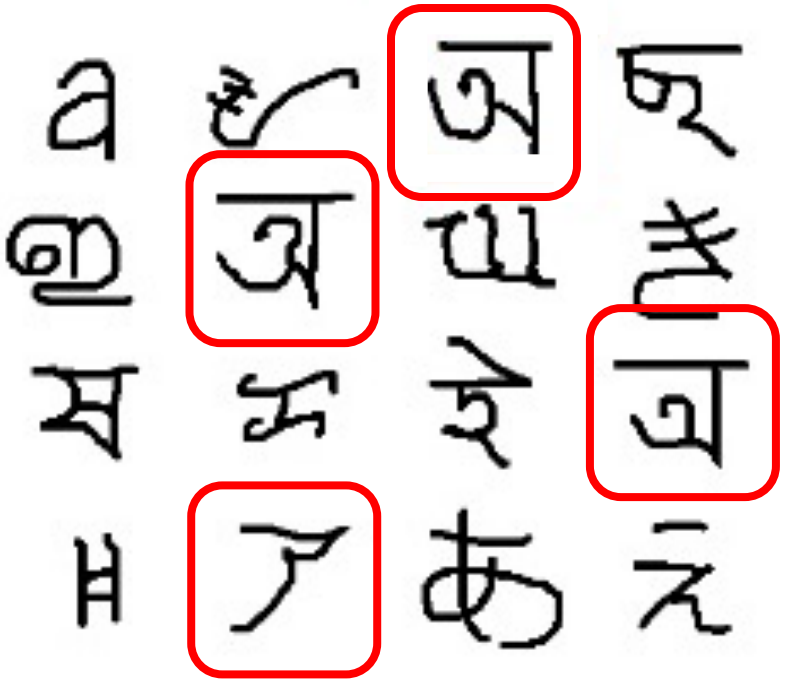
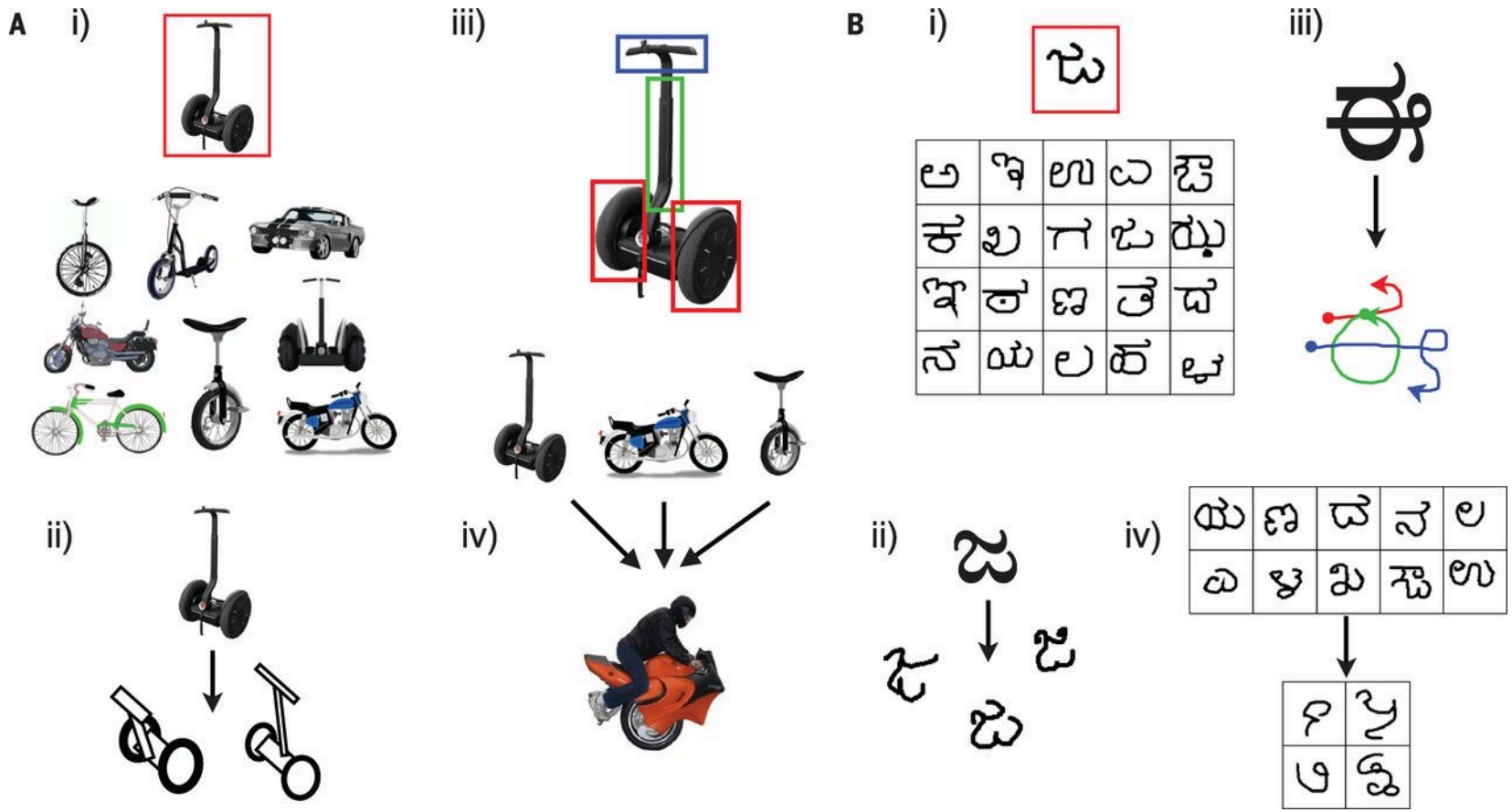


Fig. 1 People can learn rich concepts from limited data.



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

Science
AAAS

Bayesian Program Learning

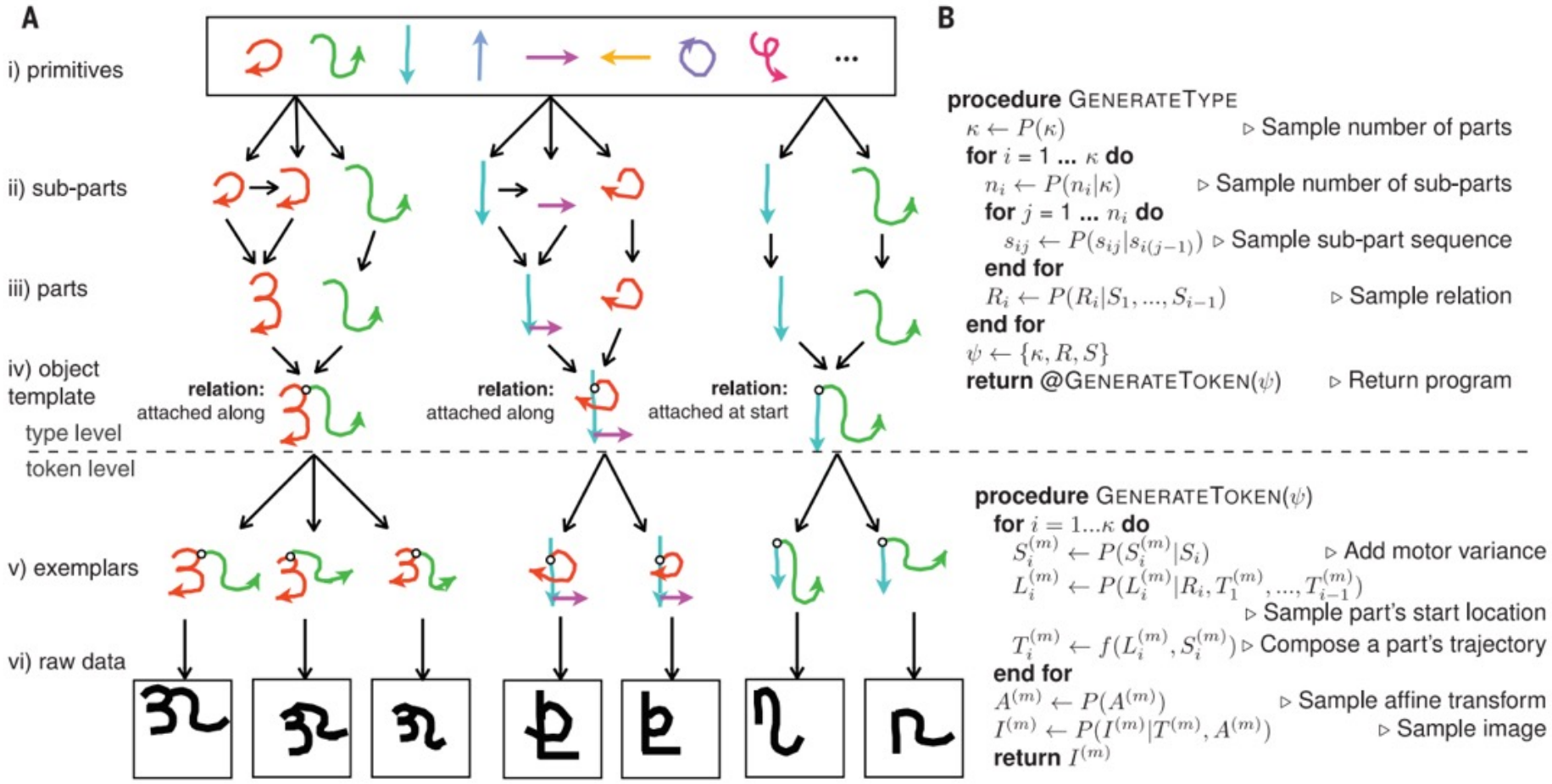
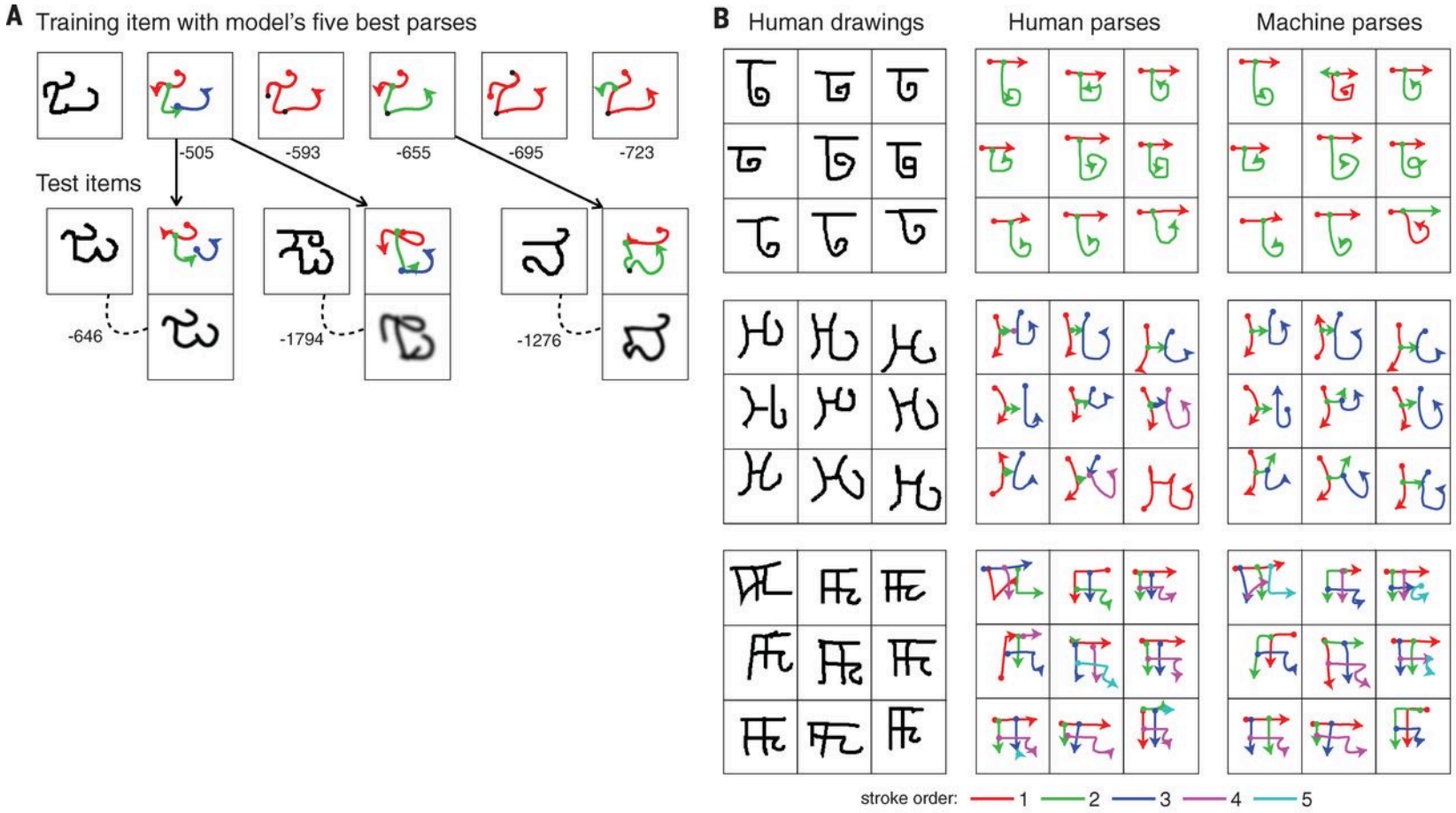
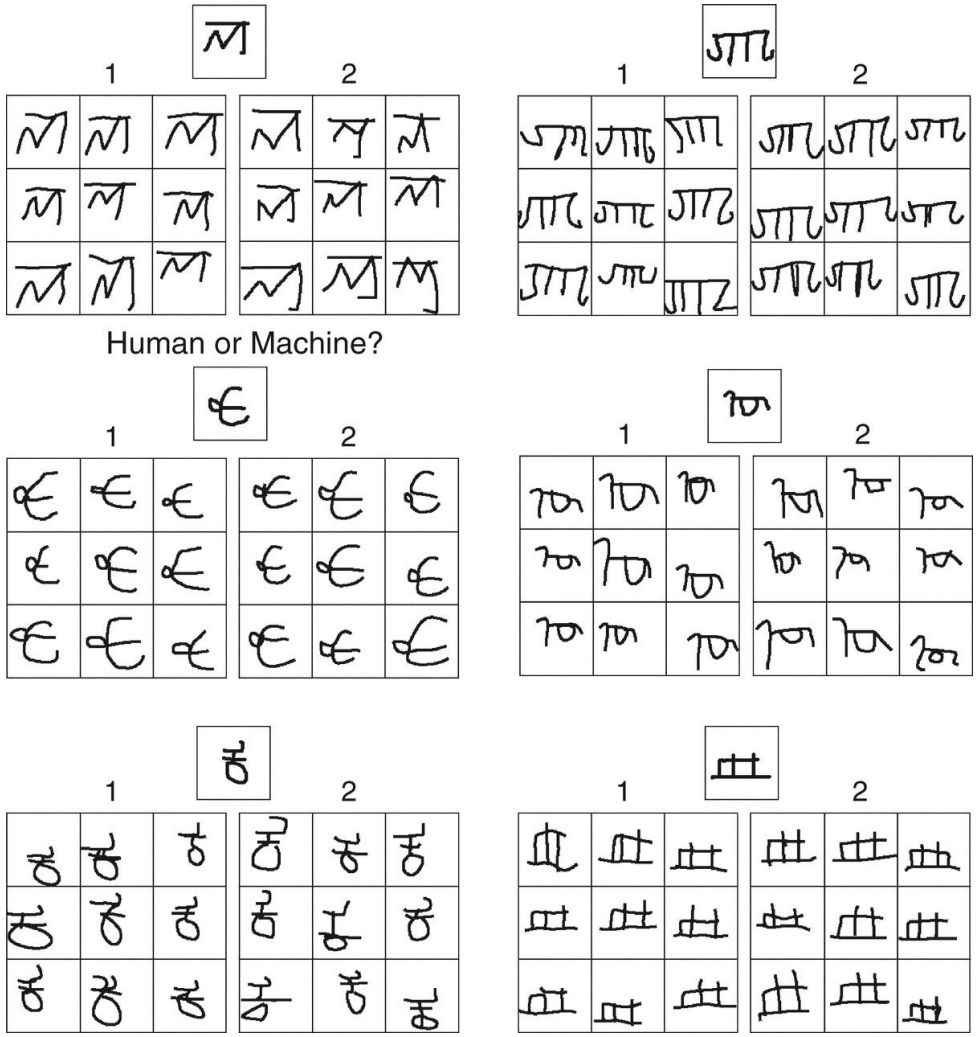


Fig. 4 Inferring motor programs from images.



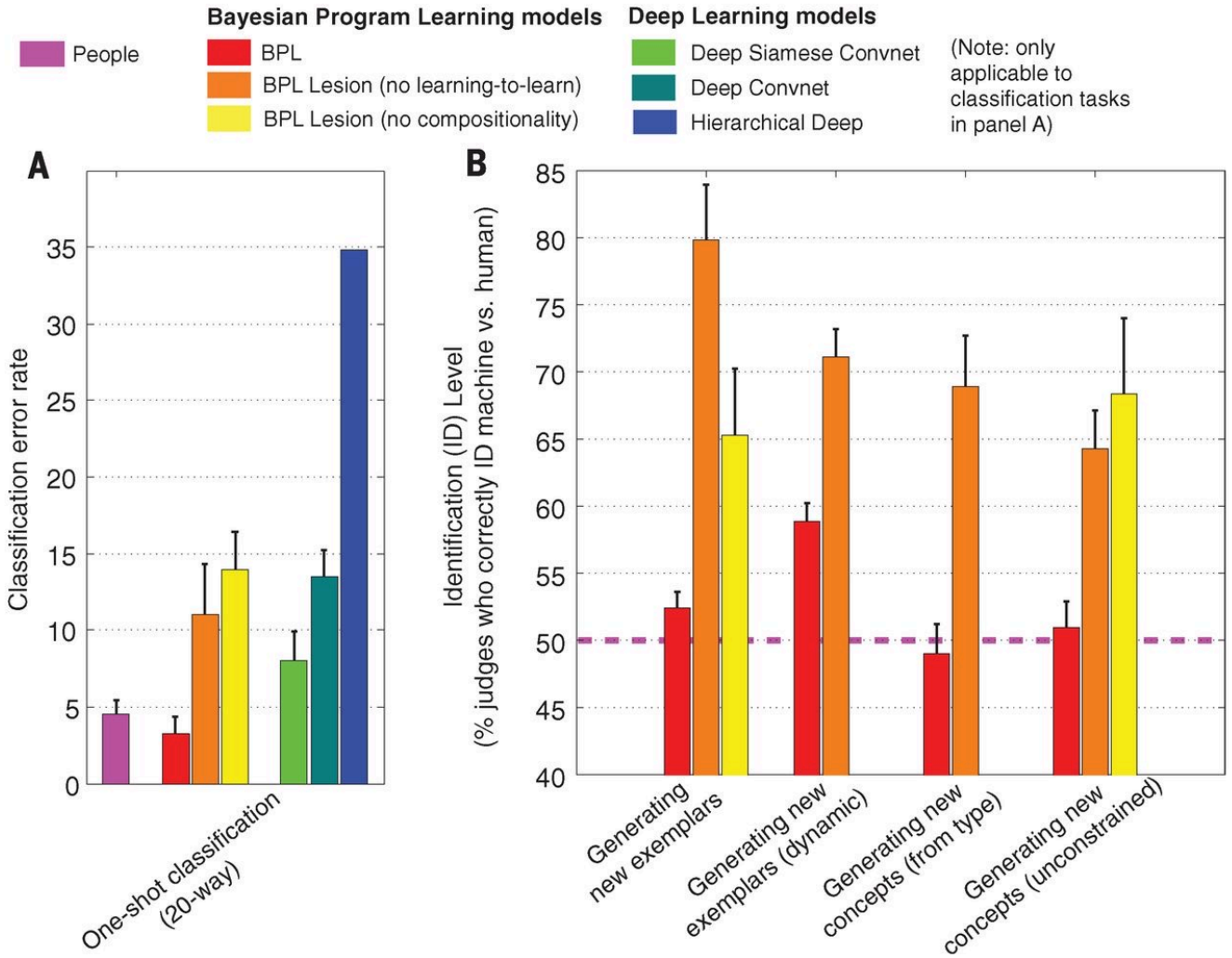
Brenden M. Lake et al. Science 2015;350:1332-1338

Fig. 5 Generating new exemplars.



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

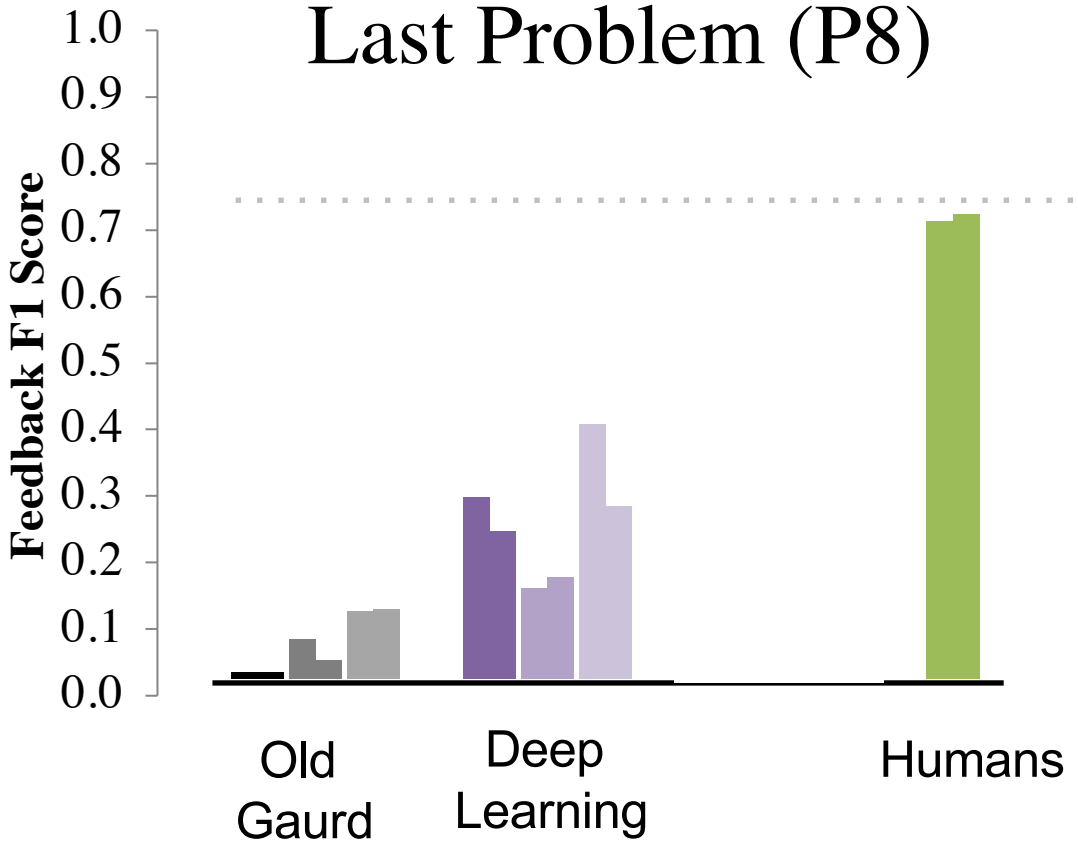
Fig. 6 Human and machine performance was compared on (A) one-shot classification and (B) four generative tasks.



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

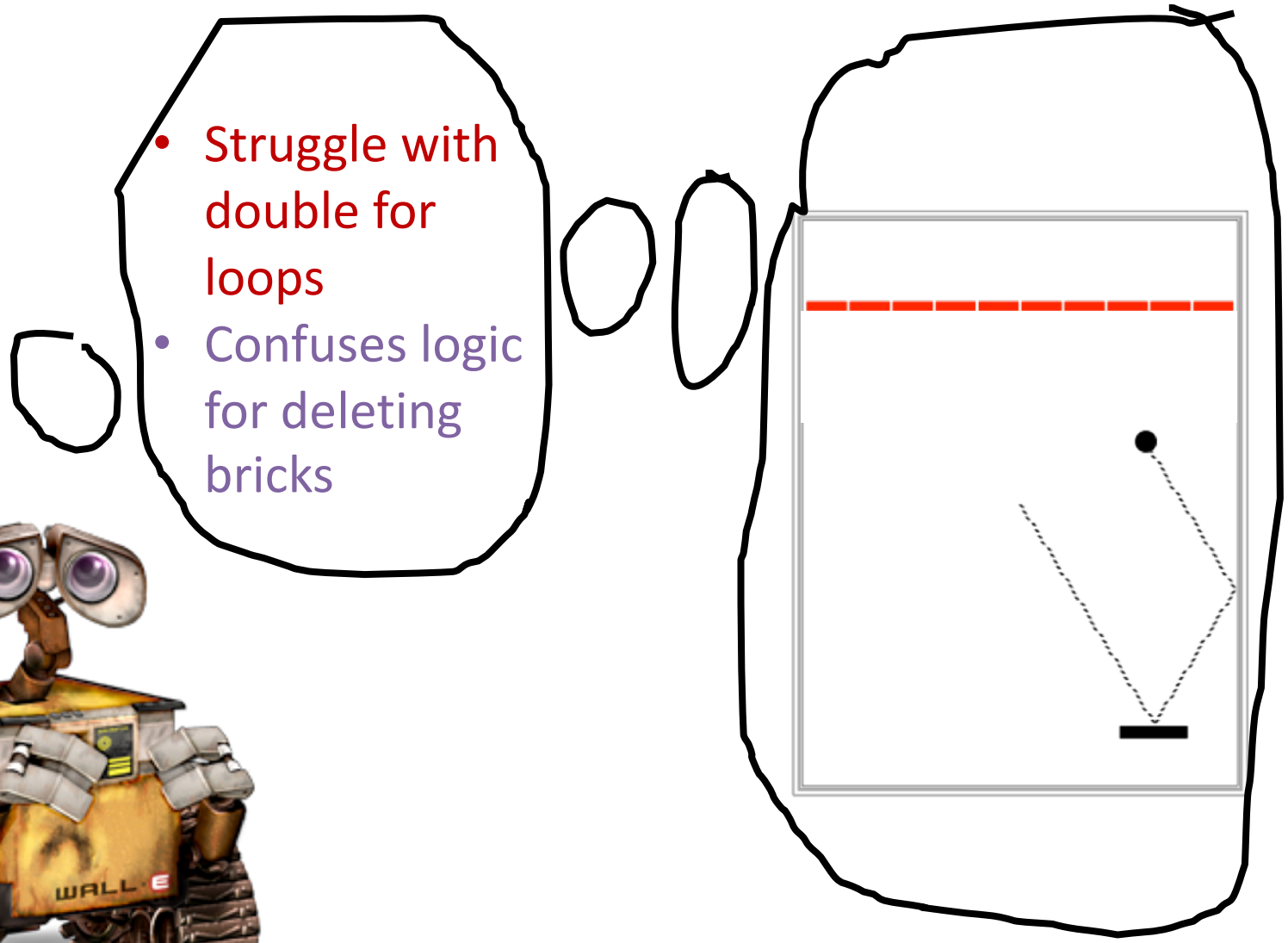
$$P(\Theta, C|\Pi)$$

Student choices

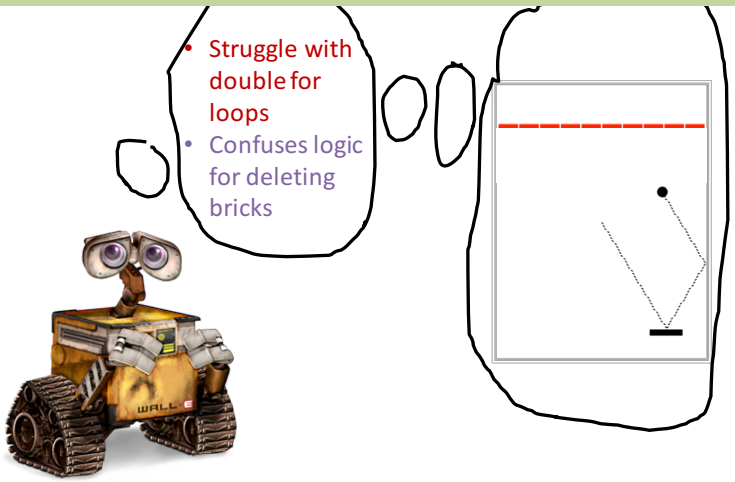
Student code

Generative Probabilistic Model?

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



Imagine Students



This is easy and exponential human impact

This is hard and linear human impact

A students "ability"

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

A students "choices"

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

The resulting code

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

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

Infer ability and choices from code

Bayesian Programming Language

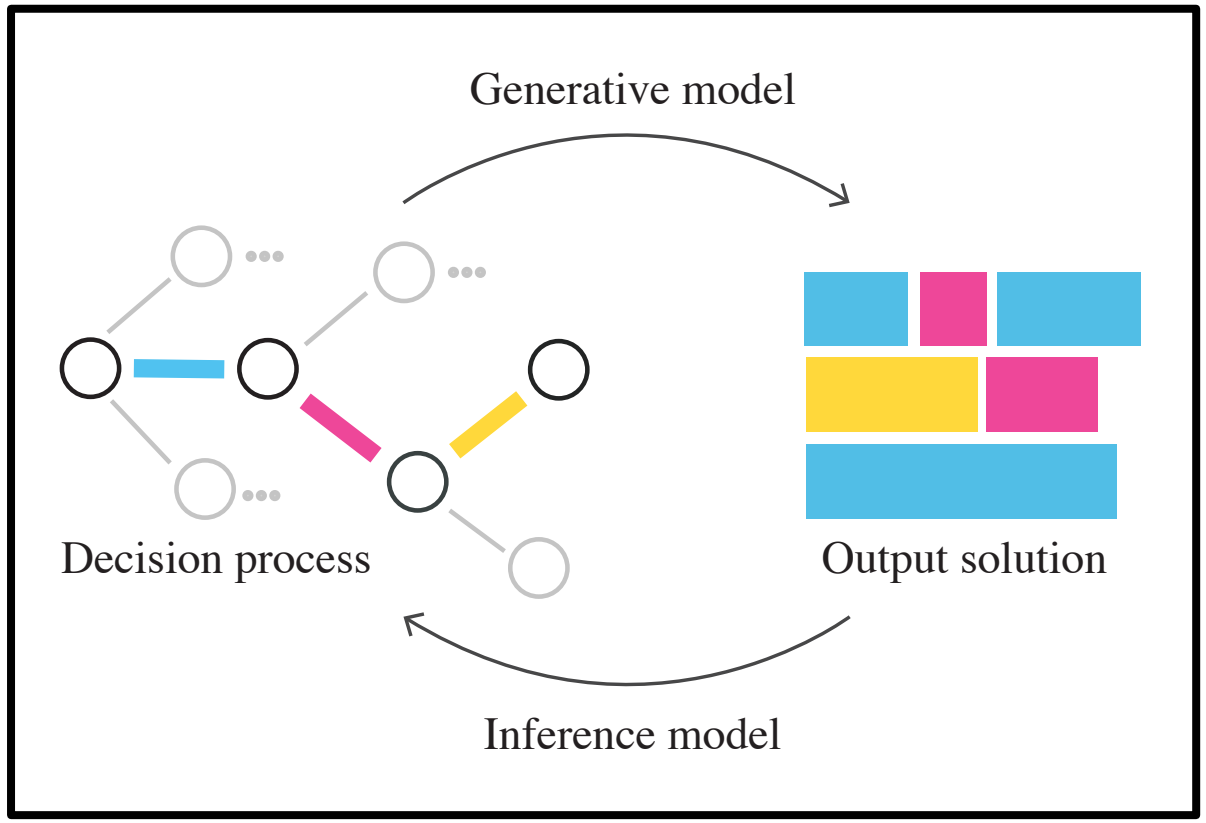
`ideaToText`

Teachers Articulate N misconceptions

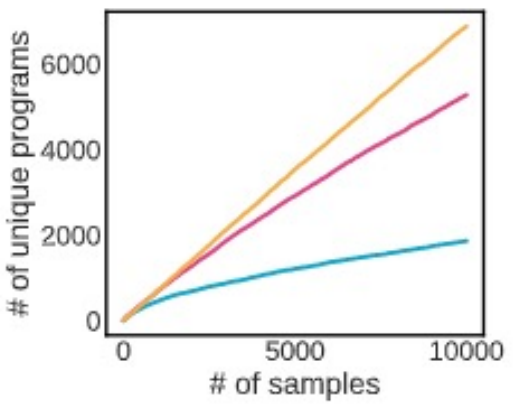
```
7
8  # This python class is a RubricSampling Decision
9  # it generates programs that print the numbers 10 -> 1
10 class Countdown(Decision):
11
12   def registerChoices(self):
13     # these are the main strategies for printing out a
14     # countdown
15     self.addRubricChoices('loop-style', {
16       'for' : θ1 ,
17       'while' : θ2,
18       'none' : θ3
19       'empty' : θ4
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

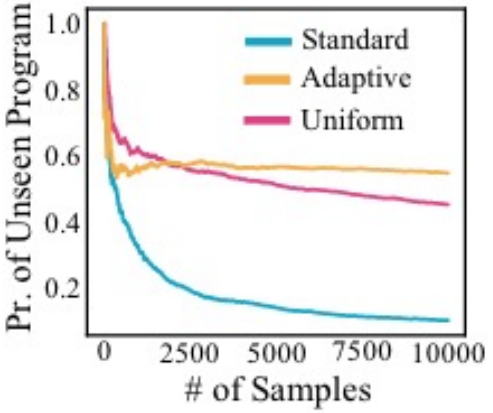
Generative Understanding



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

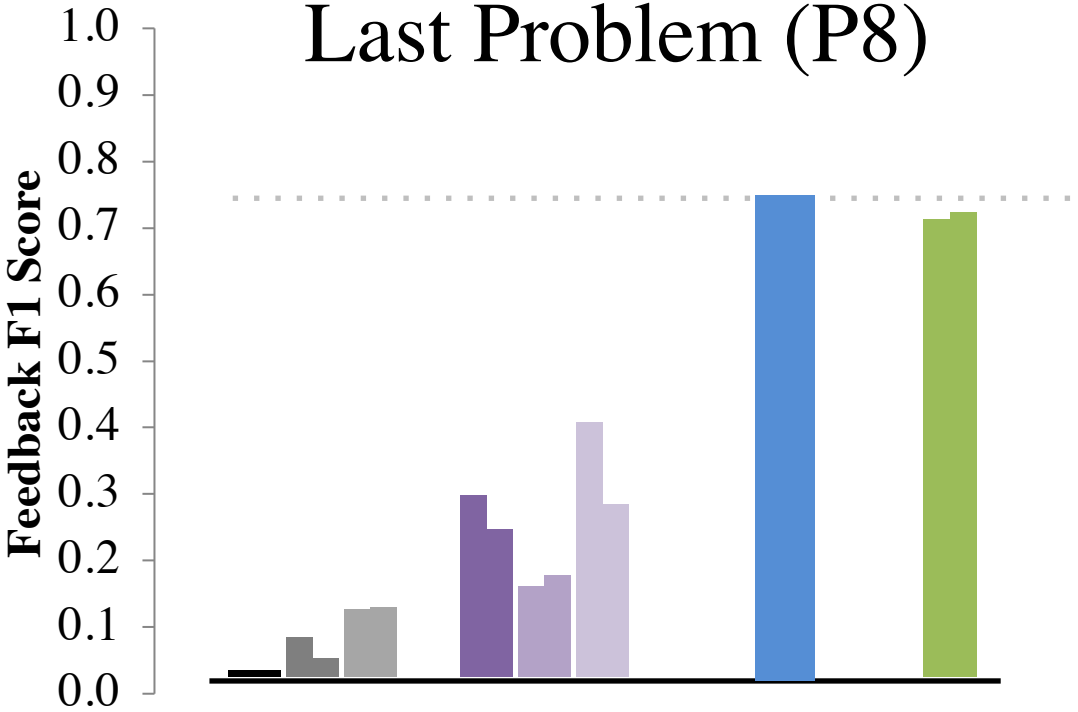
$$p_{\mathcal{G}}(x_{a_1}, \ldots x_{a_T} | y) = \prod_{t=1}^{T} p_{\mathcal{G}}(x_{a_t} | y, \mathbf{x}_{<a_t})$$

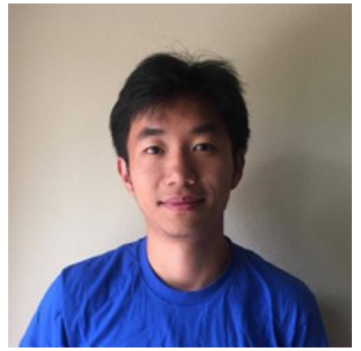
solution

Previous choices

Generative Understanding

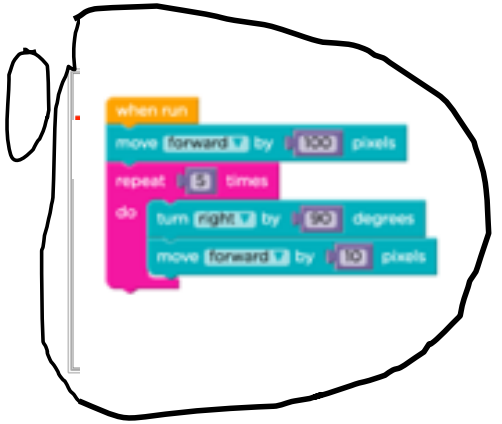
Label student code





Outstanding Student paper award, AAAI 2019

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



Not just for code

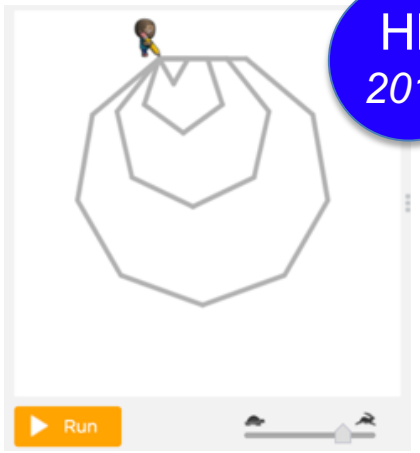
Results from early 2019

Many domains of student work

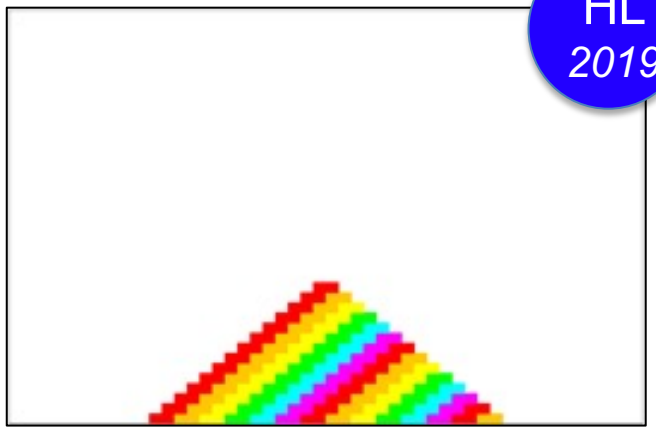
HL 2017

$$9 \times 6$$
$$= (10 - \square) \times 6$$
$$= 10 \times 6 - \square \times 6$$
$$= 60 - \square$$
$$= \square$$

HL 2019



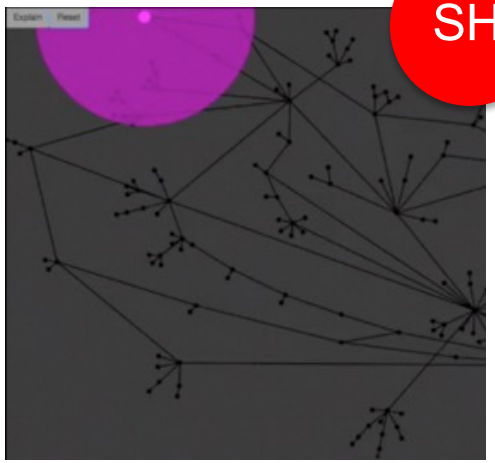
HL 2019



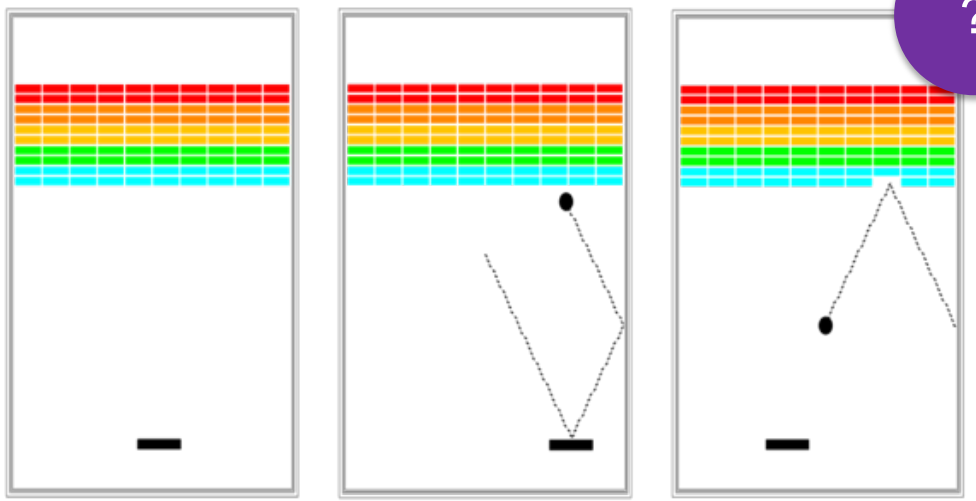
NH 2019

Why did the original colonists come to America?

SH



?



NH Near Human

HL Human Level

SH Super Human Level

So what?

What does this mean for me?!?

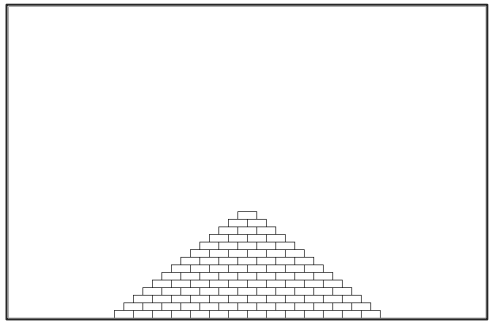
Understanding Process

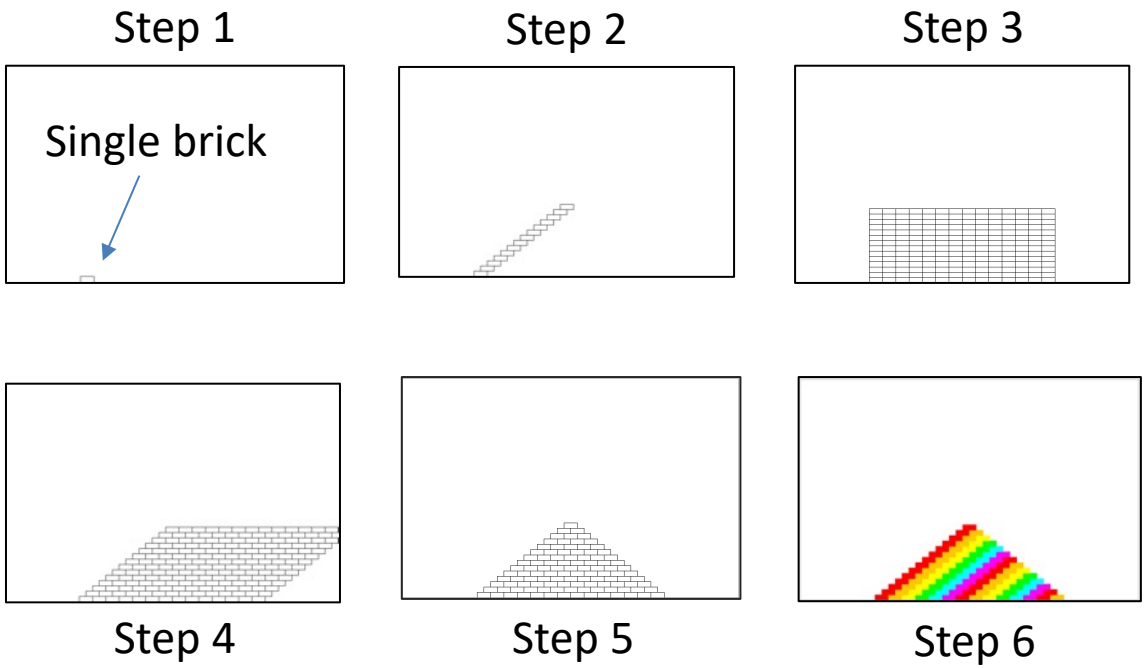
2,600 students

130,000 partial solutions

μ snapshots per student = 50

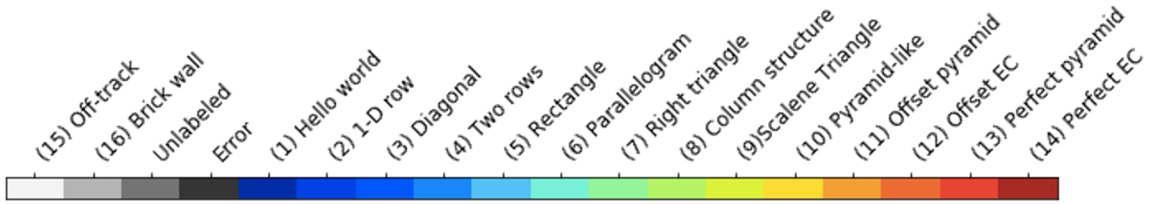
μ time per student = 2 hours



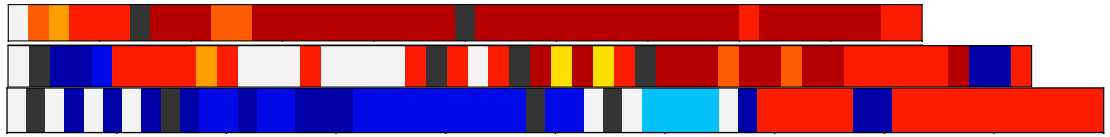


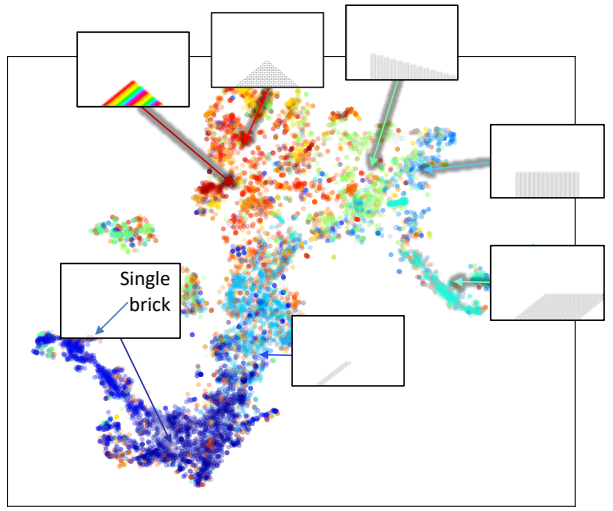
Lisa Yan, Nick McKeown, Chris Piech. Understanding students from visual output. SIGCSE 2019

Understanding Process



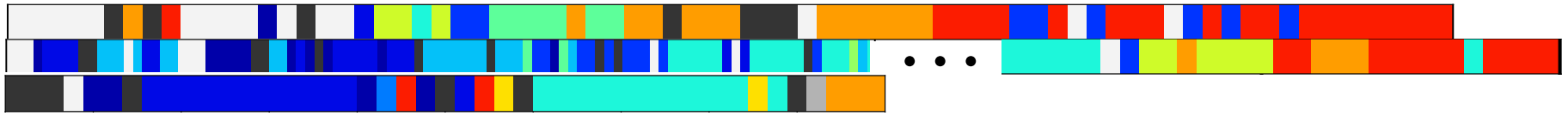
Students scoring in 99th percentile on midterm exam

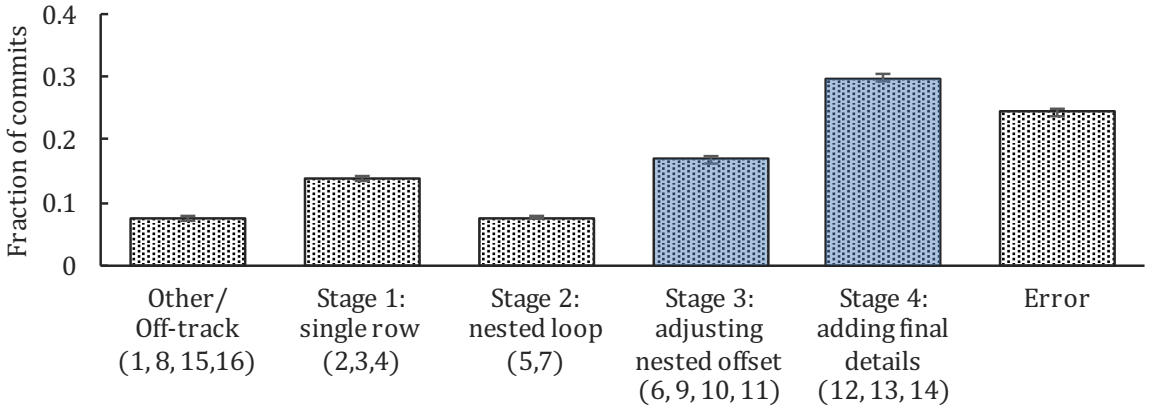




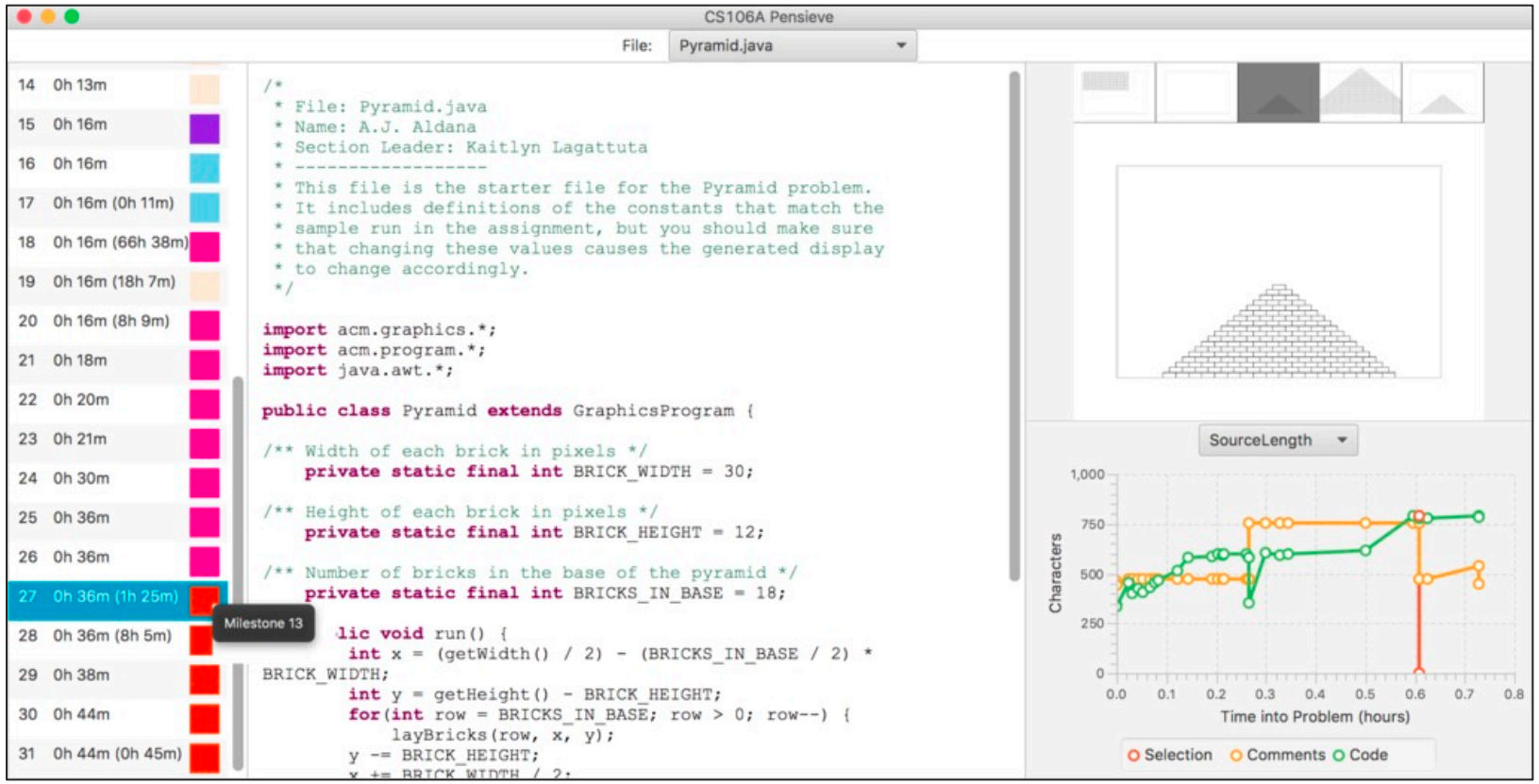
t-SNE embedding of 130,000 partial solutions

Students scoring in ≤3rd percentile on midterm exam





Understanding Process

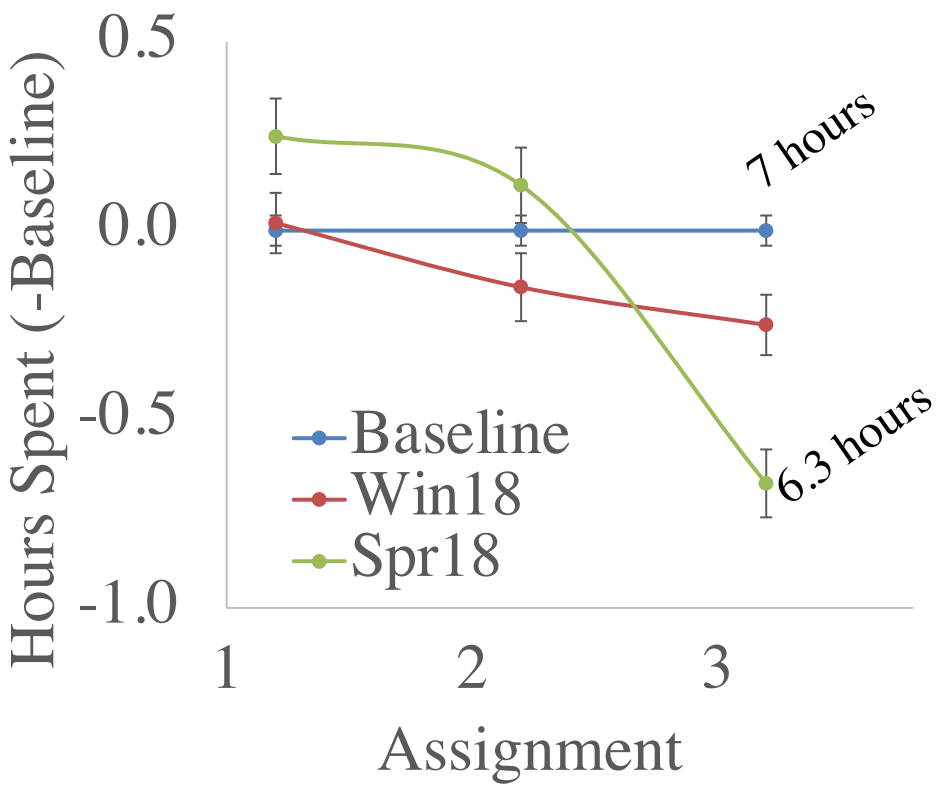


One quarter, all students were *shown* their progress:

- 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

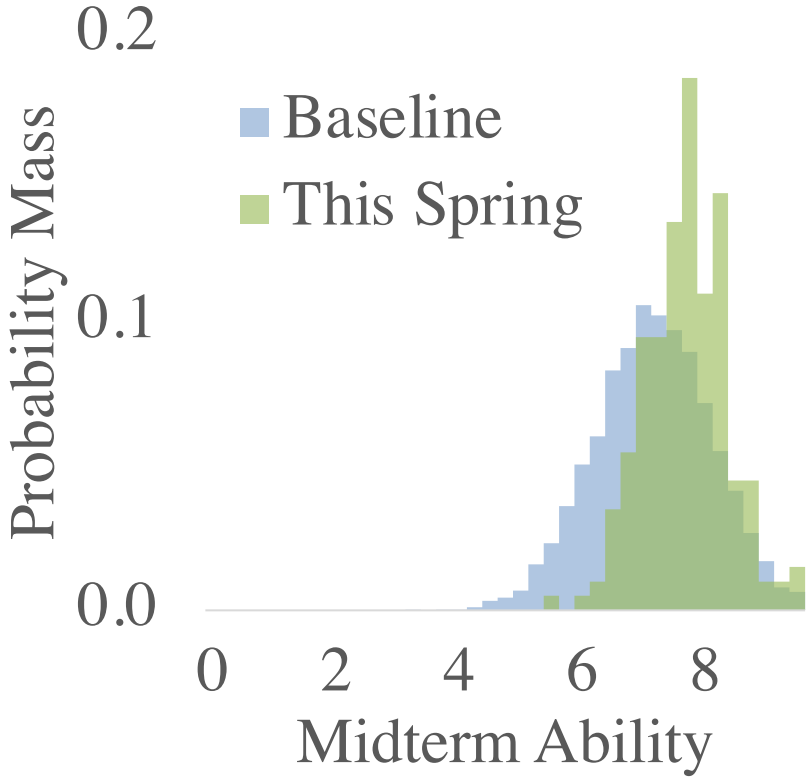


$$E[\hat{X}_3|X_1] - E[X_3] = 42 \text{ mins} \quad p < 0.00001$$

Predicted time — Actual time

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

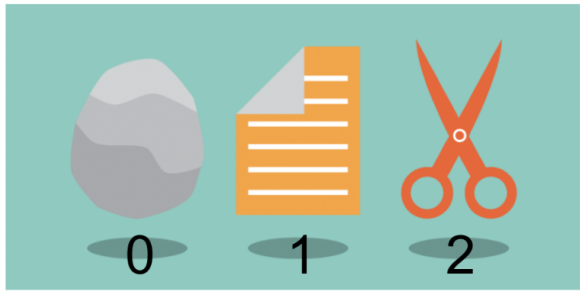
Midterm Grading Challenge

Browse Q3 | DynamoDB · AWS Console

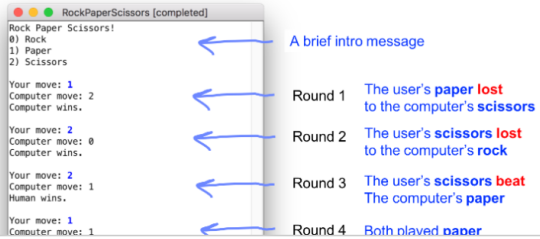
localhost:3000/#/view/cs106aMidWin19/3/abigail8/cpiech/8daa714e79a10cb

Question | Solution | Starter

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:



Student Answer

```
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

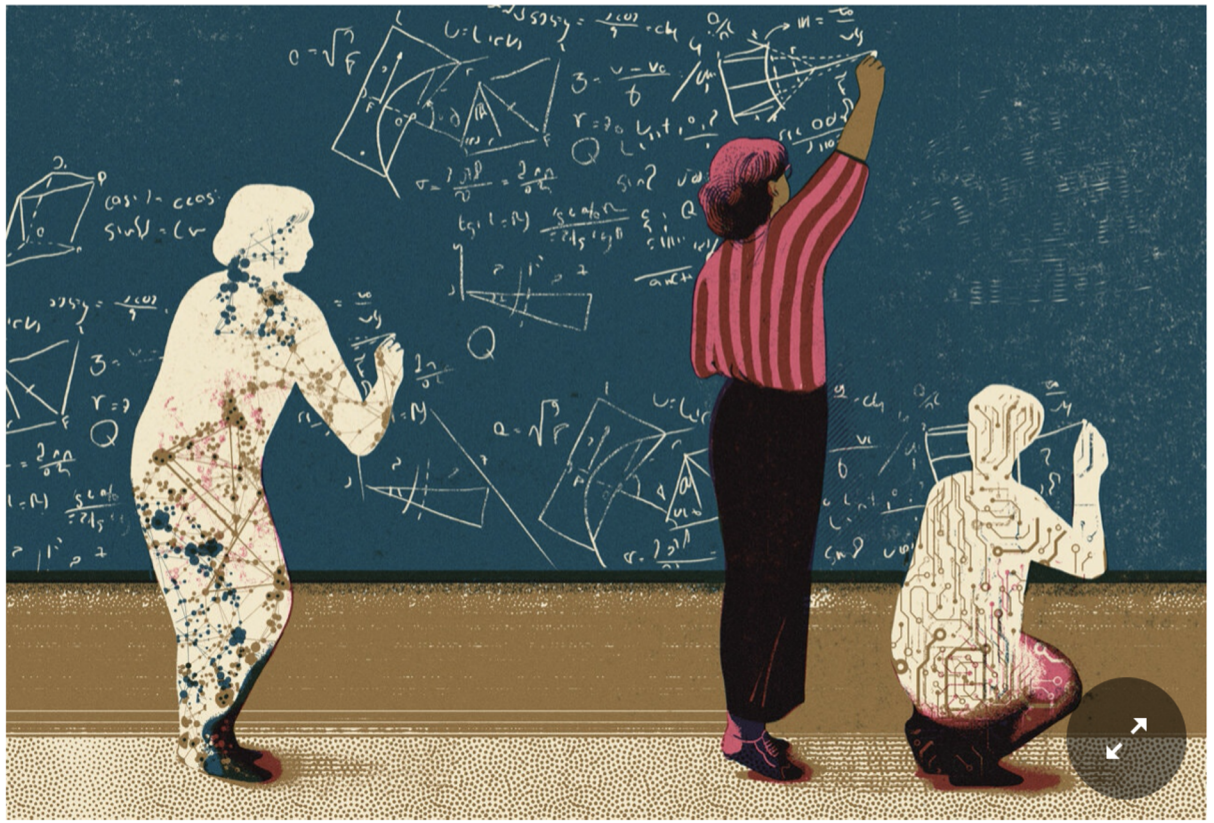
‹ Previous | Submit Grade | Next ›

The New York Times

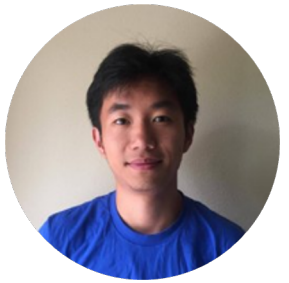
Can A.I. Grade Your Next Test?

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

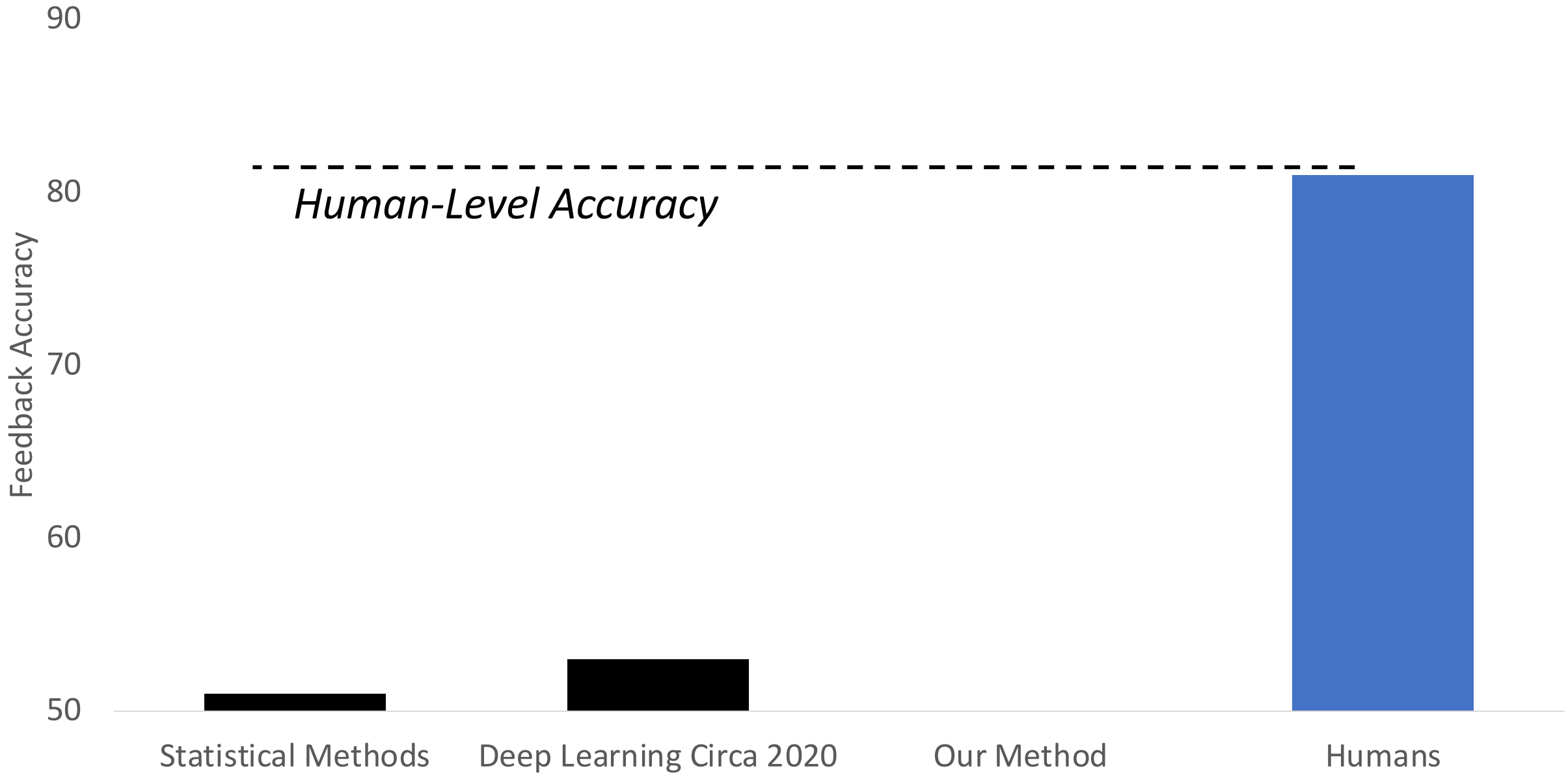
Stanford | News



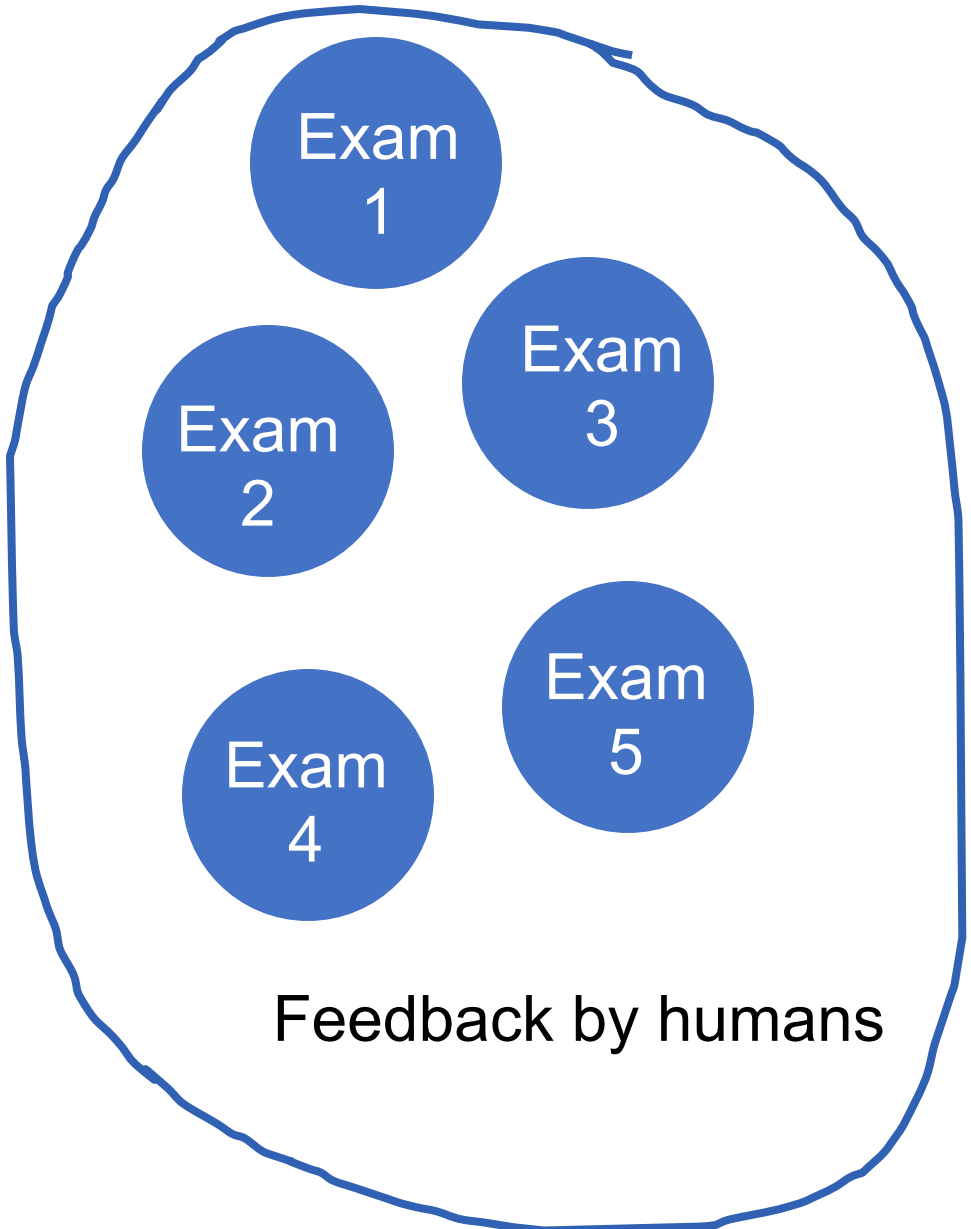
https://www.nytimes.com/2021/07/20/technology/ai-education-neural-networks.html



Rubric Level Accuracy on Few-Shot Grading a Novel Question



Give Feedback on Fresh Stanford Midterm

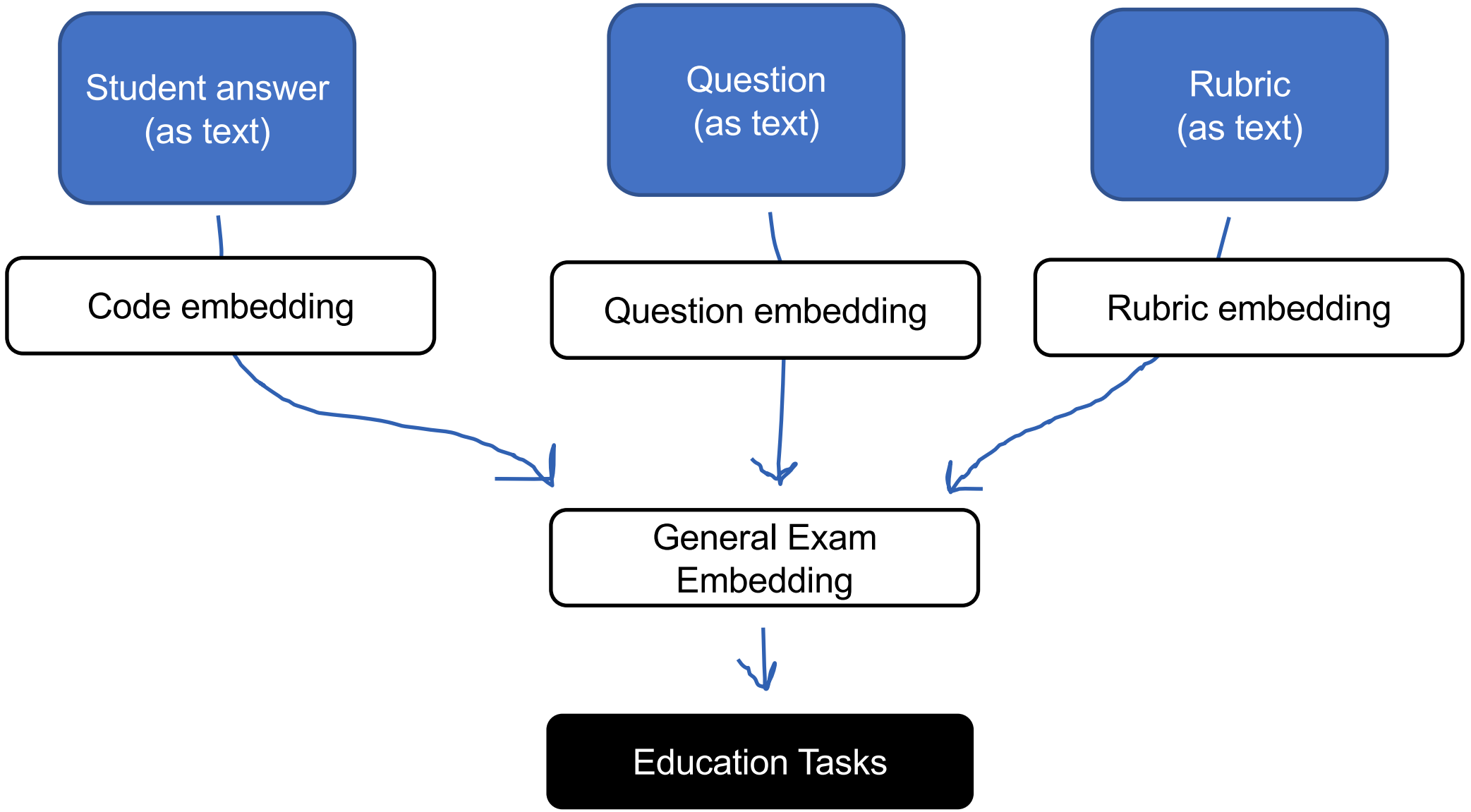


Learn how to do this

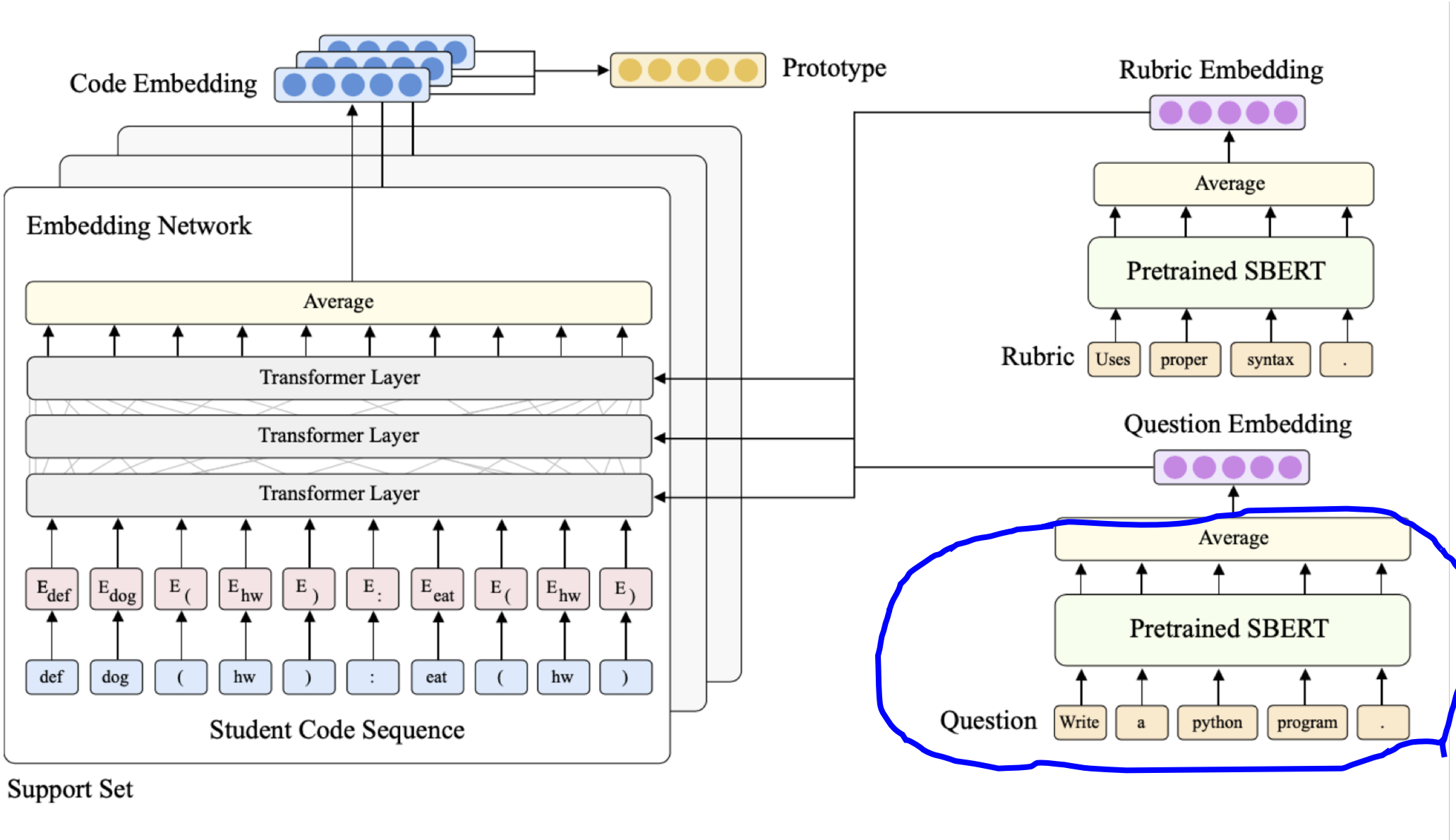
New Exam

Feedback by AI with 1 hour of human work

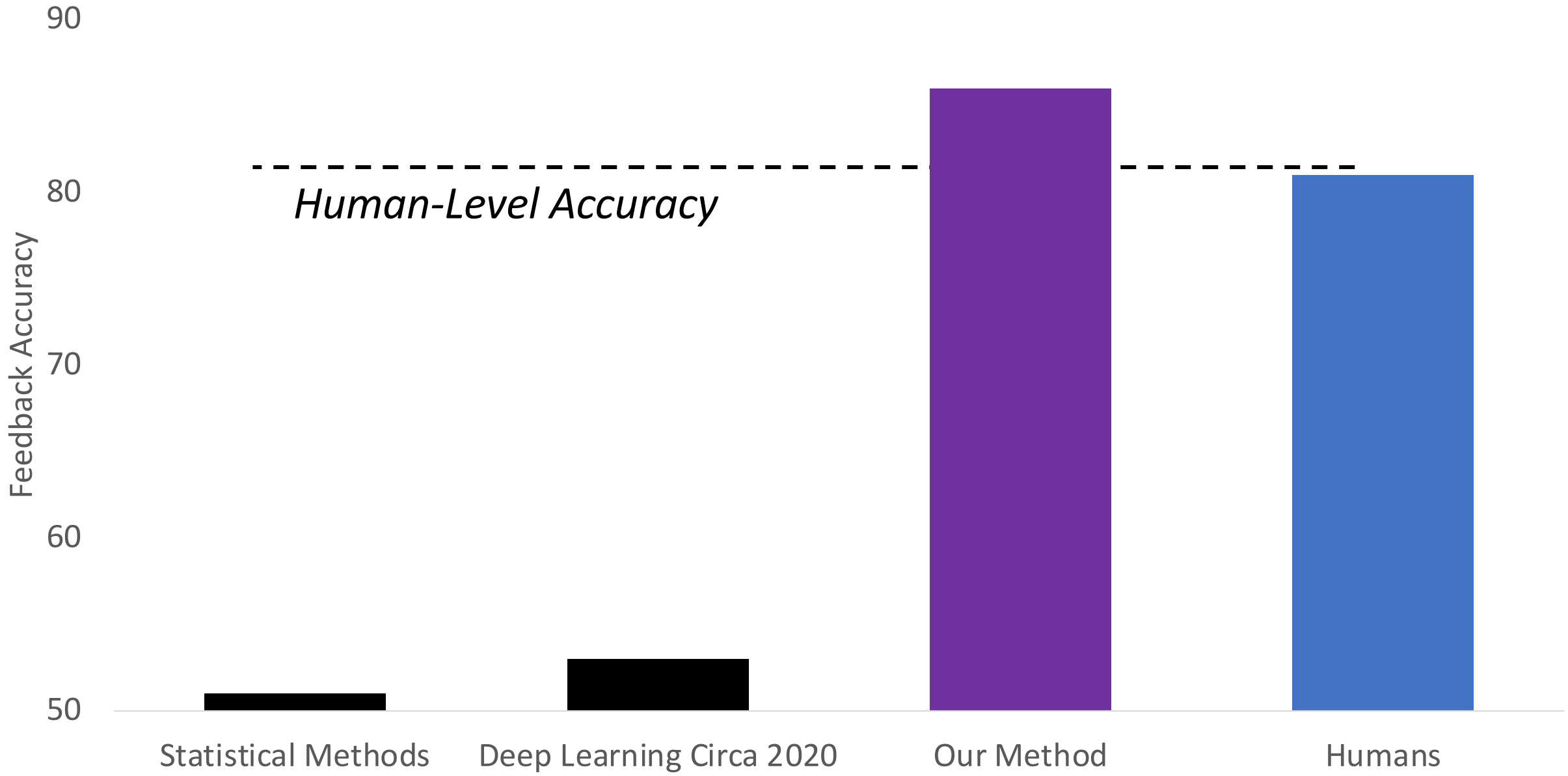
General Exam Grading Model



Invented the Proto-Transformer



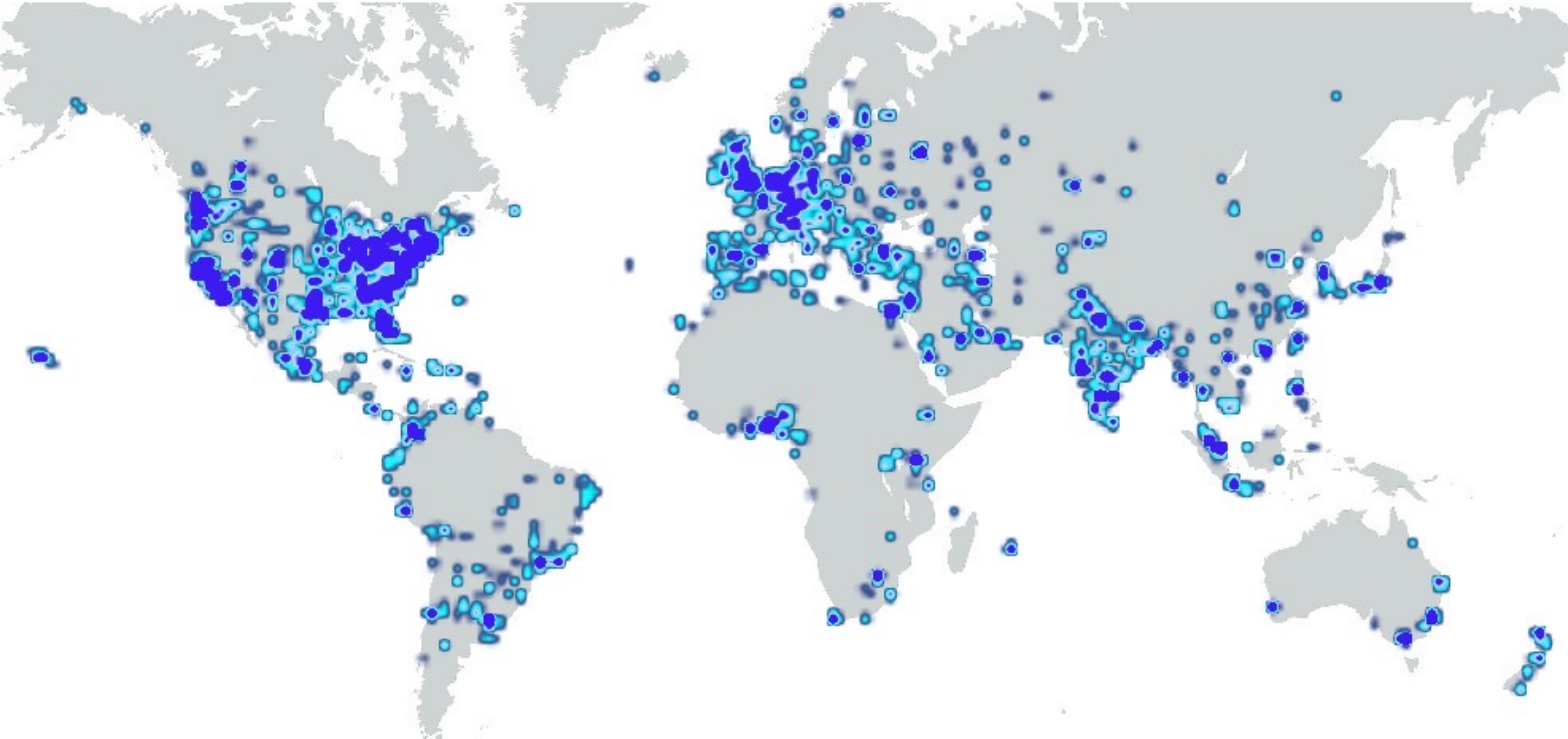
Rubric Level Accuracy on Few-Shot Grading a Novel Question



Stanford Code in Place:

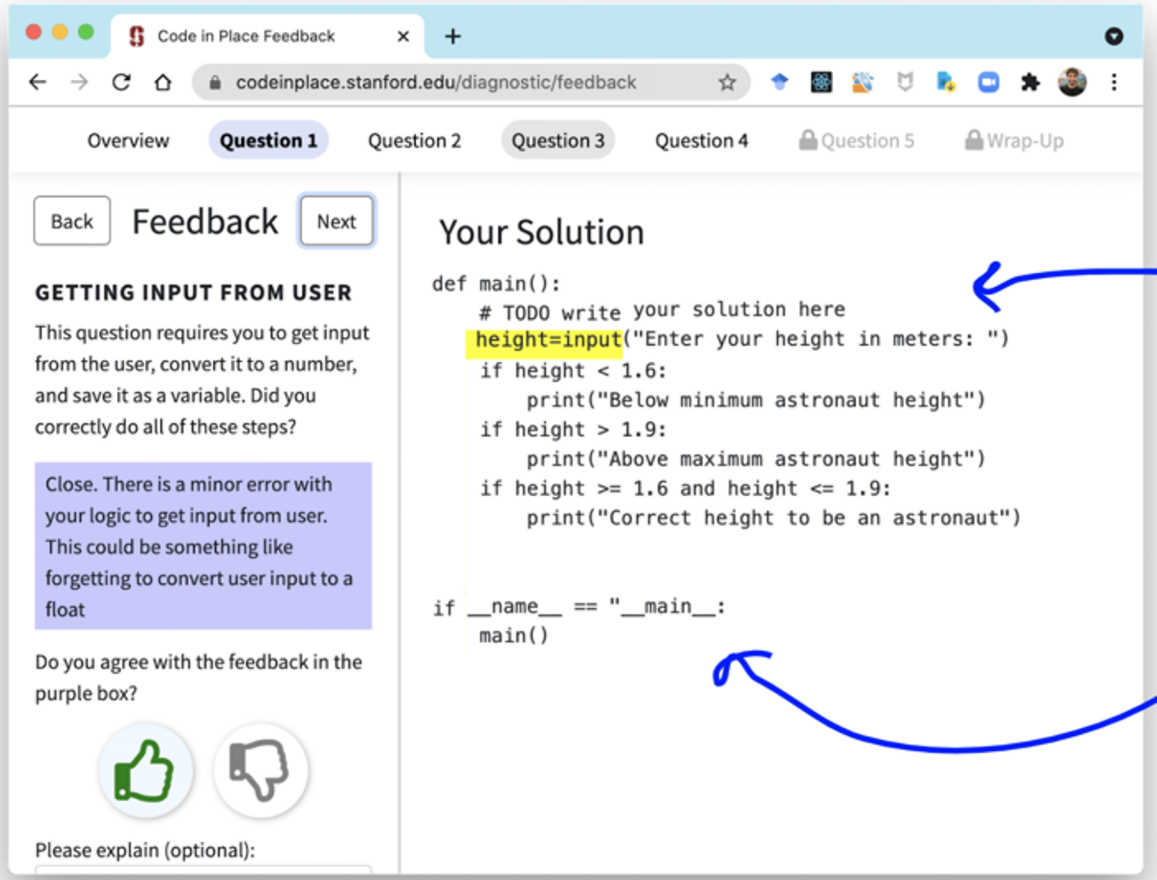
1100+ section leaders teach
12,000+ students
1/2 of CS106A
As Community Service

Featured in



Gave Feedback to 3,500 Real Students

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



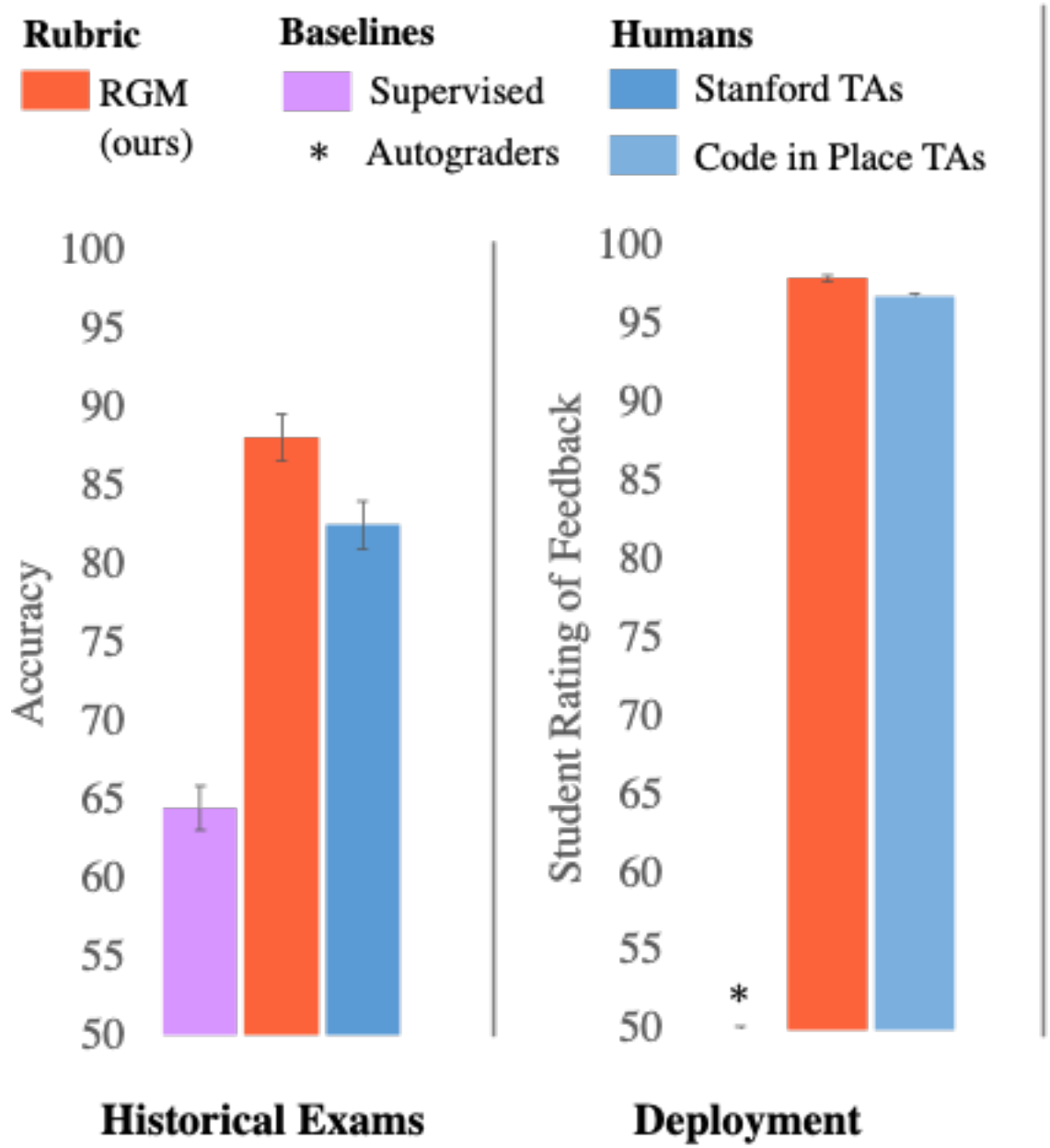
AI generated feedback

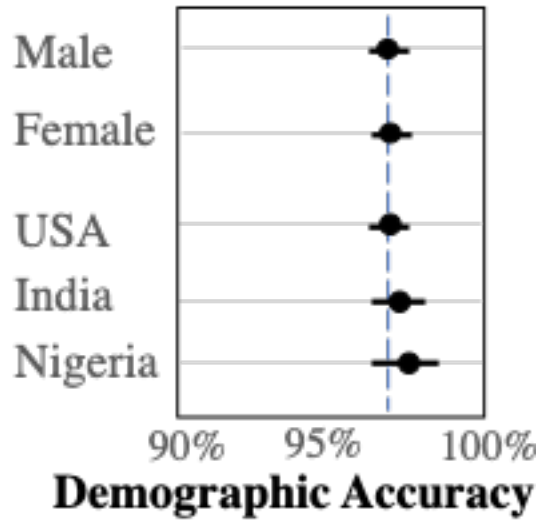
Students evaluate the feedback

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

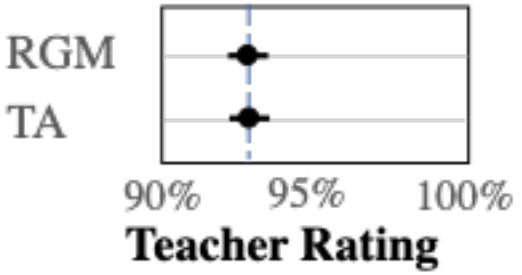
Syntax error (missing ") here would prevent auto graders from being useful.

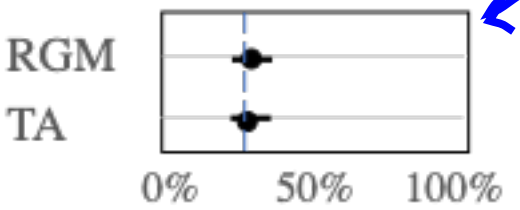
Impact and Fairness Analysis (work in progress)





calibration





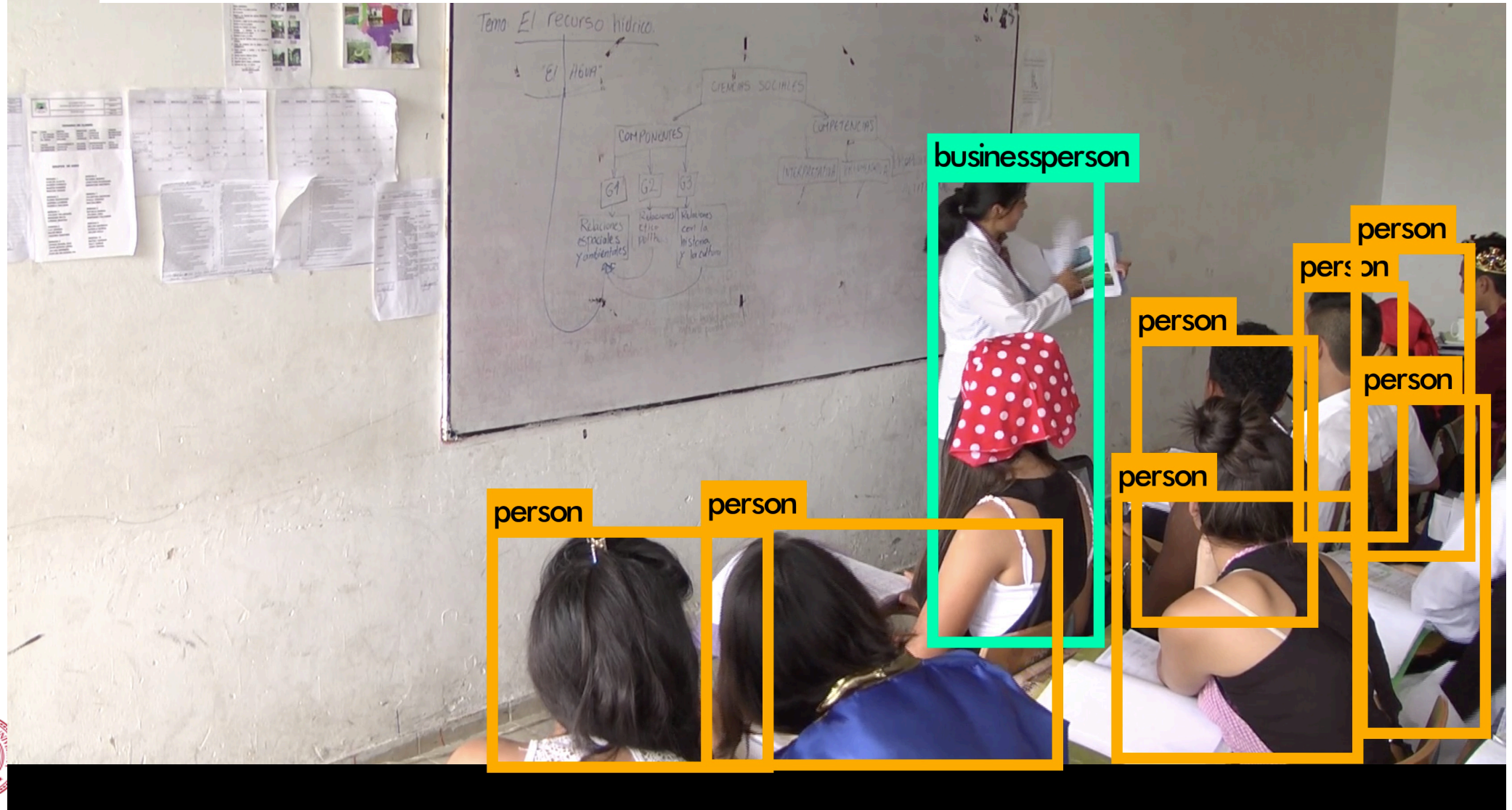
Towards parity??

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

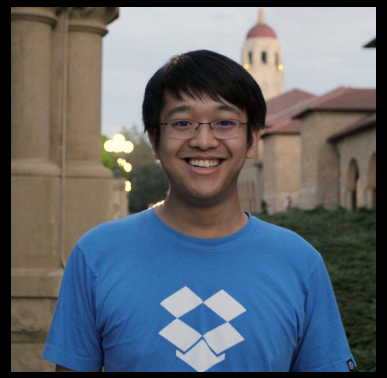
1. Teachers asked 10% **more questions**
2. Teachers "**took up**" student ideas 10% more
3. Students were significantly more likely to:

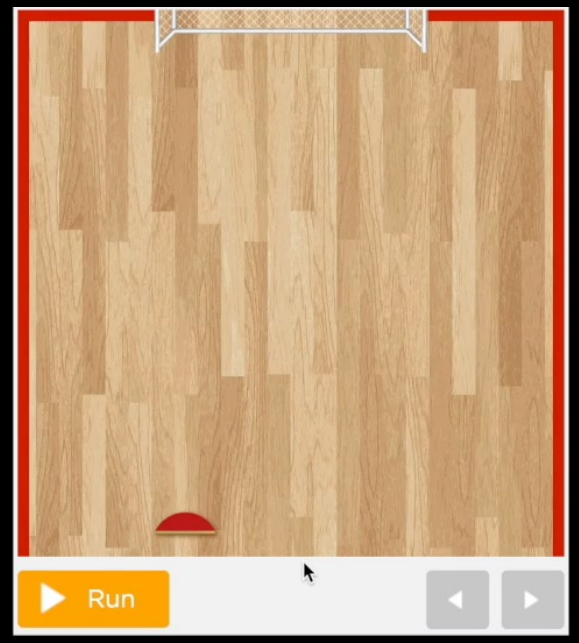
   recommend the class ($p < 0.05$),

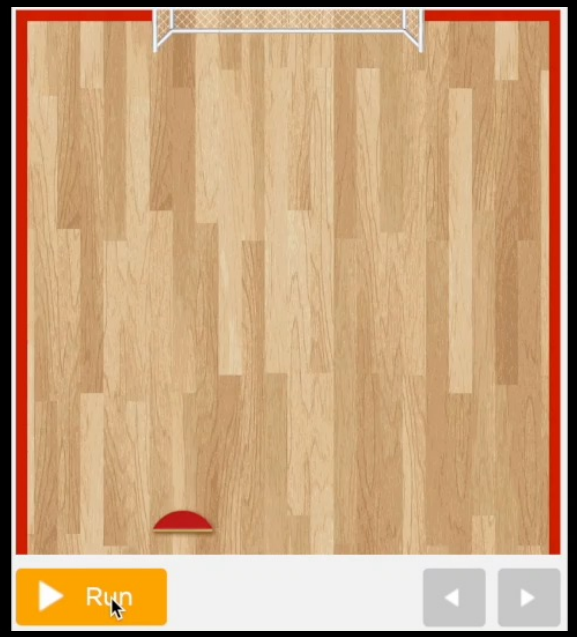
   find sections helpful ($p < 0.05$)

Play 2 Grade

Allen Nie







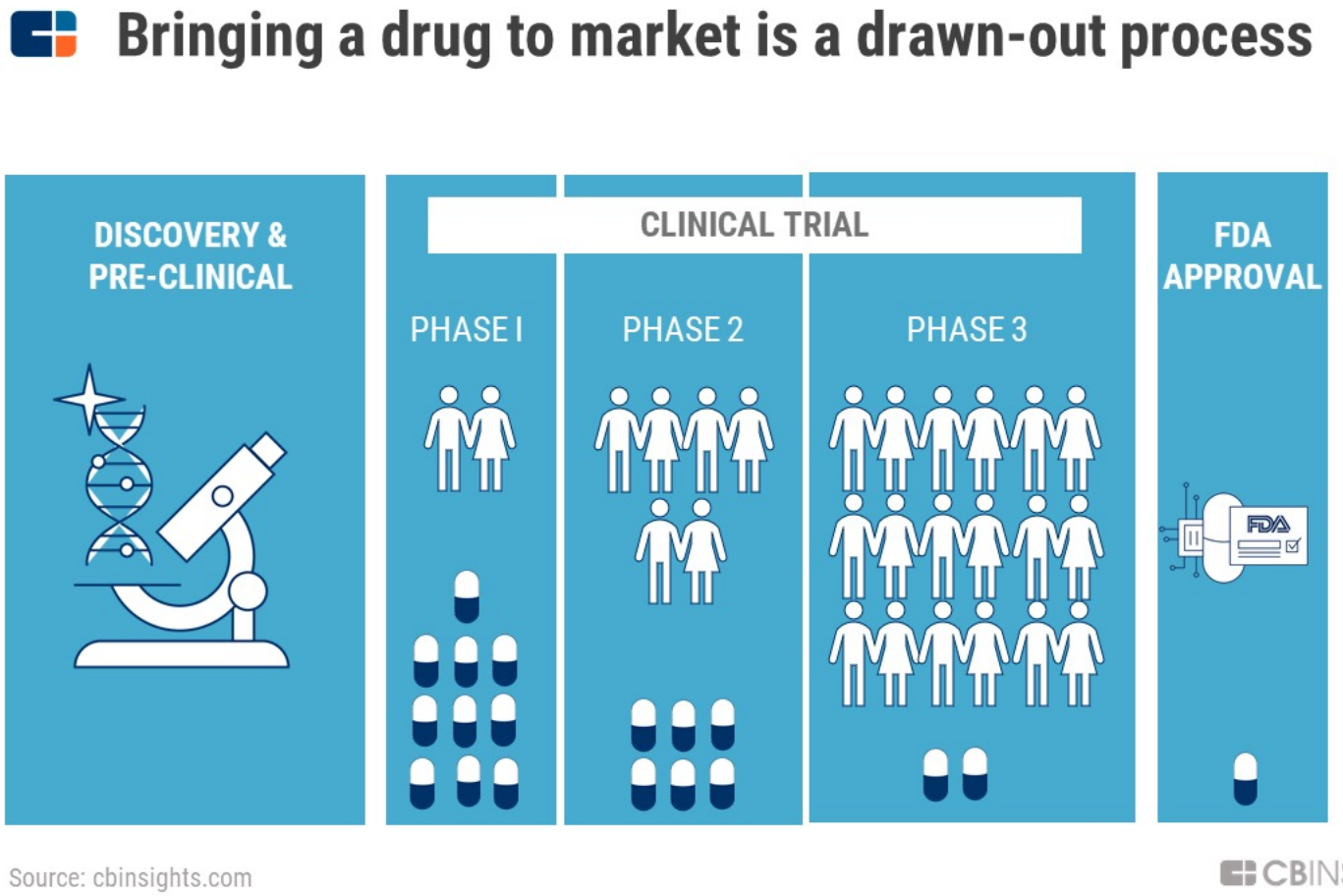
Problem:
Grade the ~1M unique student implementations of this problem on code.org

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

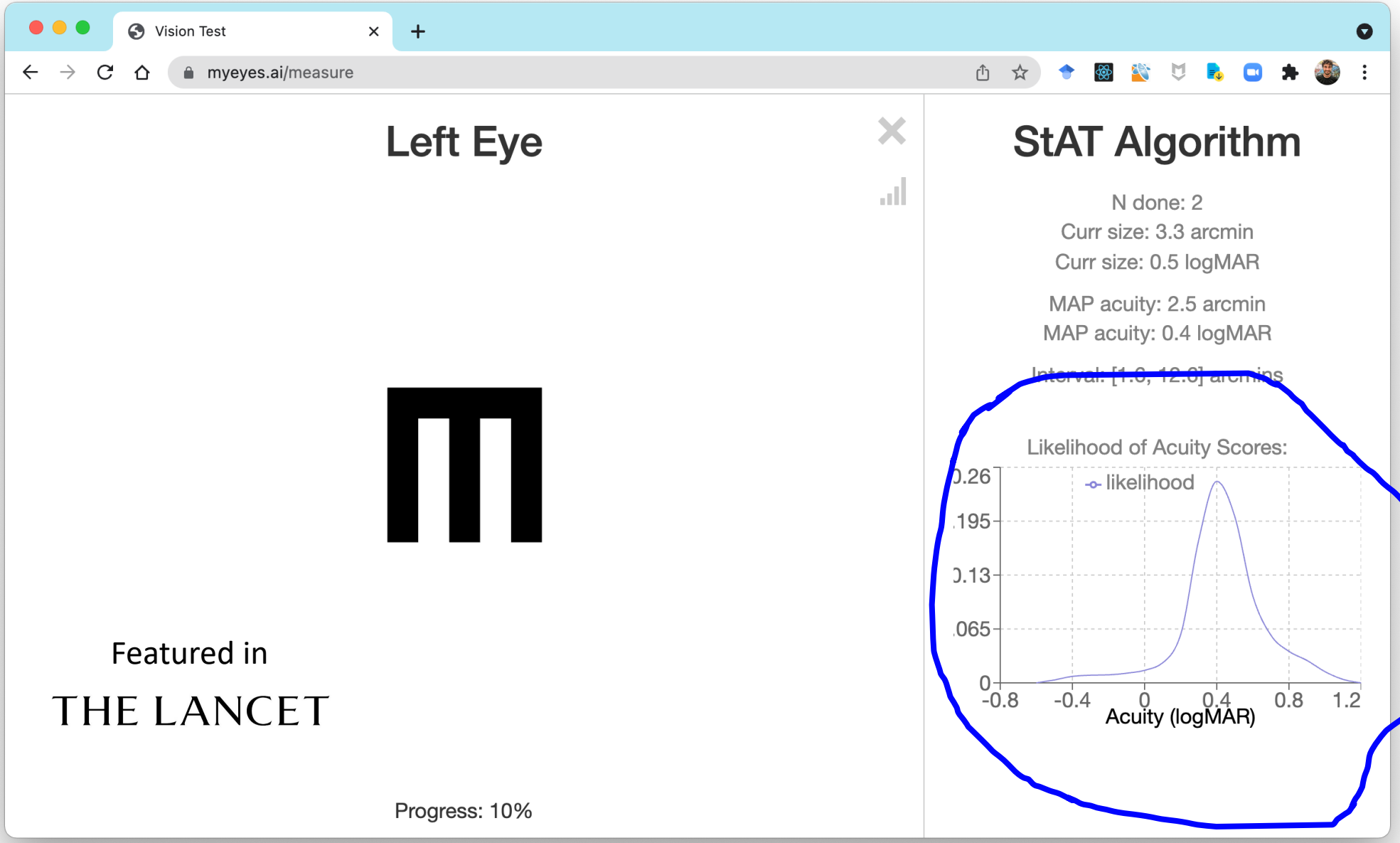
Impact:
Immediately change what sort of assignments are auto-gradable

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

More than education



More than education



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

Grades??

Application -> Theory

Understand social science, especially with small data

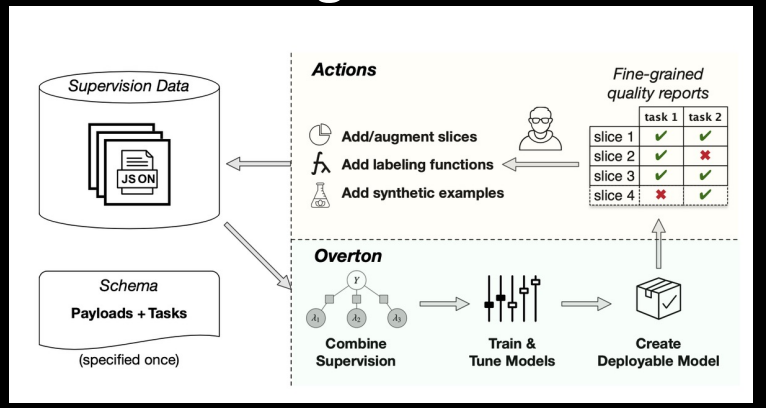
Explain why it made the choices it did

What are things that AI currently can't do?

Understand language

Design itself

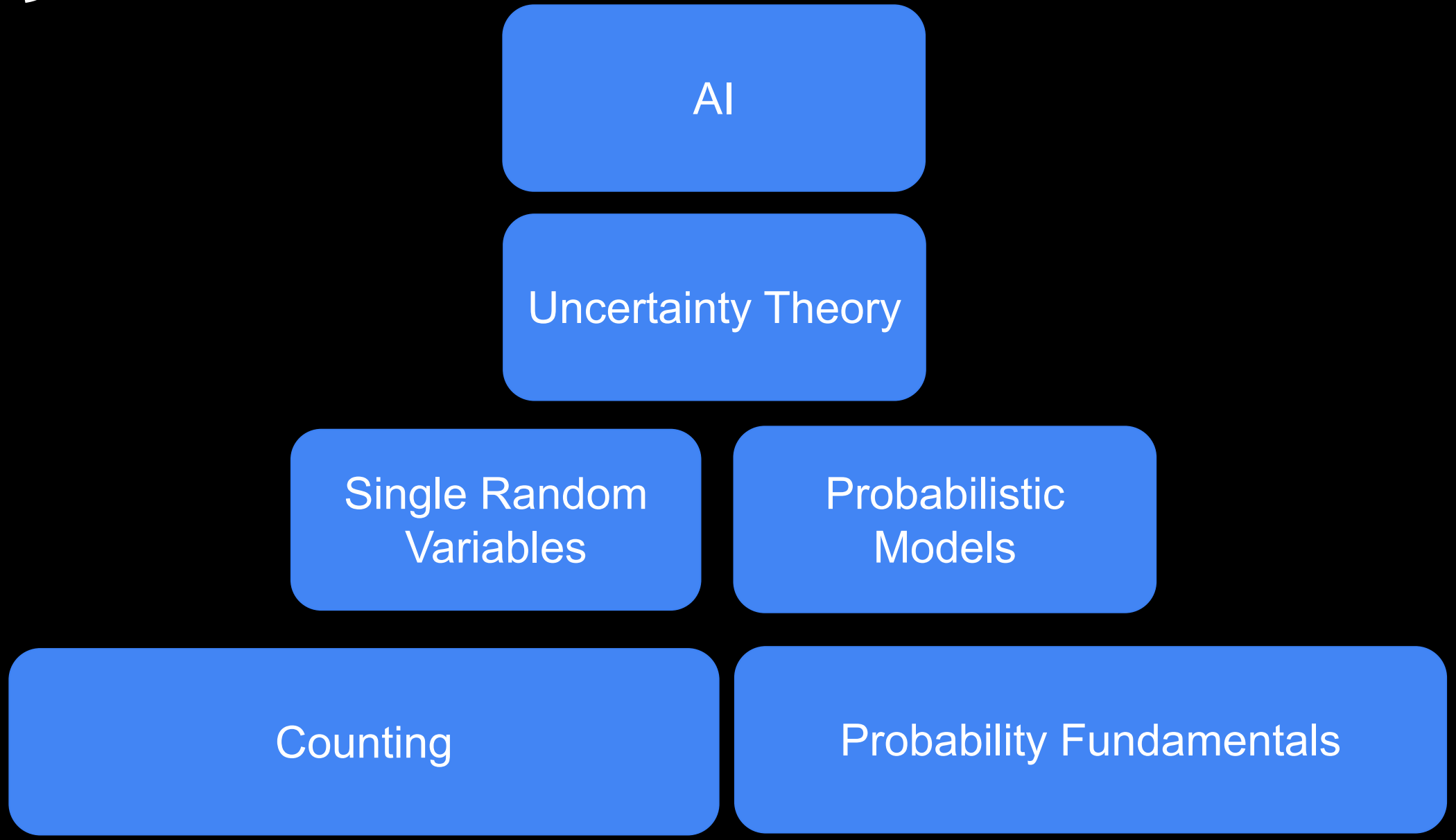
Teach humans based on what it has learned



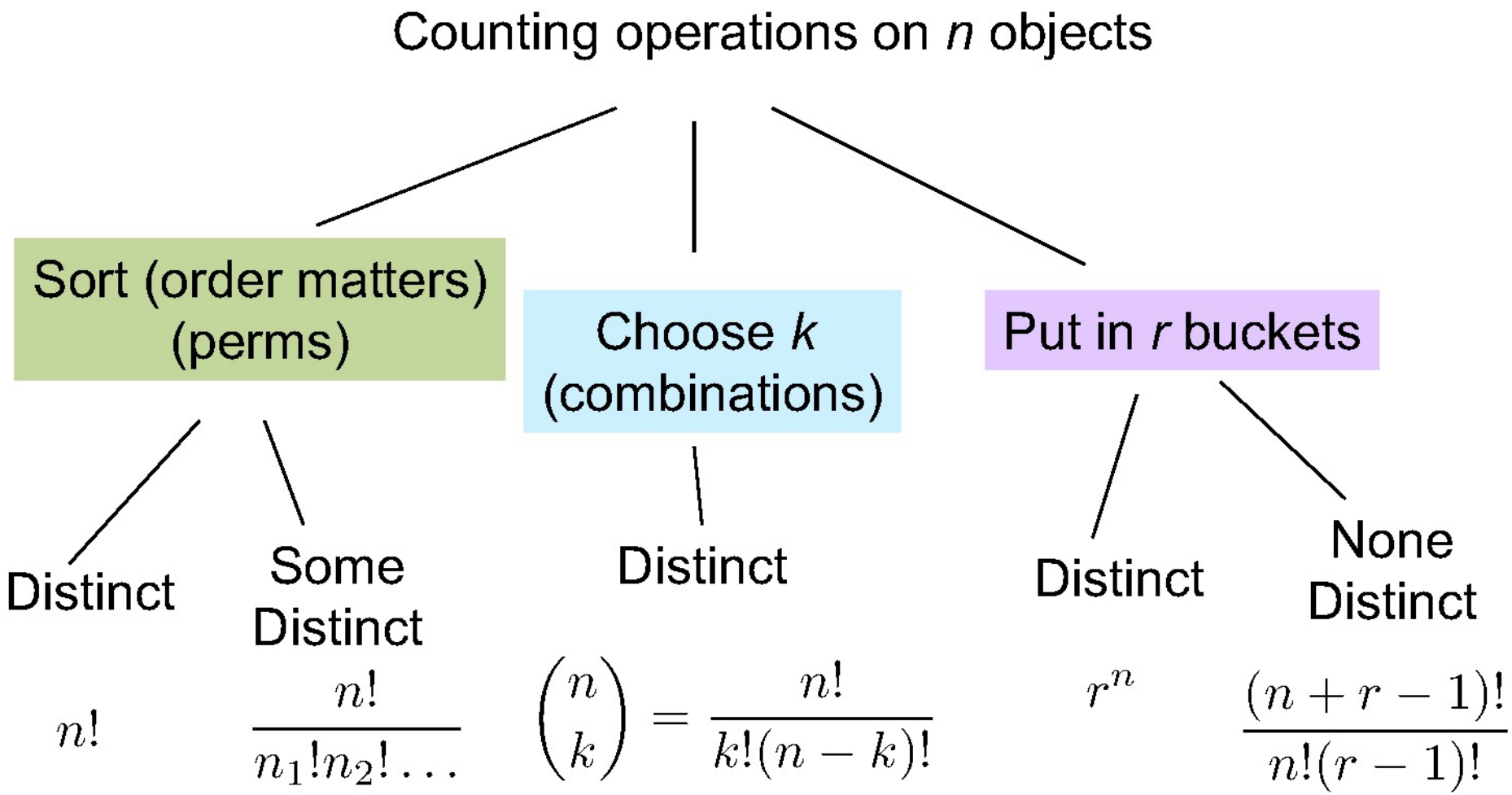
Learning Goal: Abundance of important problems

Last Class…

CS109


AI
Uncertainty Theory
Single Random Variables
Probabilistic Models
Counting
Probability Fundamentals

Counting Rules



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?

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

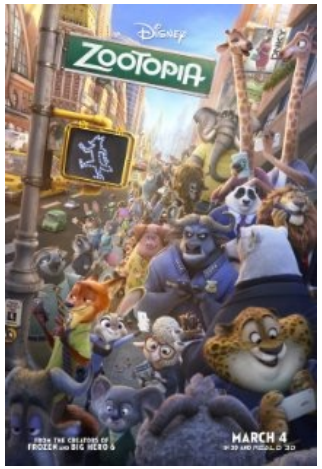
$$P(E|F) = \frac{P(EF)}{P(F)}$$ Definition of Cond. Probability

Netflix and Learn

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

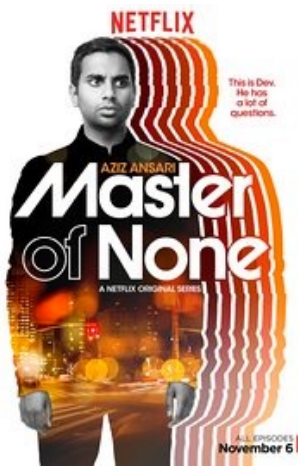
$P(E) = 0.19$

$P(E|F) = 0.14$



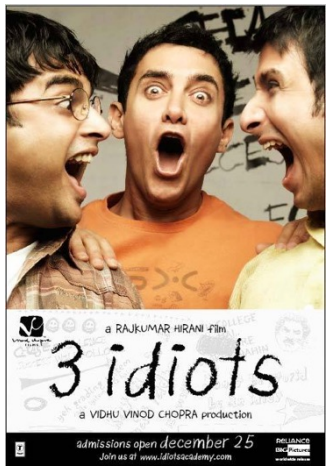
$P(E) = 0.32$

$P(E|F) = 0.35$



$P(E) = 0.20$

$P(E|F) = 0.20$



$P(E) = 0.09$

$P(E|F) = 0.72$

$P(E) = 0.20$

$P(E|F) = 0.42$

Monty Hall Problem

and Wayne Brady

Marilyn discovers the Probability Bug

Zika Test

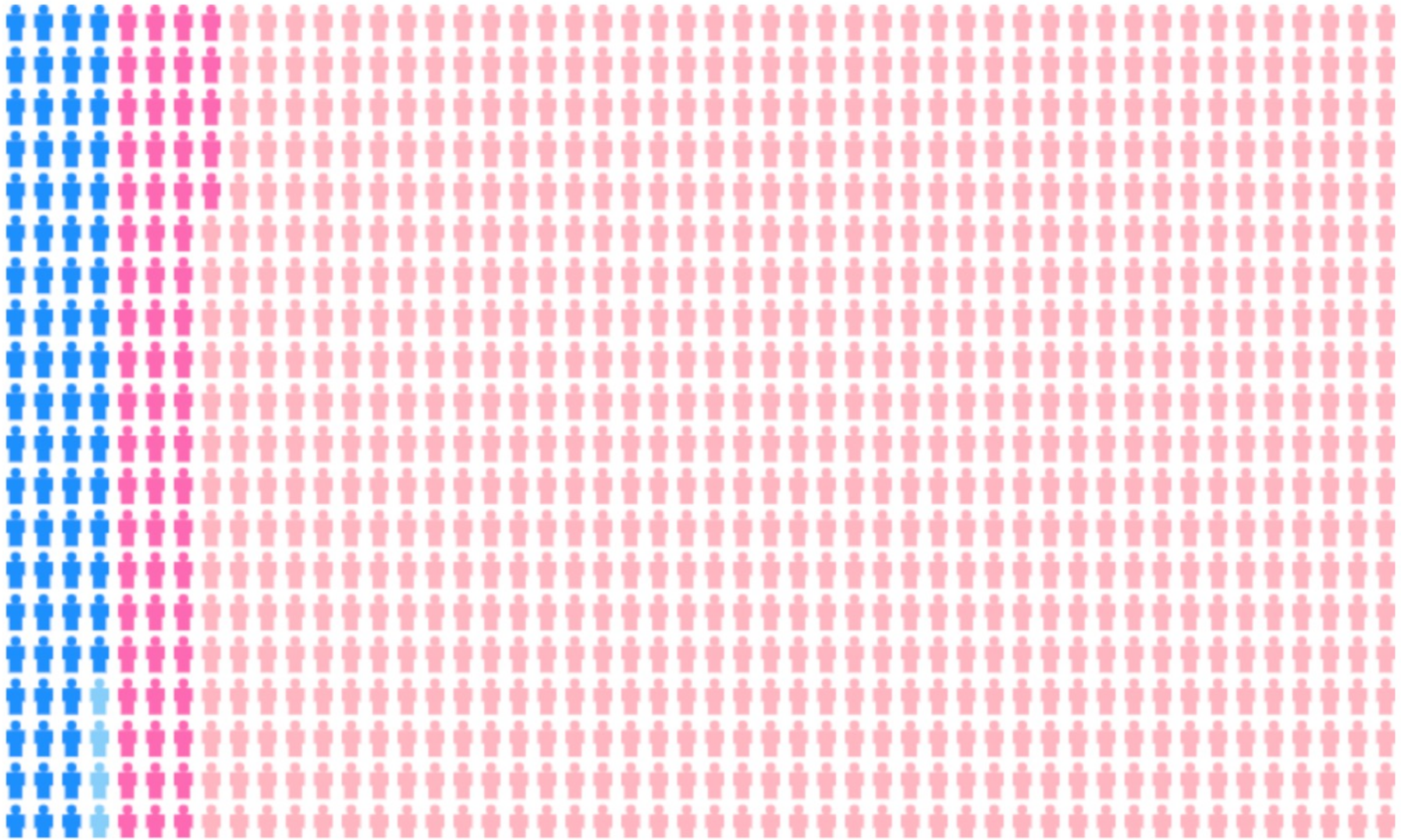
Positive Zika.

What is the probability of zika?

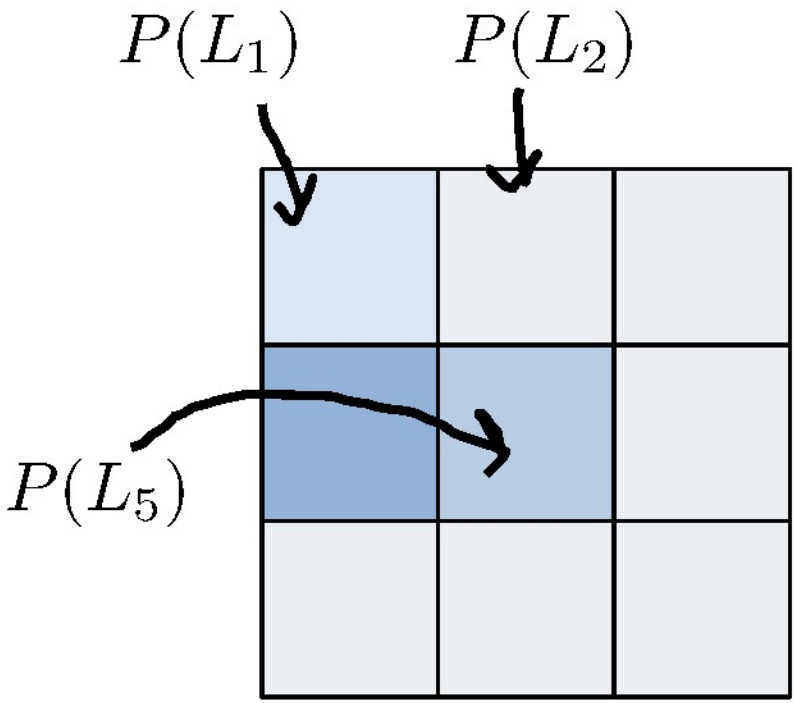
- *0.1% of people have zika*
- *90% positive rate for people with zika*
- *7% positive rate for people without zika*

The right answer is 1%

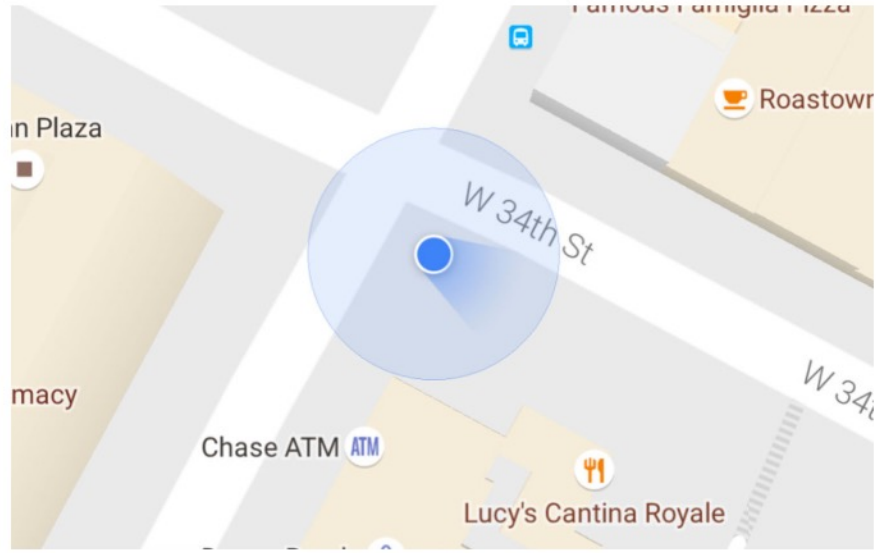
Bayes Theorem Intuition



Update Belief



Before Observation



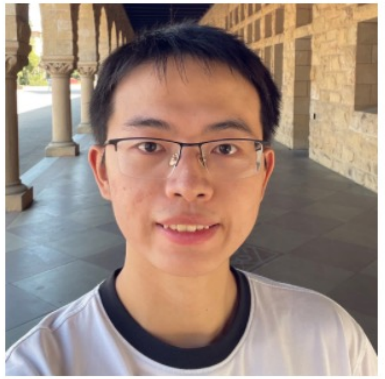
Recall our Ebola Bats

Sections and Office Hours with your TA



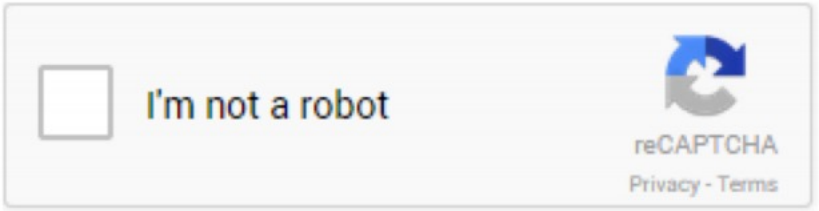


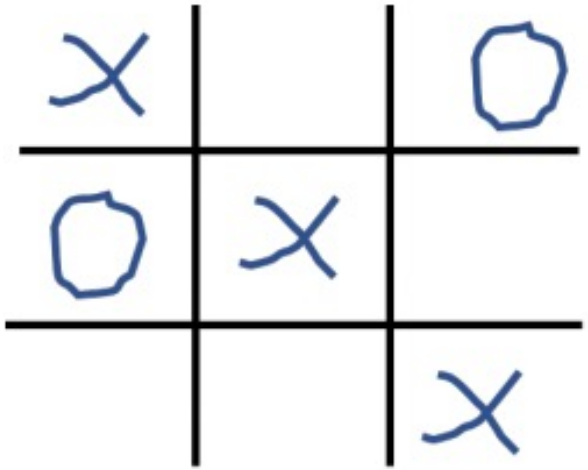




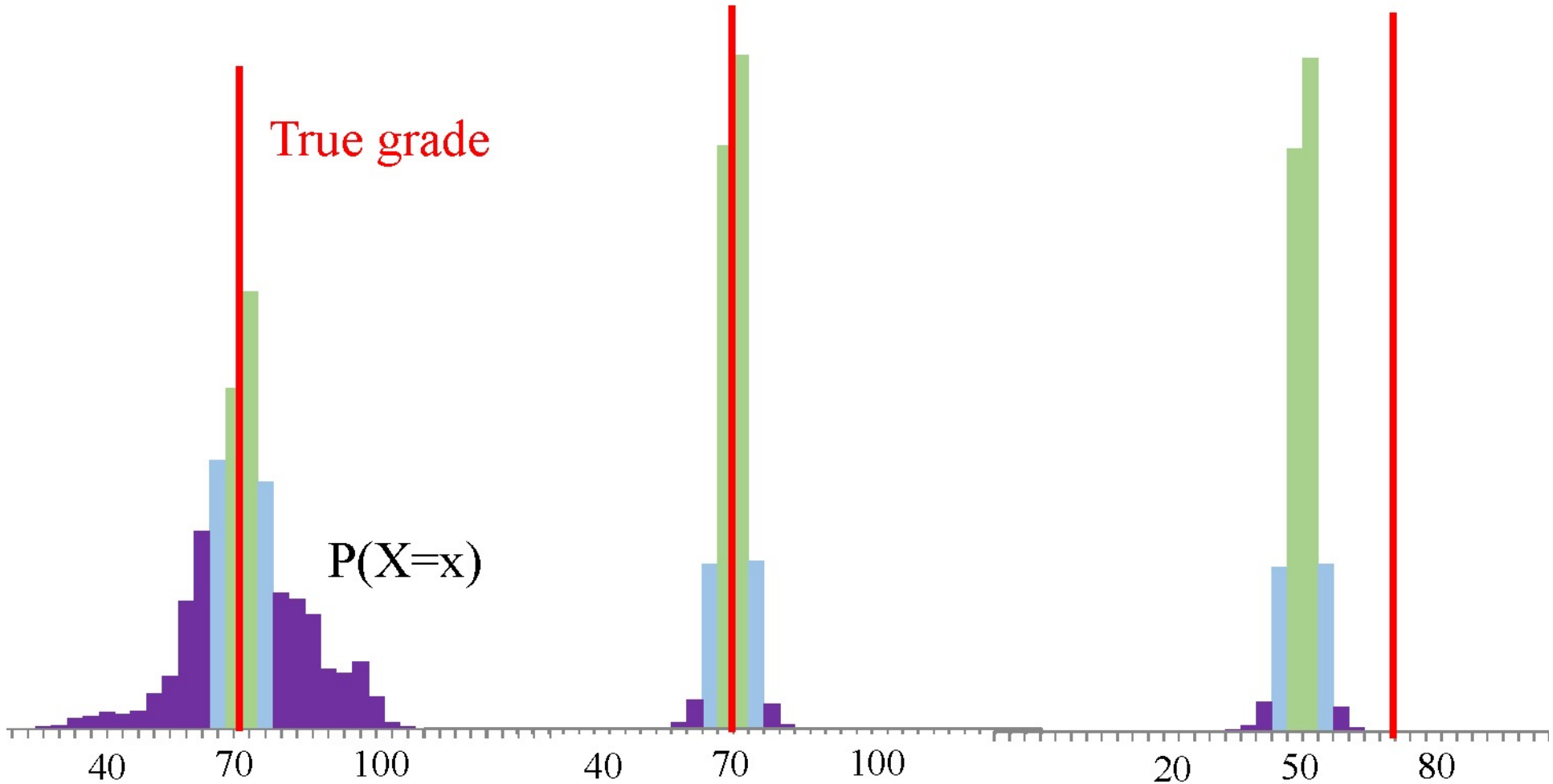
Third Year of Sections



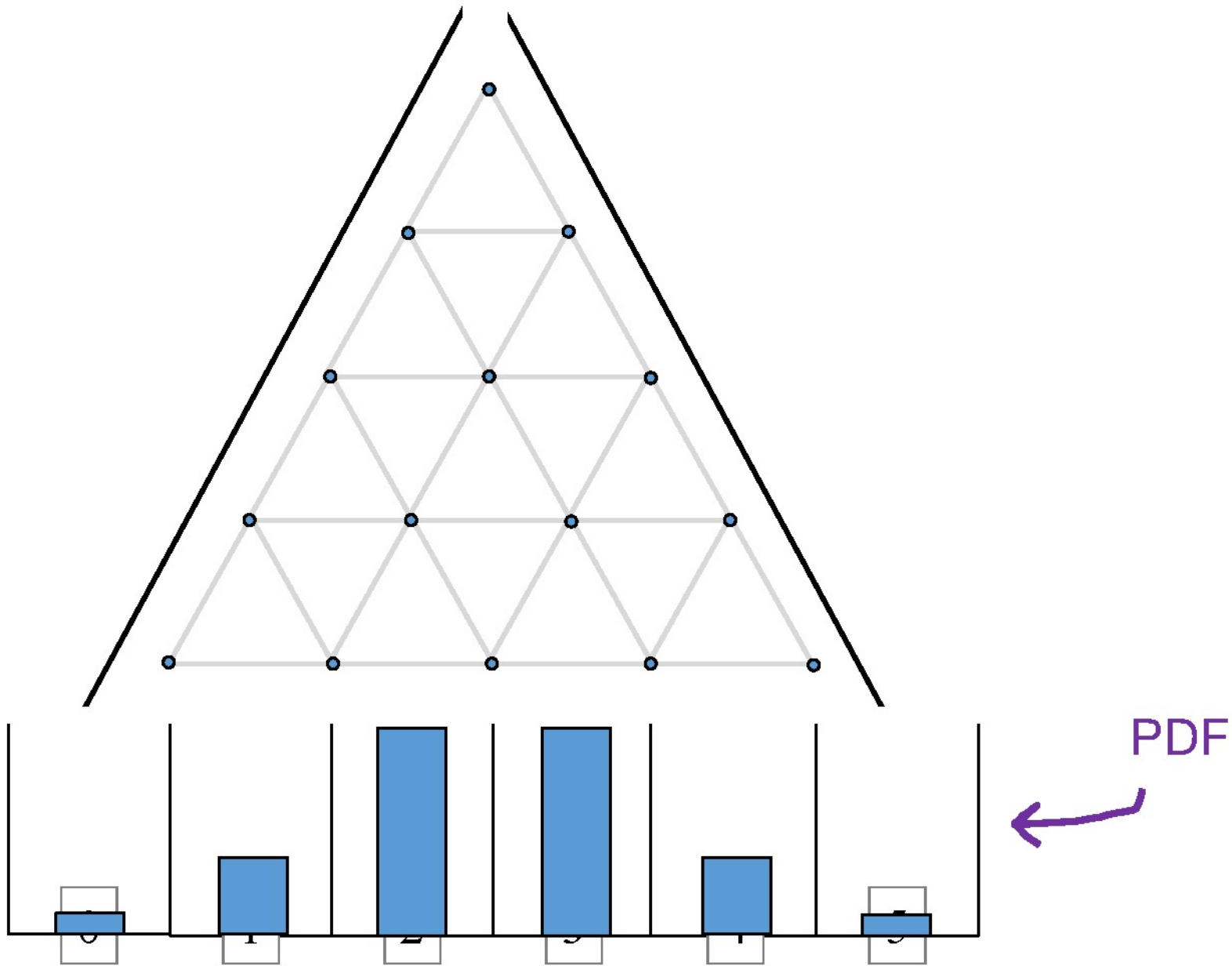


Random Variables

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



Binomial



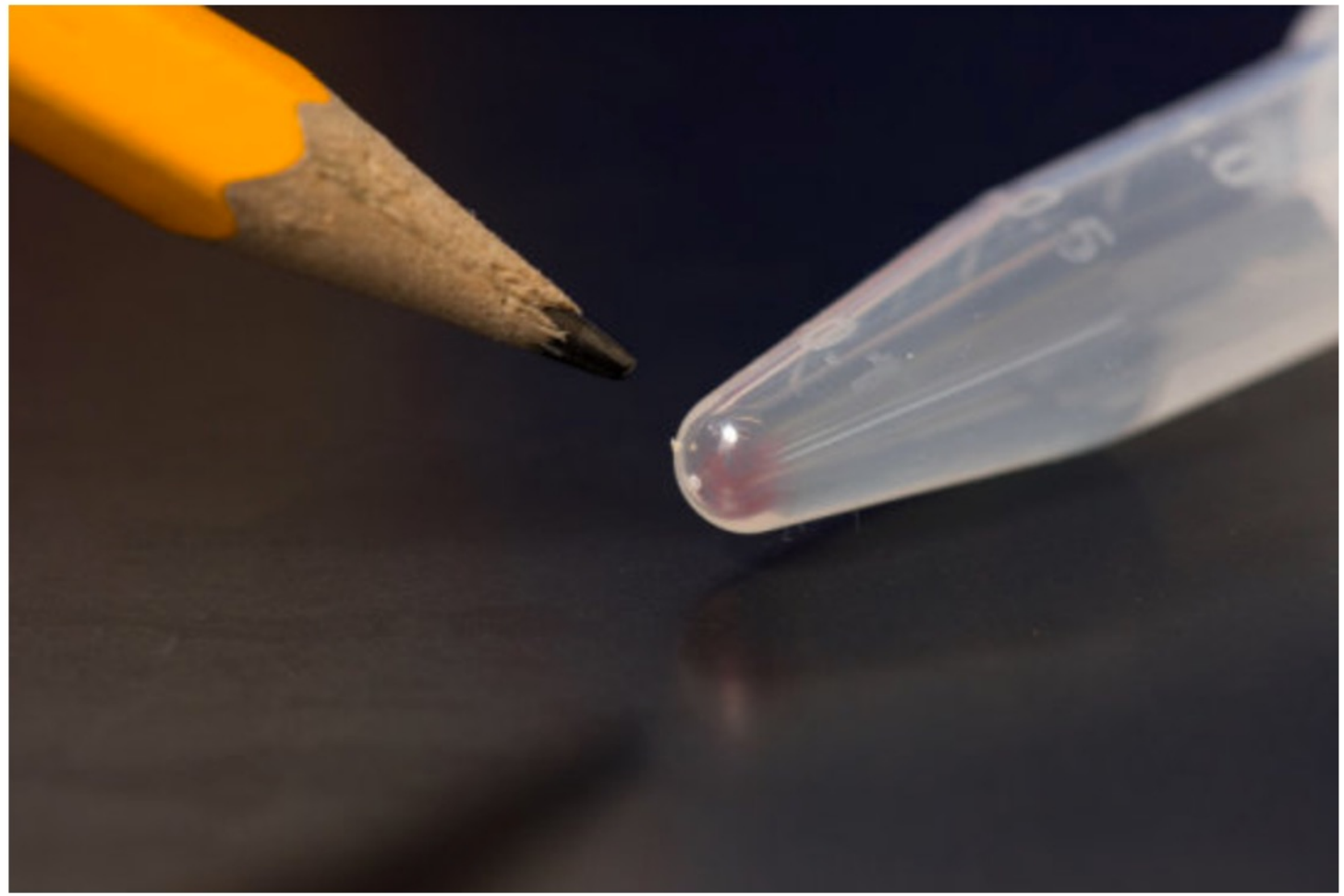

1001

Geometric

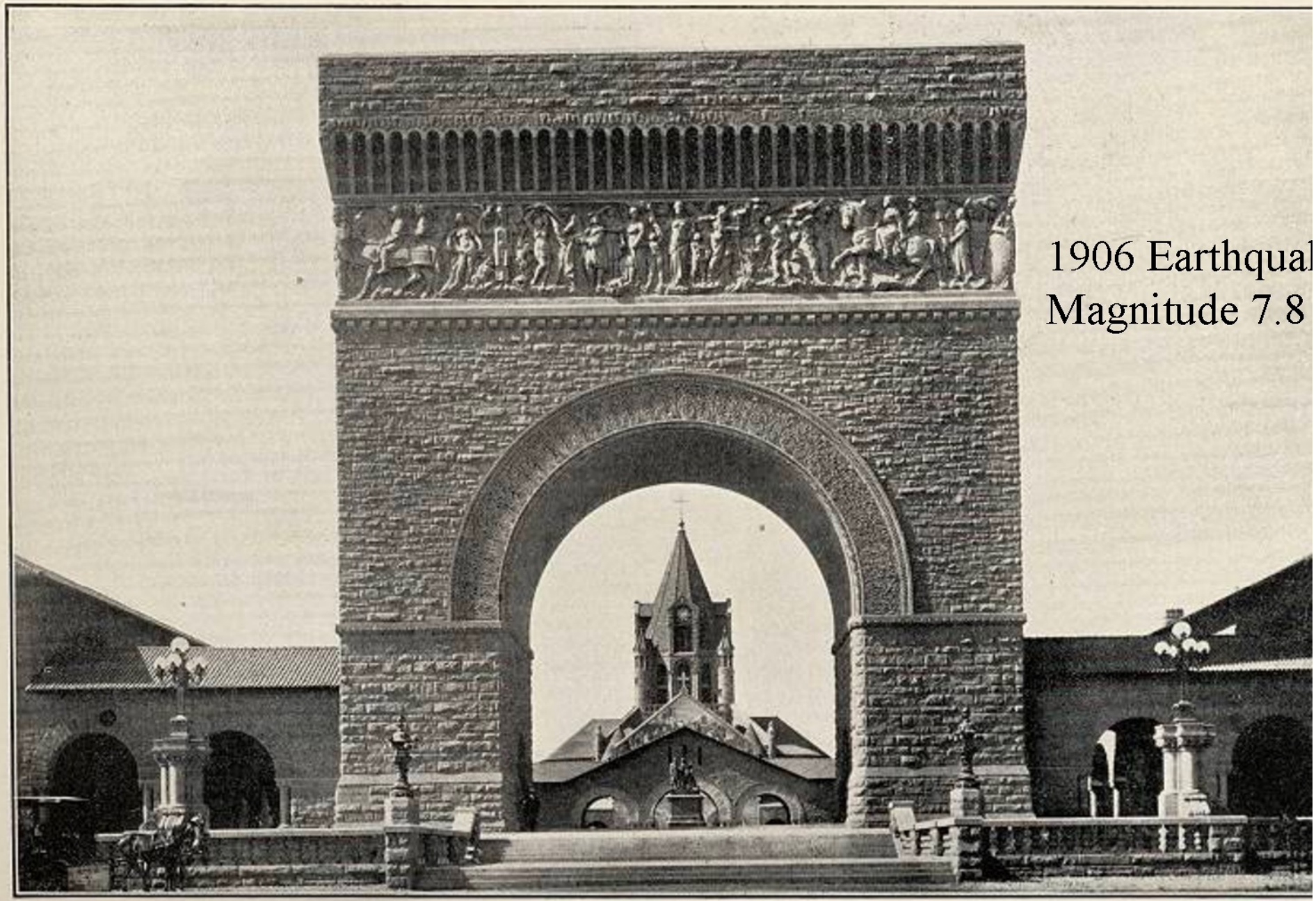
Sequence 1:

TTHHTHTTHTTTHTTTTHTTTHTTHTHHTH
HTTHHTTHHHTHHTHTTHTHTTHHTHHHT
HTHHTHHHTHTHTTHTTHHTHTHTHTTHH
TTHHTHTTHTHTHTHTHTHTHHHTHTHTH
TTHHTHTHTHTHHHTTHHTHTTTHTHHTH

Storing Data on DNA

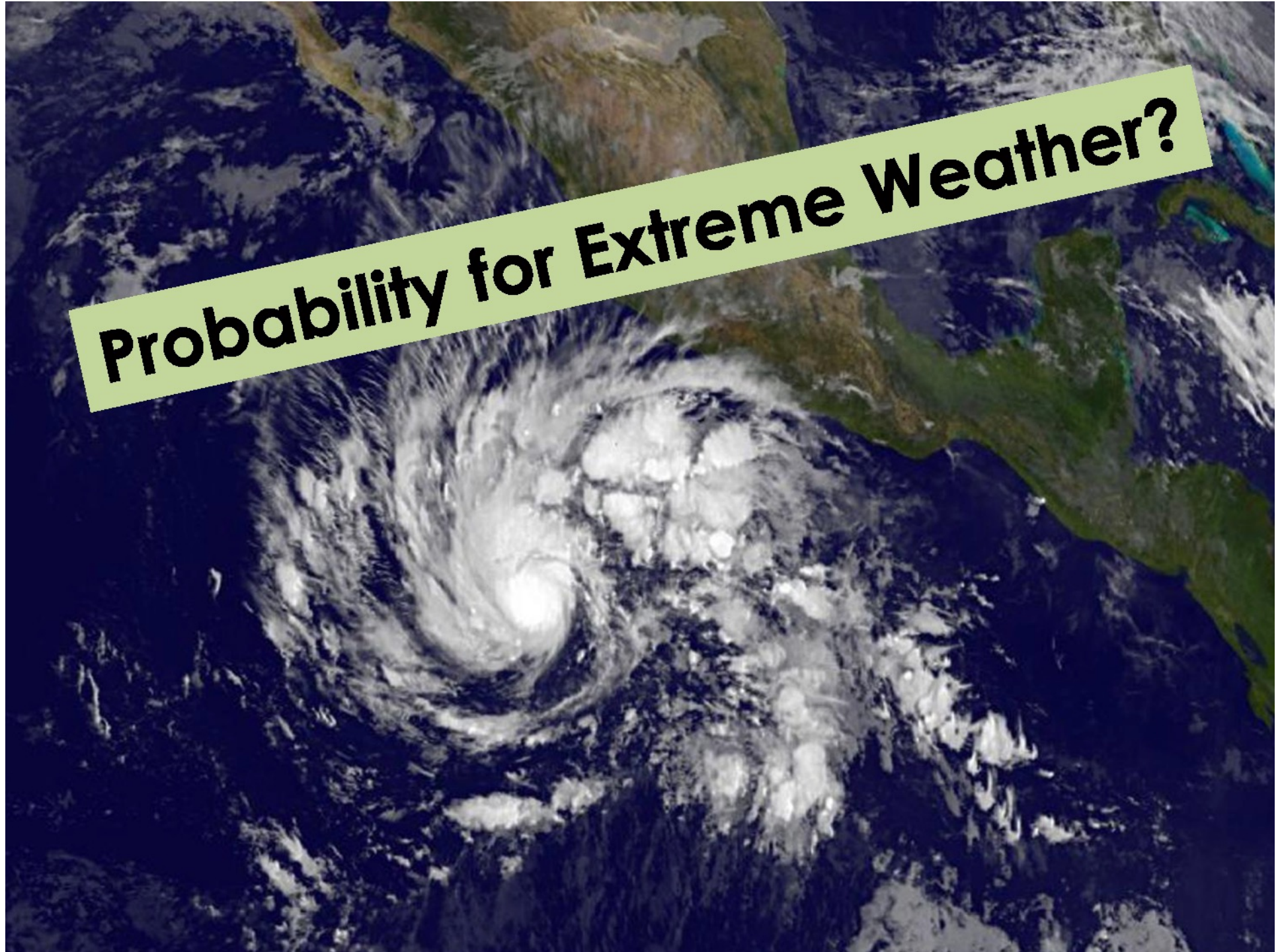


All the movies, images, emails and other digital data from more than 600 smartphones (10,000 gigabytes) can be stored in the faint pink smear of DNA at the end of this test tube.



1906 Earthqual
Magnitude 7.8

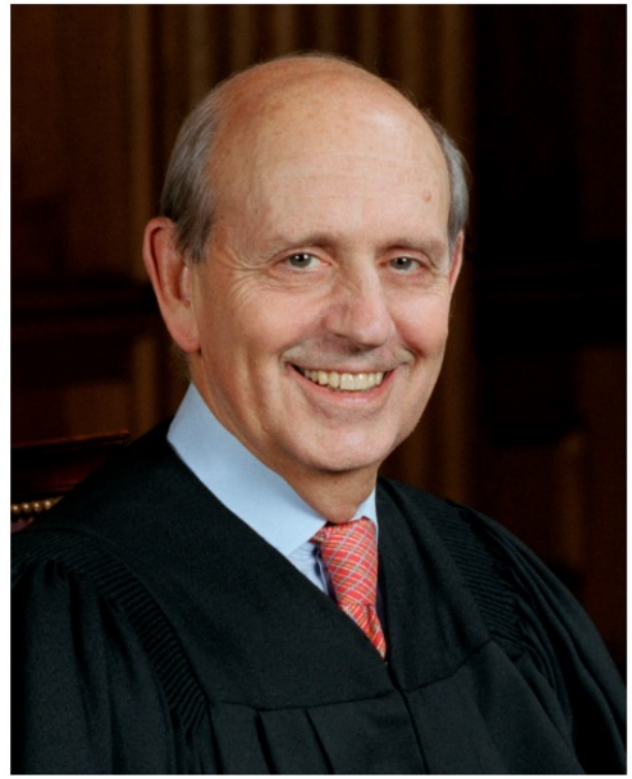
ILL. No. 65. MEMORIAL ARCH, WITH CHURCH IN BACKGROUND, STANFORD UNIVERSITY, SHOWING TYPES OF CARVED WO
WITH THE SANDSTONE.


Probability for Extreme Weather?

Bit Coin Mining

You "mine a bitcoin" if, for given data D, you find a number N such that Hash(D, N) produces a string that starts with g zeroes.

Representative Juries



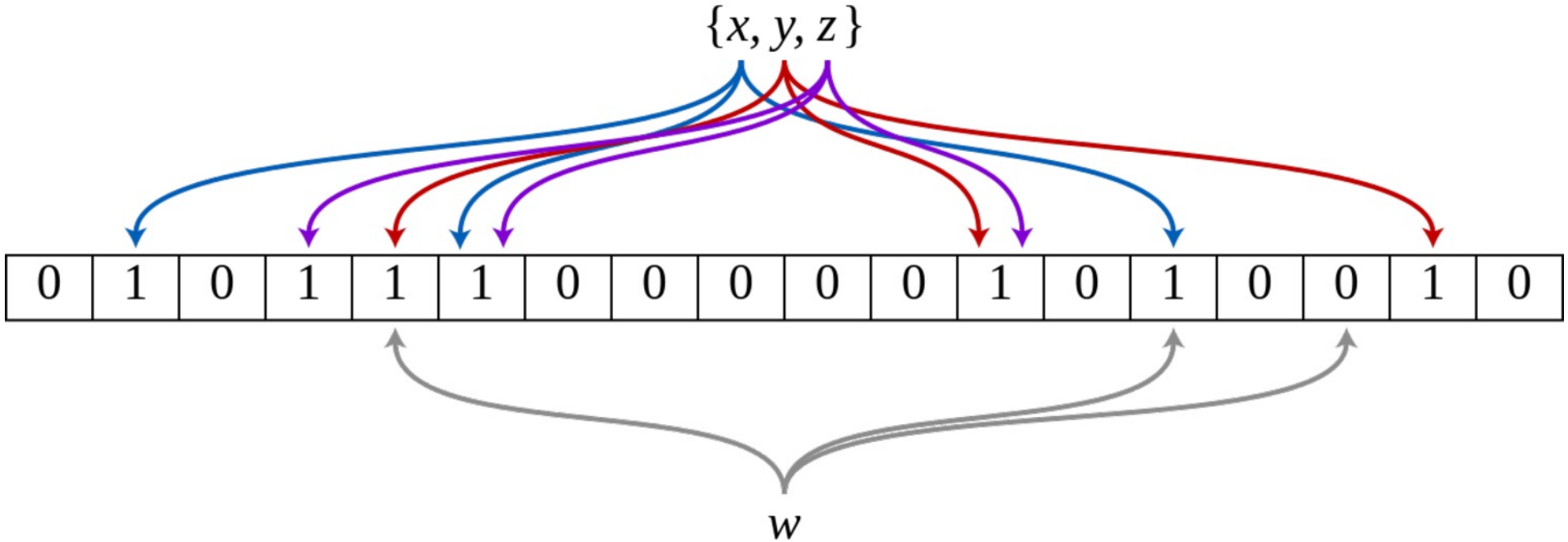
Simulate

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

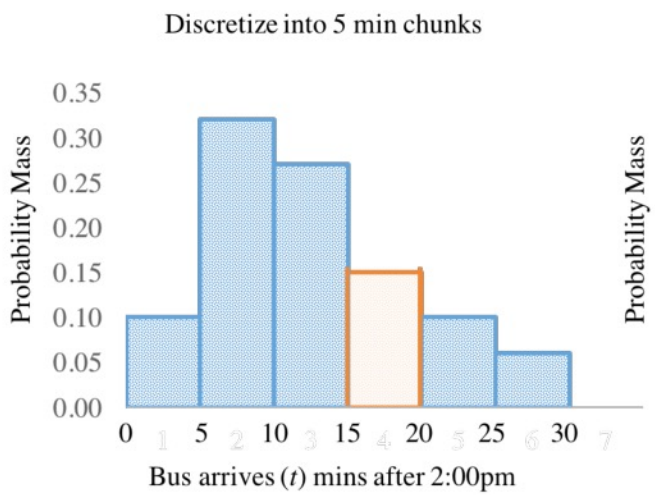


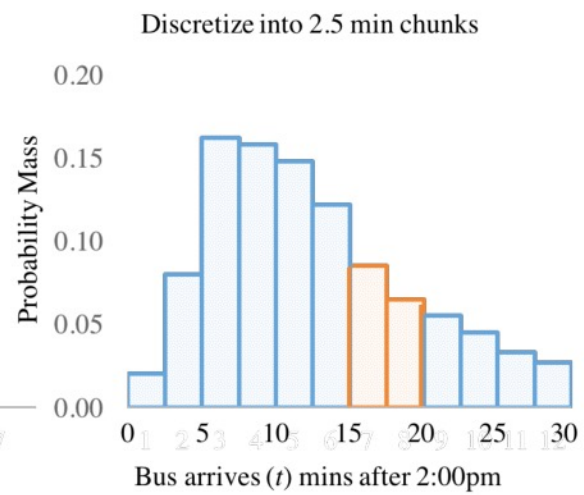
random()?

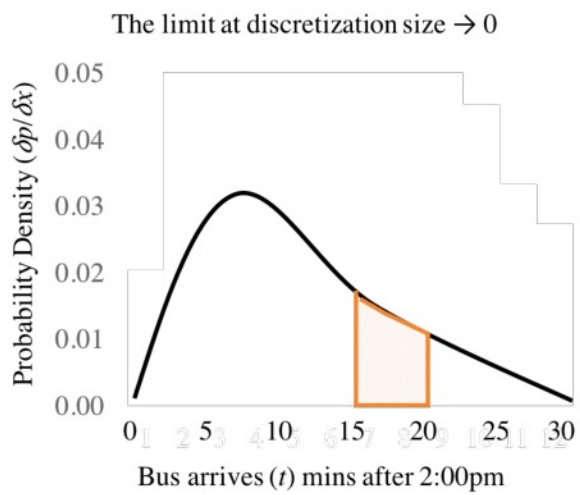
Riding the Marguerite

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

What is P(wait < 5 min)?







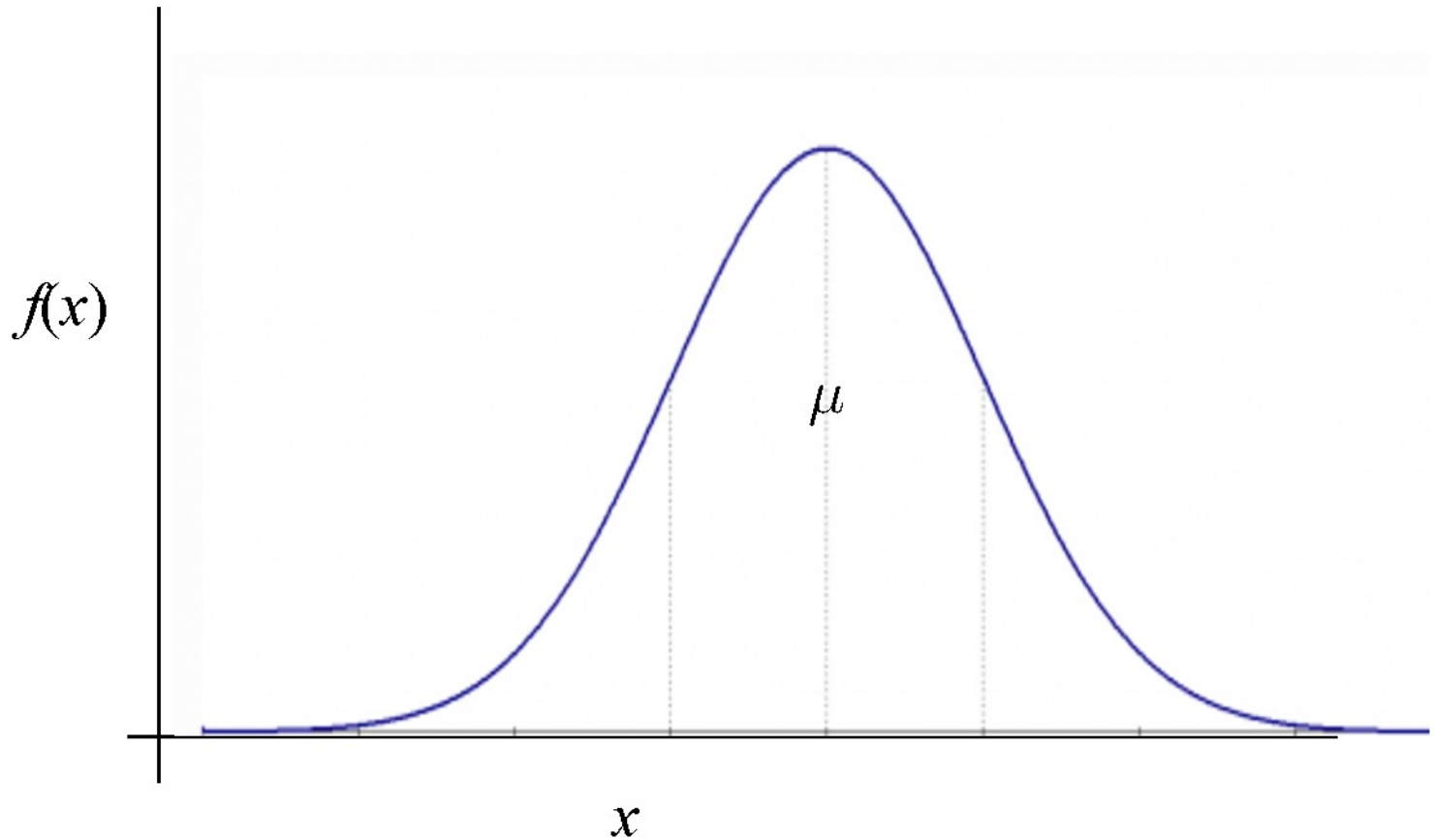
Integrals

*loving, not scary

Probability Density Function

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

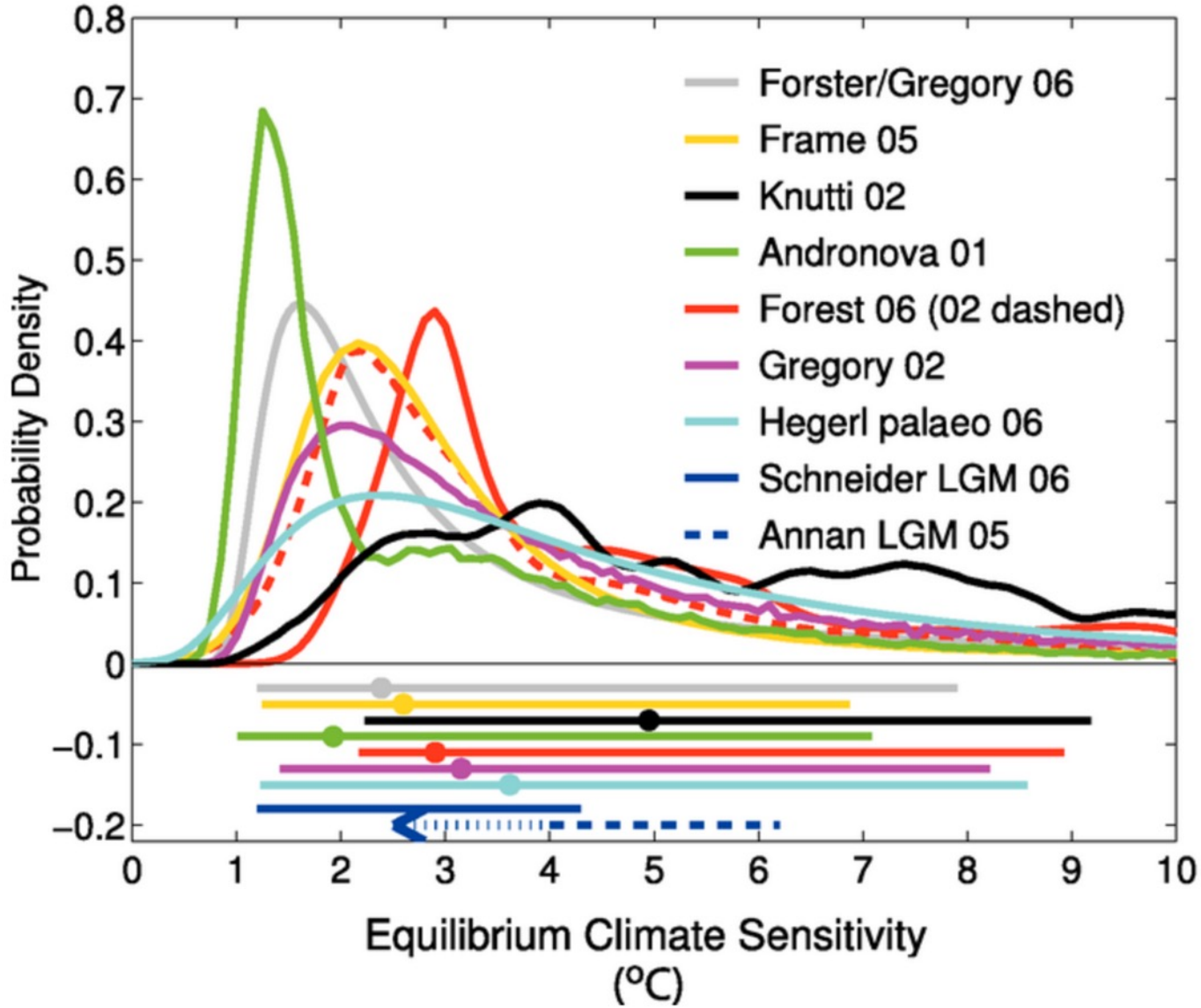
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$



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

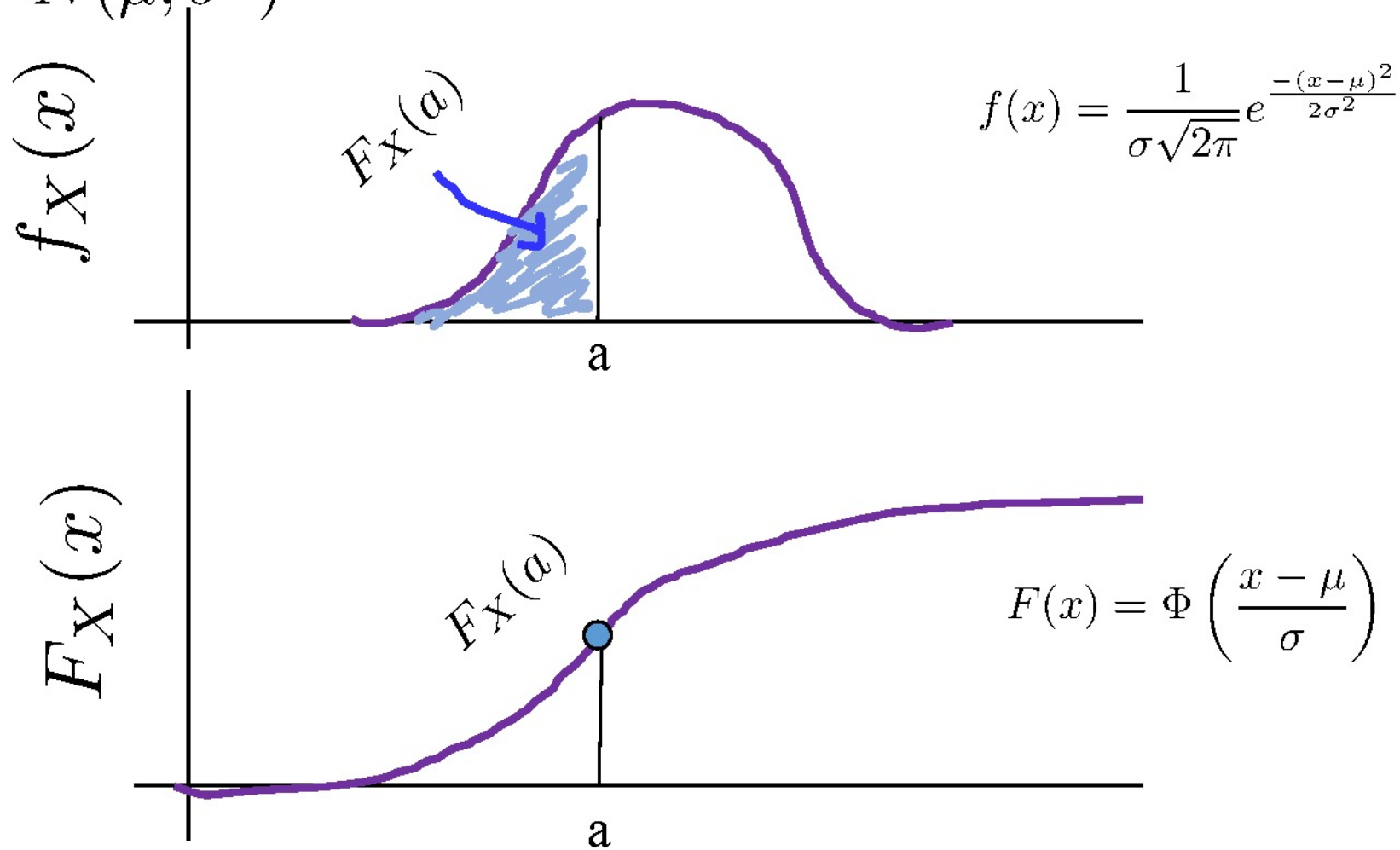
A probability!

Climate Sensitivity



PDF and CDF of a Normal

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



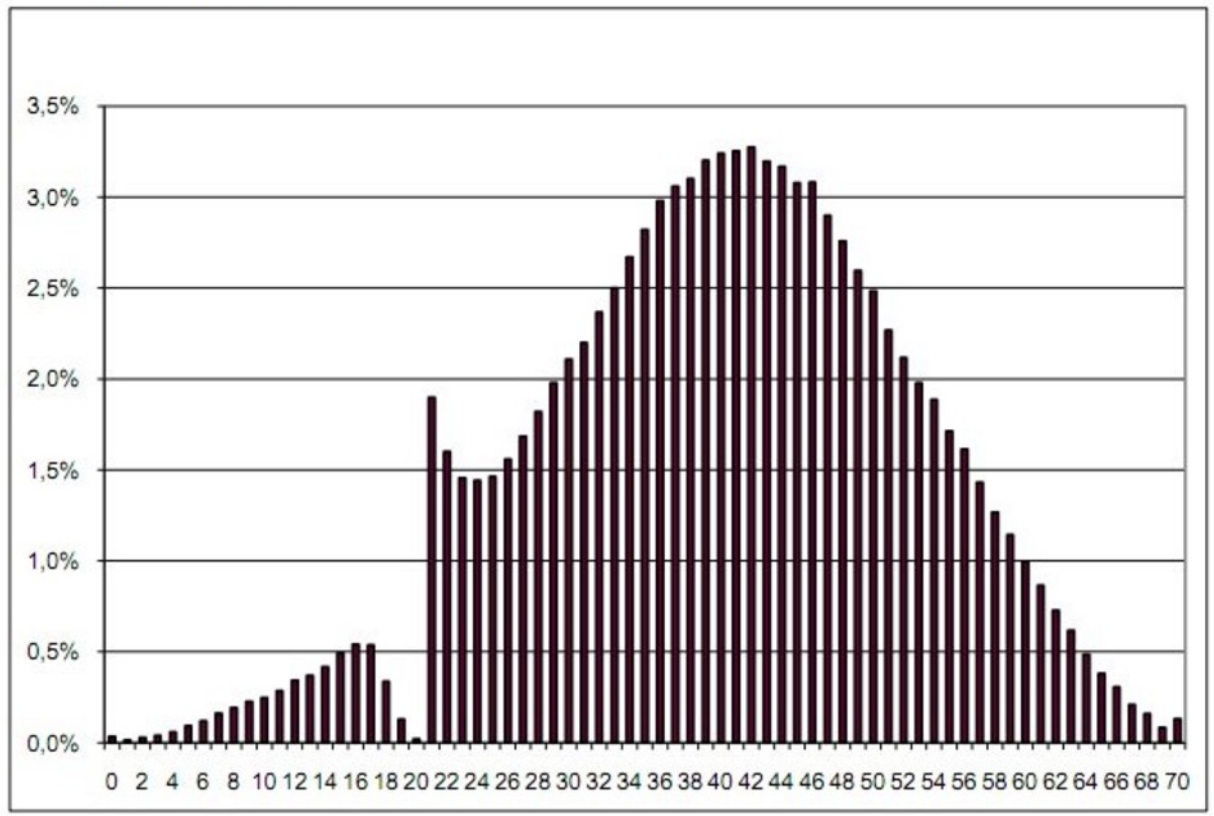
A CDF is the integral from –infinity to x of the PDF

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



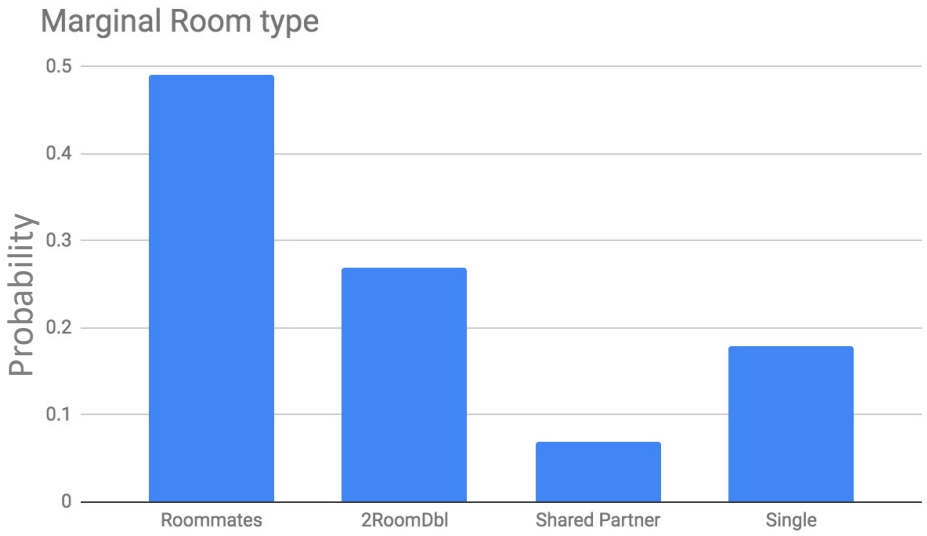
Wykres 1. Rozkład wyników na poziomie podstawowym

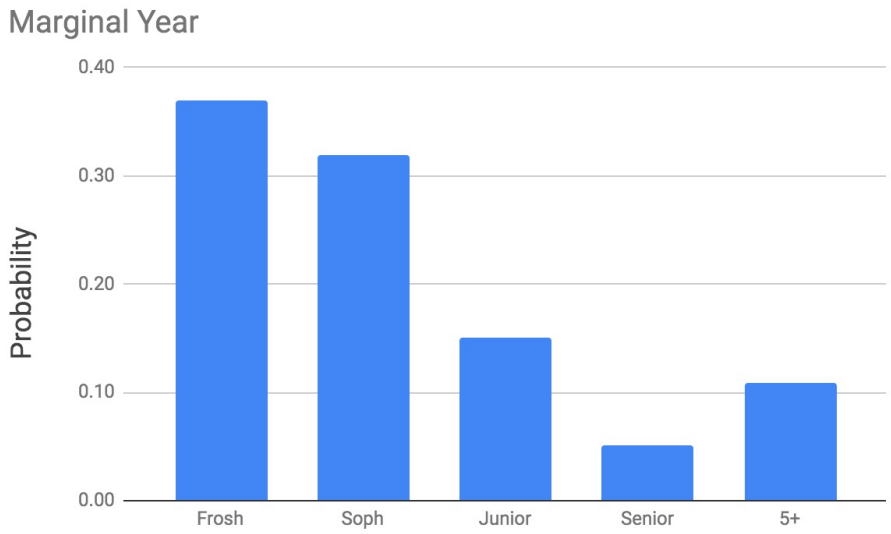
http://freakonomics.com/2011/07/07/another-case-of-teacher-cheating-or-is-it-just-altruism/comment-page-2/

Probabilistic Models

Joint Probability Table

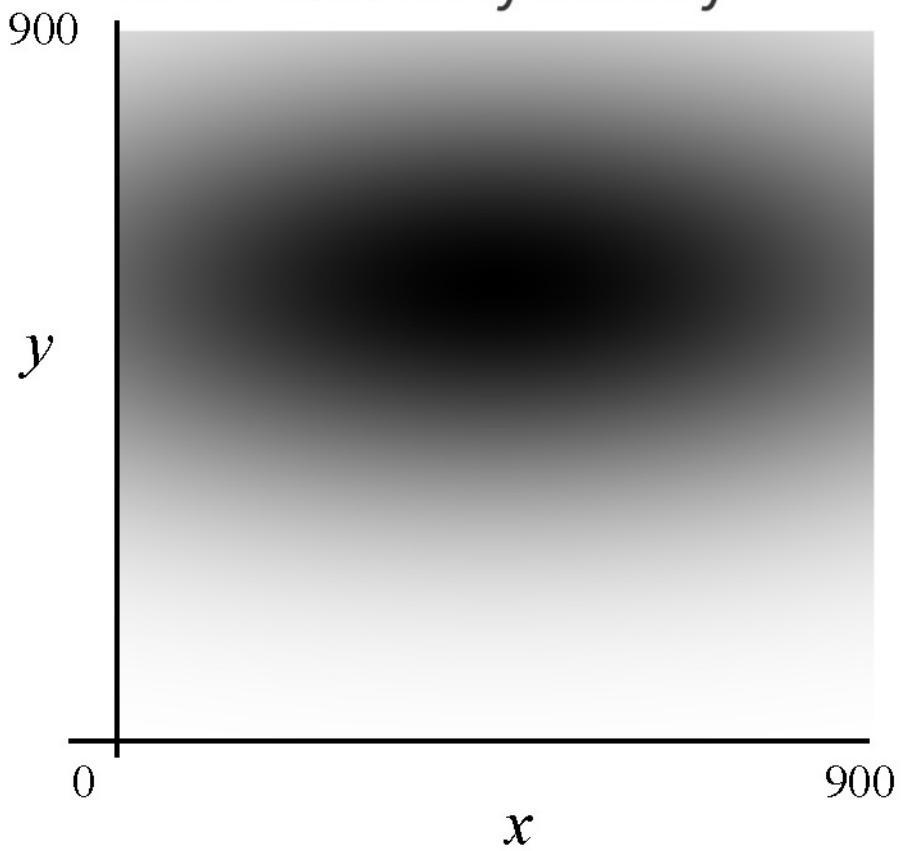
	Roommates	2RoomDbl	Shared Partner	Single	
Frosh	0.30	0.07	0.00	0.00	0.37
Soph	0.12	0.18	0.00	0.03	0.32
Junior	0.04	0.01	0.00	0.10	0.15
Senior	0.01	0.02	0.02	0.01	0.05
5+	0.02	0.00	0.05	0.04	0.11
	0.49	0.27	0.07	0.18	1.00





Joint Dart Distribution

Dart Probability Density



Dart Results

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

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

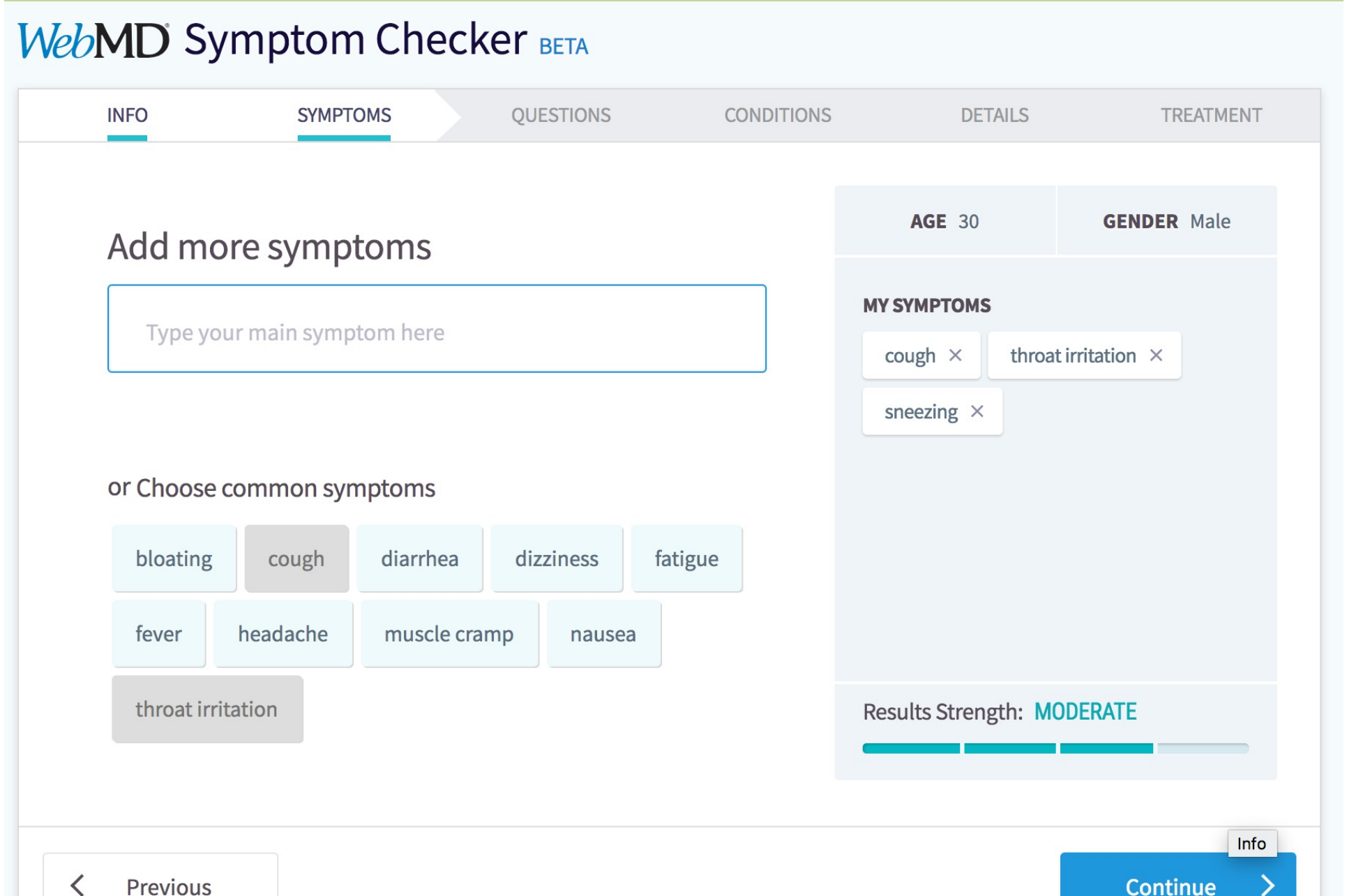
It's a Multinomial!

Probability of seeing this document | spam

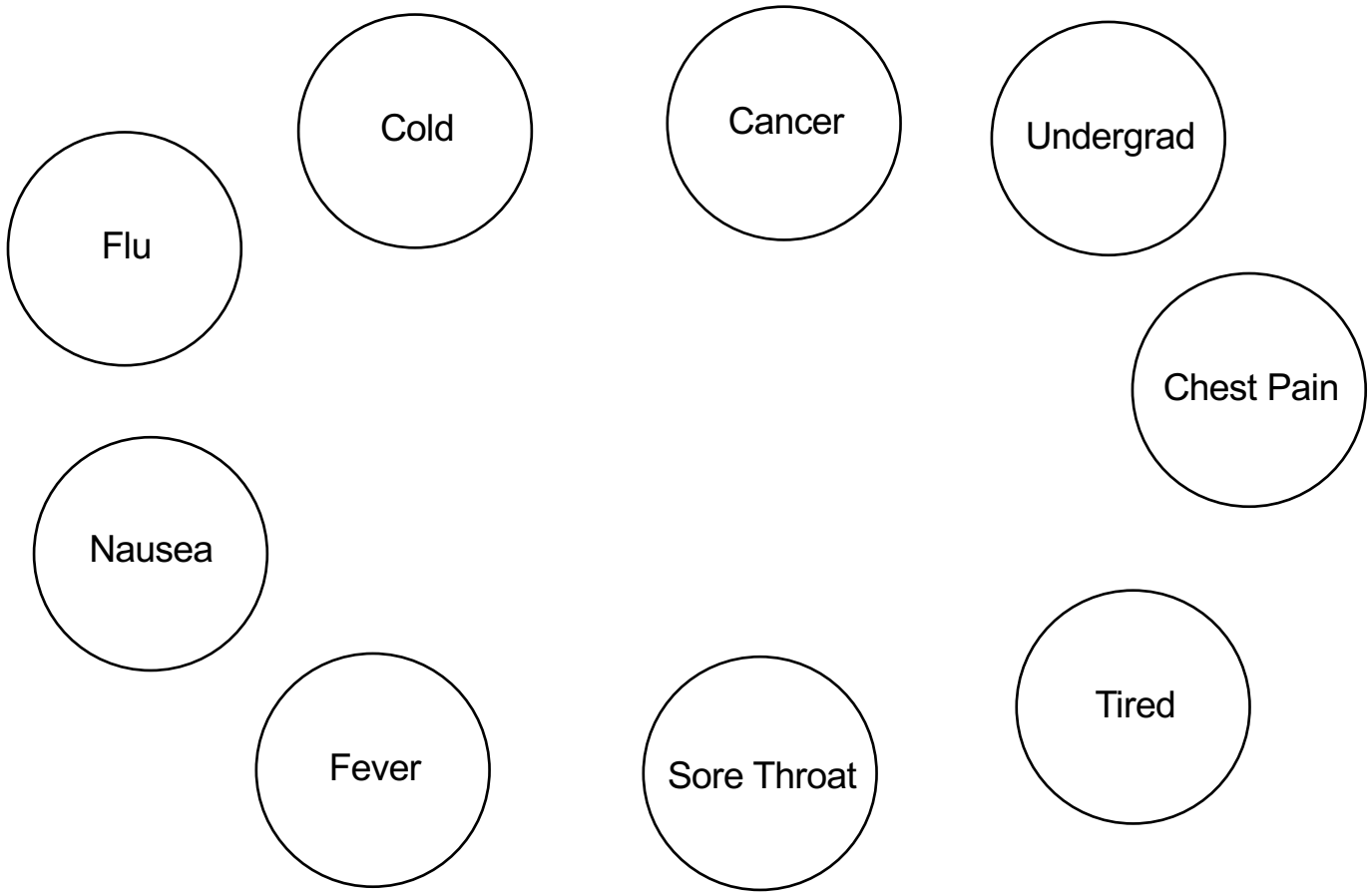
The probability of a word in spam email being viagra

General “Inference”

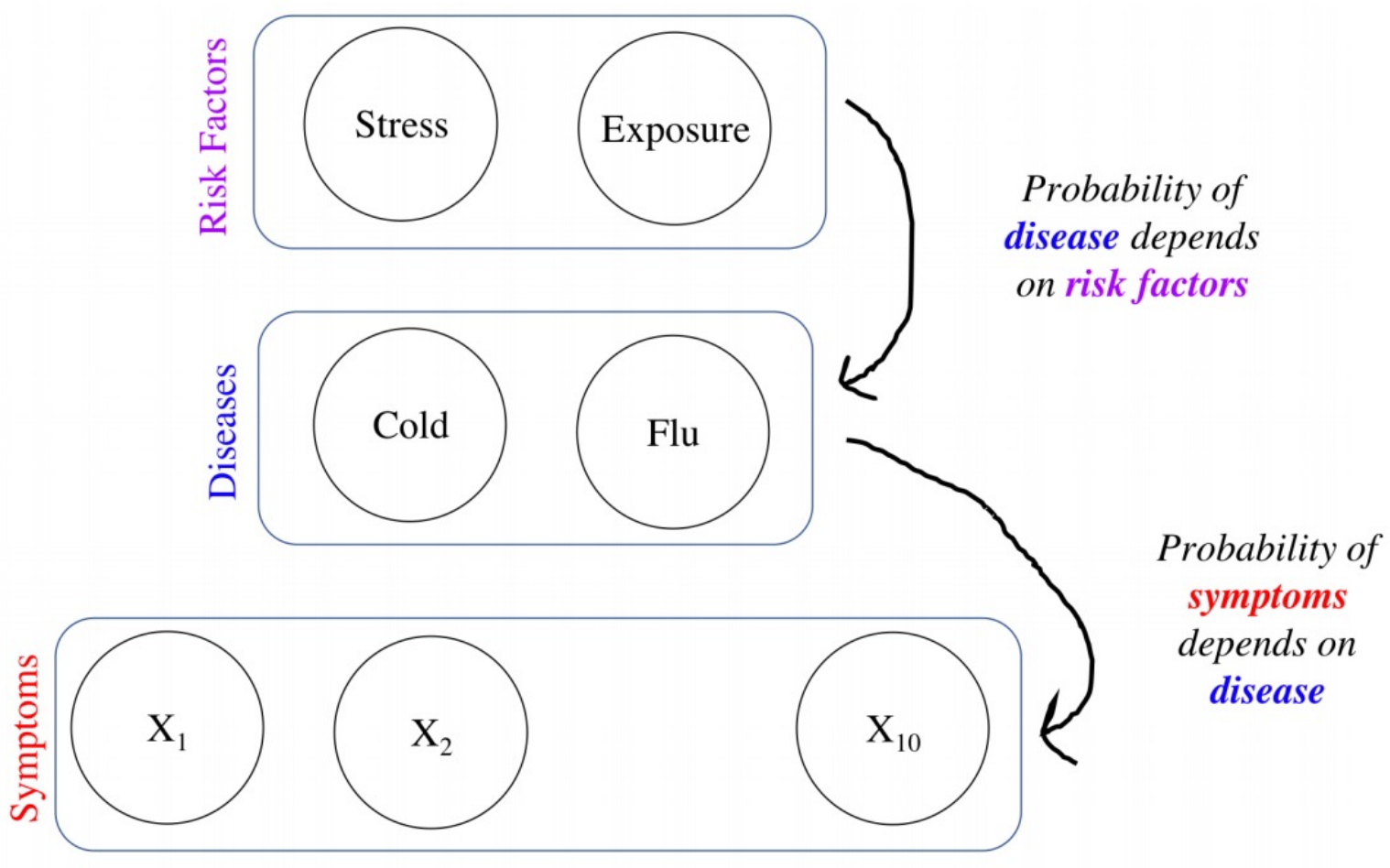
General “Inference”



Lots of Random Vars?



Bayes Nets!

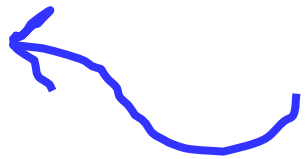

Risk Factors
Stress
Exposure
Diseases
Cold
Flu
Symptoms
X_1
X_2
X_{10}
Probability of disease depends on risk factors
Probability of symptoms depends on disease

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)
```

```
webMd — -bash — 38×22
[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:

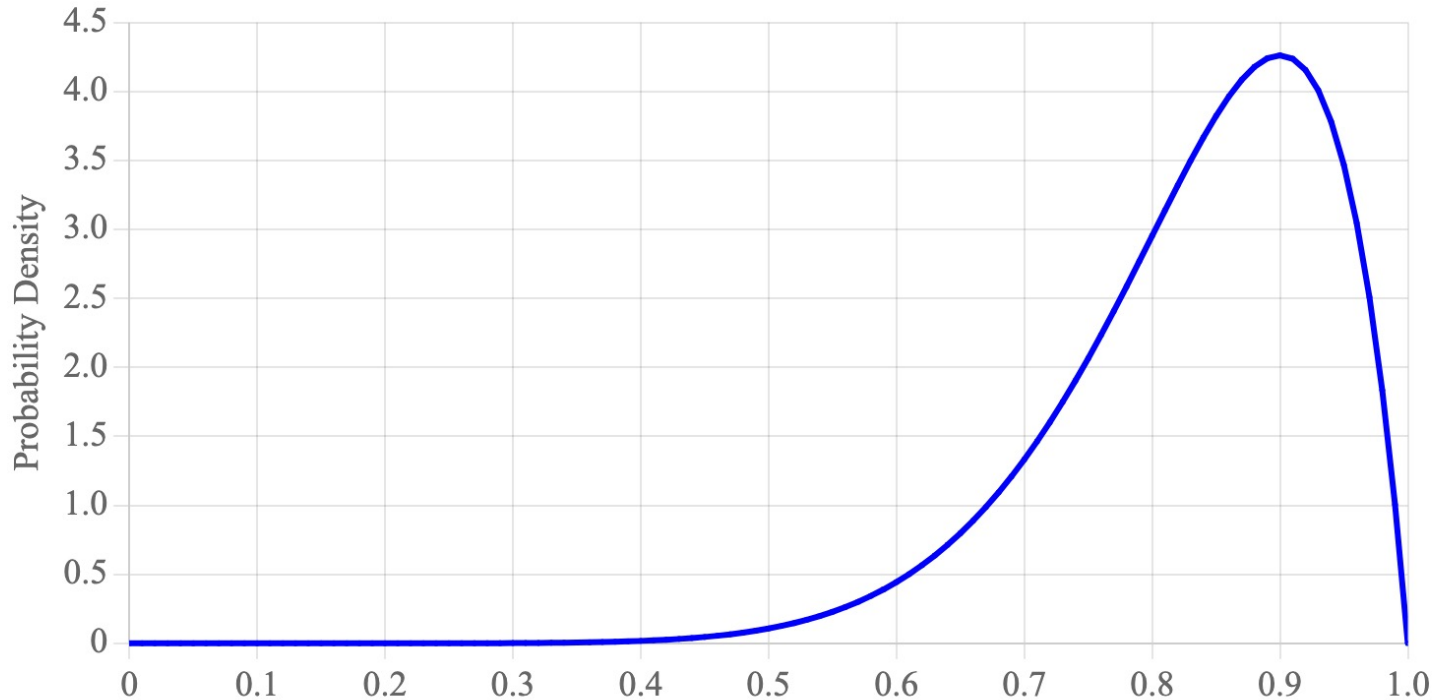


[Flu, Ugrad, Fever, Tired]

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:



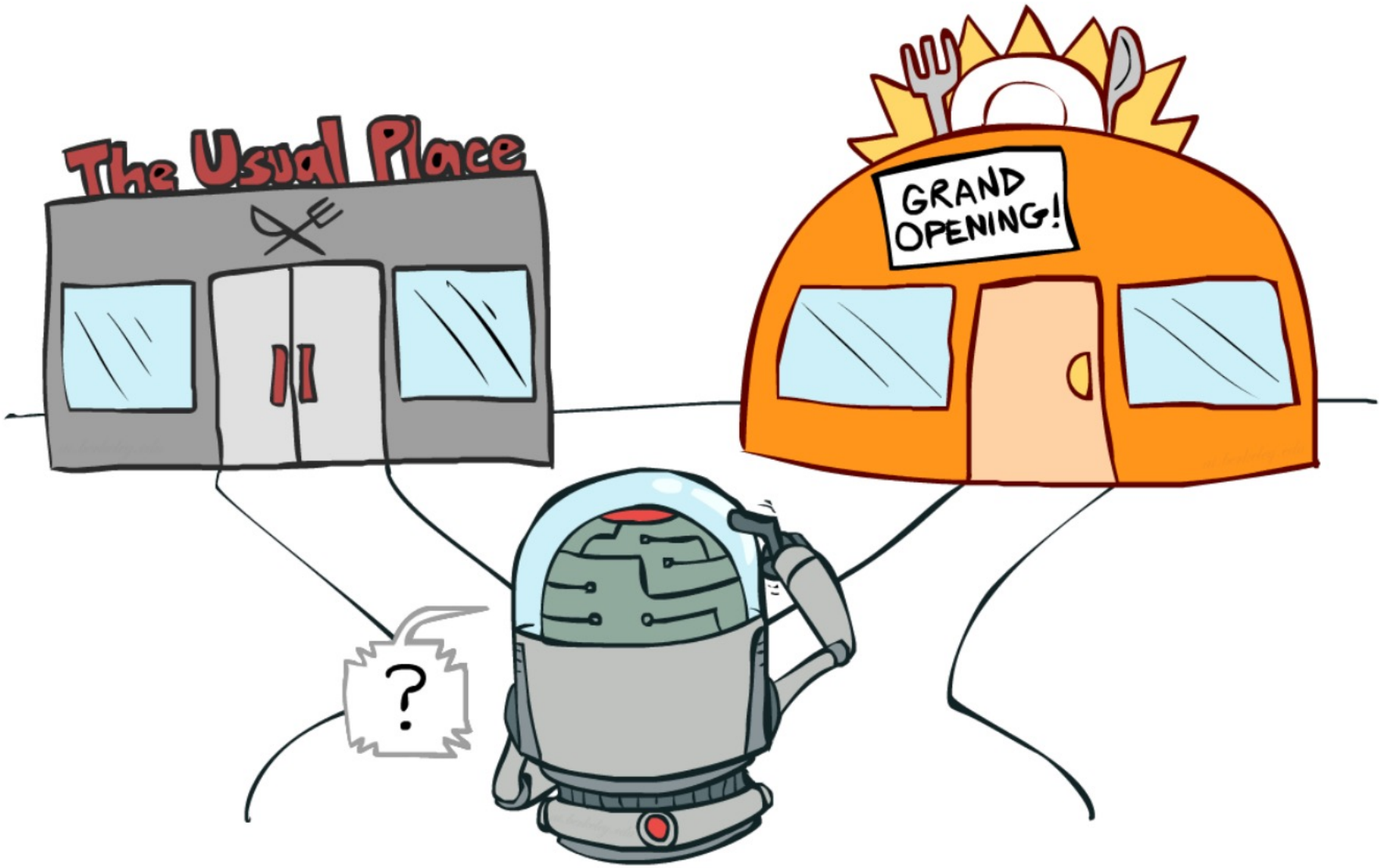
Lets Play!

Drug A

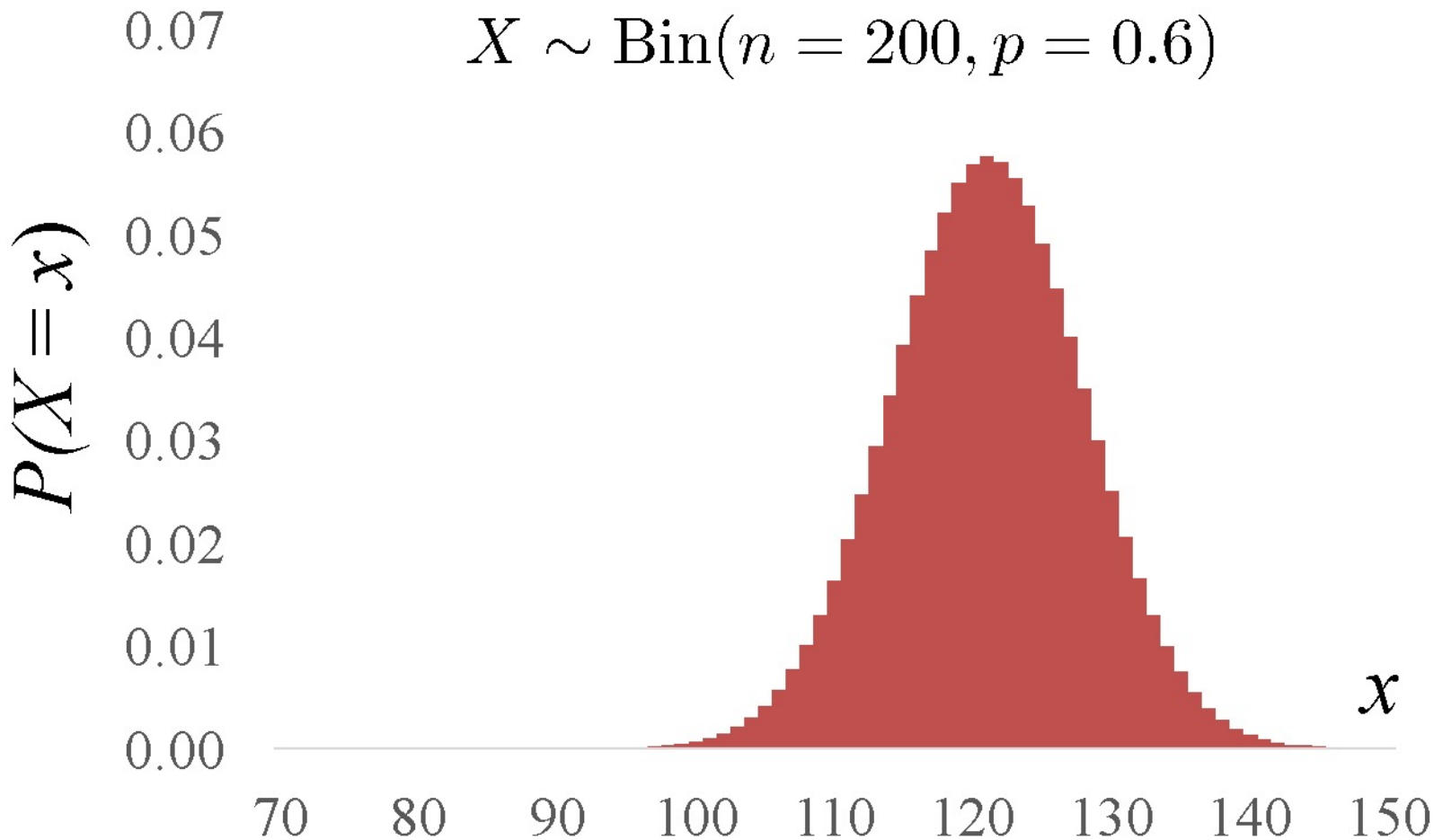
Drug B

Which one do you give to a patient?

Thompson Sampling

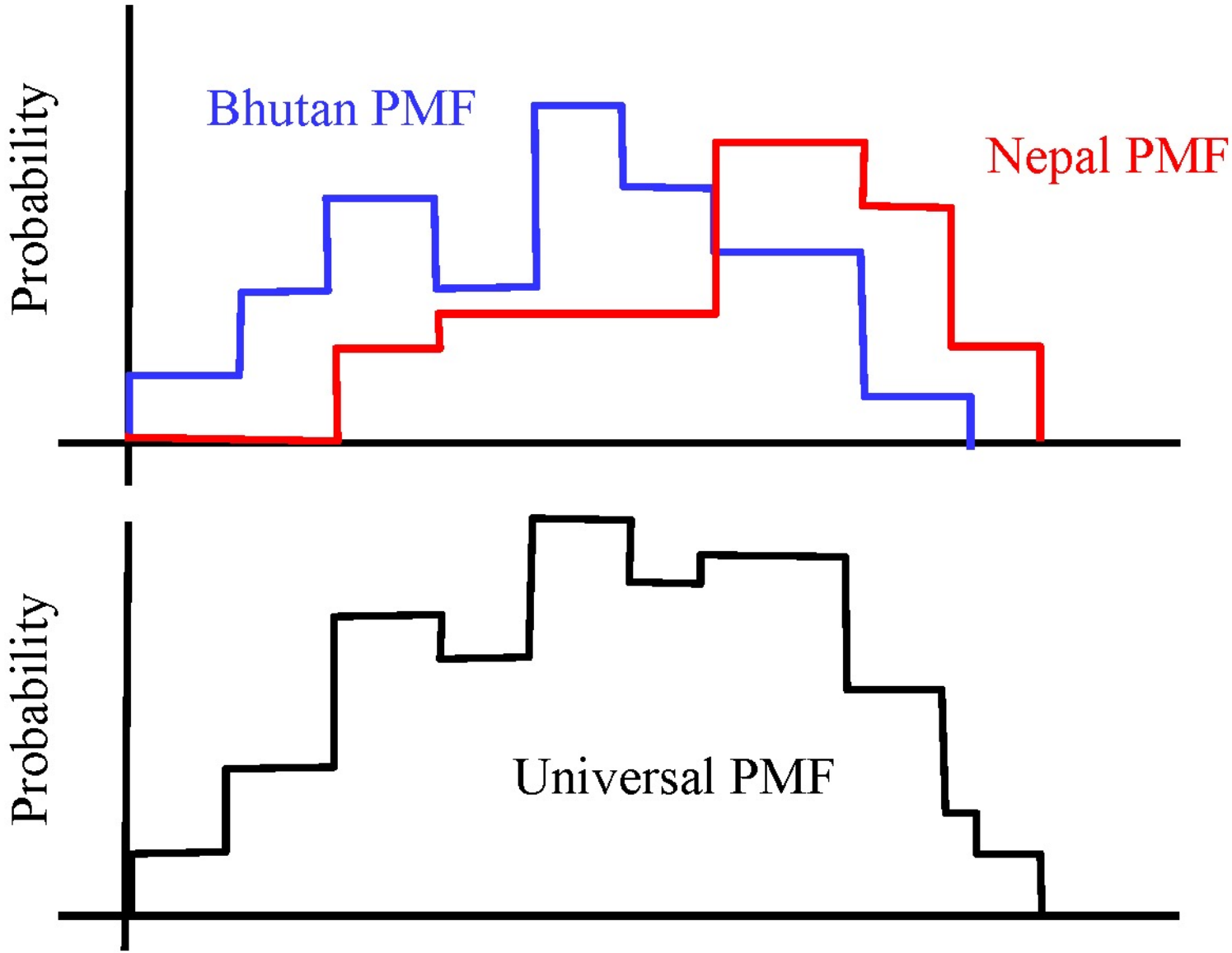


C.L.T. Explains This

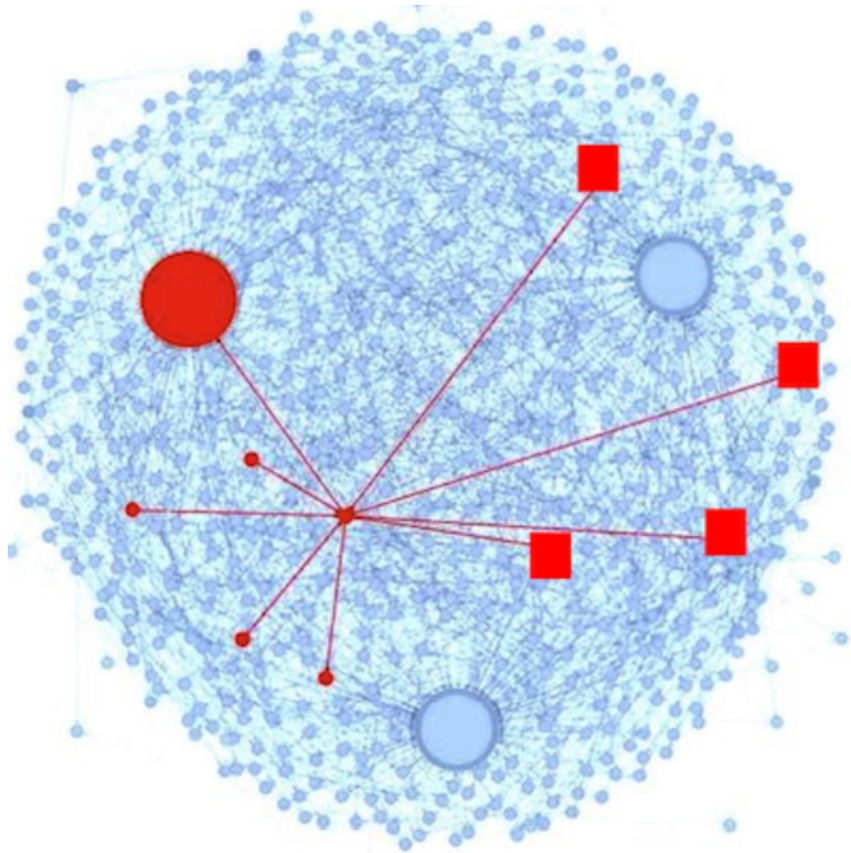


Bootstrap

Universal Sample



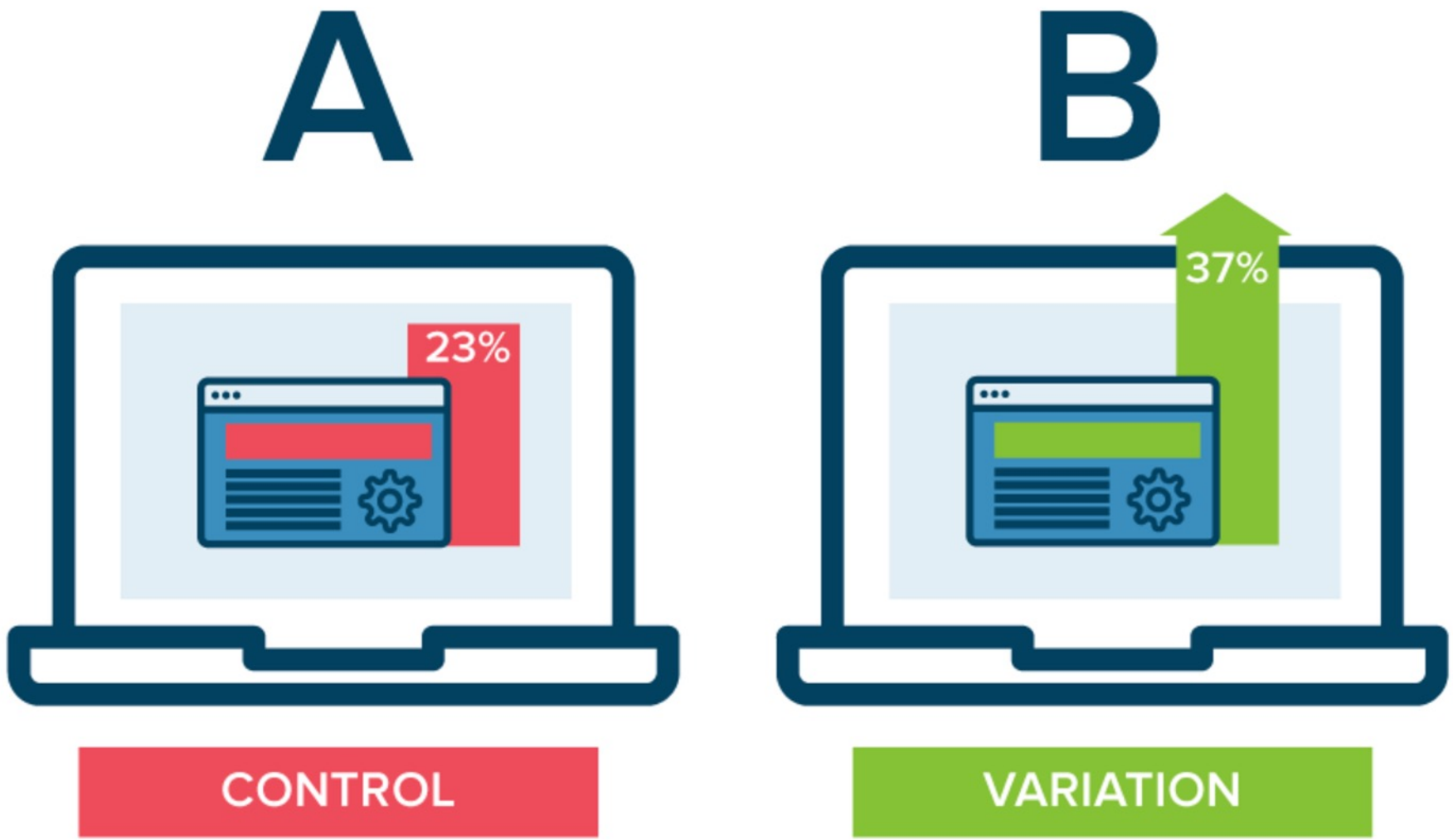
Peer Grading



Peer Grading on Coursera HCI.

31,067 peer grades for 3,607 students.

A/B Testing

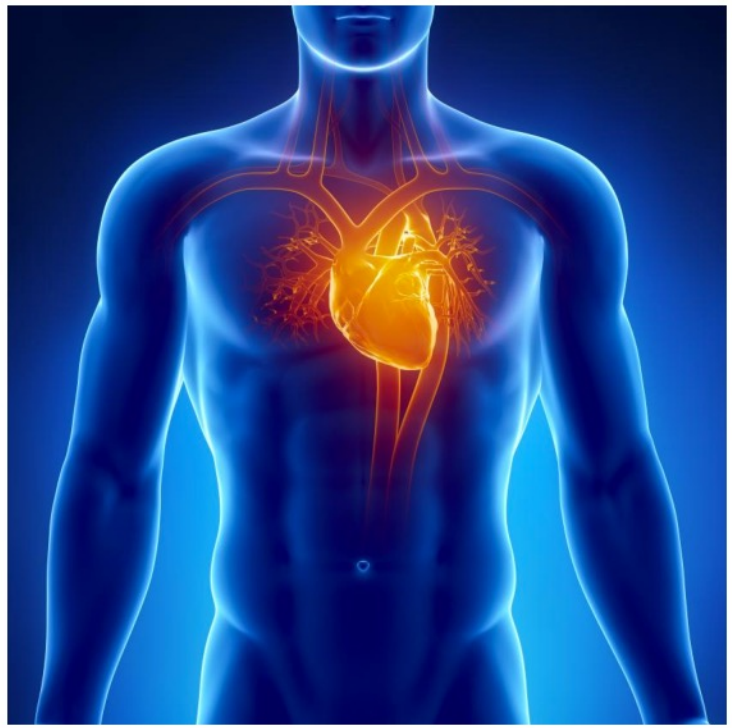


Machine Learning

MAP: Most Probable Parameter

Machine Learning

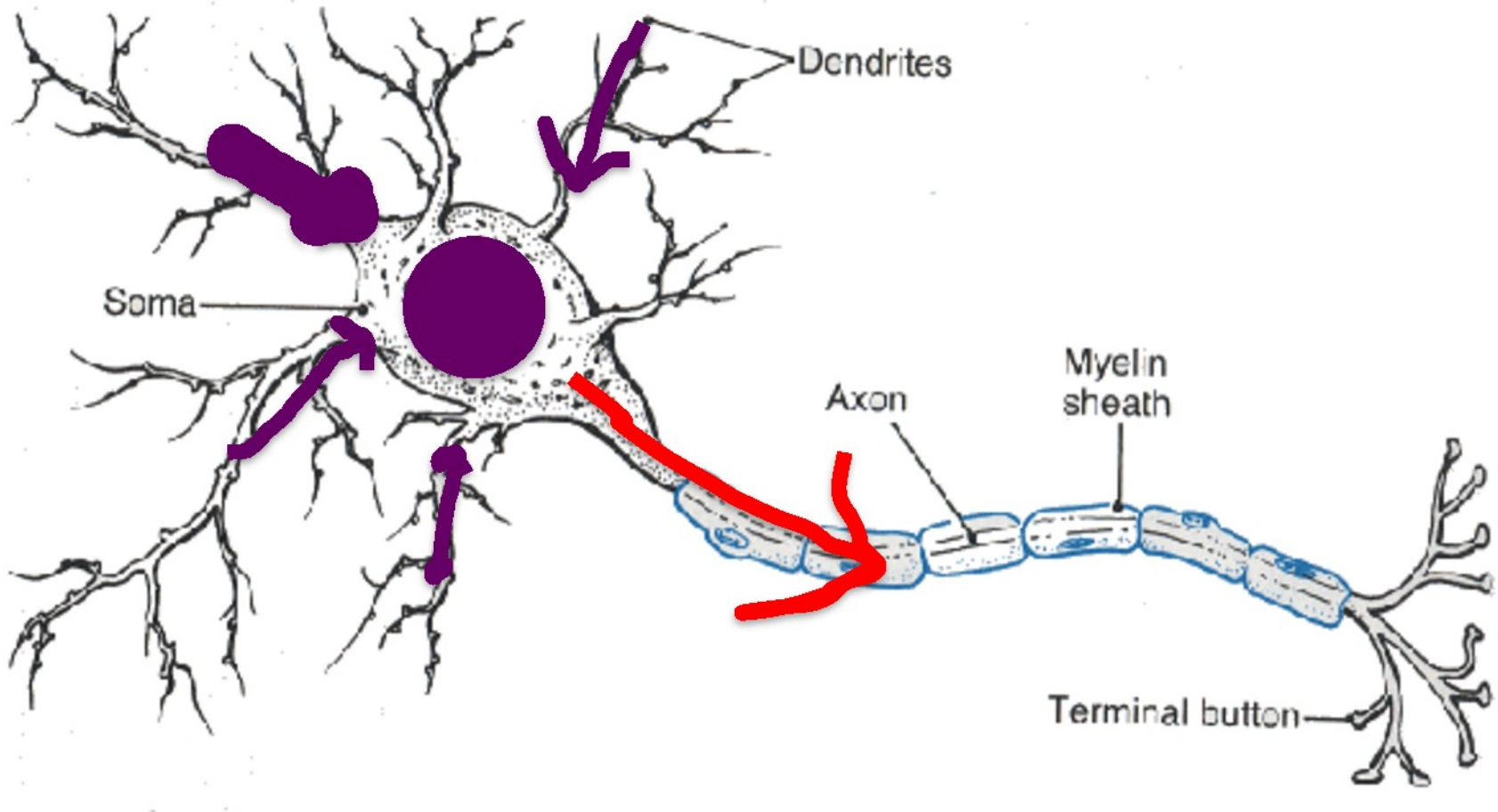
Heart



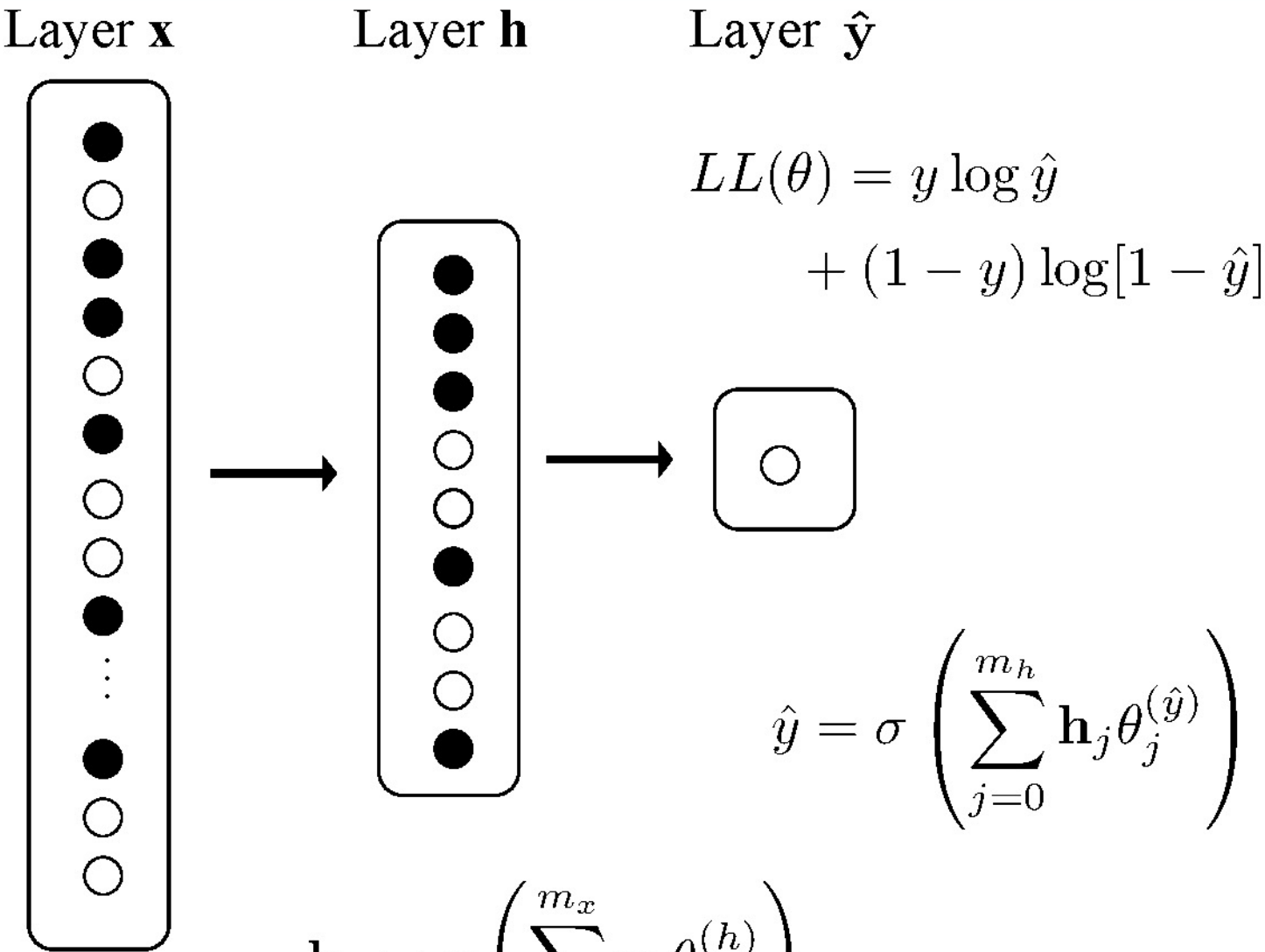
Ancestry

Netflix

Logistic Regression



Deep Learning



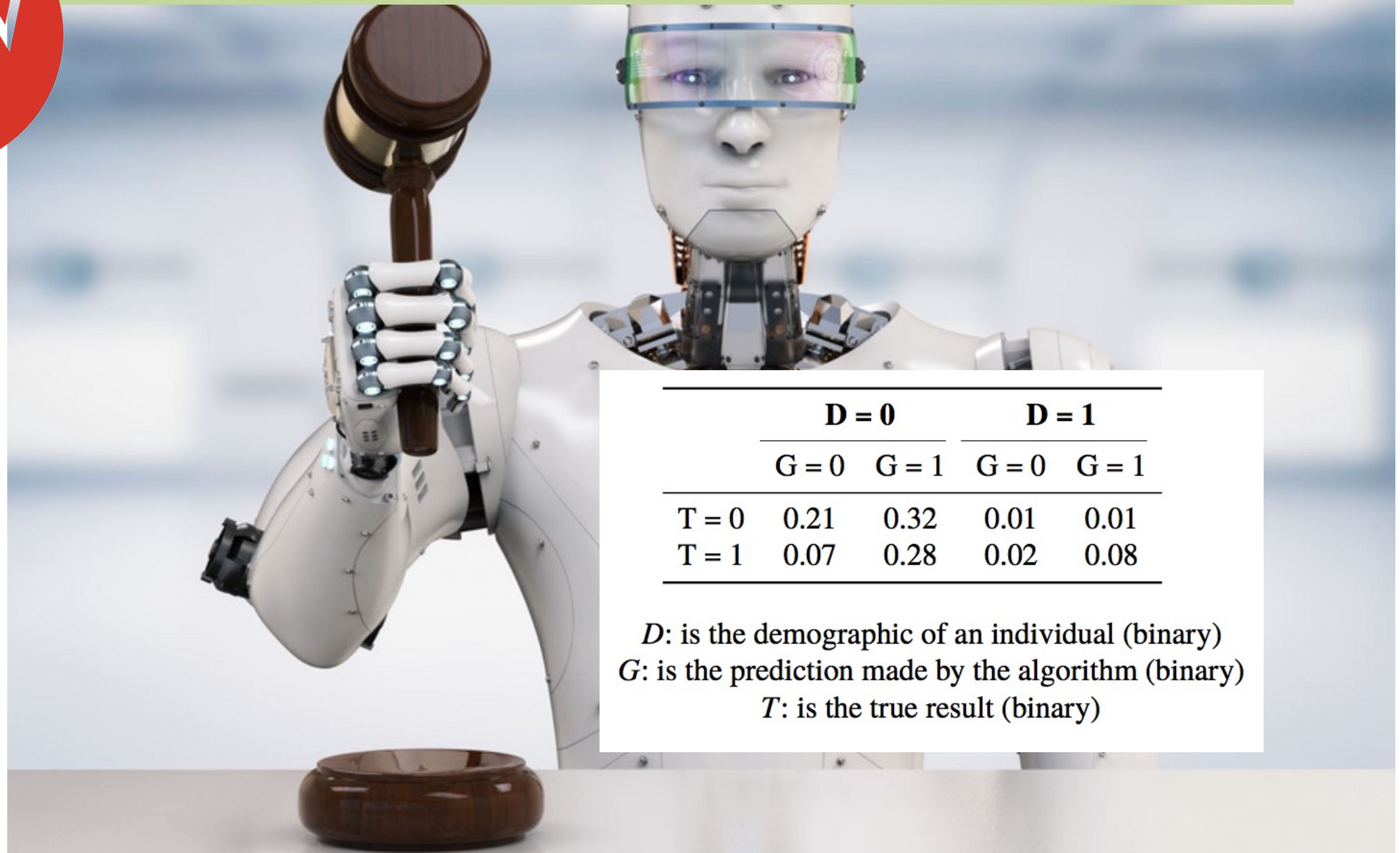
$$LL(\theta) = y \log \hat{y} + (1 - y) \log[1 - \hat{y}]$$

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

$$\mathbf{h}_j = \sigma\left(\sum_{i=0}^{m_x} \mathbf{x}_i \theta_{i,j}^{(h)}\right)$$

NEW

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)

By the numbers

~ 30 Major Keys

Naïve Bayes Assumption:

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

64 Course Reader Chapters

NEW

1 Personal Challenge

25% Participation Rate!

1 New Walker


thank you