

# AI Generated Classes

homeworks, exams, notes, books, feedback, grading, explanations, recitations, lectures, and curriculum

Iddo Drori

Associate Professor of Computer Science

Boston University (faculty, practice), MIT (visiting), Columbia University (adjunct)

# AI Generated

Homeworks  
2021

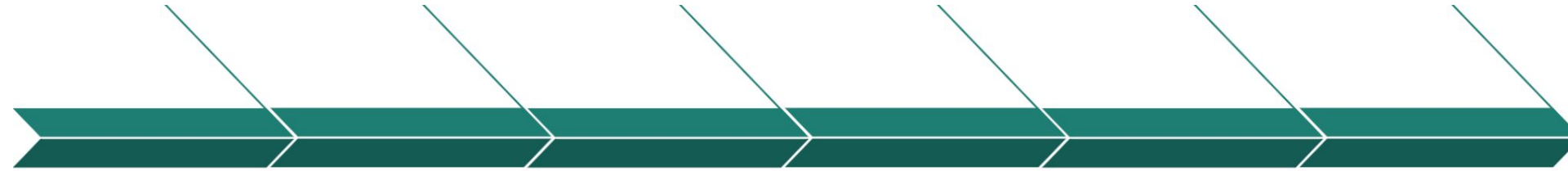
Exams  
Spring 2022

Curriculum  
2022

Books  
2023

Lectures  
Spring 2023

Classes  
Fall 2023



# Homeworks (2021)

RESEARCH ARTICLE | COMPUTER SCIENCES | 8



## A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level

Iddo Drori , Sarah Zhang, Reece Shuttleworth , Leonard Tang, Albert Lu , Elizabeth Ke, Kevin Liu, Linda Chen, Sunny Tran , Newman Cheng , Roman Wang , Nikhil Singh , Taylor L. Patti, Jayson Lynch, Avi Shporer , Nakul Verma, Eugene Wu, and Gilbert Strang  [14 Authors Info & Affiliations](#)

Edited by Jeffrey Ullman, Stanford University (Retired), Stanford, CA; received January 3, 2022; accepted June 13, 2022

August 2, 2022 | 119 (32) e2123433119 | <https://doi.org/10.1073/pnas.2123433119>

 24,859



**PNAS**

Vol. 119 | No. 32

Significance

Abstract

Data Availability

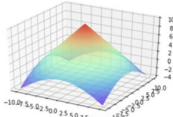
Supporting Information

References

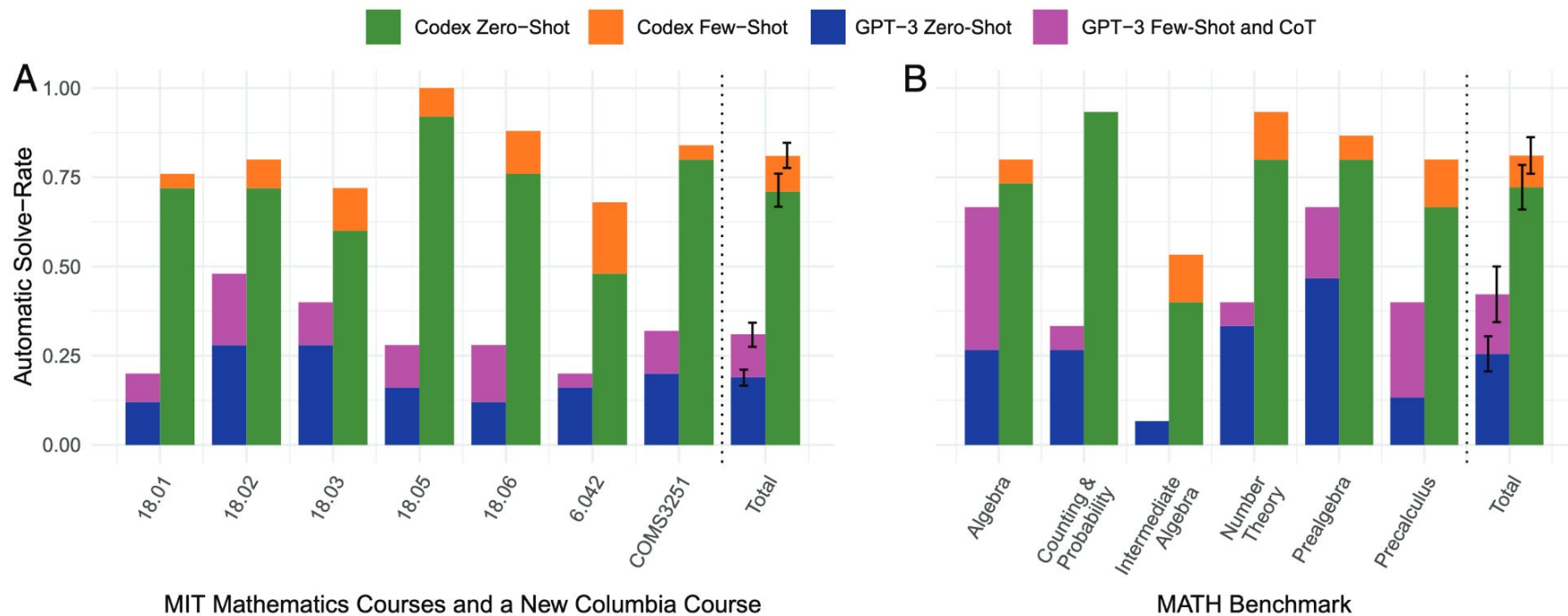
### Significance

We demonstrate that a neural network automatically solves, explains, and generates university-level problems from the largest Massachusetts Institute of Technology (MIT) mathematics courses at a human level. Our methods combine three innovations: 1) using recent neural networks pretrained on text and fine-tuned on code rather than pretrained on text; 2) few-shot learning synthesizing programs that correctly solve course problems automatically; and 3) a pipeline to solve questions, explain solutions, and generate new questions indistinguishable by students from course questions. Our work solves university-level mathematics courses and improves upon state-of-the-art, increasing automatic accuracy on randomly sampled questions on a benchmark by order of magnitude. Implications for higher education include roles of artificial intelligence (AI) in automated course evaluation and content generation.

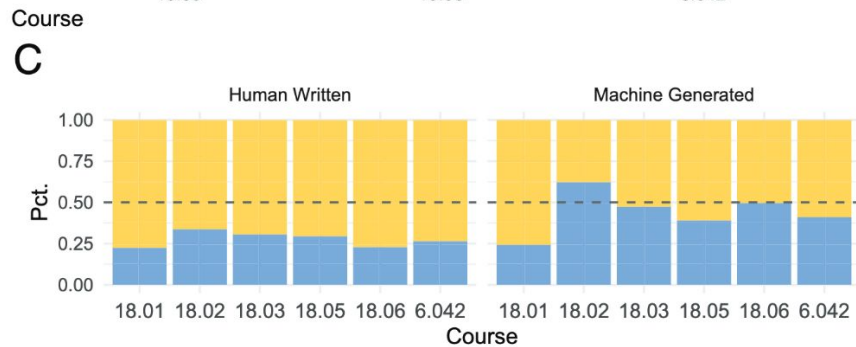
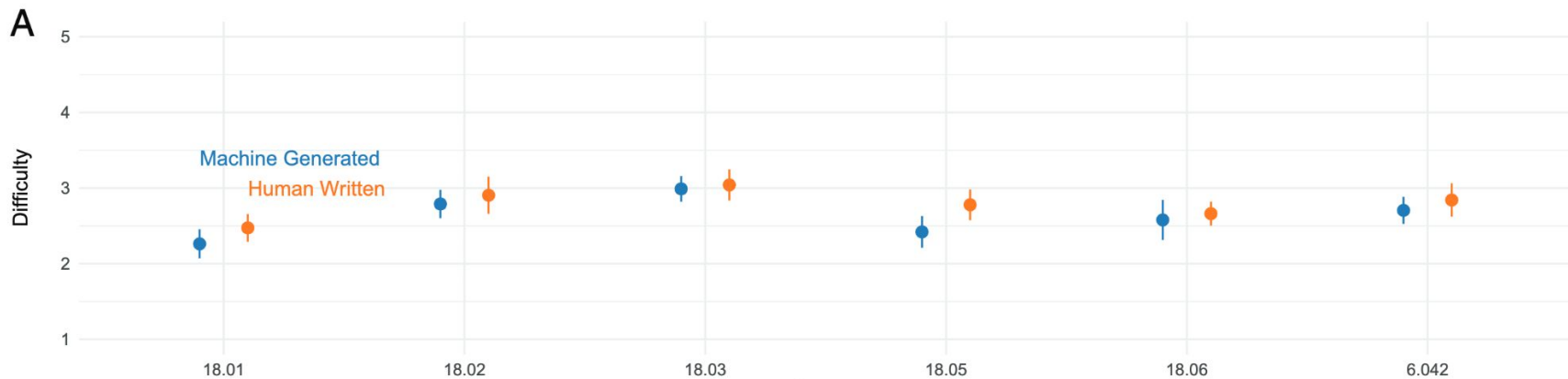
# Homeworks (2021)

ID	Course	Question	Solution
1	18.01 Single Variable Calculus	A bacteria population is 4,000 at time $t = 0$ and its rate of growth is $1,000 * 2^t$ bacteria per hour after $t$ h. What is the population after 1 h?	$4000 + \frac{1000}{\log(2)}$
2	18.02 Multivariable Calculus	Describe the graph of the function $f$ : $f(x,y) = 10 - \sqrt{x^2 + y^2}$ .	
3	18.03 Differential Equations	Find general solutions of the differential equations. If an initial condition is given, find the corresponding particular solution. Throughout, primes denote derivatives with respect to $x$ . $y' + y = 2$ , $y(0) = 0$ .	$y(x) = 2(1 - e^{-x})$
4	18.05 Introduction to Probability and Statistics	Calculate the probability of getting a three-of-a-kind poker hand.	0.021128
5	18.06 Linear Algebra	Find a combination $x_1 w_1 + x_2 w_2 + x_3 w_3$ that gives the zero vector with $x_1 = 1$ . $w_1$ is the vector (1;2;3). $w_2$ is the vector (4; 5; 6). $w_3$ is the vector (7; 8; 9).	$x_1 = 1, x_2 = -2, x_3 = 1$
6	6.042 Mathematics for Computer Science	Find a number $x \in \{0, 1, \dots, 112\}$ such that $11x \equiv 1 \pmod{113}$ .	72
7	COMS3251 Computational Linear Algebra	Given a $d$ -dimensional nonzero vector $v$ , compute the rank of the matrix $wv$ .	1
8	MATH Prealgebra	What is the greatest common factor of 84, 112, and 210?	14
9	MATH Algebra	Let $N, O$ be functions such that $N(x) = 2\sqrt{x}$ , and $O(x) = x^2$ . What is $N(O(N(O(N(O(3))))))$ ?	24
10	MATH Number Theory	How many four-digit numbers whose digits add up to 9 are divisible by 11?	0
11	MATH Counting and Probability	A standard six-sided fair die is rolled four times. The probability that the product of all four numbers rolled is a perfect square is $\frac{m}{n}$ , where $m$ and $n$ are relatively prime positive integers. Find $m + n$ .	187
12	MATH Intermediate Algebra	Given that $x^2 + y^2 = 14x + 6y + 6$ , find the largest possible value of $3x + 4y$ .	73
13	MATH Precalculus	If the six solutions of $x^6 = -64$ are written in the form $a + bi$ , where $a$ and $b$ are real, find the product of those solutions with $a > 0$ .	4

# Homeworks (2021)



# Homeworks (2021)



Rated as Not Appropriate

Rated as Appropriate

Rated as Human Written

Rated as Machine Generated

# Final Exams (2022)

Grading:	Human	Human	Human	Machine
Answers:	Human	Human	Machine	Machine
Questions:	All	Non-Image	Non-Image	Non-Image Non-Open
MIT Spring 2021	75.84	80.77	62.09	64
MIT Fall 2021	74.38	60.88	58.94	51.33
MIT Spring 2022	69.07	70.82	68.86	73.53
Mean	73.10	70.82	63.29	62.95

Drori et al, 2022

# Final Exams (2022)

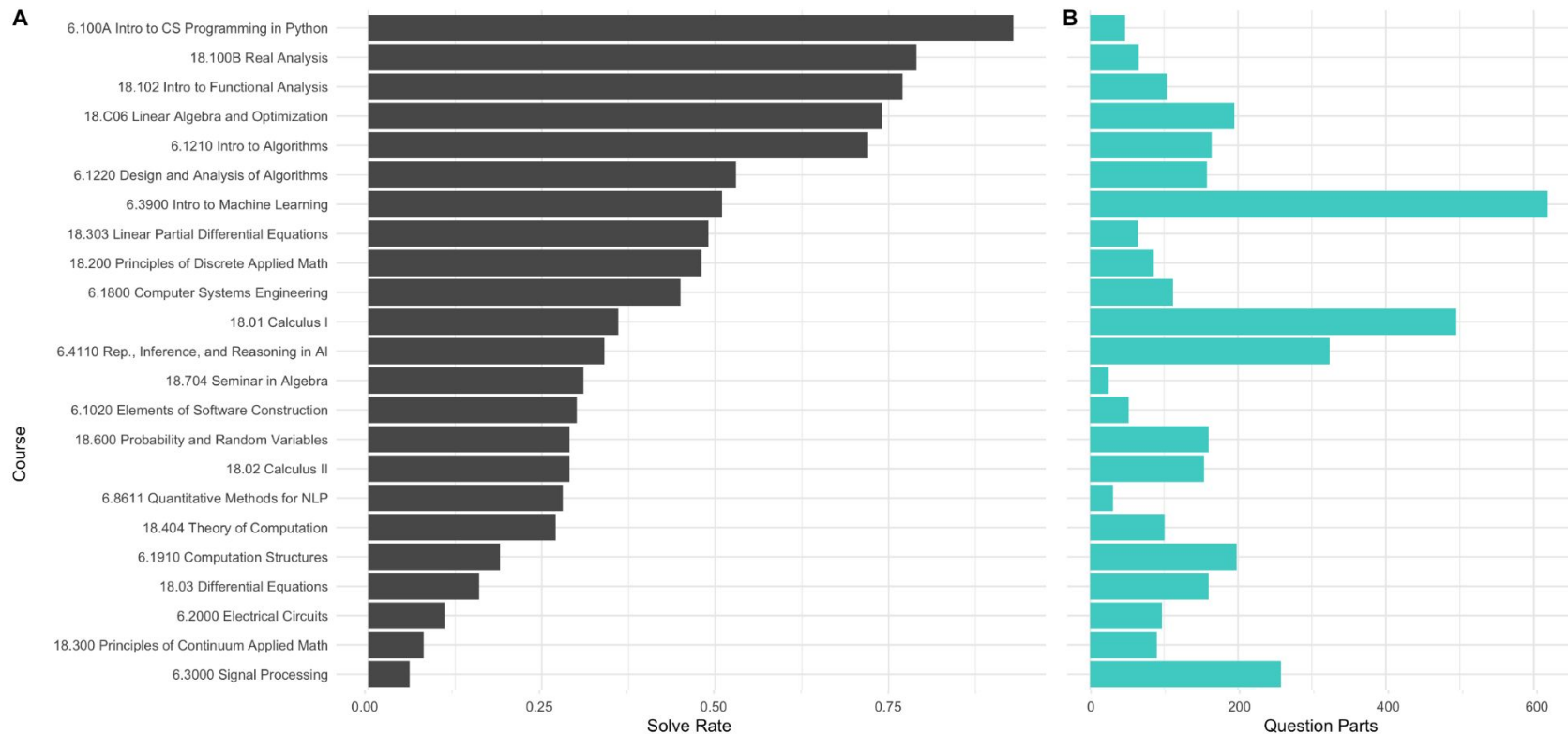
Semester	GPT-3 2 ZS	GPT-3 3 ZS	ChatGPT ZS	GPT-3 FS	GPT-3 ZS CoT	GPT-3 FS CoT	Codex ZS	Codex FS	OPT ZS
MIT Fall 2017	38.21	<b>50.00</b>	48.93	NA	22.86	NA	21.43	NA	NA
MIT Spring 2018	44.35	60.48	50.00	60.48	38.71	<b>70.97</b>	32.26	67.74	33.33
MIT Fall 2018	51.99	62.96	<b>72.50</b>	52.18	61.63	64.17	49.78	54.00	47.54
MIT Spring 2019	43.45	55.65	<b>62.14</b>	54.23	41.07	58.81	15.54	41.55	34.64
MIT Fall 2019	54.92	58.61	58.20	<b>77.05</b>	29.92	58.20	26.23	61.48	NA
MIT Spring 2021	44.33	48.31	51.26	55.81	53.45	60.21	33.62	<b>62.09</b>	33.77
MIT Fall 2021	58.94	61.53	47.24	<b>69.44</b>	50.35	54.90	18.11	42.00	24.44
MIT Spring 2022	42.78	55.29	55.07	<b>68.86</b>	32.03	53.48	51.01	65.46	60.71
Harvard Spring 2015	<b>85.71</b>	64.29	78.57	NA	<b>85.71</b>	NA	50.00	NA	21.43
Harvard Spring 2021	47.73	77.27	72.73	<b>86.36</b>	47.73	81.82	43.18	<b>86.36</b>	45.45
Cornell Spring 2017	78.91	59.90	79.03	NA	<b>80.86</b>	NA	51.30	NA	21.88
Cornell Fall 2018	36.45	57.01	<b>61.21</b>	53.27	44.39	<b>61.21</b>	42.52	56.07	28.97



# Final Exams (2022)

Topic	GPT-3 2 ZS	GPT-3 3 ZS	ChatGPT ZS	GPT-3 FS	GPT-3 ZS CoT	GPT-3 FS CoT	Codex ZS	Codex FS	OPT ZS
Regression	31.71	<b>56.67</b>	50.56	50.00	25.61	40.85	40.24	50.00	50.00
Classifiers	38.18	47.12	52.81	46.21	26.28	42.35	18.88	<b>53.74</b>	50.00
Logistic Reg.	50.00	60.71	67.86	60.00	<b>77.50</b>	<b>77.50</b>	55.00	70.00	16.67
Features	58.65	71.92	76.15	75.96	53.85	77.31	68.85	<b>81.54</b>	10.00
Loss Functions	NA	NA	NA	NA	NA	NA	NA	NA	NA
Neural Networks	48.34	60.82	67.71	60.23	44.54	<b>68.42</b>	37.82	63.45	27.27
CNNs	37.50	59.59	<b>62.05</b>	53.58	28.36	47.81	13.38	36.77	23.83
MDPs	49.19	<b>73.28</b>	47.33	52.01	46.03	54.23	24.38	38.03	28.32
RNNs	61.46	33.33	45.83	<b>71.88</b>	57.29	66.14	12.50	40.63	39.28
RL	36.09	<b>65.24</b>	55.59	42.99	36.67	50.11	28.79	45.11	24.28
Clustering	<b>100.00</b>	50.00	90.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	50.00	50.00	63.33
Decision Trees	54.70	<b>74.78</b>	69.08	71.80	32.48	51.28	46.15	54.70	55.00
Model Selection	82.93	71.12	82.10	83.74	72.76	<b>95.12</b>	67.48	69.92	21.95
Ensemble Methods	27.89	41.35	<b>69.23</b>	50.00	22.12	66.35	32.69	50.00	13.46
Bayesian Networks	<b>100.00</b>	0.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	0.00	0.00	<b>100.00</b>
HMMs	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	50.00	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>	<b>100.00</b>
Optimization	55.00	77.50	<b>87.50</b>	60.00	35.00	55.00	17.50	70.00	20.00

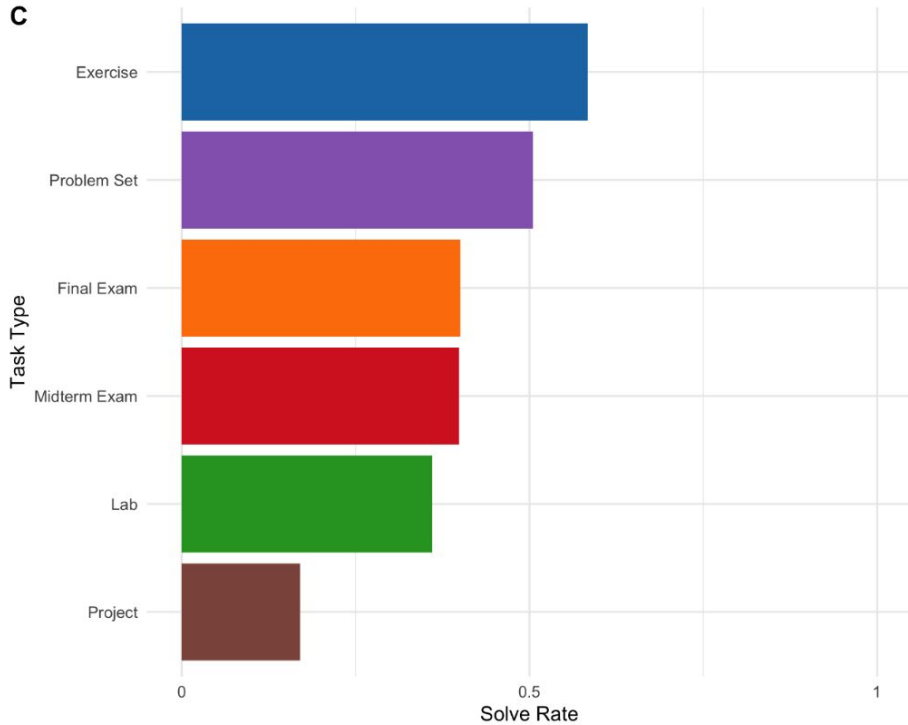
# Entire Classes (2023)



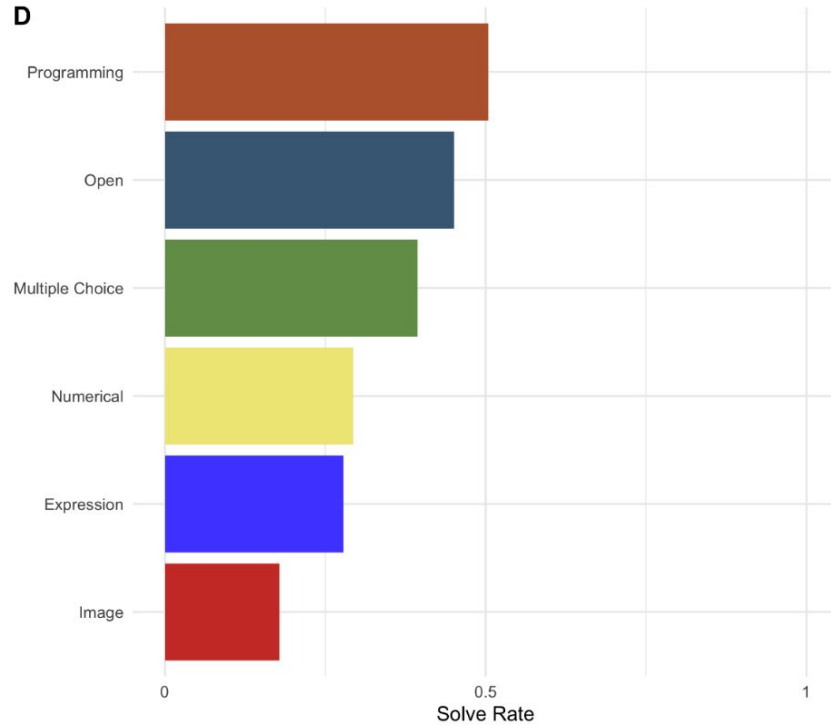
ChatGPT Solve Rate by MIT Course & # of Questions

# Entire Classes (2023)

**C**



**D**



ChatGPT Solve Rate by Task Type and Answer Type

# Math and EECS Classes (2023)

ChatGPT as a baseline by itself passes “only”  $\frac{1}{3}$  of the MIT curriculum

Our latest method automatically graduates from MIT in Mathematics and EECS across all majors with a grade A

Our novel contributions: dataset, program synthesis, self error correction, mathematical proofs, automated facial gestures.

# Homeworks and Exams embracing ChatGPT (Spring 2023)

Following are 10 questions and answers by ChatGPT. For each question identify whether the answer is correct or not. If the answer is correct, explain why. If the answer is incorrect, write the correct answer and explain.

Q2

1 Point

What is the difference between a learning algorithm and a hypothesis? Write down one possible hypothesis for a linear regression problem in which the input dimension is  $d = 3$

ChatGPT Answer: "A learning algorithm is a procedure or set of instructions used to find an approximate solution to a problem, such as finding a hypothesis that fits a given dataset. A hypothesis, on the other hand, is a statement or function that describes the relationship between inputs and outputs in a problem.

One possible hypothesis for a linear regression problem with input dimension  $d = 3$  is a linear function of the form  $h(x) = w_1x_1 + w_2x_2 + w_3x_3 + b$ , where  $w_1, w_2, w_3$  and  $b$  are parameters to be learned from the data."

Correct

Incorrect

Explanation:

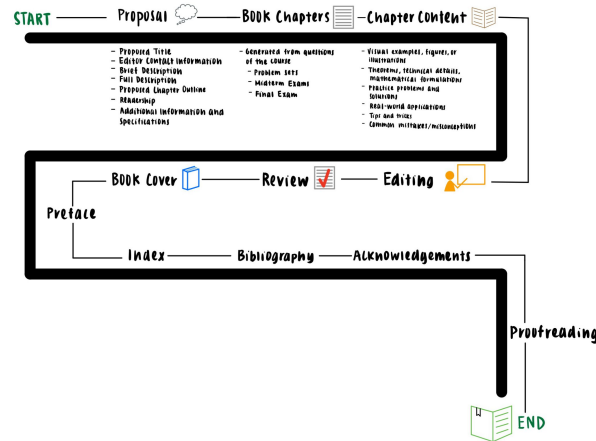
# Publishing Books (2023)

From years to months

Regular review process by established press

Correct, complete, high quality, original books

Non-trivial process: from proposal to press



Id	Course Name	Course #
1	Intro to Computer Science & Programming in Python	6.100A
2	Fundamentals of Programming	6.1010
3	Elements of Software Construction	6.1020
4	Introduction to Algorithms	6.1210
5	Design and Analysis of Algorithms	6.1220
6	Computer Systems Engineering	6.1800
7	Computation Structures	6.1910
8	Electrical Circuits: Modeling & Design of Physical Sys.	6.2000
9	Electromagnetic Waves and Applications	6.2300
10	Signal Processing	6.3000
11	Signals, Systems & Inference	6.3010
12	Introduction to Machine Learning	6.3900
13	Representation, Inference, and Reasoning in AI	6.4110
14	Computational Cognitive Science	6.4120
15	Quantitative Methods for NLP	6.8611
16	Calculus 1	18.01
17	Calculus 2	18.02
18	Differential Equations	18.03
19	Linear Algebra and Optimization	18.C06
20	Real Analysis	18.100A
21	Analysis and Manifolds	18.101
22	Seminar in Analysis	18.104
23	Principles of Discrete Applied Mathematics	18.200
24	Principles of Continuum Applied Mathematics	18.300
25	Introduction to Numerical Analysis	18.330
26	Theory of Computation	18.404
27	Probability and Random Variables	18.600
28	Algebra I	18.701
29	Algebra II	18.702
30	Introduction to Topology	18.901

# Defining New Roles and Coining Terms (2023)

Prompter, editor, publisher

Booksets: validation, testing

“Here is a 100 prompt book ready for editing by the class professor”

“Here is a meta prompt for the cover”

“This book is prompted by X, edited by Y, and published by Z”

# Example (2023)

Prompted by class teaching assistant

Edited by class instructor

“MIT quality” book

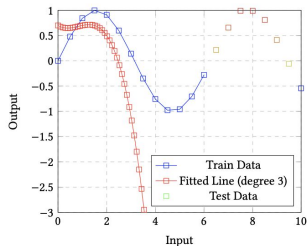
Published by Cambridge University Press

<b>1</b>	<b>Introduction to Machine Learning</b>	<b>17</b>
1.1	Overview of Machine Learning . . . . .	17
1.1.1	Types of Machine Learning . . . . .	17
1.1.2	Applications of Machine Learning . . . . .	18
1.2	Key concepts in Machine Learning . . . . .	19
1.2.1	Feature Representation . . . . .	19
1.2.2	Model Selection . . . . .	19
1.2.3	Overfitting . . . . .	19
1.2.4	Bias-Variance Tradeoff . . . . .	19
1.2.5	Evaluation Metrics . . . . .	19
1.3	Types of data in Machine Learning . . . . .	20
1.3.1	Structured Data . . . . .	20
1.3.2	Unstructured Data . . . . .	20
1.3.3	Time Series Data . . . . .	20
1.3.4	Sequential Data . . . . .	20
<b>2</b>	<b>Regression</b>	<b>23</b>
2.0.1	Types of Regressions . . . . .	23
2.0.2	Use Cases of Regression . . . . .	24
2.0.3	Assumptions of Regression . . . . .	24
2.1	Formulating the Problem . . . . .	25
2.2	Linear Regression . . . . .	26
2.2.1	Analytical Solution . . . . .	27
2.2.2	Gradient Descent . . . . .	29
2.3	Regularization . . . . .	29
2.3.1	Ridge Regression . . . . .	32
2.3.2	Lasso Regression . . . . .	33
2.3.3	Elastic Net . . . . .	35



# Example Lecture Notes (ML class, Spring 2023)

31



By adding a regularization term to the cost function, we can improve the generalization performance of the model. The green line represents the model fit with regularization, and as we can see, it fits both the training and test data well, resulting in a lower test error.

To combat overfitting, regularization is often used in regression. Regularization is a technique that adds a penalty term to the cost function to discourage the model from fitting the noise in the data and instead encourage simpler, more general models.

There are two main types of regularization used in regression: L1 regularization, also known as Lasso regularization, and L2 regularization, also known as Ridge regularization.

**L1 REGULARIZATION** adds a penalty term to the cost function proportional to the absolute value of the coefficients. The L1 regularization term is defined as:

$$\lambda \sum_{i=1}^n |\beta_i|$$

Where  $\lambda$  is the regularization parameter and  $\beta_i$  are the coefficients. The L1 regularization term encourages the model to have sparse solutions, meaning that some coefficients will be exactly equal to zero.

**L2 REGULARIZATION** adds a penalty term to the cost function proportional to the square of the coefficients. The L2 regularization term is defined as:

$$\lambda \sum_{i=1}^n \beta_i^2$$

Where  $\lambda$  is the regularization parameter and  $\beta_i$  are the coefficients. The L2 regularization term encourages the model to have small, non-zero coefficients.

32

When the L2 regularization term is added to the cost function, the resulting cost function is called Ridge Regression, and when L1 regularization term is added it is called Lasso Regression.

Regularization can be added to the cost function by simply adding the regularization term to the mean square error. For example, in Ridge Regression the cost function becomes:

$$J(\beta) = \frac{1}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)})^2 + \lambda \sum_{i=1}^n \beta_i^2$$

Where  $\lambda$  is the regularization parameter and  $\beta_i$  are the coefficients.

The regularization term acts as a constraint on the coefficients, preventing them from becoming too large and resulting in overfitting. The regularization parameter  $\lambda$  controls the strength of the regularization term and can be tuned to achieve the best trade-off between fitting the data and preventing overfitting.

## 2.3.1 Ridge Regression

Ridge Regression is a variation of linear regression that uses L2 regularization to prevent overfitting. The cost function for Ridge Regression is defined as:

$$J(\beta) = \frac{1}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)})^2 + \lambda \sum_{i=1}^n \beta_i^2$$

Where  $\lambda$  is the regularization parameter and  $\beta_i$  are the coefficients.

### DERIVING THE ANALYTICAL SOLUTION:

To derive the analytical solution for Ridge Regression, we can take the derivative of the cost function with respect to each  $\beta_i$  and set it equal to zero. By doing this we can find the normal equation as

$$\frac{\partial J(\beta)}{\partial \beta_i} = -\frac{2}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)}) x_i^{(i)} + 2\lambda \beta_i = 0$$

for  $i = 0$ ,

$$\frac{\partial J(\beta)}{\partial \beta_0} = -\frac{2}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)}) = 0$$

On solving above equations we can get the following analytical solution for Ridge Regression.

$$\beta = (X^T X + \lambda I)^{-1} X^T Y$$

Where  $X$  is the design matrix,  $Y$  is the output vector, and  $I$  is the Identity matrix.

To find the values of the coefficients that minimize the cost function, the gradient descent algorithm is used. The gradient descent algorithm is an

33

iterative algorithm that starts with initial values for the coefficients and updates them in the direction of the negative gradient of the cost function. The update rule for the coefficients in Ridge Regression is:

$$\beta_i = \beta_i - \alpha \frac{\partial J(\beta)}{\partial \beta_i}$$

Where  $\alpha$  is the learning rate. The gradient of the cost function with respect to  $\beta_i$  is:

$$\frac{\partial J(\beta)}{\partial \beta_i} = \frac{2}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)}) x_i^{(i)} - 2\lambda \beta_i$$

so, the update rule becomes:

$$\beta_i = \beta_i - \alpha \frac{2}{m} \sum_{i=1}^m (y^{(i)} - \beta_0 - \sum_{j=1}^n \beta_j x_j^{(i)}) x_i^{(i)} - 2\lambda \beta_i$$

In Ridge Regression, the L2 regularization term shrinks the coefficients towards zero, but it doesn't make them zero. The value of the regularization parameter,  $\lambda$ , determines the strength of the regularization. A high value of  $\lambda$  will result in small coefficients (close to zero) and a low value of  $\lambda$  will result in large coefficients.

It's important to note that Ridge Regression is a technique to prevent overfitting in the model by adding a bias term to the cost function, which causes the model to prefer solutions with small coefficients. As a result, the values of the coefficients become close to zero, but not zero, compared to Lasso Regression where the values of the coefficients become zero. This property of Ridge Regression makes it useful when we have a large number of features and we want to keep all of them in the model.

Additionally, Ridge Regression also helps to address the issue of multicollinearity, which occurs when there is a high correlation between independent variables. In this case, the coefficients can become unstable, leading to large variances in the model's predictions. Regularization helps to reduce the variance by shrinking the coefficients towards zero.

Another advantage of Ridge Regression is that it is computationally efficient, as it only requires the inversion of a matrix, which can be done using linear algebra libraries such as NumPy or scikit-learn. This makes it a popular choice for large datasets.

However, one limitation of Ridge Regression is that it does not perform feature selection, meaning it does not set any coefficients to zero. This can be an issue if we have a large number of features and some of them are not relevant to the model. In such cases, Lasso Regression or Elastic Net may be more suitable.

## 2.3.2 Lasso Regression

The Lasso regression is a form of linear regression that uses a regularization term known as L1 regularization. The objective function in Lasso regression

# Machine-Generated Questions, Answers, Explanations

Indistinguishable from human-written questions

Appropriate for class, controlled difficulty level, engaging

Research focus on correctness, completeness, quality, originality: quantified

Used in class homeworks and finals

# Photorealistic Avatar Lecture (ML class Spring 2023)

## Hypothesis Class

### What do we want?

Given new unseen  $x^{(i)}$ , predict  $y^{(i)}$ .

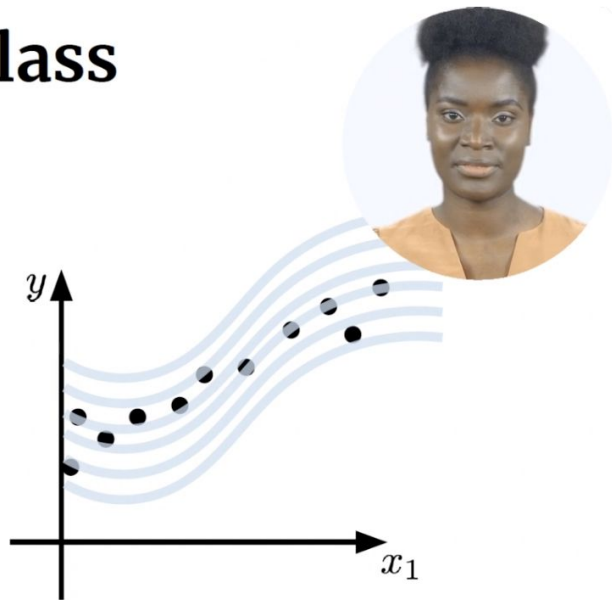
**Hypothesis**  $h : \mathbb{R}^d \rightarrow \mathbb{R}$

- hypothesis is a function

**Hypothesis class**  $\mathcal{H}$ : set of  $h$

- is parameterized  $h(x; \theta)$
- is a “family” of functions

**One important skill for a machine learning practitioner is to pick the right hypothesis class for a given problem.**



# Photorealistic Avatar Lecture (Spring 2023)

## Linear Regression Hypothesis Class



A linear regression hypothesis class when  $d=1$ :

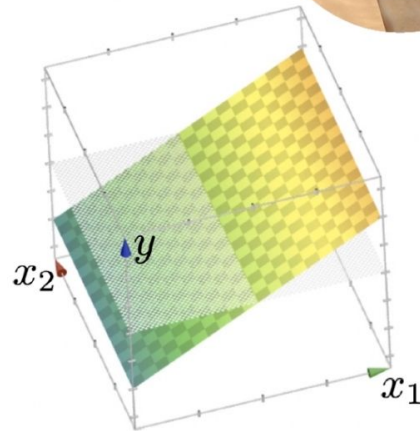
$$h(x; \theta, \theta_0) = \theta x + \theta_0$$

A linear reg. hypothesis class when  $d \geq 1$ :

$$\begin{aligned} h(x; \theta, \theta_0) &= \theta_1 x_1 + \cdots + \theta_d x_d + \theta_0 \\ &= \theta^\top x + \theta_0 \\ &\quad \text{1x2, 2x1} \end{aligned}$$

Or

$$\begin{aligned} h(x; \theta) &= \theta_1 x_1 + \cdots + \theta_d x_d + (\theta_0)(1) \\ &= \theta^\top x \\ &\quad \text{1x3, 3x1} \end{aligned}$$



Hypothesis is a  
"hyperplane"

# Photorealistic Avatar Lecture (Spring 2023)

## How Good is a Given Hypothesis?

**Generalization:** A hypothesis should provide good prediction for future data.

**Loss:** How good is a given hypothesis at one point?

- Example : squared loss

**g:** guess,  
**a:** actual

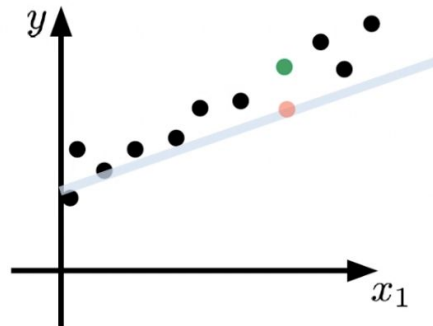
$$L(g, a) = (g - a)^2$$

- Example: asymmetric loss

$$L(g, a) = \begin{cases} (g - a)^2 & \text{if } g > a \\ 2(g - a)^2 & \text{if } g \leq a \end{cases}$$

**Training Error:**  $\mathcal{E}_n(h) = \frac{1}{n} \sum_{i=1}^n L(h(x^{(i)}), y^{(i)})$

**Test Error** ( $n'$  new points):  $\mathcal{E}(h) = \frac{1}{n'} \sum_{i=n+1}^{n+n'} L(h(x^{(i)}), y^{(i)})$



# Photorealistic Avatars

Reconstruction

Photorealistic

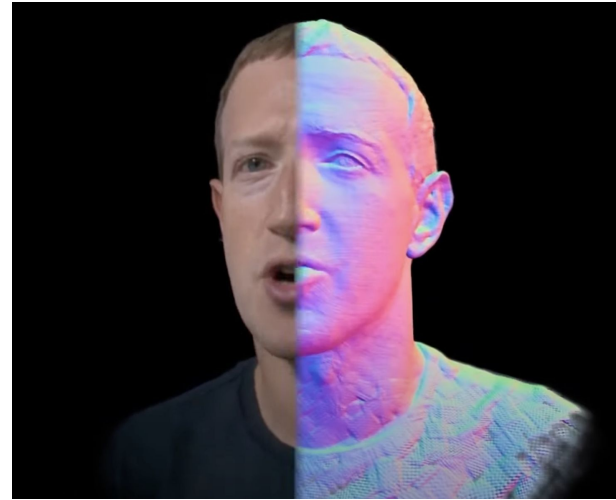


Figure source: Meta avatar codec

# Photorealistic Avatar

Uncanny valley

Example: 40% Terence Tao, 40% Esther Perel, 20% you



Figure source: Masterclass

# Photorealistic Avatar Lecture (Fall 2023)

Automatic text, speech, gestures

Pass uncanny valley of photo realism

Real-time rendering for live interaction with students

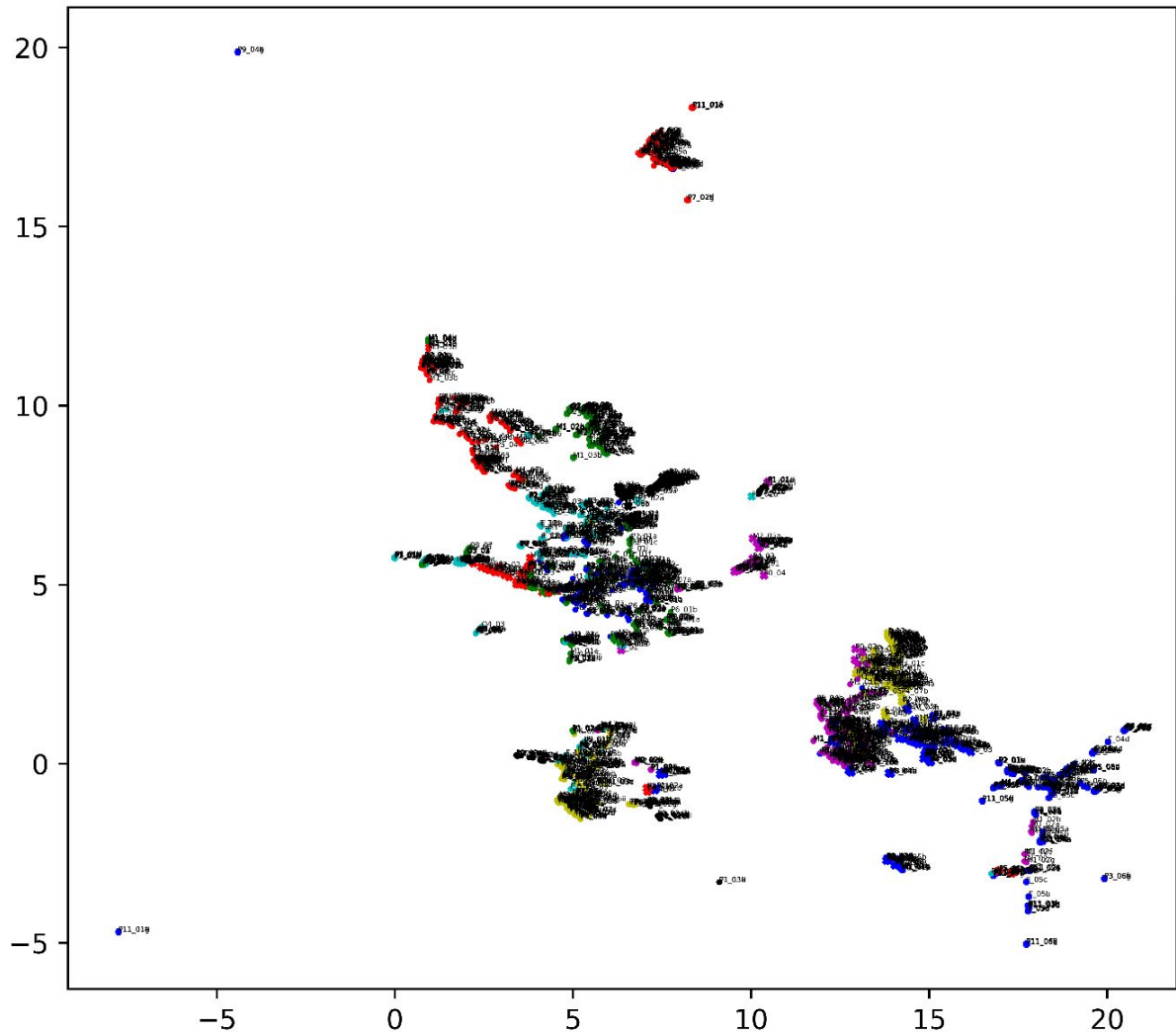


# Curriculum Analysis and Generation from Data (2022)

Prerequisites based on data

Equivalent classes based on data

Set of classes and concepts required for graduating



- 6.3000 Signal Processing
- 18.C06 Linear Algebra and Optimization
- 6.1010 Fundamentals of Programming
- 6.1200 Mathematics for Computer Science
- 18.03 Differential Equations
- 6.1910 Computation Structures
- 18.404 Theory of Computation
- 6.1800 Computer Systems Engineering
- 6.1220 Design and Analysis of Algorithms
- 6.1210 Introduction to Algorithms
- 18.600 Probability and Random Variables
- 18.02 Calculus II
- 6.4120 Computational Cognitive Science
- 6.100A Introduction to Computer Science Programming in Python
- 6.1020 Elements of Software Construction
- 6.2000 Electrical Circuits Modeling and Design of Physical Systems
- 6.8611 Quantitative Methods for Natural Language Processing

# AI Generated Curriculum

Chose class contents and topics

Example: generate a class that is 40% economics, 40% optimization, 20% you

# Current Status

Machine-generated, correctness evaluation homeworks delivered in Spring class (on gradescope)

Machine-generated, lecture notes delivered in Spring class (on Ed discussion)

Validation bookset in publication process (at Cambridge University press)

Conference presentation given by photorealistic avatar (AAAI)

Leading team of 25 students and colleagues (MIT, Harvard, Stanford, Cornell, Columbia, BU, Cambridge University Press) actively working on these topics in Spring 2023.

# Spring 2023 Team

Iddo Drori, MIT, Columbia University, Boston University

Yann Hicke, Cornell University

Sarah Zhang, MIT

Sage Simhon, MIT

Zad Chin, Harvard

Annie Wang, MIT

Alice Zhang, MIT

Eugenia Feng, MIT

Kirsi Rajagopal, MIT

Samuel Florin, MIT

Harsh Sharma, Boston University

Keith Tyser, Boston University

Andrei Marginean, MIT

Saisamrit Surbehera, Columbia University

Nikhil Singh, MIT

Leonard Tang, Harvard

Lauren Cowles, Cambridge University Press

Gilbert Strang, MIT

Tonio Buonasisi, MIT

Madeleine Udell, Stanford

Armando Solar Lezama, MIT

# 10 Most Recent Related Publications

## **Prospects and perils of writing books in Mathematics and Computer Science using AI**

Iddo Drori, Sarah J. Zhang, Sage Simhon, Yann Hicke, Zad Chin, Keith Tyser, Harsh Sharma, Kirsi Rajagopal, Alice Zhang, Annie Wang, Eugenia Feng, Nikhil Singh, Lauren Cowles, Tonio Buonassisi, Madeleine Udell, Gilbert Strang, Armando Solar-Lezama  
In progress

## **Automatically graduating from MIT Mathematics and EECS at a human level**

Iddo Drori, Sarah J. Zhang, Sage Simhon, Keith Tyser, Sarah Zhang, Reece Shuttleworth, Pedro Lantigua, Arvind Raghavan, Zad Chin, Saisamrit Surbehera, Leonard Tang, Yann Hicke, Avi Shporer, Nakul Verma, Tonio Buonassisi, Armando Solar-Lezama  
Under review

## **Dataset for graduating from MIT Math & EECS and curriculum analysis**

Sarah J. Zhang, Sage Simhon, Yann Hick, Zad Chin, Annie Wang, Kirsi Rajagopal, Alice Zhang, Eugenia Feng, Kieth Tyser, Harsh Sharma, Nikhil Singh, Tonio Buonassisi, Armando Solar-Lezama, Iddo Drori  
Under review

## **Text to graphics by program synthesis with error correction on precise, procedural, and simulation tasks**

Arvind Raghavan, Zad Chin, Alexander E. Siemenn, Vitali Petsiuk, Saisamrit Surbehera, Yann Hicke, Edward Chien, Ori Kerret, Tonio Buonassisi, Kate Saenko, Armando Solar-Lezama, Iddo Drori  
Under review

## **From human days to machine seconds: Automatically answering and generating machine learning final exams**

Sarah J. Zhang, Keith Tyser, Sage Simhon, Sarah Zhang, Reece Shuttleworth, Zad Chin, Pedro Lantigua, Saisamrit Surbehera, Gregory Hunter, Derek Austin, Yann Hicke, Leonard Tang, Sathwik Karnik, Darnell Granberry, Iddo Drori  
Under review

## **A dataset for learning university STEM courses at scale and generating questions at a human level**

Iddo Drori, Sarah Zhang, Zad Chin, Reece Shuttleworth, Albert Lu, Linda Chen, Bereket Birbo, Michele He, Pedro Lantigua, Sunny Tran, Gregory Hunter, Bo Feng, Newman Cheng, Roman Wang, Yann Hicke, Saisamrit Surbehera, Arvind Raghavan, Alexander Siemenn, Nikhil Singh, Jayson Lynch, Avi Shporer, Nakul Verma, Tonio Buonassisi, Armando Solar-Lezama  
Educational Advances in Artificial Intelligence (EAAI), 2023.

## **Human evaluation of text-to-image models on a multi-task benchmark**

Vitali Petsiuk, Alexander Siemenn, Saisamrit Surbehera, Zad Chin, Kieth Tyser, Gregory Hunter, Arvind Raghavan, Yann Hicke, Bryan Plummer, Ori Kerret, Tonio Buonassisi, Kate Saenko, Armando Solar-Lezama, Iddo Drori  
NeurIPS Workshop on Human Evaluation of Generative Models, 2022.

## **A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level**

Iddo Drori, Sarah Zhang, Reece Shuttleworth, Leonard Tang, Albert Lu, Elizabeth Ke, Kevin Liu, Linda Chen, Sunny Tran, Newman Cheng, Roman Wang, Nikhil Singh, Taylor L. Patti, Jayson Lynch, Avi Shporer, Nakul Verma, Eugene Wu, Gilbert Strang  
Proceedings of the National Academy of Sciences (PNAS), 119(32), 2022.

## **Solving Probability and Statistics problems by probabilistic program synthesis at human level and predicting solvability**

Leonard Tang, Elizabeth Ke, Nikhil Singh, Bo Feng, Derek Austin, Nakul Verma, Iddo Drori  
International Conference on Artificial Intelligence in Education (AIED), 2022.

## **Solving machine learning problems**

Sunny Tran, Pranav Krishna, Ishan Pakuwal, Prabhakar Kafle, Nikhil Singh, Jayson Lynch, Iddo Drori  
Asian Conference on Machine Learning (ACML), 2021.

## **Best paper award winner**

# The Science of Deep Learning

2022 textbook dlbook.org

Translated to Chinese and Korean

