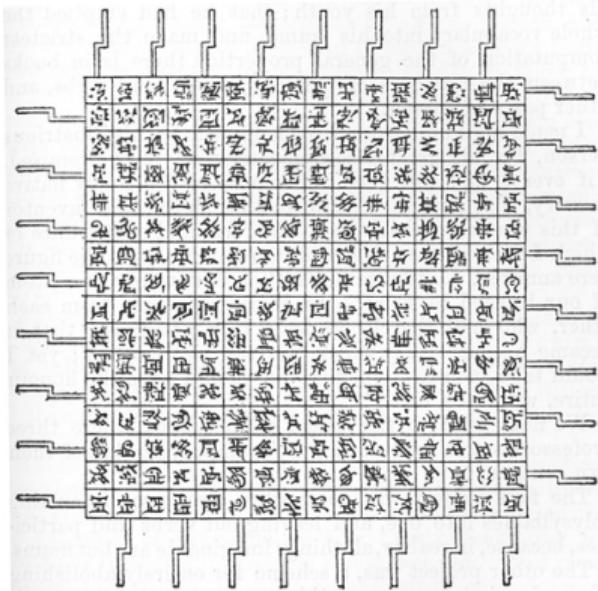


# Artificial Intelligence Basics

Francis J. DiTraglia

# **“The Frame” – Gulliver’s Travels**

*It was twenty feet square, placed in the middle of the room. The superficies was composed of several bits of wood, about the bigness of a die, but some larger than others. They were all linked together by slender wires. These bits of wood were covered, on every square, with paper pasted on them; and on these papers were written all the words of their language, in their several moods, tenses, and declensions; but without any order.*



# **“The Frame” – Gulliver’s Travels**

*The professor then desired me “to observe; for he was going to set his engine at work.” The pupils, at his command, took each of them hold of an iron handle, whereof there were forty fixed round the edges of the frame; and giving them a sudden turn, the whole disposition of the words was entirely changed.*



ChatGPT 4 ▾



OP You

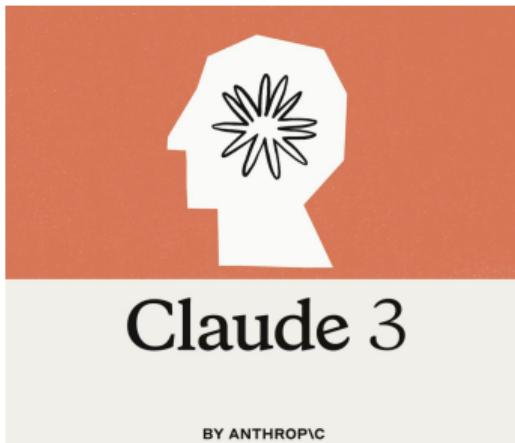
Write a humorous paragraph in the style of "Gulliver's Travels" by Jonathan Swift, describing how Gulliver--in his new role as a McKinsey consultant--visits a company that is sorely in need of management consulting.

ChatGPT

Upon my appointment to the celebrated rank of Consultant with McKinsey, it was decreed that I should undertake an expedition to the Corporation of Gigantomax, a company so vast in its confusion and disorder that the clerks themselves were oft reported to become irretrievably lost between their desks and the water cooler. Upon my arrival, I was met with such a scene of bureaucratic disarray and procedural pandemonium that it would have confounded even the Lilliputians, who could tie a man down with nothing but a multitude of tiny threads. The executives, great towering figures in their own esteem, wandered like lost sheep in the wilderness of their own making, unable to decide even the simplest course without a flock of consultations, and with charters so muddled one might suspect they had been drafted in the dark. Thus, armed with naught but my wits and a sheaf of spreadsheets, I set forth to untangle this Gordian knot, hoping to restore some semblance of order before all was lost to the anarchy of mismanagement.



 OpenAI  
GPT-4



# Lecture Outline

1. What is Artificial Intelligence?
2. Large Language Models
3. Case Study: Personalization at Netflix

# What is Artificial Intelligence?



Lieutenant Commander Data:  
*Star Trek the Next Generation*

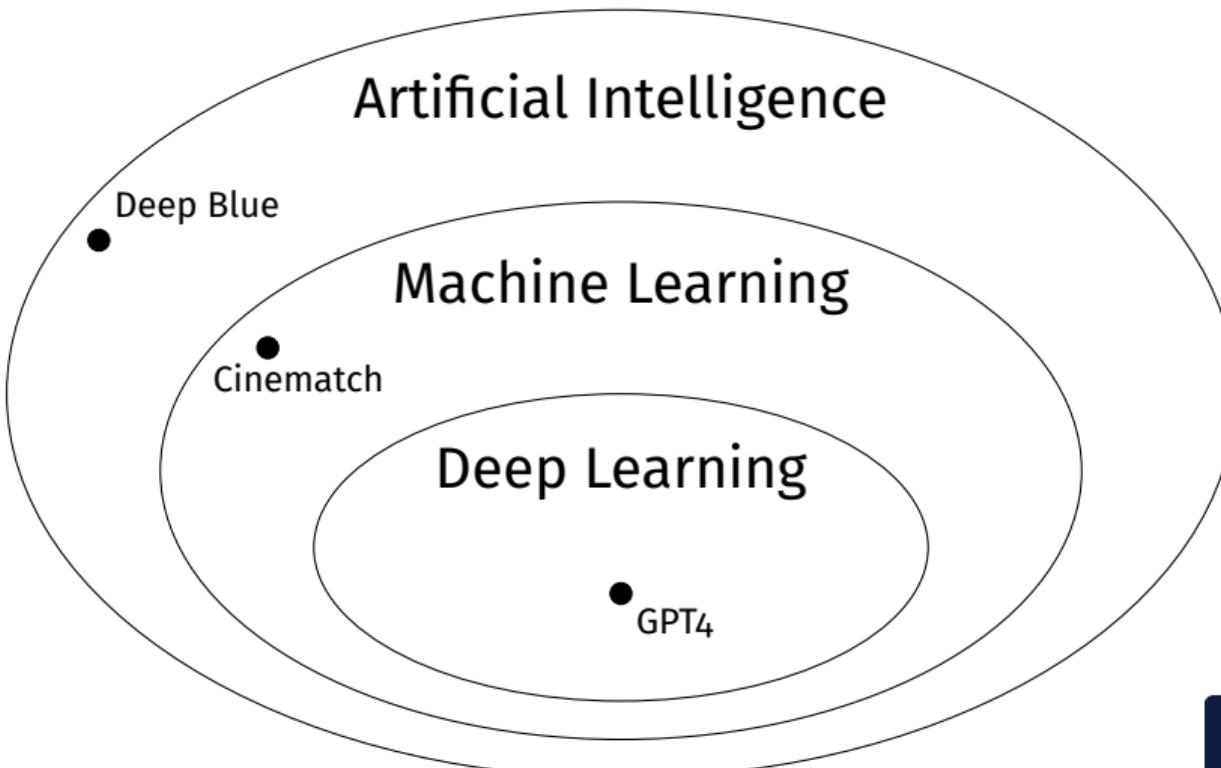
# What is Artificial Intelligence?



Lieutenant Commander Data:  
*Star Trek the Next Generation*

*AI is a field that develops algorithms and builds software to automate complex tasks: things like pattern recognition, decision-making, and problem solving.*

# Artificial Intelligence: The Big Picture



# **Symbolic Artificial Intelligence**

Humans write explicit rules that tell a computer how to translate input data into decisions.

## Model of Expert Performance

---

### Win

If there is a row, column, or diagonal with two of my pieces and a blank space,  
Then play the blank space (thus winning the game).

### Block

If there is a row, column, or diagonal with two of my opponent's pieces and a  
blank space,  
Then play the blank space (thus blocking a potential win for my opponent).

### Fork

If there are two intersecting rows, columns, or diagonals with one of my pieces  
and two blanks, and  
If the intersecting space is empty,  
Then move to the intersecting space (thus creating two ways to win on my next turn).

### Block Fork

If there are two intersecting rows, columns, or diagonals with one of my  
opponent's pieces and two blanks, and  
If the intersecting space is empty,  
Then  
If there is an empty location that creates a two-in-a-row for me (thus  
forcing my opponent to block rather than fork),  
Then move to the location.  
Else move to the intersection space (thus occupying the location that my  
opponent could use to fork).

### Play Center

If the center is blank,  
Then play the center.

### Play Opposite Corner

If my opponent is in a corner, and  
If the opposite corner is empty,  
Then play the opposite corner.

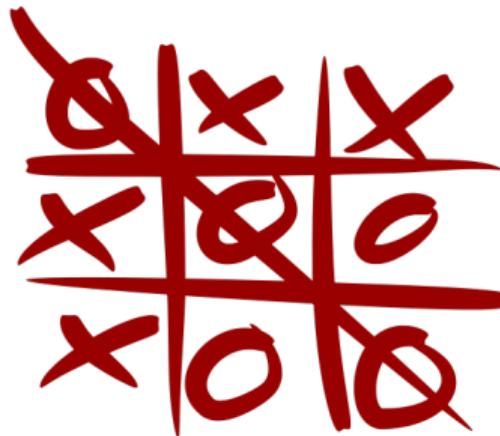
### Play Empty Corner

If there is an empty corner,  
Then move to an empty corner.

### Play Empty Side

If there is an empty side,  
Then move to an empty side.

---

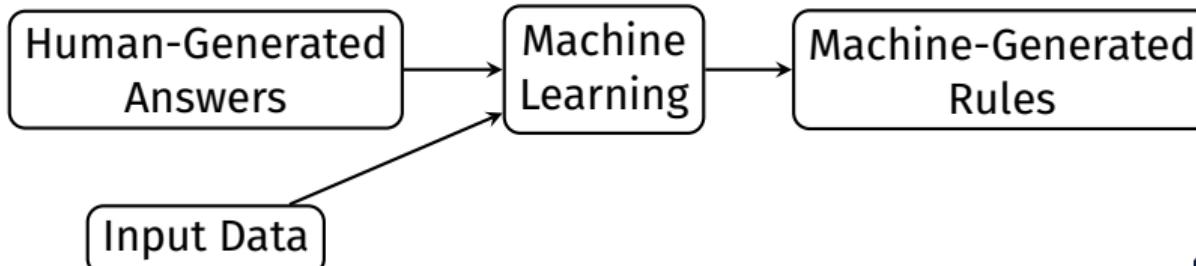
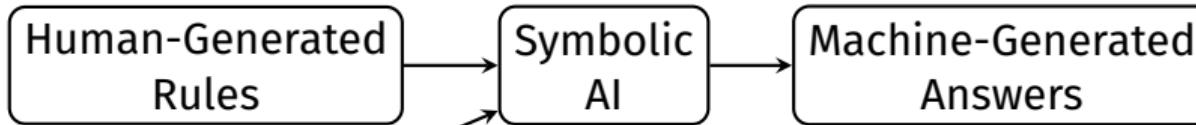


# Ada Lovelace on Babbage's “Analytical Engine”

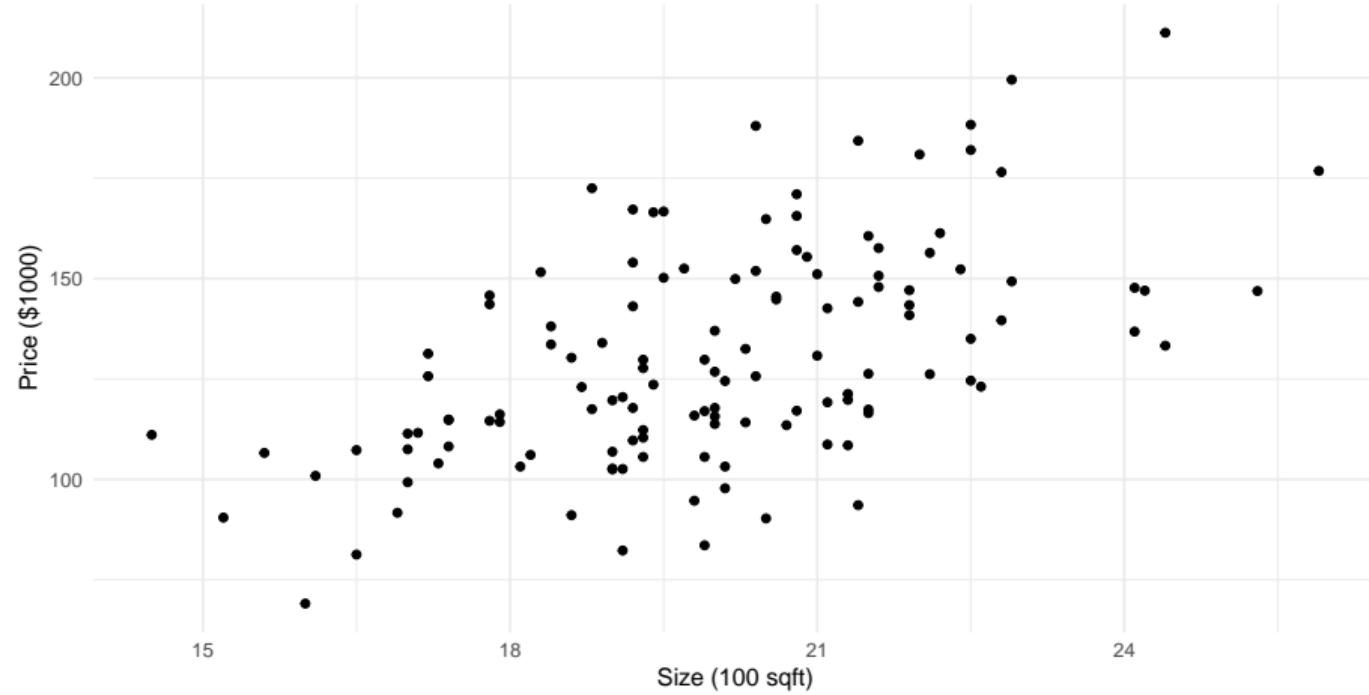
*The Analytical Engine has no pretensions whatever to **originate** any thing. It can do whatever we **know how to order it** to perform. It can follow analysis; but it has no power of **anticipating** any analytical relations or truths. Its province is to assist us in making **available** what we are already acquainted with.*



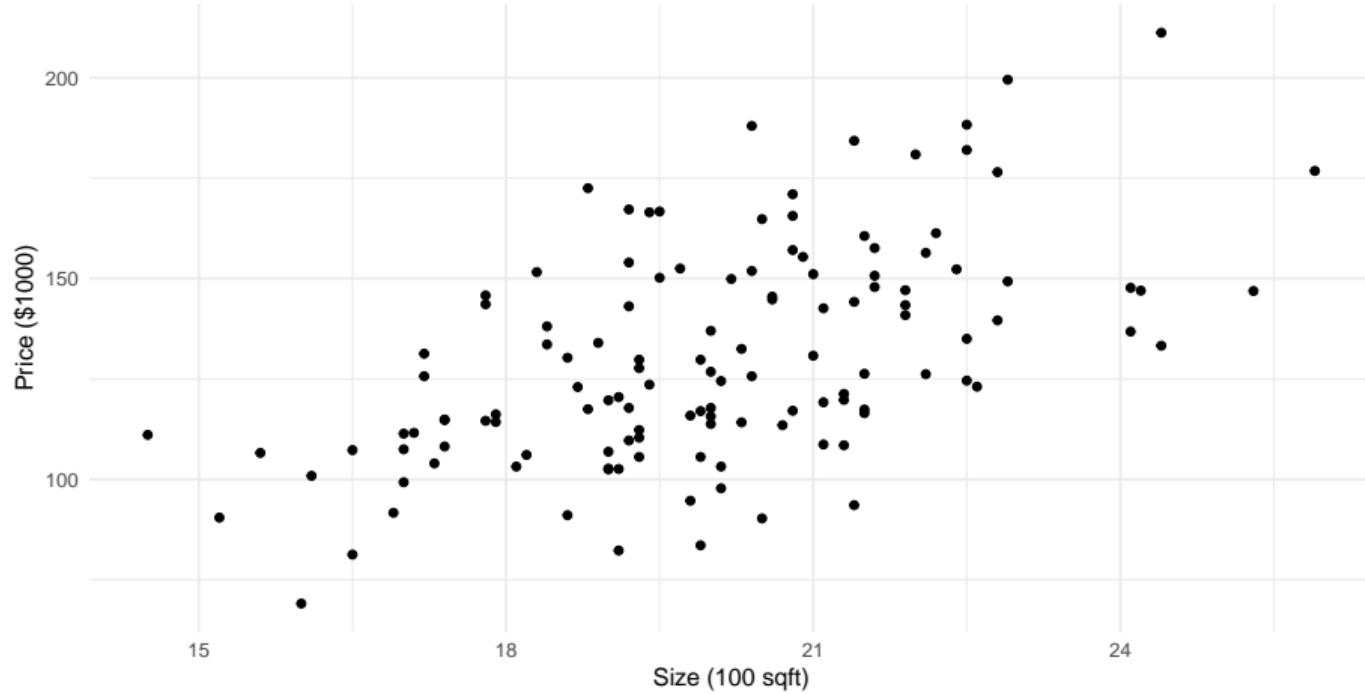
# Symbolic AI versus Machine Learning



## Predicting House Prices



## Predicting House Prices

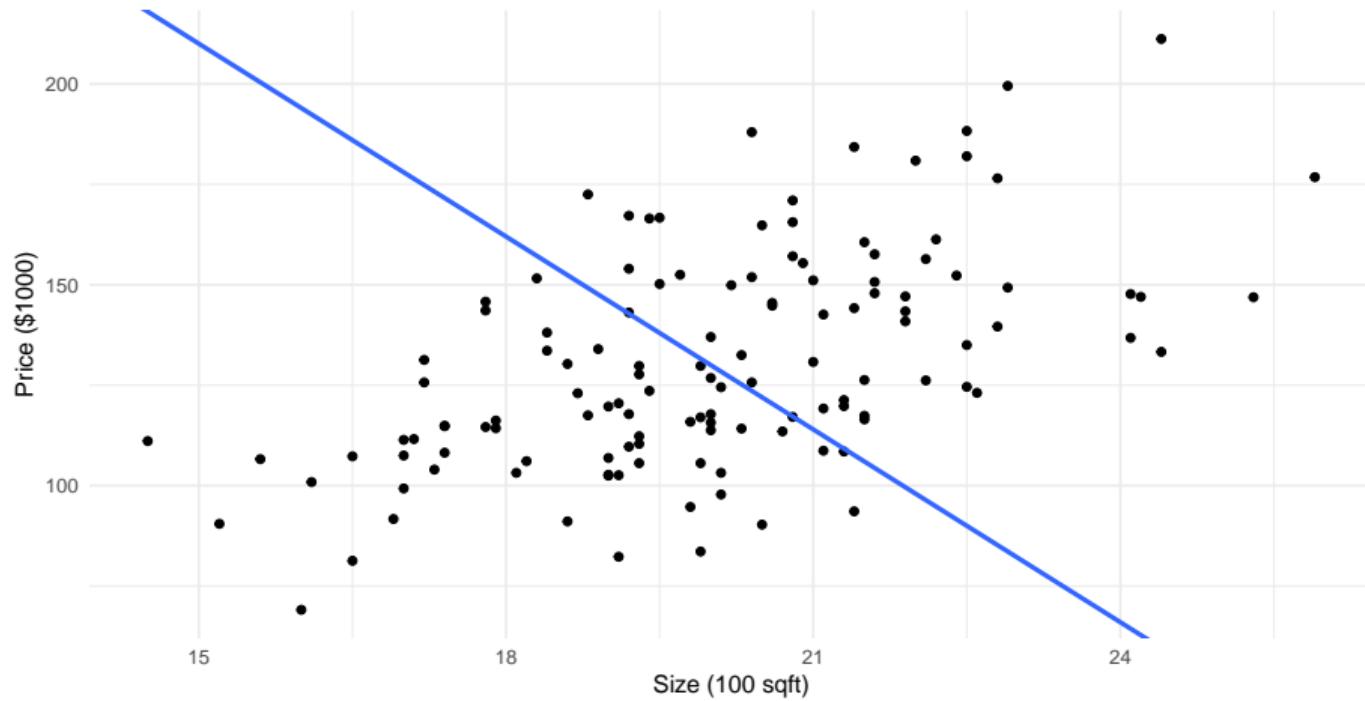


$\text{Predicted Price} = a + b \times \text{Size}$

# Root Mean Squared Error (RMSE)

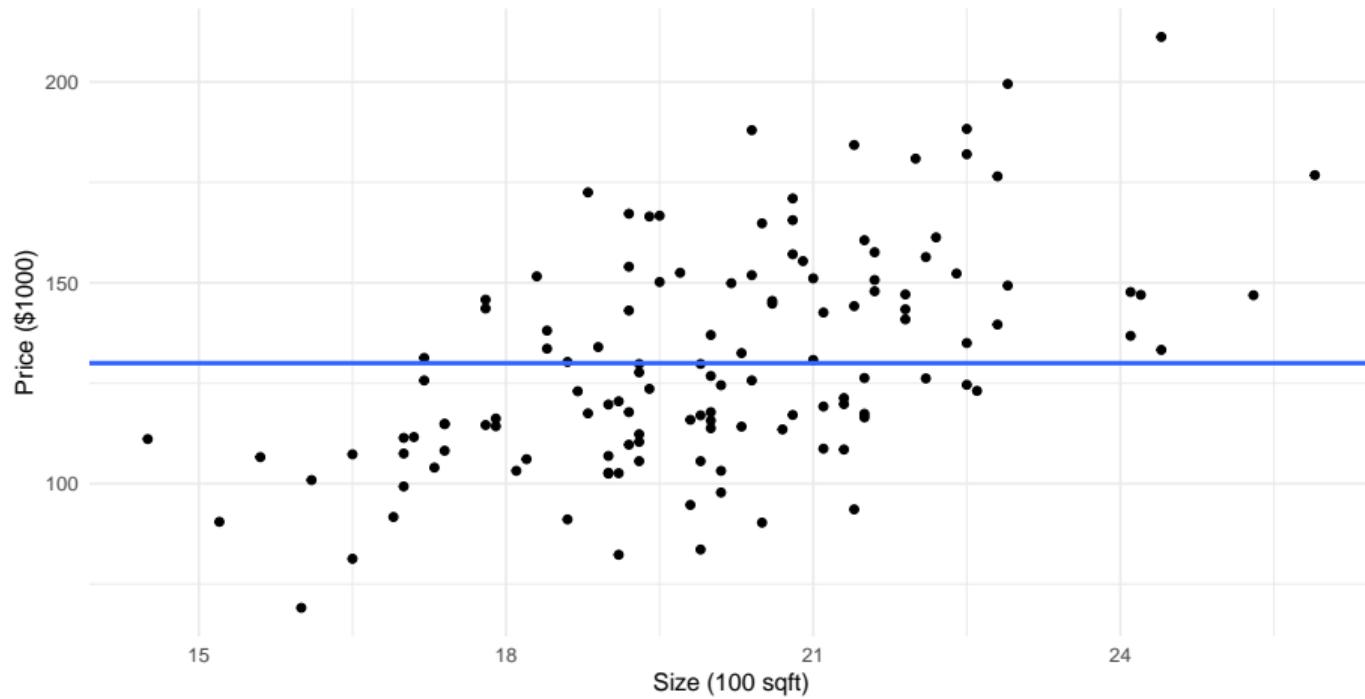
- On average, how close are our predictions to the truth?
- $\text{RMSE} = 0$  means perfect predictions; otherwise  $\text{RMSE} > 0$
- Units are identical to those of what we're predicting.

Intercept = 450.0, Slope = -16.0, RMSE = 53



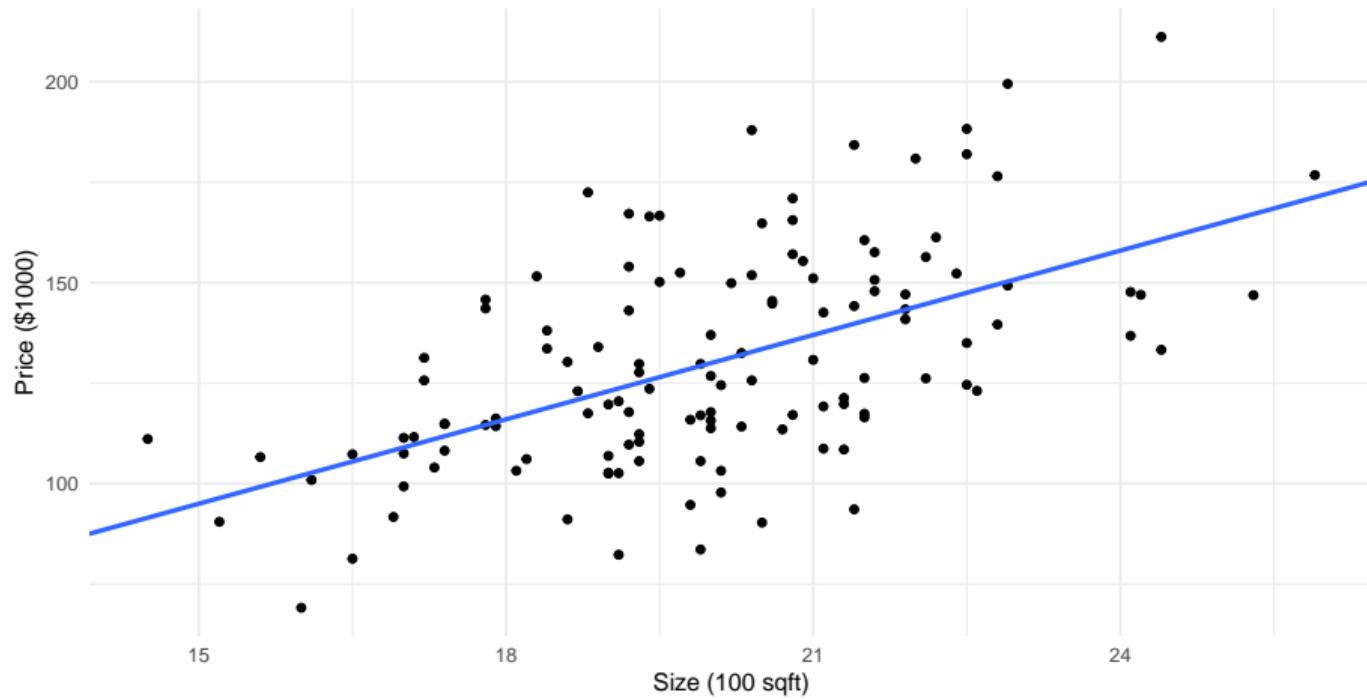
Predicted Price =  $450 - 16 \times \text{Size}$

Intercept = 130.0, Slope = 0.0, RMSE = 27



Predicted Price = 130

Intercept = -10.0, Slope = 7.0, RMSE = 22

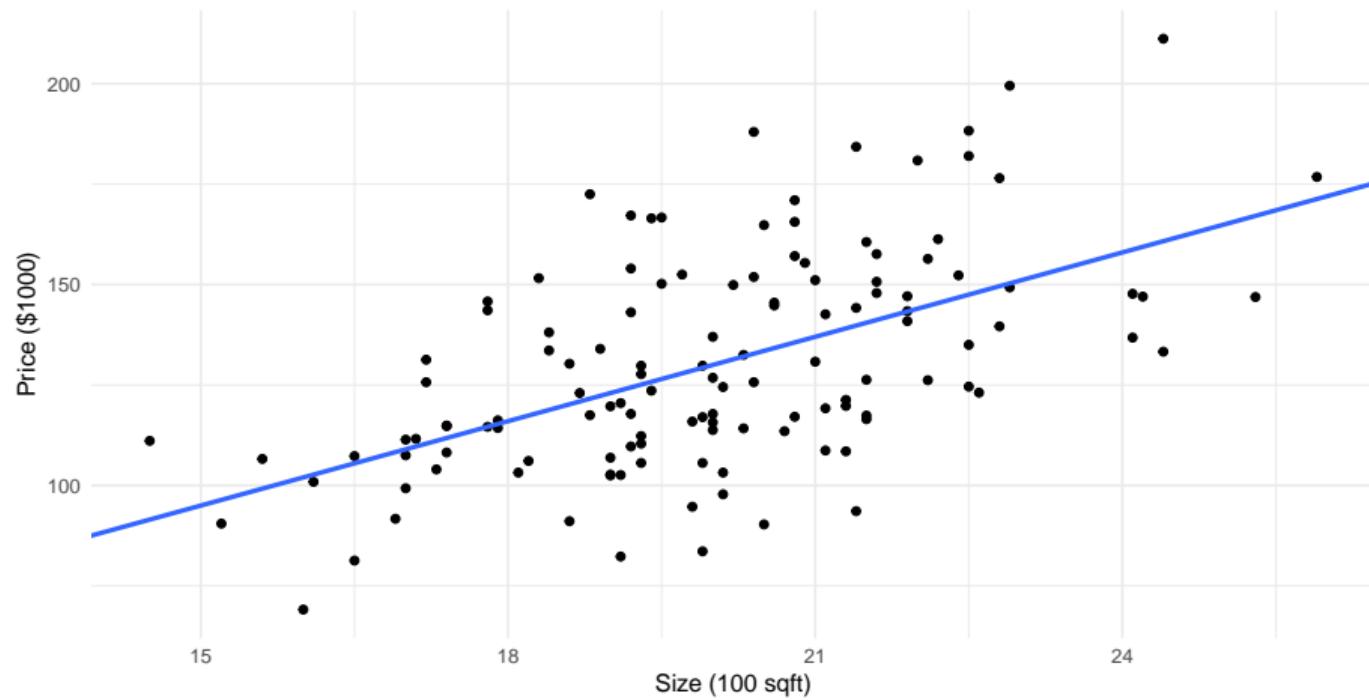


Predicted Price =  $-10 + 7 \times \text{Size}$

# Machine Learning

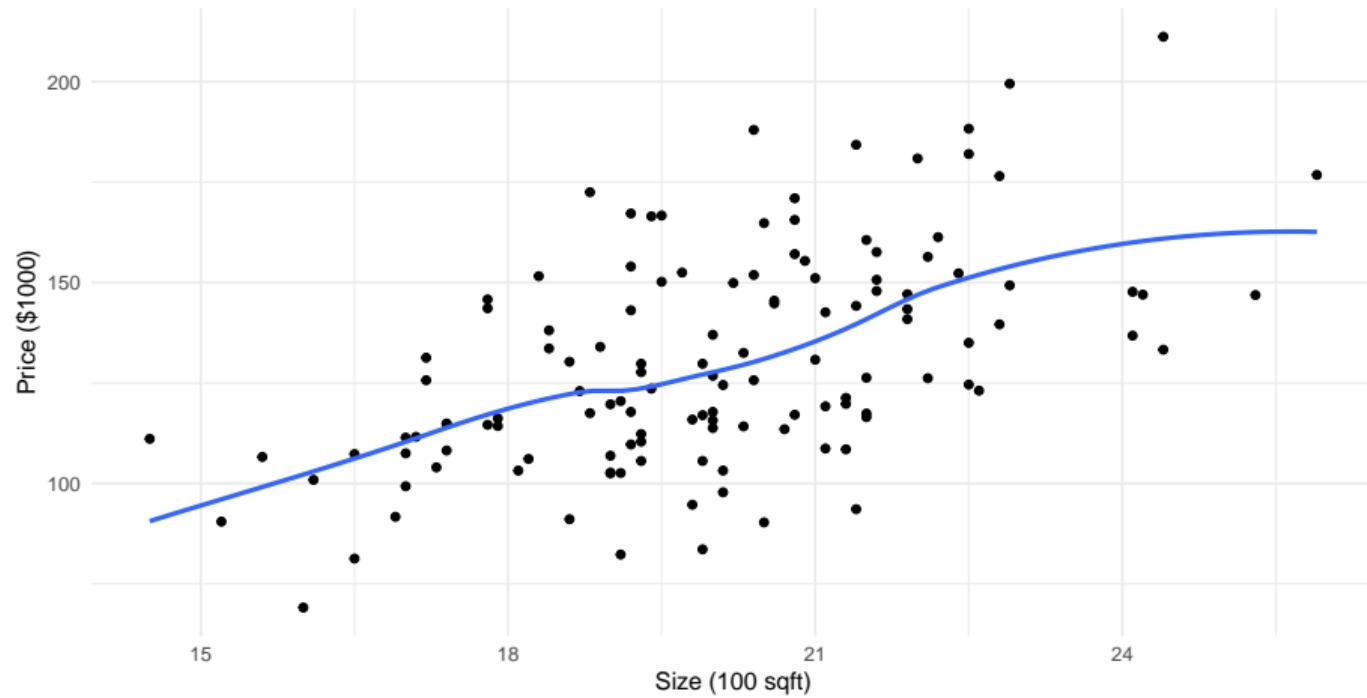
1. Obtain training data: predictors X and target Y.
2. Choose family of models with adjustable parameters.
3. Choose loss function: measure of “how close” predictions are to truth.
4. Fit/Train the model: adjust parameters to minimize loss on the training data.
5. Use the fitted model to predict the target in *new* situations

Intercept = -10.0, Slope = 7.0, RMSE = 22



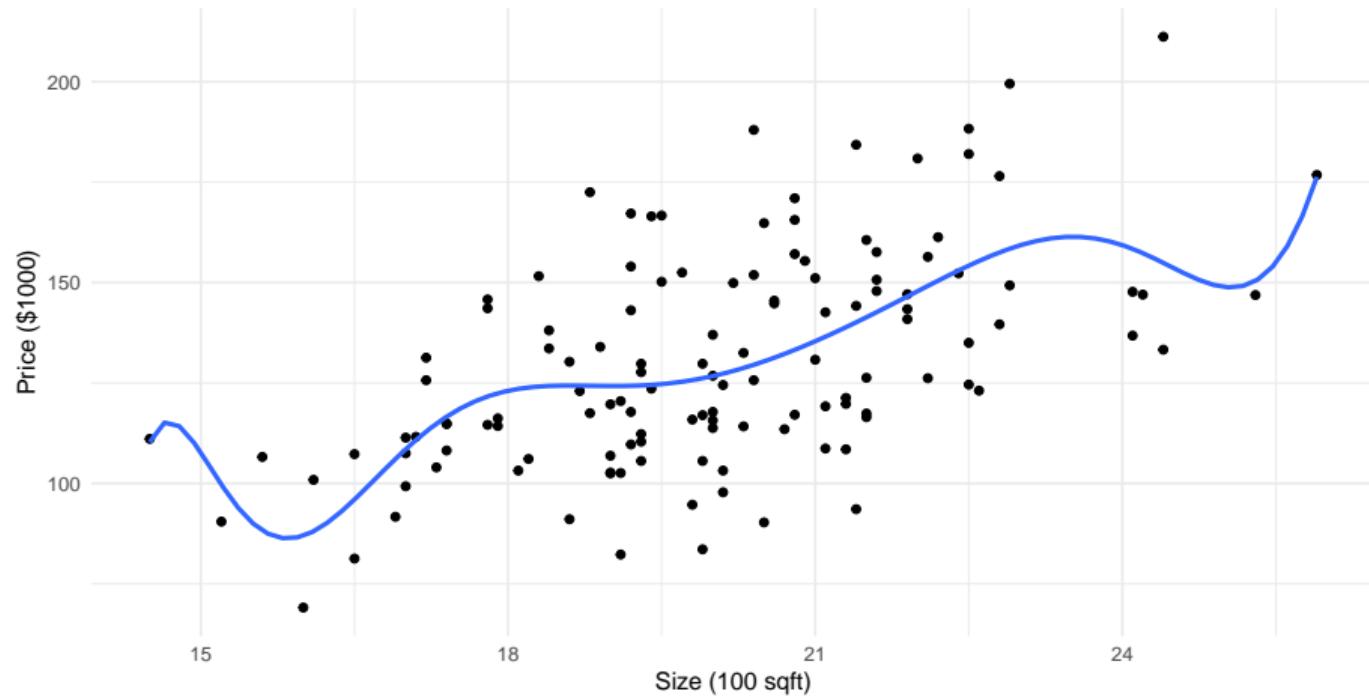
Underfitting: using a model that isn't flexible enough.

### A More Flexible Model of House Prices



Underfitting: using a model that isn't flexible enough.

This Model May Be Too Flexible



Overfitting: using a model that's too flexible.

# Training / Testing Split

- Randomly divide your data into two pieces.
- 80% goes into the **training set**: use these to choose parameter values
- 20% goes into the **test set**: hold these in reserve
- After fitting your model on the training set, check to see how well it predicts the data in the test set.
- This gives a reliable measure of how well your model will predict unseen data.

# Generative Artificial Intelligence

Combine relationships and patterns from the training data with an element of chance to **generate new data**.

# Key Point about Large Language Models

At their core, LLMs are machine learning models that are *really good* at predicting the next word in a sequence of text.

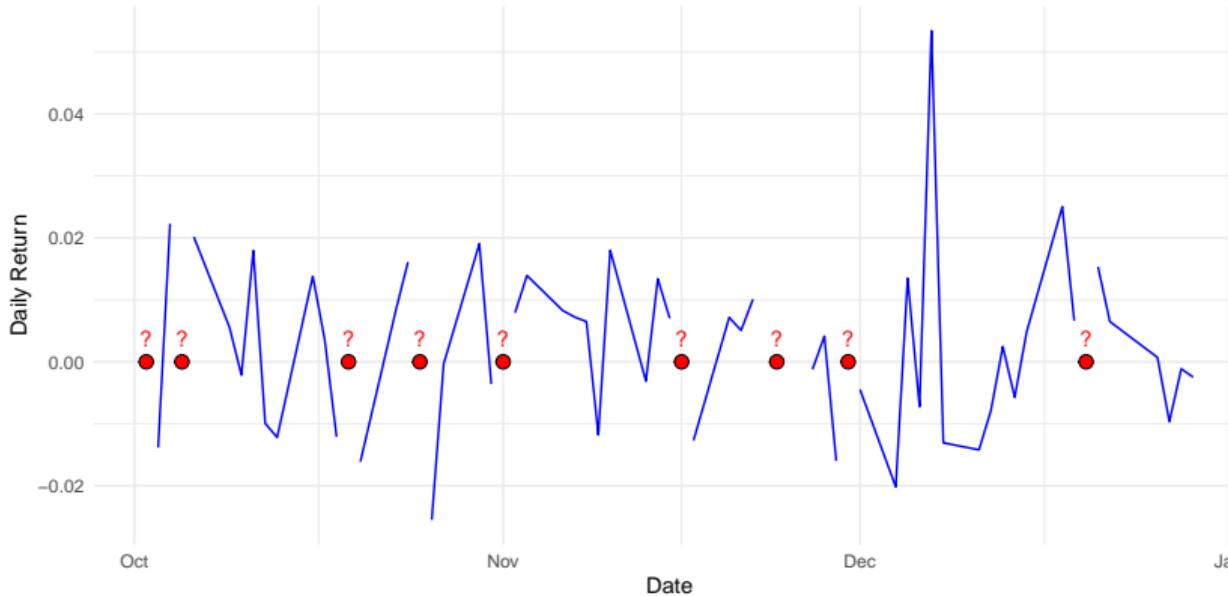
# Human-generated Text is Highly Predictable

The empire of Blefuscu is an island situated to the north-east of Lilliput, from which it is parted only \*y a channel of e\*ght hundred yards wide. I had not\*yet seen it, and upon this notice of an intended invasion, I\*avoided appearing on that side of the \*oas\*, for fea\* of\*being disco\*\*\*ed\* by so\*e of th\* enemy's ships, w\*o had received n\* inte\*li\*ence \*f me; all intercourse\*b\*\*we\*n th\* two emp\*res having bee\* s\*strict\*y \*or\*id\*en duri\*\* th\*\*war,\*\*p\*n pa\*n of d\*ath, a\*d an\*\*m\*\*rg\* laid\*by our emper\*r \*pon all\*vessels\*wh\*\*so\*ver. \*\*commun\*cat\*\* to\*his maje\*ty\*a project\*I \*ad formed of se\*zing t\*e enem\*'s\*w\*o\*e fleet; wh\*\*, as \*u\*\*scouts a\*\*ured \*s, l\*y at anch\*r\*in\*the h\*r\*our, r\*ady to \*ail wi\*h t\*\* f\*rst\*fair\*win\*.

# Stock Returns Are Not Predictable

Daily Stock Returns for Google in Q4 2023

15% of Values Deleted at Random



# But LLMs Predict *Tokens*, Not *Words*!

**Token:** word / part of a word / punctuation / combination

# Recall: House Price Example

- Training set: houses with known price and size.
- Model Family: all lines relating size to price
- Parameters: slope and intercept.
- Choose parameters to minimize RMSE.

# What does GPT stand for?

Generative Pre-trained Transformer

# Language Models are Few-Shot Learners

Tom B. Brown<sup>\*</sup>      Benjamin Mann<sup>\*</sup>      Nick Ryder<sup>\*</sup>      Melanie Subbiah<sup>\*</sup>  
 Jared Kaplan<sup>†</sup>      Prafulla Dhariwal      Arvind Neelakantan      Pranav Shyam      Girish Sastry  
 Amanda Askell      Sandhini Agarwal      Ariel Herbert-Voss      Gretchen Krueger      Tom Henighan  
 Rewon Child      Aditya Ramesh      Daniel M. Ziegler      Jeffrey Wu      Clemens Winter  
 Christopher Hesse      Mark Chen      Eric Sigler      Mateusz Litwin      Scott Gray  
 Benjamin Chess      Jack Clark      Christopher Berner  
 Sam McCandlish      Alec Radford      Ilya Sutskever      Dario Amodei

OpenAI

## Abstract

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3’s few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

<sup>\*</sup>Equal contribution

<sup>†</sup>Johns Hopkins University, OpenAI

## Language Models are Few-Shot Learners

Tom B. Brown*	Benjamin Mann*	Nick Ryder*	Melanie Subbiah*	
Jared Kaplan†	Prafulla Dhariwal	Arvind Neelakantan	Pranav Shyam	Girish Sastry
Amanda Askell	Sandhini Agarwal	Ariel Herbert-Voss	Gretchen Krueger	Tom Henighan
Rewon Child	Aditya Ramesh	Daniel M. Ziegler	Jeffrey Wu	Clemens Winter
Christopher Hesse	Mark Chen	Eric Sigler	Mateusz Litwin	Scott Gray
Benjamin Chess		Jack Clark		Christopher Berner
Sam McCandlish	Alec Radford	Ilya Sutskever		Dario Amodei
				OpenAI

### Abstract

Recent work has demonstrated substantial gains on many NLP tasks and benchmarks by pre-training on a large corpus of text followed by fine-tuning on a specific task. While typically task-agnostic in architecture, this method still requires task-specific fine-tuning datasets of thousands or tens of thousands of examples. By contrast, humans can generally perform a new language task from only a few examples or from simple instructions – something which current NLP systems still largely struggle to do. Here we show that scaling up language models greatly improves task-agnostic, few-shot performance, sometimes even reaching competitiveness with prior state-of-the-art fine-tuning approaches. Specifically, we train GPT-3, an autoregressive language model with 175 billion parameters, 10x more than any previous non-sparse language model, and test its performance in the few-shot setting. For all tasks, GPT-3 is applied without any gradient updates or fine-tuning, with tasks and few-shot demonstrations specified purely via text interaction with the model. GPT-3 achieves strong performance on many NLP datasets, including translation, question-answering, and cloze tasks, as well as several tasks that require on-the-fly reasoning or domain adaptation, such as unscrambling words, using a novel word in a sentence, or performing 3-digit arithmetic. At the same time, we also identify some datasets where GPT-3’s few-shot learning still struggles, as well as some datasets where GPT-3 faces methodological issues related to training on large web corpora. Finally, we find that GPT-3 can generate samples of news articles which human evaluators have difficulty distinguishing from articles written by humans. We discuss broader societal impacts of this finding and of GPT-3 in general.

\*Equal contribution

†Johns Hopkins University, OpenAI

Author contributions listed at end of paper.

## House Price Regression

- Parameters: 2 (slope and intercept)
- Training Data: 178 observations
- Computing: milliseconds / one processor

## GPT3

- Parameters: 175 billion (Transformer)
- Training Data: hundreds of billions of words
- Computing: months / millions of processors

# Predicting the last word in each “chunk”

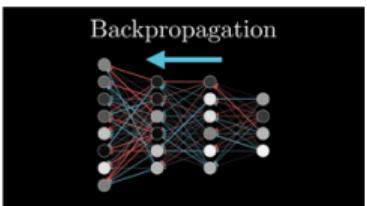
## Chunks

1. It was about twelve at noon, and a servant brought in dinner
2. the company were, the farmer and his wife, three children and an old grandmother
3. The mistress sent for her maid for a small dram cup which held about two gallons

## Predictions

1. It was about twelve at noon, and a servant brought in lunch ×
2. the company were, the farmer and his wife, three children and an old grandmother ✓
3. The mistress sent for her maid for a small dram cup which held about two ounces ×

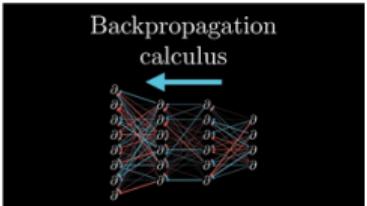
<https://3blue1brown.com>



### What is backpropagation really doing?

An overview of backpropagation, the algorithm behind how neural networks learn.

Chapter 4    Nov 2, 2017



### Backpropagation calculus

The math of backpropagation, the algorithm by which neural networks learn.

Chapter 5    Nov 2, 2017



### But what is a GPT? Visual intro to Transformers | Deep learning, chapter 5

A visual introduction to transformers. This chapter focusses on the overall structure, and word...

Apr 1, 2024



### Visualizing Attention, a Transformer's Heart | Chapter 6, Deep Learning

Demystifying attention, the key mechanism inside transformers and LLMs.

Apr 7, 2024

The most important thing to remember about AI is

### **Top 5 next word probabilities:**

that	0.7305
the	0.0423
how	0.0325
not	0.0258
to	0.0190

**Randomly selected word:** **that**

The most important thing to remember about AI is **that**

### **Top 5 next word probabilities:**

fact	0.1216
ability	0.0471
importance	0.0320
way	0.0282
power	0.0171

**Randomly selected word:** **the**

The most important thing to remember about AI is **that the**

### **Top 5 next word probabilities:**

brain	0.1073
mind	0.0841
-	0.0489
element	0.0489
being	0.0333

**Randomly selected word:** **human**

The most important thing to remember about AI is **that the human**

### **Top 5 next word probabilities:**

to 0.5788

for 0.3454

. 0.0279

not 0.0090

, 0.0086

**Randomly selected word:** **need**

The most important thing to remember about AI is **that the human need**

### **Top 5 next word probabilities:**

control 0.0442

autonomy 0.0316

a 0.0304

self 0.0304

empathy 0.0295

**Randomly selected word:** **for**

The most important thing to remember about AI is **that the human need for**

### **Top 5 next word probabilities:**

- . 0.4219
- over 0.1727
- and 0.0992
- , 0.0958
- of 0.0543

**Randomly selected word:** **control**

The most important thing to remember about AI is **that the human need for control**

### **Top 5 next word probabilities:**

the	0.1664
our	0.1110
its	0.1002
what	0.0562
it	0.0363

**Randomly selected word:** **over**

The most important thing to remember about AI is **that the human need for control over**

### **Top 5 next word probabilities:**

world	0.1505
environment	0.1225
machine	0.0494
system	0.0336
machines	0.0254

**Randomly selected word:** **the**

The most important thing to remember about AI is **that the human need for control over the**

### **Top 5 next word probabilities:**

- . 0.5376
- , 0.1225
- and 0.0479
- 's 0.0232
- that 0.0218

**Randomly selected word:** **machine**

The most important thing to remember about AI is **that the human need for control over the machine**

### **Top 5 next word probabilities:**

The 0.0779

:

0.0621

It 0.0562

This 0.0466

We 0.0465

**Randomly selected word:** .

The most important thing to remember about AI is that the human need for control over the machine.

### Top 5 next word probabilities:

is	0.3467
means	0.0814
needs	0.0445
requires	0.0393
can	0.0305

Randomly selected word: **This**

# Pre-training versus Fine-tuning

## Pre-training

- Build a highly quality next-word prediction algorithm.
- When you're done: not good at answering questions.
- Slow, expensive, data intensive

## Fine-tuning

- “Tweak” pre-trained model parameters to get a helpful assistant.
- New training data: high-quality human-generated Questions/Answers
- Quick, cheap, small training dataset ( 100,000)

tell me about Francis DiTraglia



Searched 6 sites

Francis J. DiTraglia is an Associate Professor of Economics at the University of Oxford, where he is also a Tutor and Fellow at Lady Margaret Hall. His academic background includes a PhD in Economics from the University of Cambridge and an MA in Economics from the University of Pennsylvania. Before his tenure at Oxford, he was an Assistant Professor of Economics at the University of Pennsylvania ([Oxford Economics](#)) ([Lady Margaret Hall Oxford](#)).

DiTraglia's research focuses on applied econometrics, particularly in causal inference, instrumental variables, measurement error, spillovers, and Bayesian inference. His applied work spans various topics, including machine learning for predicting lead poisoning prevalence, experimental studies of pawnbroking in Mexico City, and economic modeling of paramilitary violence in Colombia ([DiTraglia](#)) ([Oxford Economics](#)).

In addition to his research, DiTraglia is actively involved in teaching. At Oxford, he teaches courses such as Empirical Research Methods, Treatment Effects, and Behavioral and Experimental Economics. His previous teaching experience at the University of Pennsylvania includes courses in statistics, econometrics, and causal inference ([DiTraglia](#)).

DiTraglia is also a visiting researcher at the Federal Reserve Bank of Chicago and has published numerous working papers and articles on econometrics and public policy ([DiTraglia](#)) ([NBER](#)). For more detailed information about his work, you can visit his [personal website](#).

Where did he go to college?



Francis J. DiTraglia attended the University of Pennsylvania for his undergraduate studies, where he earned his BA. He later pursued graduate studies at the University of Cambridge, where he obtained his PhD in Economics ([Lady Margaret Hall Oxford](#)) ([Ditraglia](#)).

are you sure about that?

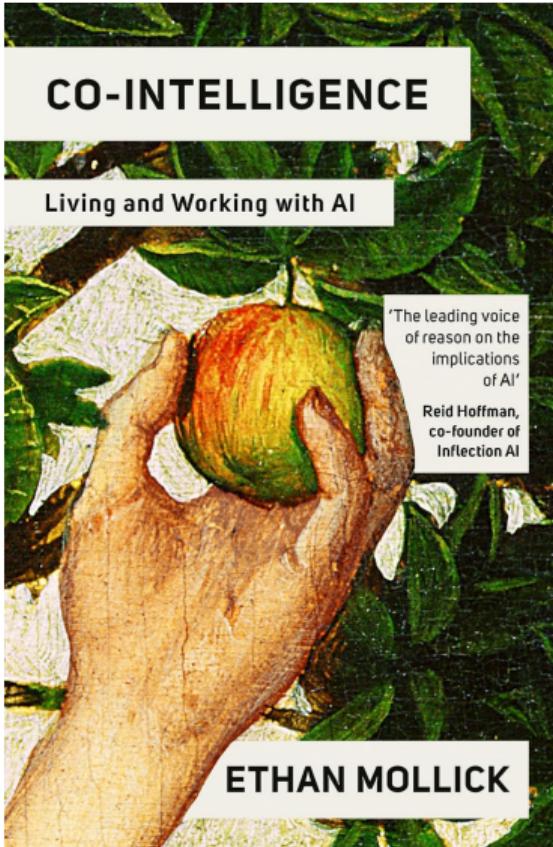


Searched 7 sites

Francis J. DiTraglia completed his undergraduate studies at the University of Pennsylvania, where he earned a BA. He then pursued graduate studies at the University of Cambridge, where he obtained his PhD in Economics ([Ditraglia](#)) ([SAS Economics Department](#)).

Currently, he is an Associate Professor of Economics at the University of Oxford and a Tutor and Fellow at Lady Margaret Hall. His research interests include causal inference, instrumental variables, and Bayesian inference, with applied work in areas such as lead poisoning prediction, pawnbroking in Mexico City, and paramilitary violence in Colombia ([Oxford Economics](#)) ([Lady Margaret Hall Oxford](#)). Additionally, DiTraglia has taught various economics and econometrics courses at both the University of Oxford and the University of Pennsylvania ([Ditraglia](#)).





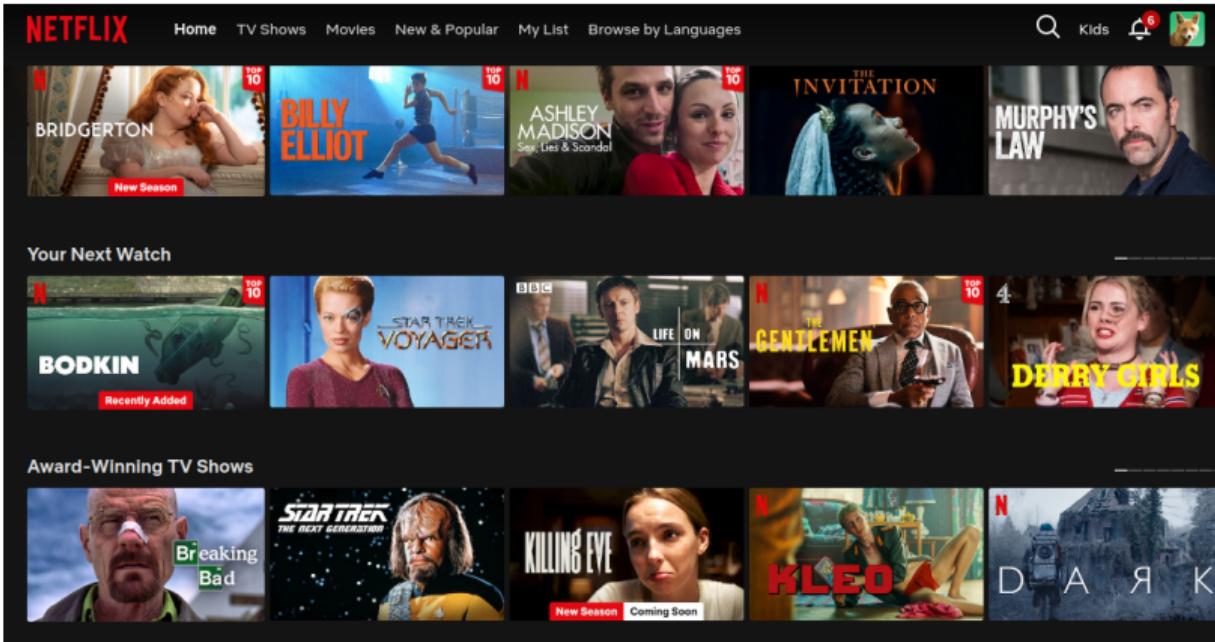
*Large Language Models are not knowledge bases. Instead, they are probabilistic models of knowledge bases.*

– Tom Dietterich

*Think of LLMs as mostly inscrutable artifacts; develop correspondingly sophisticated evaluations.*

– Andrej Karpathy

# What I See When I Log Into Netflix



# Further Reading

The screenshot shows a Medium article page. At the top, there are navigation links for 'Write', 'Sign up', and 'Sign in'. The main title 'A Brief History of Netflix Personalization' is displayed in bold black text. Below the title, a short summary reads: 'From its startup in 1998 to today, a detailed history of the strategy, metrics, and experiments Netflix executes to develop a personalized experience focused on delivering its members movies they love.' The author's profile picture and name 'Gibson Biddle - Follow' are shown, along with a note that it's a '21 min read' published on 'Jun 1, 2021'. Below the summary is a large image of a hand holding a remote control pointing at a television screen that displays the word 'NETFLIX'.

Gibson Biddle (2021)

The screenshot shows a Medium article page from the 'Netflix Technology Blog'. The title 'Netflix Recommendations: Beyond the 5 stars (Part 1)' is prominently displayed. Below the title, it says 'Published in Netflix TechBlog · 9 min read · Apr 6, 2012'. The author is listed as 'Xavier Amatriain and Justin Basilico (Personalization Science and Engineering)'. The main content starts with: 'In this two-part blog post, we will open the doors of one of the most valued Netflix assets: our recommendation system. In Part 1, we will relate the Netflix Prize to the broader recommendation challenge, outline the external components of our personalized service, and highlight how our task has evolved with the business. In Part 2, we will describe some of the data and models that we use and discuss our approach to algorithmic innovation that combines offline machine learning experimentation with online AB testing. Enjoy... and remember that we are always looking for more star talent to add to our great team, so please take a look at our jobs page.'

## The Netflix Prize and the Recommendation Problem

In 2006 we announced the [Netflix Prize](#), a machine learning and data mining competition for movie rating prediction. We offered \$1 million to whoever improved the accuracy of our existing system called Cinematch by 10%. We

The screenshot shows a paper titled 'All Together Now: A Perspective on the NETFLIX PRIZE' by Robert M. Bell, Yehuda Koren, and Chris Volinsky. The paper features a large image of a DVD disc and a film strip. The text begins with: 'What the Netflix Prize was all about is (believe it or not) the Netflix Prize itself: what our team did to win the Netflix Prize. Our group had worked for many years on building profiles of customers and their movie preferences, and we were comfortable with large data sets, but our team was not. One member was absentminded, and who doesn't like movies? We thought it would be a fun project.'

**Fundamental Analysis Challenge**

The Netflix Prize challenge concerns the prediction of movie ratings. The researchers released a training set consisting of data from about 500,000 customers and 10,000 movies. This amounted to more than 100 million ratings. The task was to use these data to predict the missing ratings, or, equivalently, to build a hold-out set of 3 million ratings. These models, known as collaborative filtering models, learn to predict the rating that each member of the whole group is likely to make individual predictions.

Most of the characteristics are abstract and difficult to understand. For example, the data contains no explicit information describing setting, plot, characters, cast, and many more subtle features such as the time of day, the day of the week, and the Movie Genome Project (more on come later). Given this lack of representation, thousands of people have tried to improve upon this model. It is likely to be one of the most famous datasets ever used in machine learning.

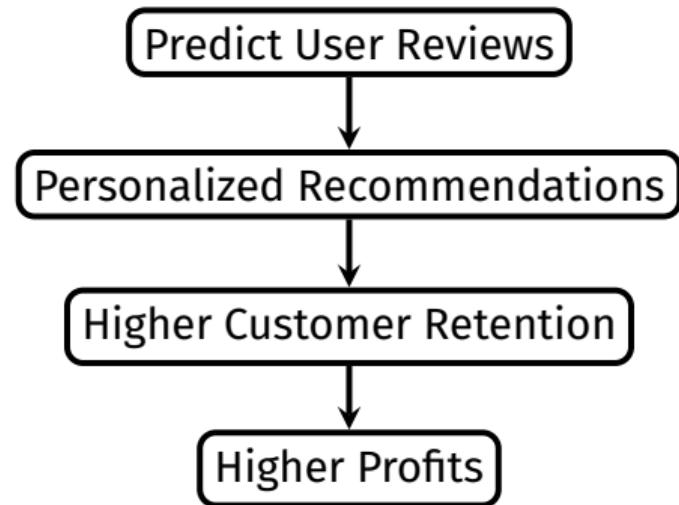
On the other hand, complex models can easily overfit the data, especially small details rather than the big picture—especially when data are scarce.

Bell, Koren & Volinsky (2010)

# The Business Case for Personalization

*The high-level hypothesis: personalization would improve retention by making it easy for members to find movies they'll love.*

*– Gibson Biddle*



# Cinematch: “Fill In” the Missing Movie Ratings

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4		
User B	4	2	5	3	3
User C	3	5		4	4
User D		3	4	2	1

# Users A and B Have Similar Taste in Movies

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4		
User B	4	2	5	3	3
User C	3	5		4	4
User D		3	4	2	1

# Maybe D Would Also Give 5 Stars to Movie 1?

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4		
User B	4	2	5	3	3
User C	3	5		4	4
User D	5?	3	4	2	1

**Training Set: Black    Test Set: Red    Predictions: Blue**

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4	2	3
User B	4	2	5	3	3
User C	3	5	5	4	4
User D	4	3	4	2	1

# Training Set: Black   Test Set: Red   Predictions: Blue

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4	2	3
User B	4	2	5	3	3
User C	3	5	5	4	4
User D	4	3	4	2	1

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4	3.7?	2.5?
User B	4	2	5	3	3
User C	3	5	4.1?	4	4
User D	4.3?	3	4	2	1

# Training Set: Black   Test Set: Red   Predictions: Blue

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4	2	3
User B	4	2	5	3	3
User C	3	5	5	4	4
User D	4	3	4	2	1

	Movie 1	Movie 2	Movie 3	Movie 4	Movie 5
User A	5	3	4	3.7?	2.5?
User B	4	2	5	3	3
User C	3	5	4.1?	4	4
User D	4.3?	3	4	2	1

$$\text{RMSE} = \sqrt{\frac{1}{4} [(4 - 4.3)^2 + (5 - 4.1)^2 + (2 - 3.7)^2 + (3 - 2.5)^2]} \approx 1$$

# The Netflix Prize

NETFLIX

## Netflix Prize

**COMPLETED**

Home Rules Leaderboard Update

### Leaderboard

Showing Test Score. [Click here to show quiz score](#)

Display top 20 leaders.

Rank	Team Name	Best Test Score	% Improvement	Best Submit Time
Grand Prize - RMSE = 0.8567 - Winning Team: BellKor's Pragmatic Chaos				
1	BellKor's Pragmatic Chaos	0.8567	10.06	2009-07-26 18:18:28
2	The Ensemble	0.8567	10.06	2009-07-26 16:36:22
3	Grand Prize Team	0.8582	9.90	2009-07-10 21:24:40
4	Oreka Solutions and Vendelay United	0.8588	9.84	2009-07-10 01:12:31
5	Vendelay Industries?	0.8591	9.81	2009-07-10 00:32:20
6	PragmaticTheory	0.8594	9.77	2009-06-24 12:06:56
7	BellKor in BigChaos	0.8601	9.70	2009-05-13 06:14:09
8	Dace_	0.8612	9.59	2009-07-24 17:18:43
9	Feeds2	0.8622	9.48	2009-07-12 13:11:51
10	BioChaos	0.8623	9.47	2009-04-07 12:33:59
11	Oreka Solutions	0.8623	9.47	2009-07-24 00:34:07
12	BellKor	0.8624	9.46	2009-07-26 17:19:11
Progress Prize 2008 - RMSE = 0.8627 - Winning Team: BellKor in BigChaos				
13	xiaoliang	0.8642	9.27	2009-07-15 14:53:22
14	Gravity	0.8643	9.26	2009-04-22 18:31:32
15	Ces	0.8651	9.18	2009-06-21 19:24:53
16	Invisible Ideas	0.8653	9.15	2009-07-15 15:53:04
17	Just a guy in a garage	0.8662	9.06	2009-05-24 10:02:54
18	JDennis Su	0.8666	9.02	2009-03-07 17:16:17
19	Craig Carmichael	0.8666	9.02	2009-07-25 16:00:54
20	armehill	0.8668	9.00	2009-03-21 16:20:50
Progress Prize 2007 - RMSE = 0.8723 - Winning Team: KorBell				
Cinematch score - RMSE = 0.9525				

## Prize

Beat Cinematch RMSE by 10%, win \$1 million

## Training Set

≈ 100 million ratings from 500k users

## Test Set

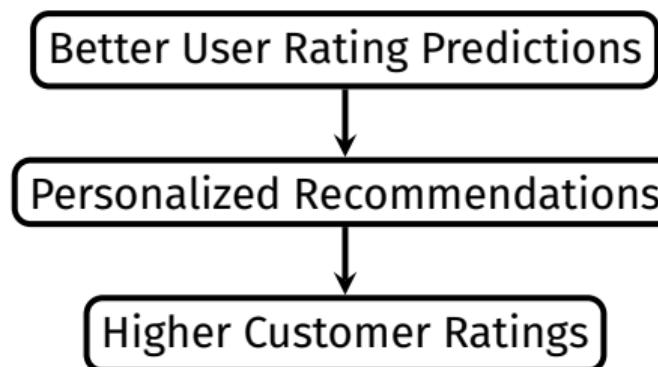
3 million movie ratings (same users)

# Higher Ratings Don't Cause Retention

*But when Netflix executed the new algorithm in a large-scale A/B test, there was no measurable retention difference.*

*While there was evidence that average ratings increased, nothing in the Netflix dataset suggested that higher average ratings improved retention. **It turns out that movie ratings do not equal movie enjoyment.***

– Gibson Biddle



# Lady Lovelace gets the last word.

*It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine. In considering any new subject, there is frequently a tendency, first, to **overrate** what we find to be already interesting or remarkable; and, secondly, by a sort of natural reaction, to **undervalue** the true state of the case, when we do discover that our notions have surpassed those that were really tenable.*

