

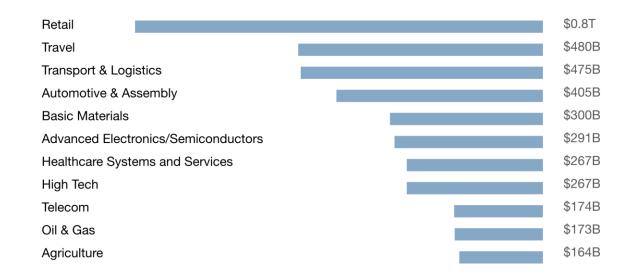
Introduction

Introduction

AI value creation by 2033

\$13-22 trillion

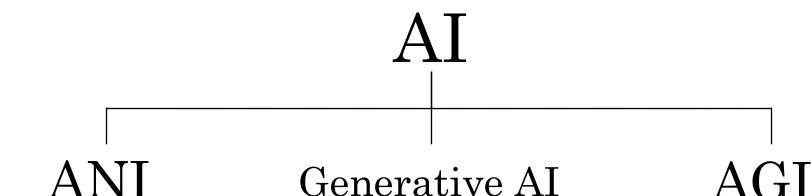
(includes \$3-4 trillion from generative AI)



[Source: McKinsey]



Demystifying AI



(artificial narrow intelligence)

E.g., smart speaker, self-driving car, web search, AI in farming and factories (generative artificial intelligence)

E.g., ChatGPT, Bard, Midjourney, DALL-E

(artificial general intelligence)

Do anything a human can do

What you'll learn

- What is AI?
 - Machine Learning
 - Data
 - What makes an AI company
 - What machine learning can and cannot do
 - Optional: Intuitive explanation of Deep Learning
- Building AI projects
- Building AI in your company
- AI and society





Machine Learning

Supervised Learning

Input

Output

		•
Input (A)	Output (B)	Application
email	spam? (0/1)	spam filtering
audio ->>	text transcripts	speech recognition
English ->>	Chinese	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info ->	position of other cars	Self-driving car
image of phone ->	defect? (0/1)	visualinspection
sequence of words ->	the next word	chatbot

How large language models (LLMs) work

LLMs are built by using supervised learning ($A\rightarrow B$) to repeatedly predict the next word.

My favorite drink is lychee bubble tea.

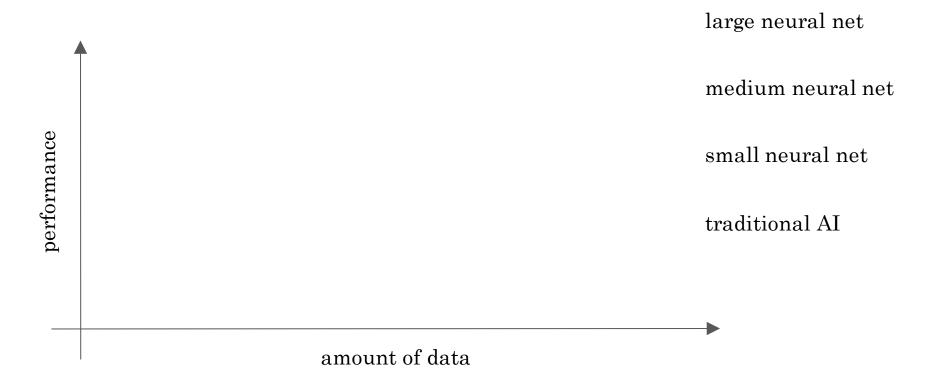
Input (A)	Output (B)
My favorite drink	is
My favorite drink is	lychee
My favorite drink is lychee	bubble
My favorite drink is lychee bubble	tea



When we train a very large AI system on a lot of data (hundreds of billions of words), we get a Large Language Model like ChatGPT.

		•
Input (A)	Output (B)	Application
email	spam? (0/1)	spam filtering
audio ->>	text transcripts	speech recognition
English ->>	Chinese	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info ->	position of other cars	Self-driving car
image of phone ->	defect? (0/1)	visualinspection
sequence of words ->	the next word	chatbot

Why Now?





What is data

Example of a table of data (dataset)

house (square feet)	# of bedrooms	price (1000\$)
523 645	1	100 150
708 1034	$\frac{1}{2}$	200 300
$\begin{array}{c c} 1034 \\ 2290 \end{array}$	3 4	350
2545	4	440

image	label	
	cat	
	not cat	
	cat	
8 8	not cat	

"Google cat"

Acquiring data

- Manual labeling







not cat



cat



not cat

- From observing user behaviors

machine	temperature	pressure	machine
	(°C)	(psi)	fault
17987	60	7.65	N
34672	100	25.50	N
08542	140	75.50	Y
98536	165	125.00	Y

- Download from websites / partnerships

Use and mis-use of data

Don't throw data at an AI team and assume it will be valuable.



Data is messy

- Garbage in, garbage out
- Data problems
 - Incorrect labels
 - Missing values
- Multiple types of data

images, audio, text

house	# of	price
(square feet)	bedrooms	(1000\$)
523	1	100
645	1	0.001
708	unknown	200
1034	3	unknown
unknown	4	350
2545	unknown	440





The terminology of AI

Machine learning vs. data science

Home prices

size	# of	# of	newly	price
(square feet)	bedrooms	bathrooms	renovated	(1000\$)
523 645 708 1034 2290 2545	1 1 2 3 4 4	2 3 1 3 4 5	N N N Y N	100 150 200 300 350 440

Running AI system (e.g., websites / mobile app)

Homes with 3 bedrooms are more expensive than homes with 2 bedrooms of a similar size.

Newly renovated homes have a 15% premium.

Machine learning vs. data science

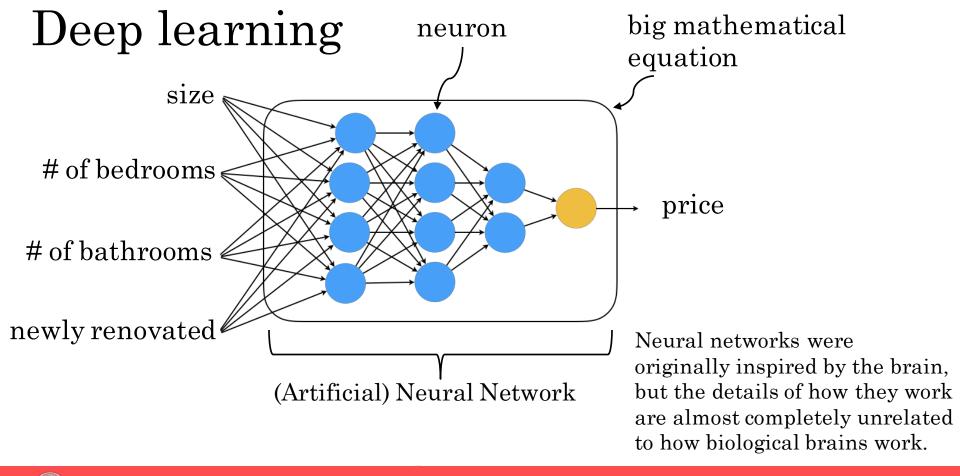
Machine learning

Data science

"Field of study that gives computers the ability to learn without being explicitly programmed."

Science of extracting knowledge and insights from data.

-Arthur Samuel (1959)



AI has many tools

- Machine learning and data science
- Deep learning / neural network
- Other buzzwords: Generative
 AI, unsupervised learning, reinforcement learning, graphical models, planning, knowledge graph, ...



What makes an AI company?

A lesson from the rise of the Internet

Internet Era

Shopping mall + website ≠ Internet company

- A/B testing
- Short iteration time
- Decision making pushed down to engineers and other specialized roles

AI era

Any company + deep learning ≠ AI company

- Strategic data acquisition
- Unified data warehouse
- Pervasive automation
- New roles (e.g., MLE) and division of labor

AI Transformation

- 1. Execute pilot projects to gain momentum
- 2. Build an in-house AI team
- 3. Provide broad AI training
- 4. Develop an AI strategy
- 5. Internal and external communication



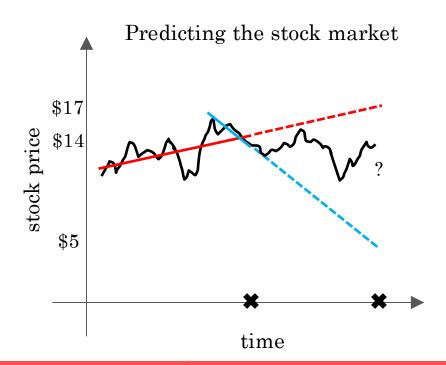
What machine learning can and cannot do

Supervised Learning

	Input (A)	Output (B)	Application
·	email	spam? (0/1)	spam filtering
	> audio	text transcripts	speech recognition
	English	Chinese	machine translation
	ad, user info	click? (0/1)	online advertising
	image, radar info	position of other cars	Self-driving car
	image of phone	defect? (0/1)	visual inspection
	sequence of words	the next word	chatbot

Anything you can do with 1 second of thought, we can probably now or soon automate.

What machine learning today can and cannot do



What makes an ML problem easier

1. Learning a "simple" concept

2. Lots of data available



More examples of what machine learning can and cannot do

Self-driving car

Can do





Cannot do







hitchhiker



bike turn left signal

- 1. Data
- 2. Need high accuracy

X-ray diagnosis











Can do

Diagnose pneumonia from ~10,000 labeled images

Cannot do

Diagnose pneumonia from 10 images of a medical textbook chapter explaining pneumonia

Strengths and weaknesses of machine learning

ML tends to work well when:

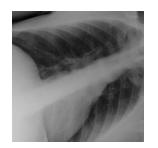
- 1. Learning a "simple" function
- 2. There is lots of data available

ML tends to work poorly when:

- 1. Learning complex functions from small amounts of data
- 2. It is asked to perform on new types of data that it learned from





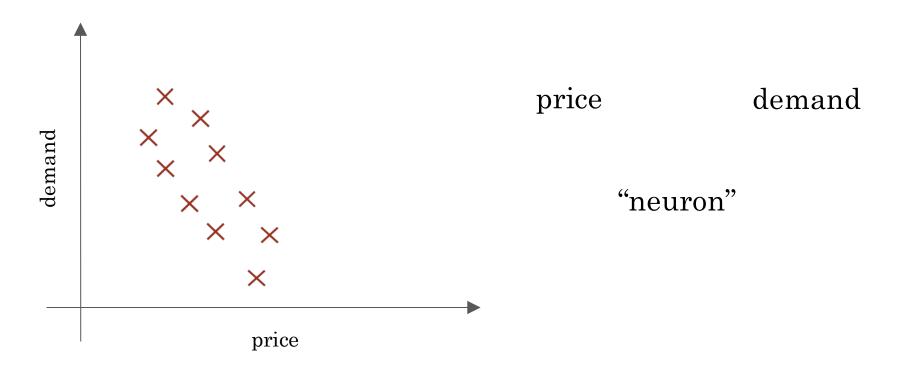






Non-technical explanation of deep learning I (optional)

Demand prediction



Demand prediction

price

shipping cost

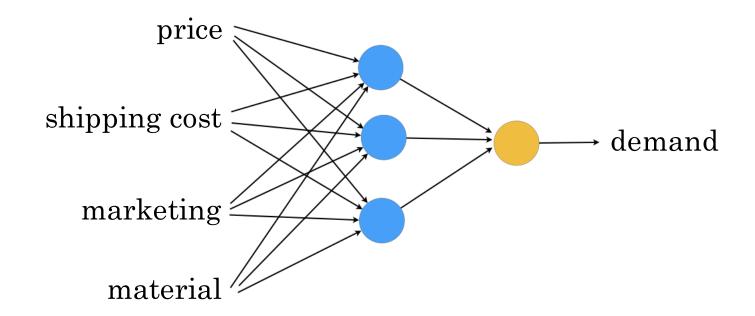
marketing

material

demand



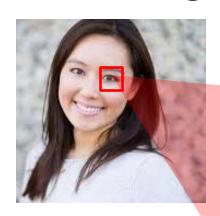
Demand prediction





Non-technical explanation of deep learning II (optional)

Face recognition



30	32	22	12	10	10	12	33	35	30
12	11	12	234	170	176	13	15	12	12
234	222	220	230	200	222	230	234	56	78
190	220	186	112	110	110	112	180	30	32
49	250	250	250	4	2	254	200	44	6
55	250	250	250	3	1	250	245	25	3
189	195	199	150	110	110	182	190	199	55
200	202	218	222	203	200	200	208	215	222
219	215	220	220	222	214	215	210	220	220
220	220	220	220	221	220	221	220	220	222

Face recognition

