

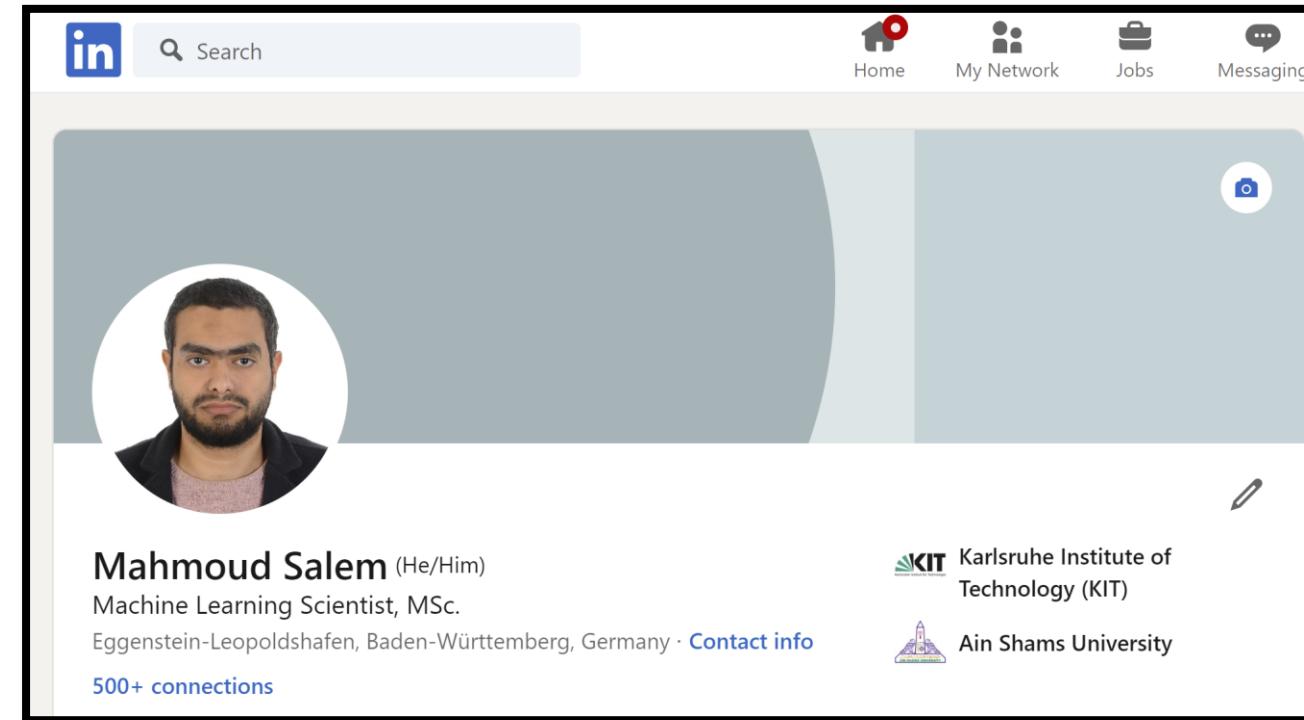
Welcome to Embedded AI

Mahmoud Salem, MSc, Dipl.-Eng.

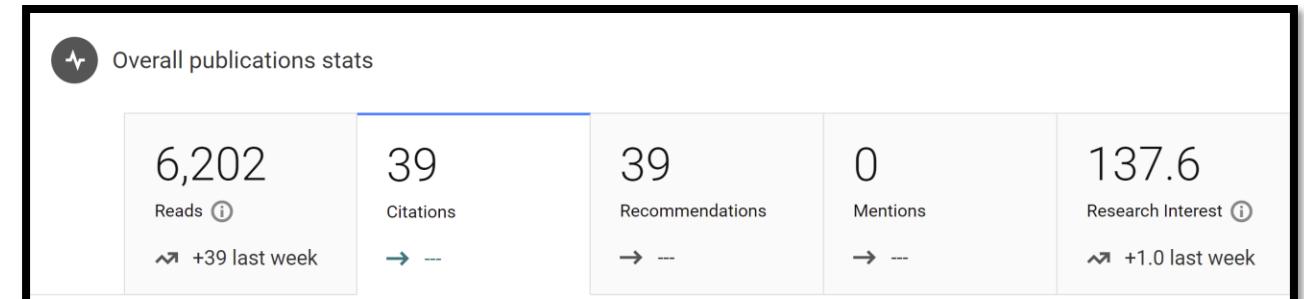
Profile Snapshot:

Mahmoud Salem, MSc, Dipl.-Eng.

- Senior ML and Robotics Scientist at KIT, Germany
- Computer Vision Consultant at EL2Labs, USA
- Former Computer Vision Scientist in MSD, Malaysia
- BSc in Electronics and Communication Engineering
- MSc in Computer Engineering
- Best Paper Awards, Rome, 2018, ACHI
- Best Paper Awards, Budapest, 2019, KES-MSD19
- More than 10 papers in ML, Computer Vision and IoT
- More than 20 project in EU, Taiwan, and Gulf in area of I4.0, IoT, Computer Vision
- 30+ Citations and 6000+ Reads
- Find more in LinkedIn and ResearchGate profile



A screenshot of Mahmoud Salem's LinkedIn profile. At the top, there's a navigation bar with icons for Home, My Network, Jobs, and Messaging. Below the header is a large circular profile picture of a man with dark hair and a beard. To the right of the profile picture, the name "Mahmoud Salem" is displayed with the pronouns "(He/Him)". Below the name, it says "Machine Learning Scientist, MSc." and "Eggenstein-Leopoldshafen, Baden-Württemberg, Germany". There is a "Contact info" link and a "500+ connections" button. On the far right, there are logos for "Karlsruhe Institute of Technology (KIT)" and "Ain Shams University".

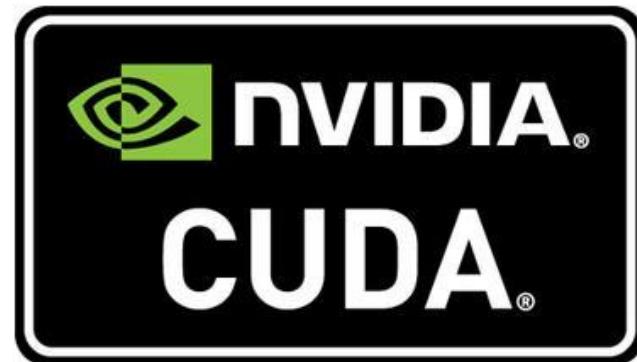
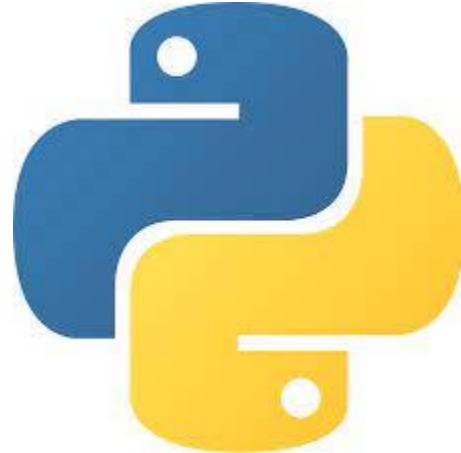


<https://www.linkedin.com/in/mahmoud-salem-ms/>

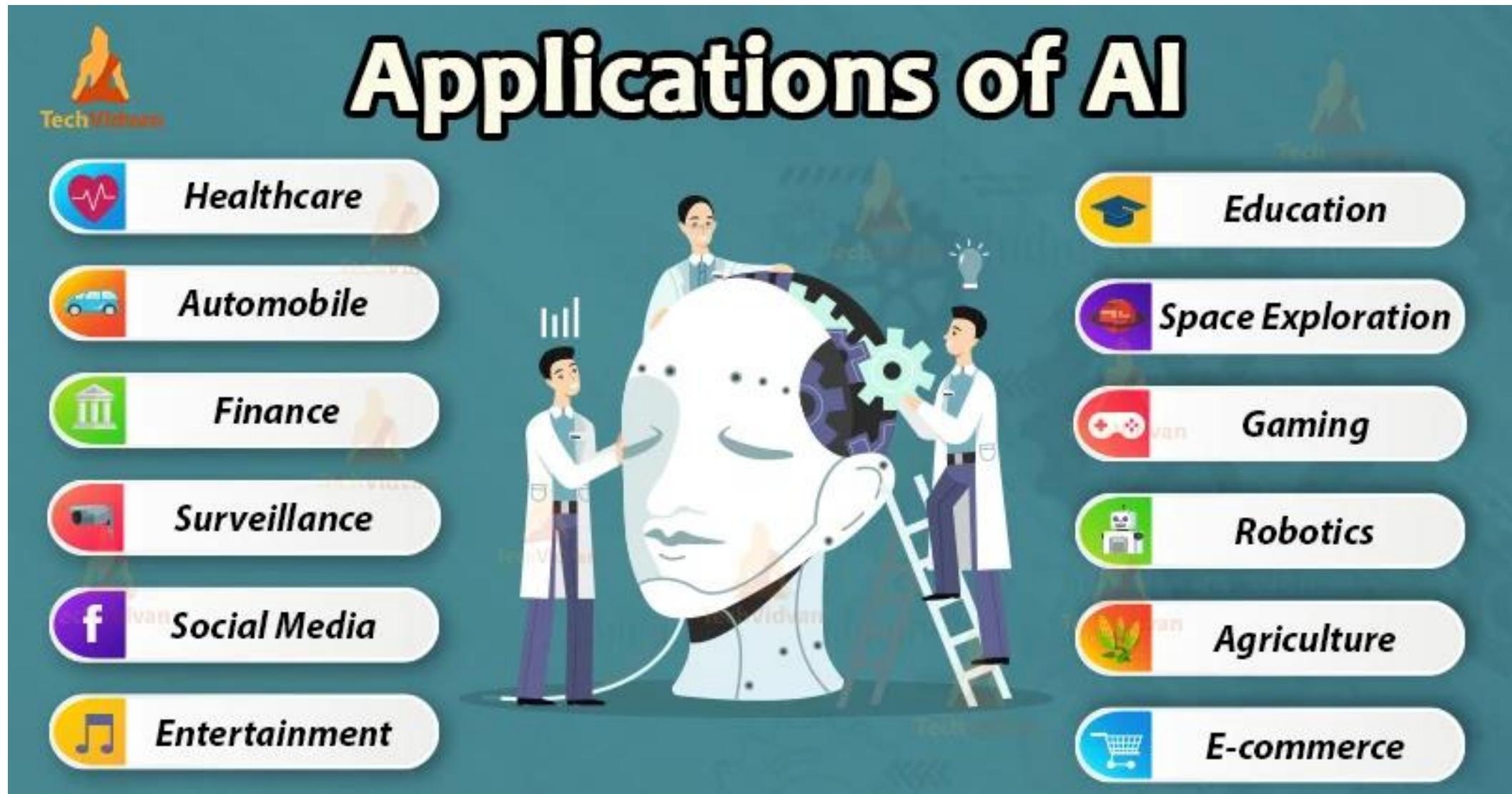
<https://www.researchgate.net/profile/Mahmoud-Salem-2>

Agenda:

- Programming
- Machine Learning Module
- Deep Learning Module
- Embedded AI



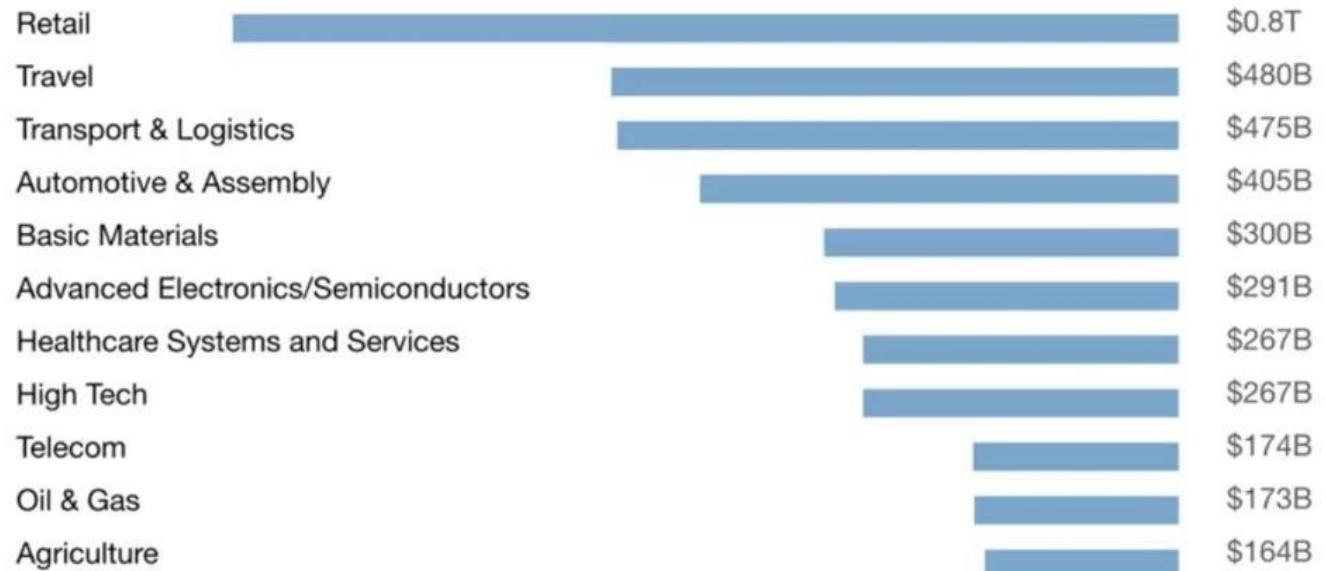
The AI revolution



What value AI is creating?

AI value creation
by 2030

\$13
trillion



[Source: McKinsey Global Institute.]

What AI can do?

LAWYERS



HR MANAGERS



MARKETERS



TEACHING ASSISTANTS



REPORTERS & EDITORS



TRADERS



ACCOUNTANTS & AUDITORS



COMPLIANCE OFFICERS



INVESTMENT MANAGERS



CRM & SALES CLERKS



RESEARCHERS & CONSULTANTS



SOFTWARE DEVELOPERS



Vision is really hard

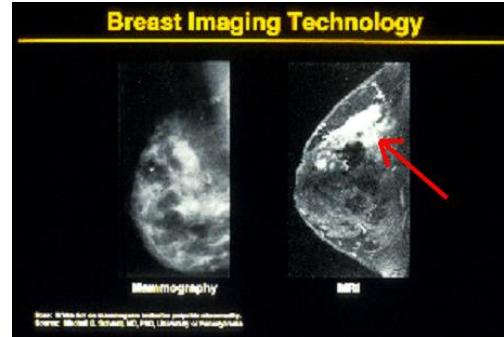
- Vision is an amazing feat of natural intelligence
 - Visual cortex occupies about 50% of Macaque brain
 - More human brain devoted to vision than anything else



Why computer vision matters



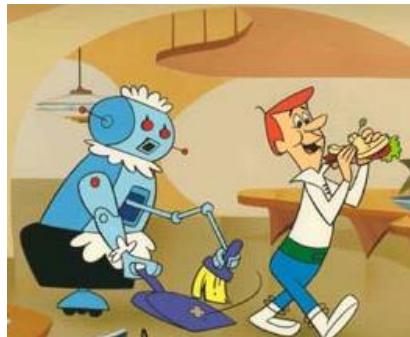
Safety



Health



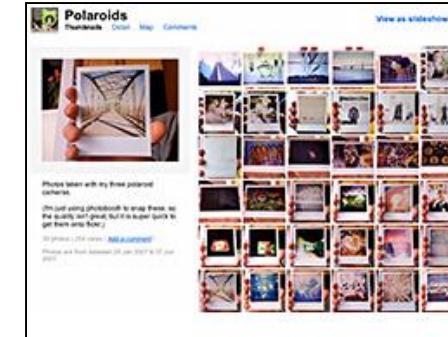
Security



Comfort



Fun

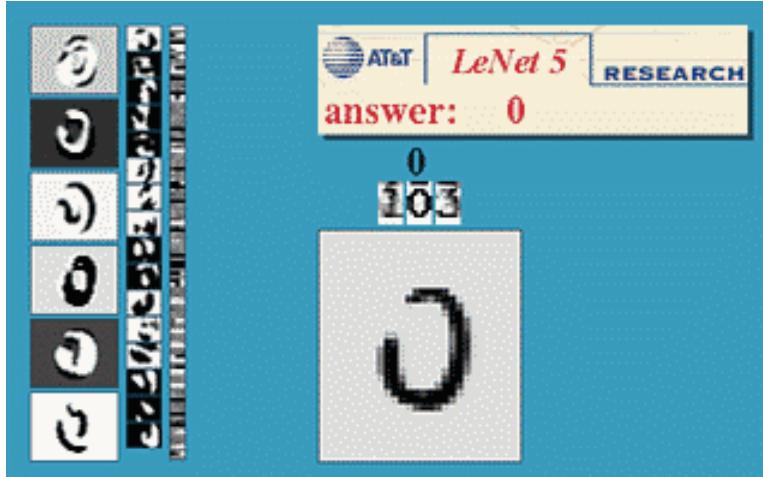


Access

Optical character recognition (OCR)

Technology to convert scanned docs to text

- If you have a scanner, it probably came with OCR software



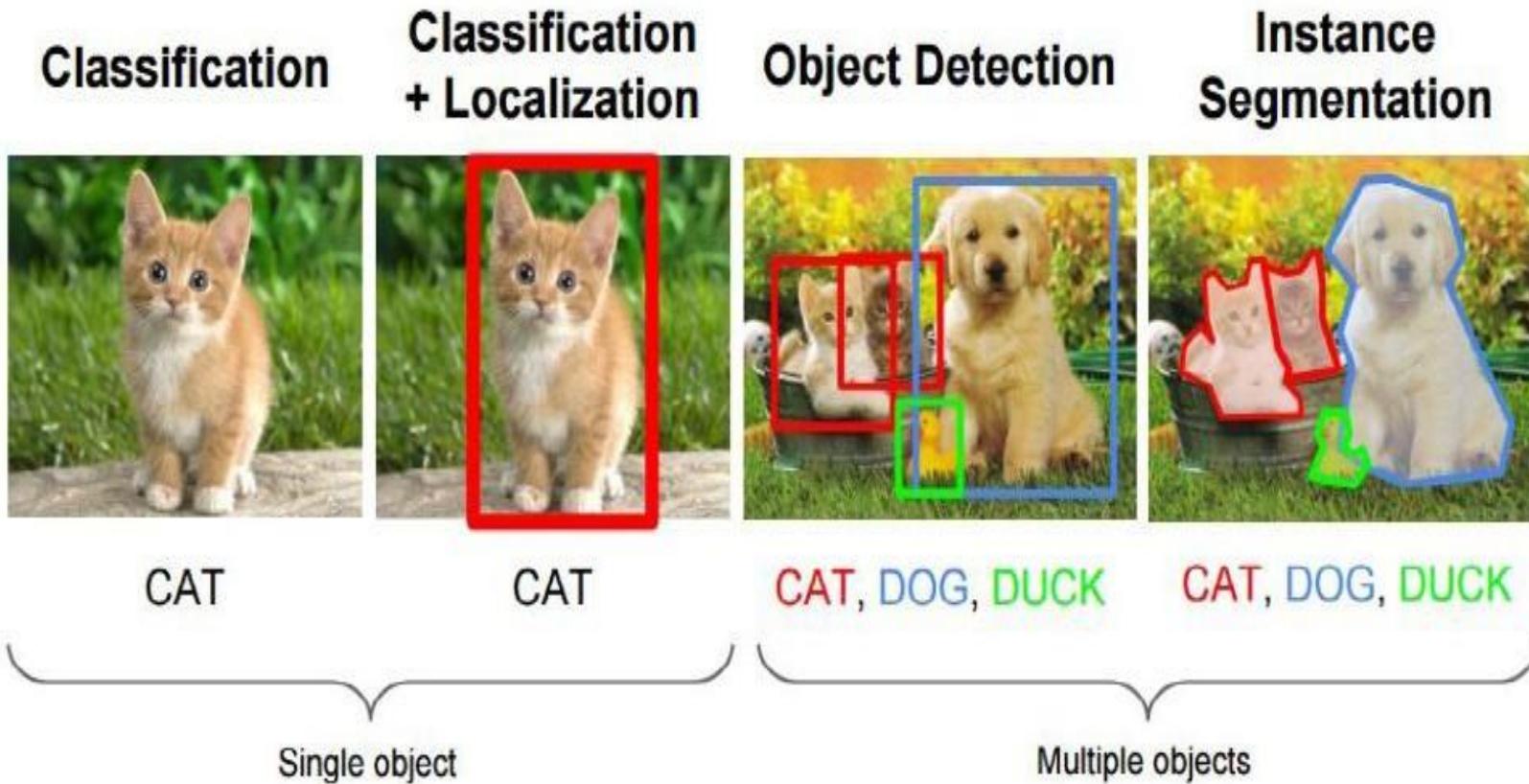
Digit recognition, AT&T labs
<http://www.research.att.com/~yann/>

Source: <https://www.slideshare.net/ShilpaSharma175/computer-vision-basics-242537412>



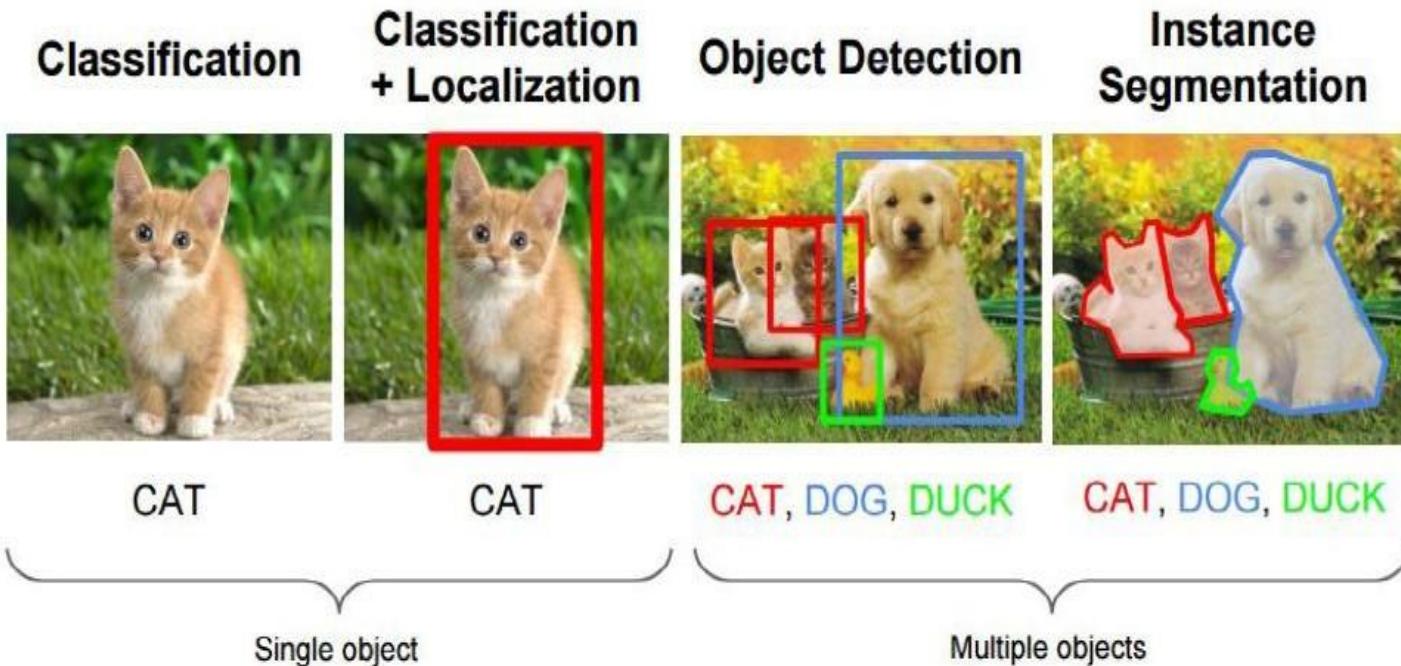
License plate readers
http://en.wikipedia.org/wiki/Automatic_number_plate_recognition

Computer vision possible tasks

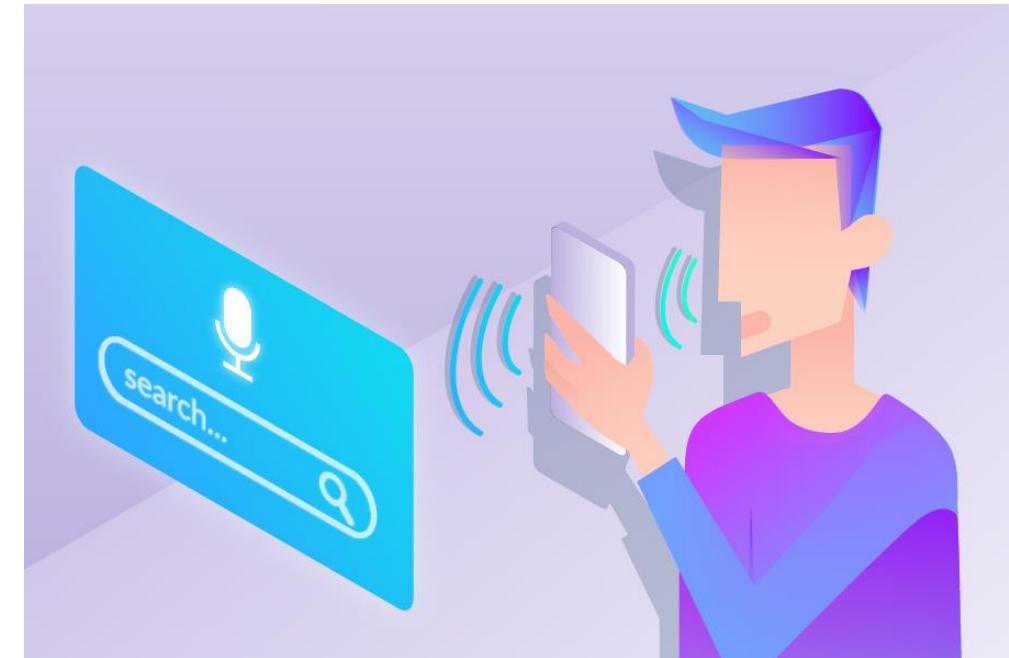


Source: Fei-Fei Li, Andrej Karpathy & Justin Johnson (2016) cs231n, Lecture 8 - Slide 8, *Spatial Localization and Detection* (01/02/2016). Available:
http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

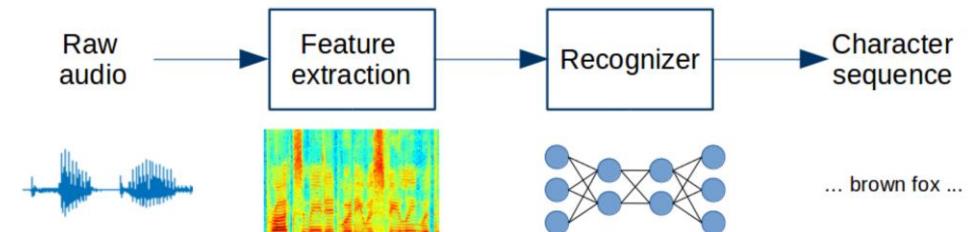
What AI can do?



Source: Fei-Fei Li, Andrej Karpathy & Justin Johnson (2016) cs231n, Lecture 8 - Slide 8, *Spatial Localization and Detection* (01/02/2016). Available:
http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf



Source: <https://blackcliffmedia.com/blog/part-2-voice-recognition-the-greatest-marketing-disruptor/>



Source: <https://www.mdpi.com/2079-9292/9/7/1157/htm>

What is AI Limitation?



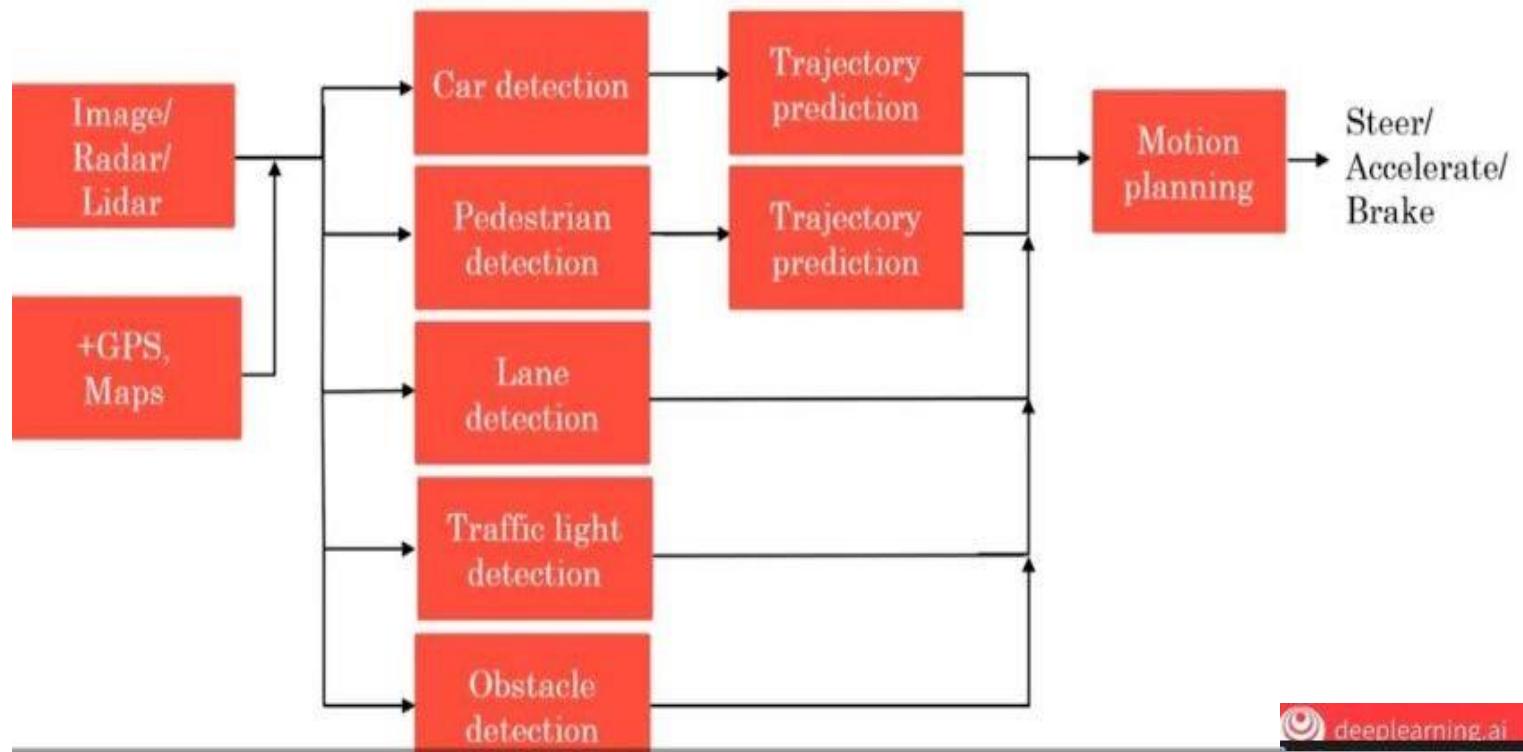
Source: <https://tomorrow.city/a/incapable-yes-artificial-intelligence-cant-do-these-things>

Workflow of an AI project

AD pipeline

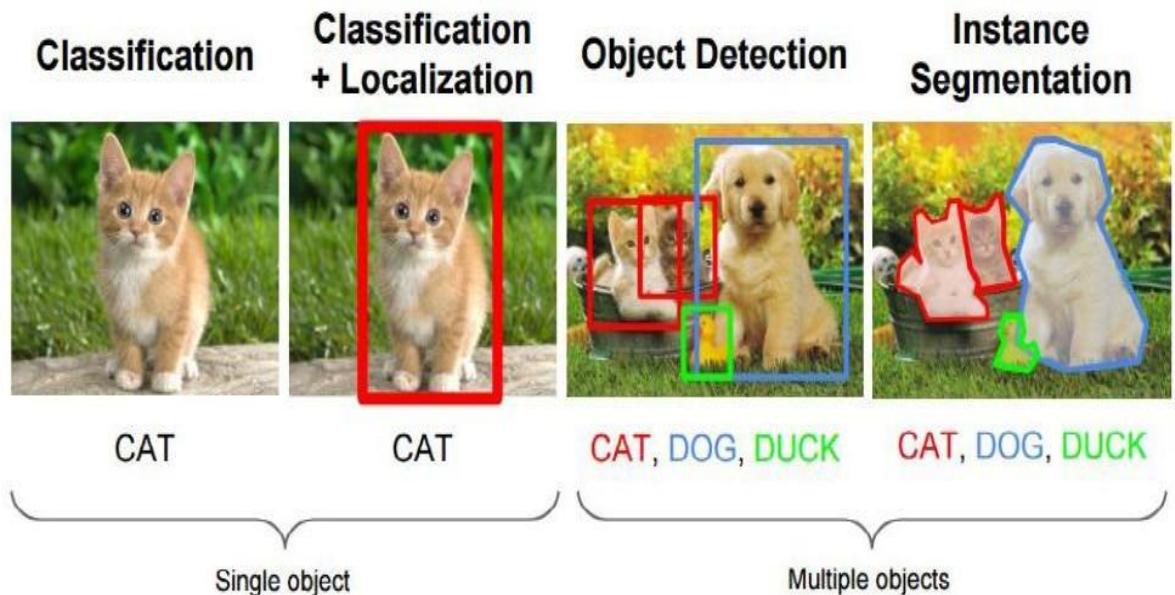
Each is an AI project: A--> B

1. Data collection
2. Data Preprocessing
3. Train
 - a. Many iterations
 - b. Hyper parameters optimization
4. Deployment



How to get your dataset?

1. Raw data collection, where?
 - a. From users experience (early deployment)
 - b. Internet of Things
 - c. Automotive
2. Annotation = Labeling
 - a. Manual
 - b. Automatic → Observe behaviors
 - c. Semi-automatic



Source: Fei-Fei Li, Andrej Karpathy & Justin Johnson (2016) cs231n, Lecture 8 - Slide 8, *Spatial Localization and Detection* (01/02/2016). Available:
http://cs231n.stanford.edu/slides/2016/winter1516_lecture8.pdf

Data Treatment tips and tricks

Integrate AI teams

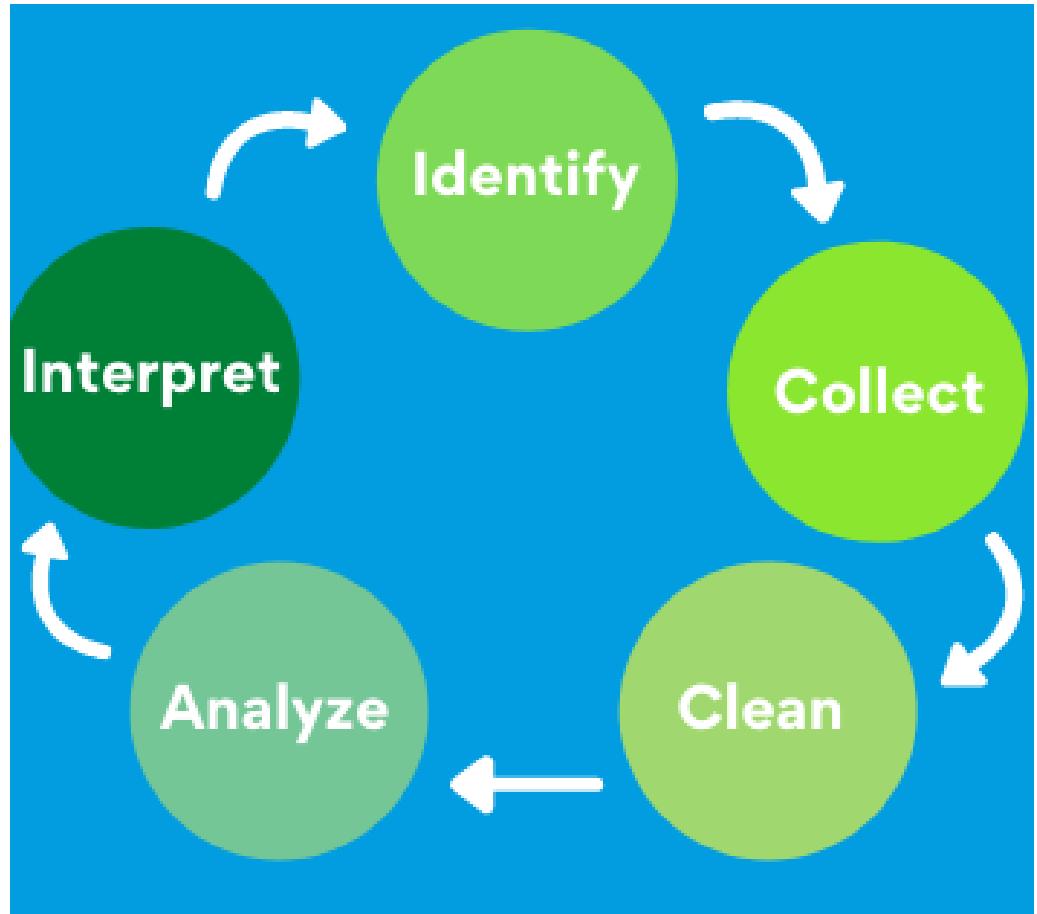
- Quick feedback
- Set data collection requirements and quality

Quick Deployment

- Do not wait years to collect all the data
- Quick feedback

How big?

- The bigger the better → MB, GB, TB, PB
- Problem dependent
- Rule of thumb, 5k/class → if classification



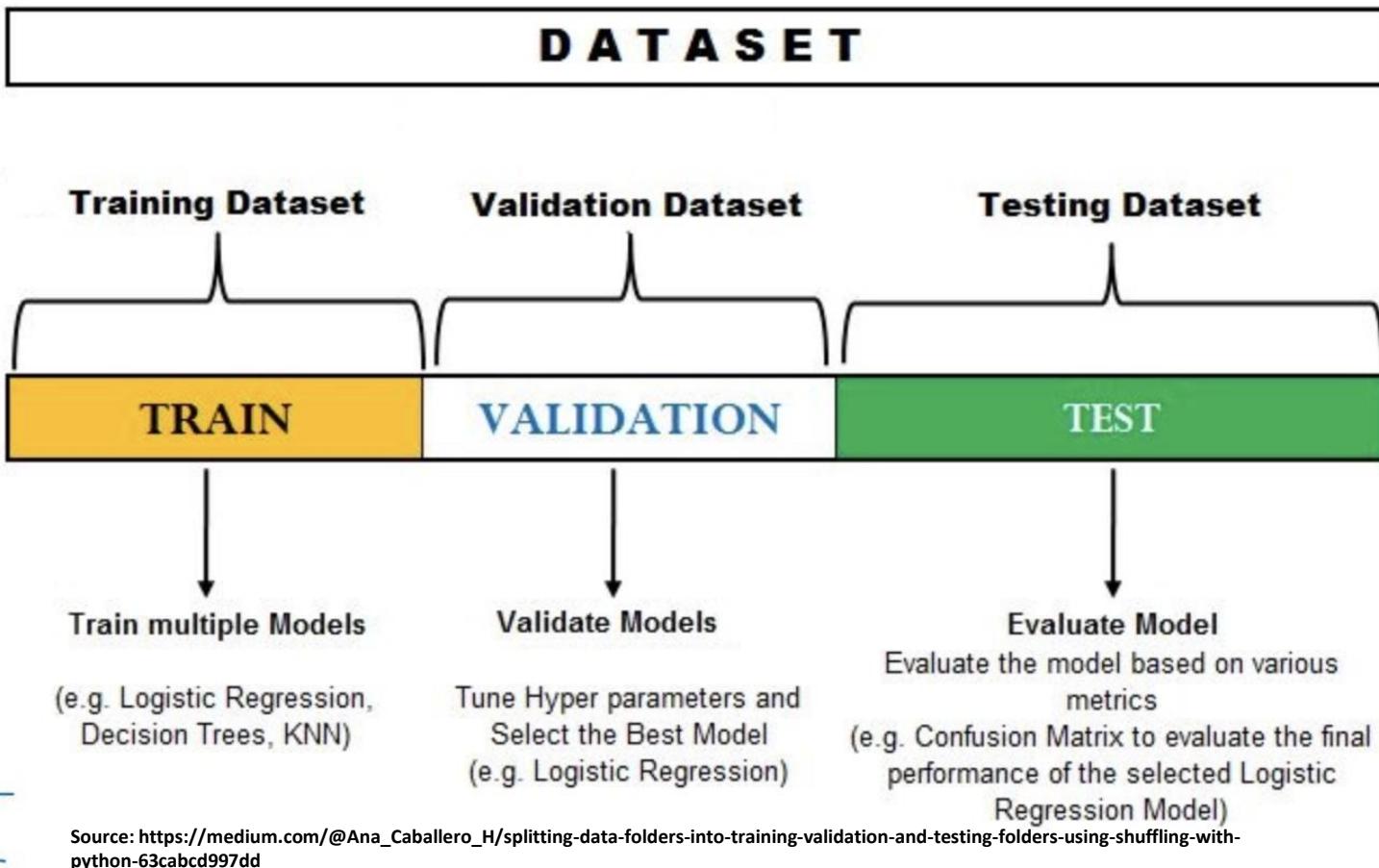
Source: <https://www.datapine.com/blog/data-analysis-methods-and-techniques/>

Datasets splitting?

- Training data
- Testing data
- Validation data

Test set	ok	defect	ok
	ok		defect
	ok		ok defect?
	defect		ok

- Limitations of ML
- Insufficient data ←
- Mislabeled data ←
- Ambiguous labels ←



100% is not possible

- Garbage in - Garbage out
 - o Incorrect labels
 - o Missing values
 - o Ambiguous labels
- ML limitations
 - o Statistical methods
 - o Traditional ways fails due to many cases



Terminologies in Computer Vision

Dog



Image Classification:

recognize an object in an image.

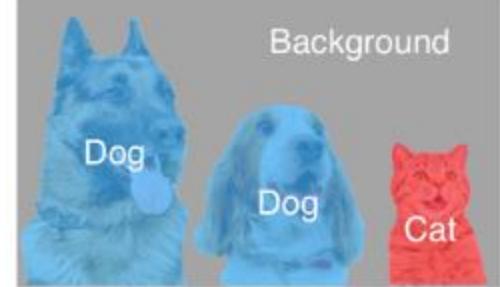
Dog



Object Detection:

detect multiple objects with their bounding boxes in an image.

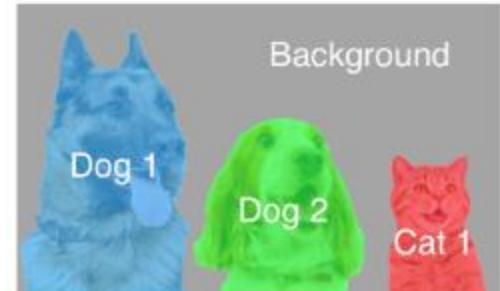
Background



Semantic Segmentation:

associate each pixel of an image with a categorical label.

Background



Instance Segmentation:

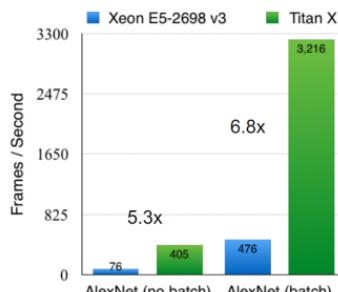
associate each pixel of an image with an instance label.

AI Technical tools

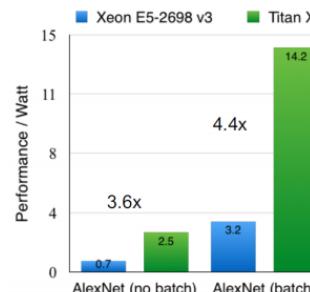
- Machine Learning frameworks
 1. TensorFlow
 2. Pytorch
 3. Keras
 4. Etc
- GPU vs CPU

Comparing CPU and GPU - server class

Xeon E5-2698 and Tesla M40

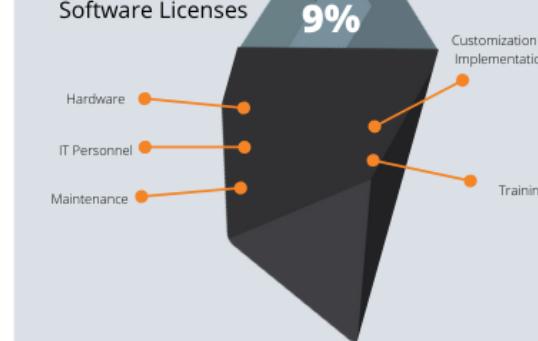


Source: <https://www.nowakultura-warmia.org/gpu-vs-cpu-machine-learning>

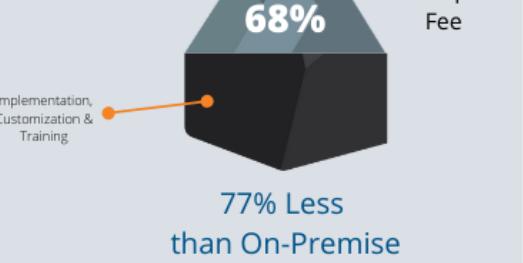


Looking Beneath the Surface

On Premise Software



Cloud Computing



"Customers can spend up to four times the cost of their software licenses per year to own and manage their applications." -Gartner
"The End of Software"

Limitations of AI

Bias

Gender bias

Machine Learning can amplify bias.



Language

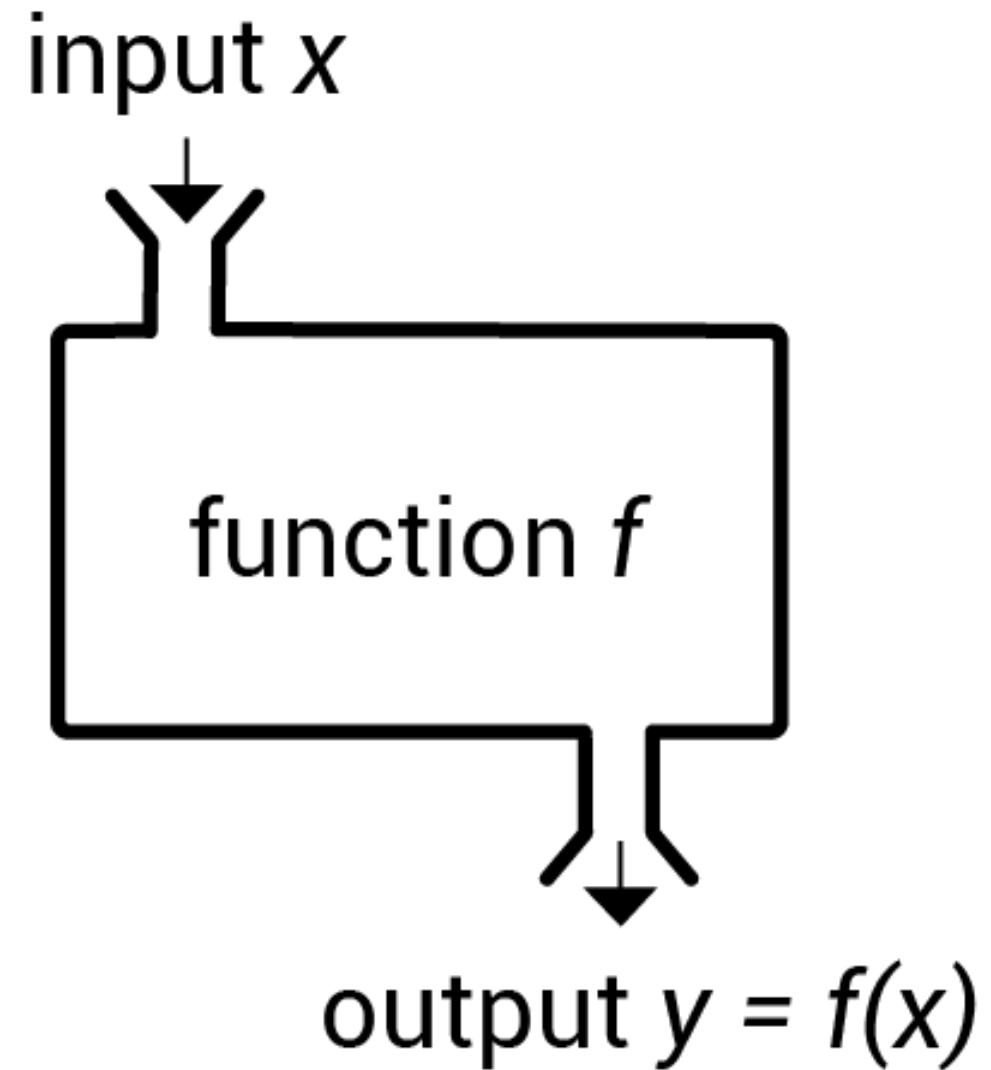
The screenshot shows a machine translation interface comparing English, Turkish, and Spanish. It highlights how gender bias is amplified in the target languages.

Top Translation: She is a doctor.
O bir doktor.
He is a nurse.
O bir hemşire.

Bottom Translation: O bir doktor.
O bir hemşire
He is a doctor.
She is a nurse.

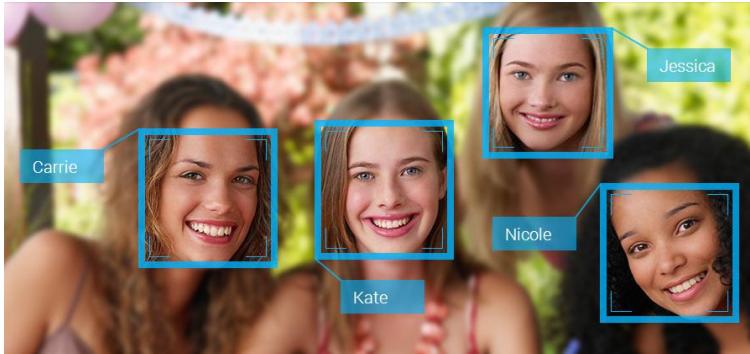


What AI can do?

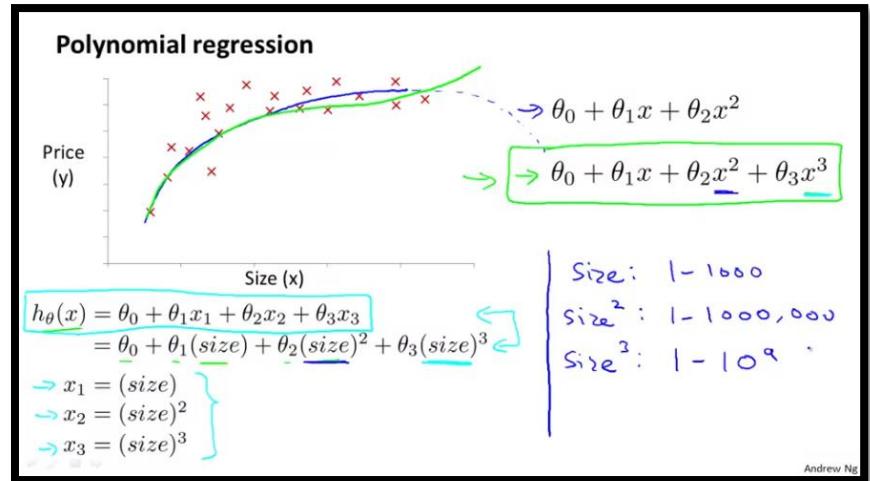
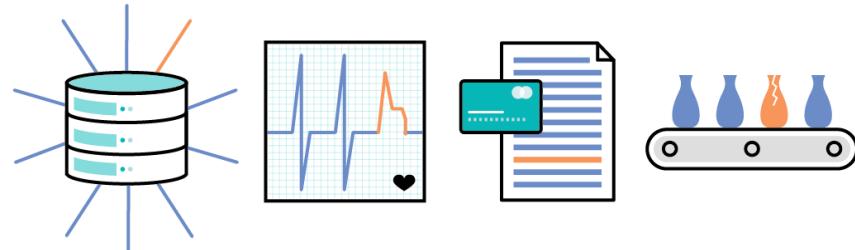
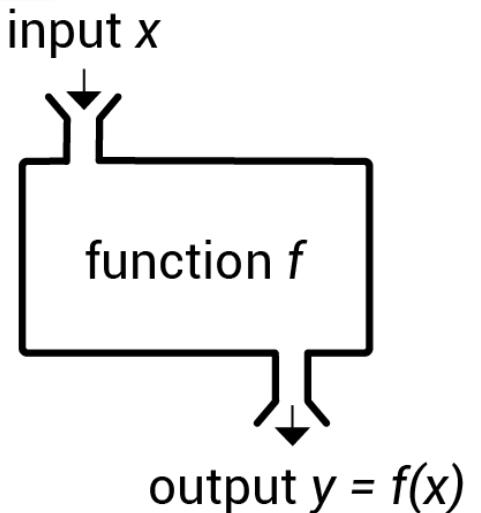


What is AI is best at today?

Supervised Learning:



Price	Floor space	Rooms	Lot size	Appartment	Row house	Corner house	Detached
250000	71	4	92	0	1	0	0
209500	98	5	123	0	1	0	0
349500	128	6	114	0	1	0	0
250000	86	4	98	0	1	0	0
419000	173	6	99	0	1	0	0
225000	83	4	67	0	1	0	0
549500	165	6	110	0	1	0	0
240000	71	4	78	0	1	0	0
340000	116	6	115	0	1	0	0



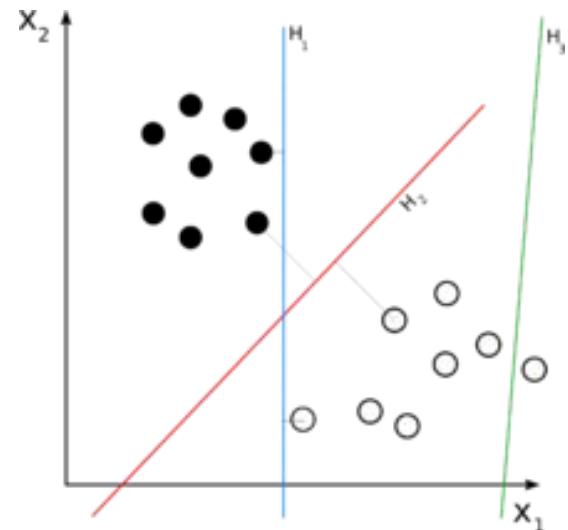
ML: How to learn w's?

Assume only 2 q's:

Basically → Search: ax_1+bx_2+c (**Model**) → $w_1=a$,
 $w_2=b$, $w_0=c$ (bias, to be clear later)

Try many lines, and get the one that separates the
Data better → How good? **Loss**

Practically → Smarter methods (**Optimizer**) are
used better than brute force or random search!



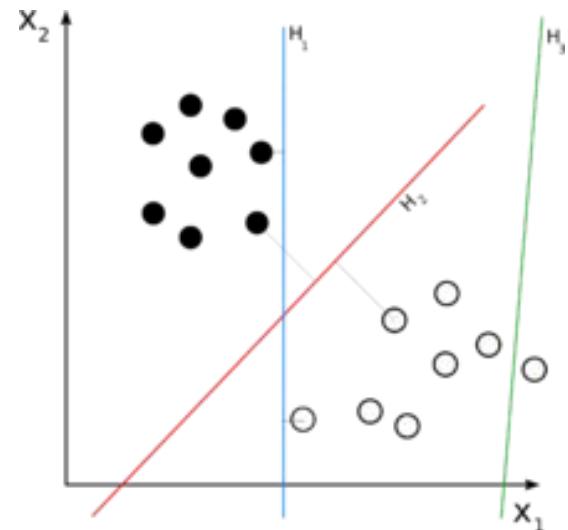
ML: How to learn w's?

Assume only 2 q's:

Basically → Search: ax_1+bx_2+c (**Model**) → $w_1=a$,
 $w_2=b$, $w_0=c$ (bias, to be clear later)

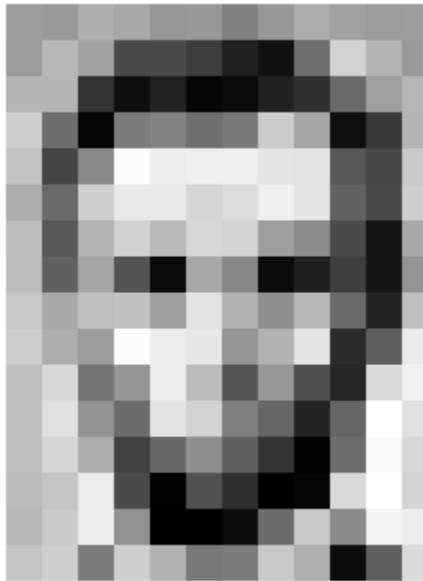
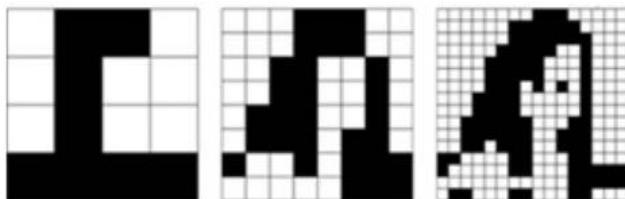
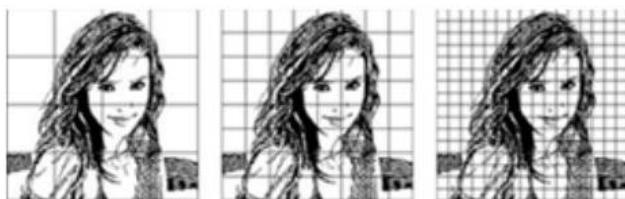
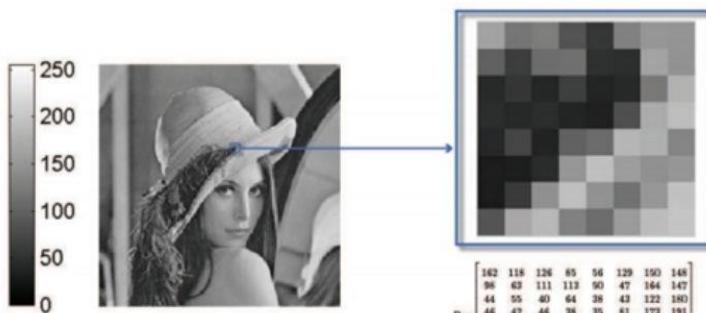
Try many lines, and get the one that separates the
Data better → How good? **Loss**

Practically → Smarter methods (**Optimizer**) are
used better than brute force or random search!



How can computer see?

https://colab.research.google.com/drive/1lwhBkAvgBG0QiBwGPJb2FbOXtz_sQyBW?authuser=1#scrollTo=kJbwGQ5QhYXa



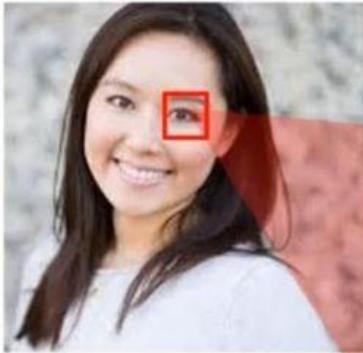
157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	105	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	257	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	199	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	105	36	190
205	174	158	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	35	101	255	224
190	214	173	66	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	199	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	105	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	85	150	79	38	218	241
190	224	147	108	227	210	127	102	35	101	255	224
190	214	173	66	103	143	95	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

Source: <https://medium.com/swlh/computer-vision-how-can-our-computers-see-and-make-sense-out-of-what-they-see-f6dd777aed07>

Source: <https://santandergto.com/en/computer-vision-vs-human-vision-this-is-how-computers-see/>

Semantic gap - What the computer can see?



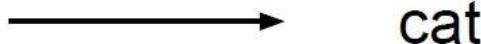
30	32	22	12	10	10	12	33	35	30
12	11	12	234	170	176	13	15	12	12
234	222	220	230	200	222	230	234	56	78
190	220	186	112	110	110	112	180	30	32
49	250	250	250	4	2	254	200	44	6
55	250	250	250	3	1	250	245	25	3
189	195	199	150	110	110	182	190	199	55
200	202	218	222	203	200	200	208	215	222
219	215	220	220	222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

Image Classification: A core task in Computer Vision

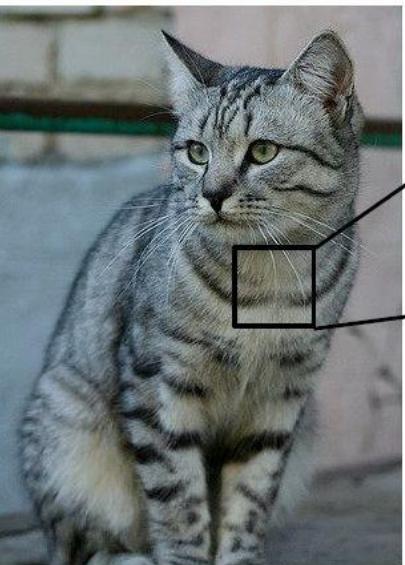


This image by Nikita is
licensed under CC-BY 2.0

(assume given set of discrete labels)
{dog, cat, truck, plane, ...}



The Problem: Semantic Gap



This image by Nikita is
licensed under CC-BY 2.0.

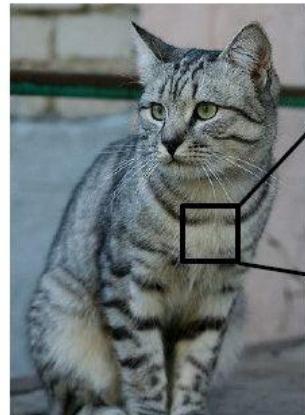
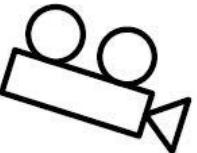
```
[165 112 108 111 104 99 105 99 96 103 112 119 104 97 93 87]  
[ 91 98 102 106 104 79 98 103 99 105 123 136 110 105 94 85]  
[ 76 85 90 105 128 105 87 96 95 99 115 112 105 103 99 85]  
[ 99 81 81 93 128 131 127 104 95 98 102 98 96 97 101 94]  
[105 91 61 64 69 91 81 85 101 107 109 98 75 84 96 95]  
[114 108 85 55 55 69 64 54 64 87 112 129 98 74 84 91]  
[133 137 147 103 65 83 88 65 52 54 74 84 102 93 85 82]  
[128 137 144 140 109 95 85 70 62 65 63 63 60 73 86 101]  
[125 133 148 137 119 121 117 94 65 79 80 65 54 64 72 98]  
[127 125 131 147 133 127 126 131 111 96 89 75 61 64 72 84]  
[115 114 109 123 158 148 131 118 113 109 100 92 74 65 72 78]  
[ 89 93 90 97 100 147 131 118 113 114 113 106 106 95 77 80]  
[ 63 77 86 81 77 79 102 123 117 115 117 125 125 130 115 87]  
[ 62 65 82 89 78 71 88 101 124 126 119 101 107 114 131 119]  
[ 63 65 75 86 89 71 62 81 128 138 135 105 81 98 110 116]  
[ 87 65 71 87 106 95 69 45 76 138 126 107 92 94 105 112]  
[118 97 82 86 117 123 115 66 41 51 95 93 89 95 102 107]  
[164 146 112 88 82 120 124 104 76 48 45 56 88 101 102 109]  
[157 170 157 128 93 86 114 132 112 97 69 55 78 82 99 94]  
[130 128 134 151 139 100 109 118 121 134 114 87 65 53 69 86]  
[128 112 96 117 158 144 120 115 104 107 102 93 87 81 72 79]  
[123 107 96 86 83 112 153 149 122 109 104 75 88 107 112 99]  
[122 121 102 80 82 86 94 117 145 148 153 102 58 70 92 107]  
[122 154 148 103 71 56 78 83 93 103 119 139 102 61 69 84]
```

What the computer sees

An image is just a big grid of
numbers between [0, 255]:

e.g. 800 x 600 x 3
(3 channels RGB)

Challenges: Viewpoint variation



[1] 105 112 108 111 204 99 106 99 96 103 112 119 104 97 93 87]
[2] 91 98 102 106 204 79 98 103 99 105 123 136 110 105 94 85]
[3] 90 83 81 85 93 128 131 127 109 95 98 102 99 96 93 103 94]
[4] 89 83 81 85 93 128 131 127 109 95 98 102 99 96 93 103 94]
[5] 106 93 61 63 69 60 68 88 85 101 107 106 98 75 84 66 95]
[6] 114 106 85 85 85 60 64 54 64 87 112 120 98 74 84 91]
[7] 122 117 107 107 107 65 66 62 63 64 65 66 67 68 69 70 71]
[8] 128 137 144 148 149 65 66 68 62 65 63 63 69 73 86 101]
[9] 125 133 148 137 119 121 117 91 65 75 88 65 54 64 72 98]
[10] 127 125 131 147 133 127 126 131 111 99 89 75 61 64 72 94]
[11] 129 135 141 142 131 119 119 119 100 99 99 75 61 64 72 94]
[12] 109 93 98 97 289 142 131 151 113 154 113 160 106 65 77 89]
[13] 63 77 86 81 77 79 102 121 117 115 117 125 125 136 115 87]
[14] 62 65 82 89 78 71 88 101 124 126 119 101 107 111 112 119]
[15] 63 65 82 89 78 71 88 101 124 126 119 101 107 111 112 119]
[16] 67 65 73 87 100 95 69 45 79 138 126 107 92 90 105 112]
[17] 65 73 87 100 95 69 45 79 138 126 107 92 90 105 112
[18] 97 92 82 86 117 123 116 69 41 51 95 93 93 95 102 107
[19] 97 92 82 86 117 123 116 69 41 51 95 93 93 95 102 107
[20] 146 112 88 82 128 134 103 76 49 45 66 88 101 102 103]
[21] 128 134 141 142 131 119 119 119 100 99 99 75 61 64 72 94]
[22] 128 134 141 139 108 109 118 121 134 114 87 65 53 69 86]
[23] 128 122 96 111 158 144 128 115 104 107 102 93 97 81 72 79]
[24] 123 101 96 89 83 112 112 141 122 108 104 75 89 107 112 99]
[25] 121 101 96 89 83 112 112 141 122 108 104 75 89 107 112 99]
[26] 122 104 146 103 71 86 78 43 93 103 119 139 102 61 69 84]

All pixels change when
the camera moves!

This image by Nikita is
licensed under CC-BY 2.0.

Challenges: Illumination



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)

Challenges: Deformation



[This image by Umberto Salvagnin](#)
is licensed under CC-BY 2.0



[This image by Umberto Salvagnin](#)
is licensed under CC-BY 2.0



[This image by sare bear](#)
is licensed under CC-BY 2.0



[This image by Tom Thai](#)
is licensed under CC-BY 2.0

Challenges: Occlusion



[This image is CC0 1.0 public domain](#)



[This image is CC0 1.0 public domain](#)



[This image by jonsson is licensed under CC-BY 2.0](#)

Semantic gap in text - What computers can read?

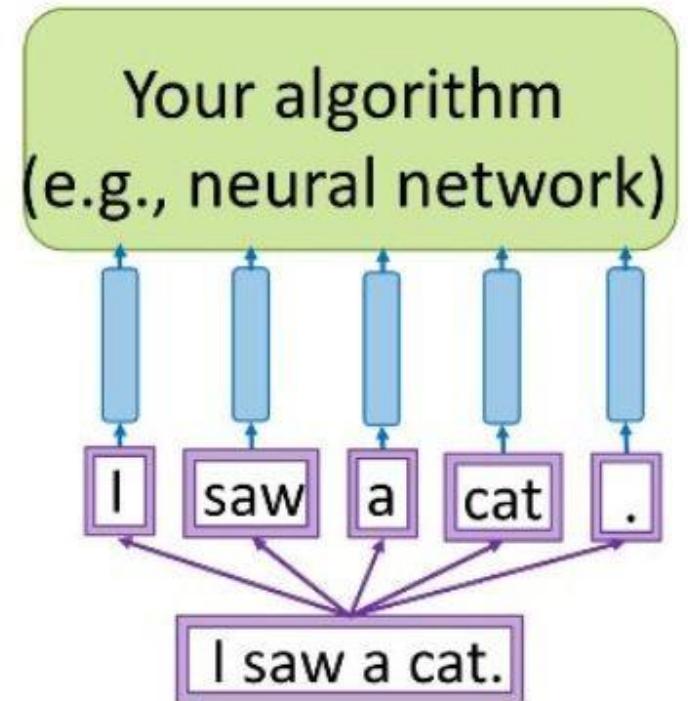
In images, we have matrix representation

In text → ??

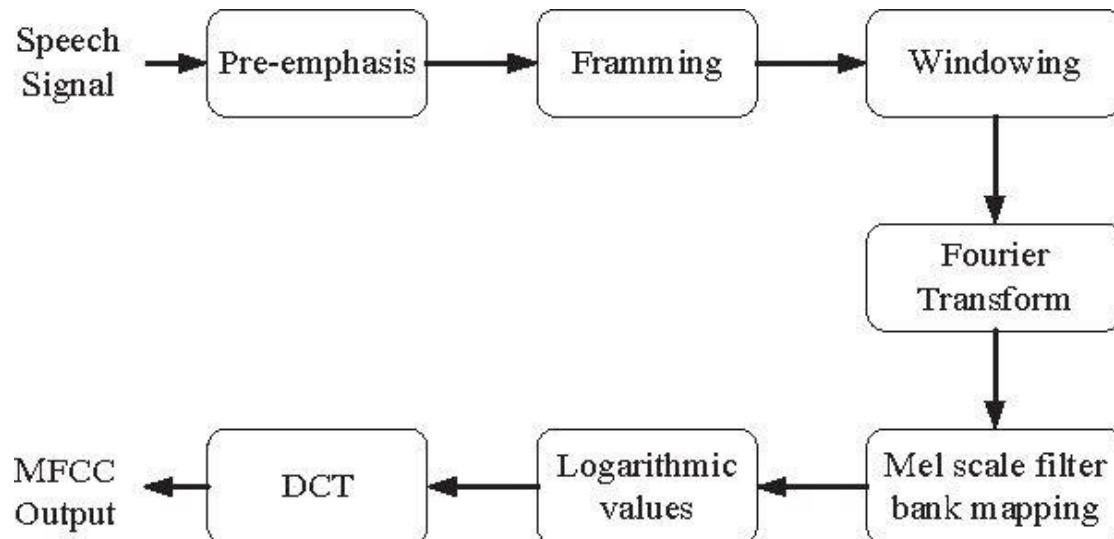
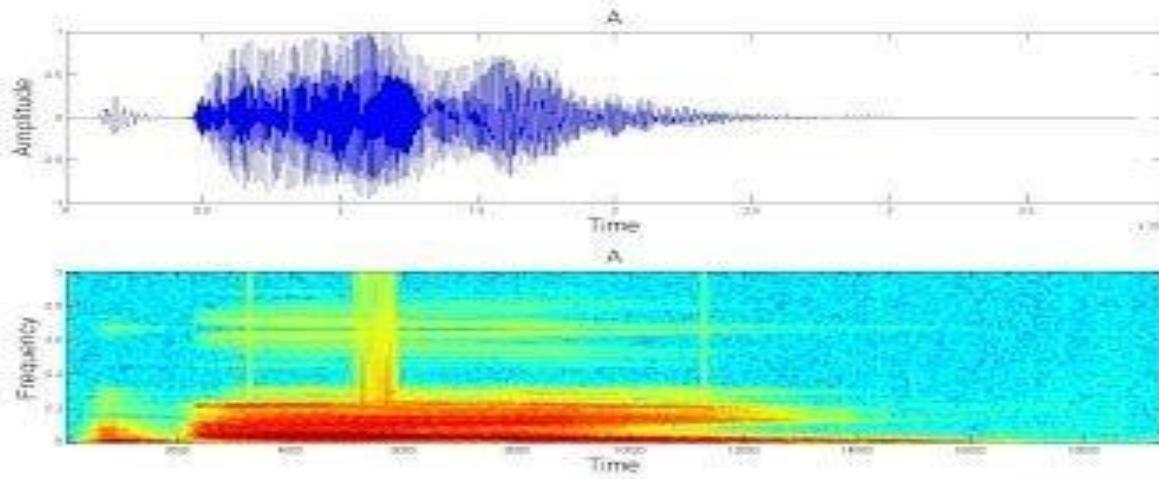
→ Naive = BoW against vocab

Raw Text		Bag-of-words vector	
it	2		
they	0		
puppy	1		
and	1		
cat	0		
aardvark	0		
cute	1		
extremely	1		
...	...		

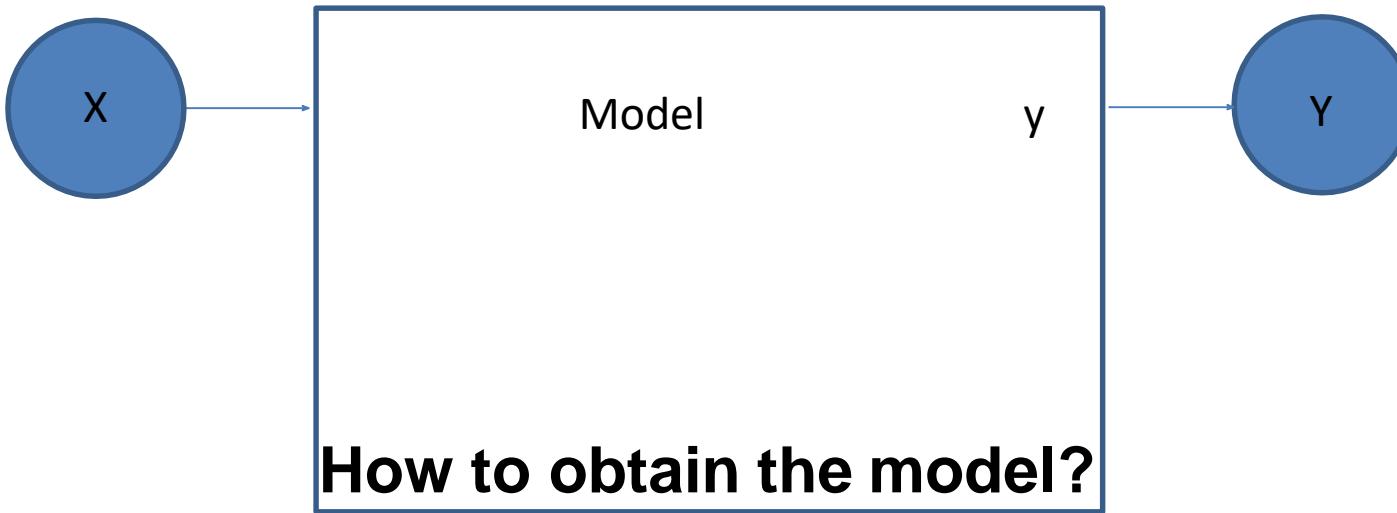
it is a puppy and it
is extremely cute



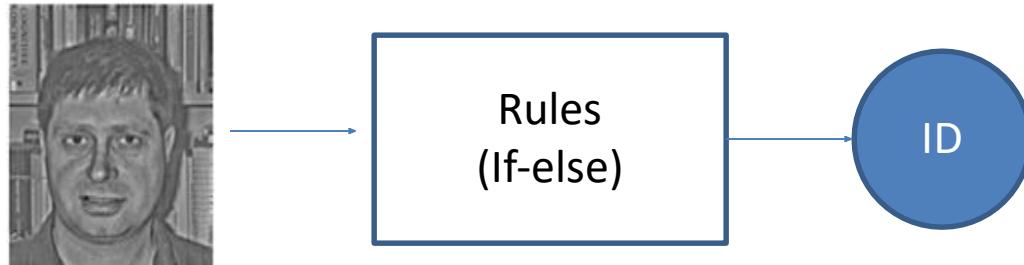
Semantic gap in speech - What computers can hear?



How to close the Gap?

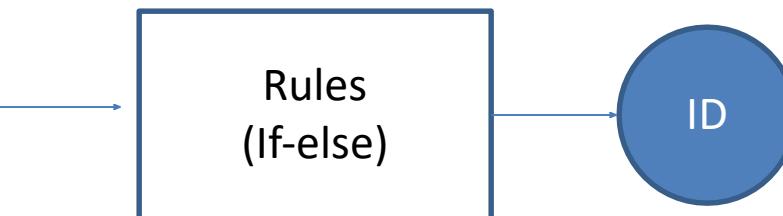


Face recognition with traditional AI



Rule based AI

Face recognition with traditional AI



Did you handle Clutter?



Did you handle scale?



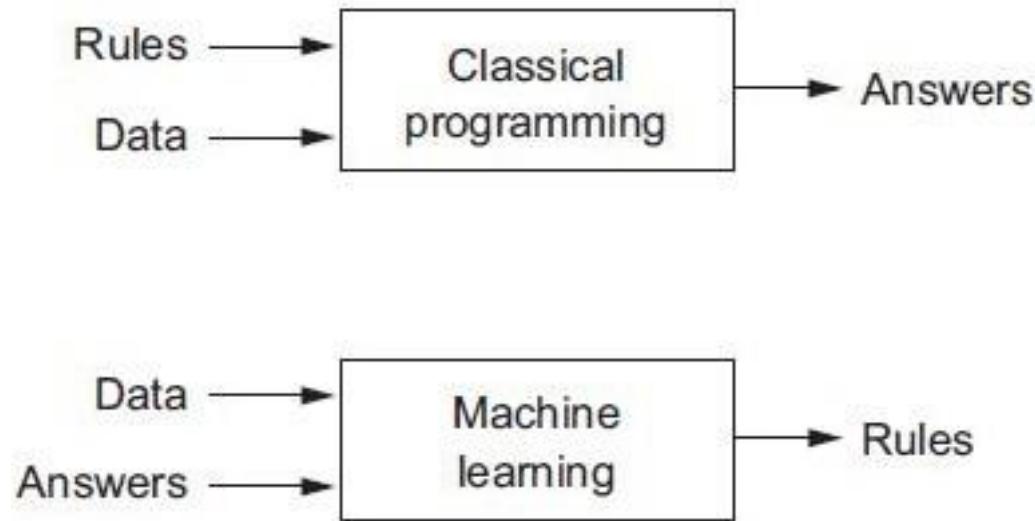
Did you handle pose?

Did you handle colors?

Too many cases!!

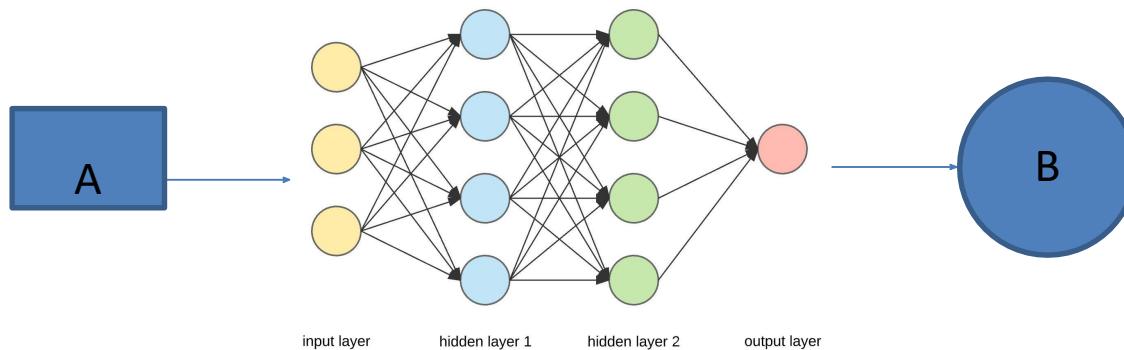
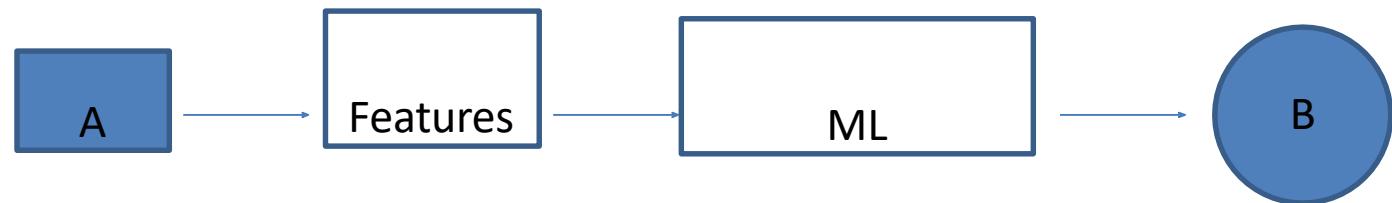
Let's Learn the rule!

Rule based AI vs. ML

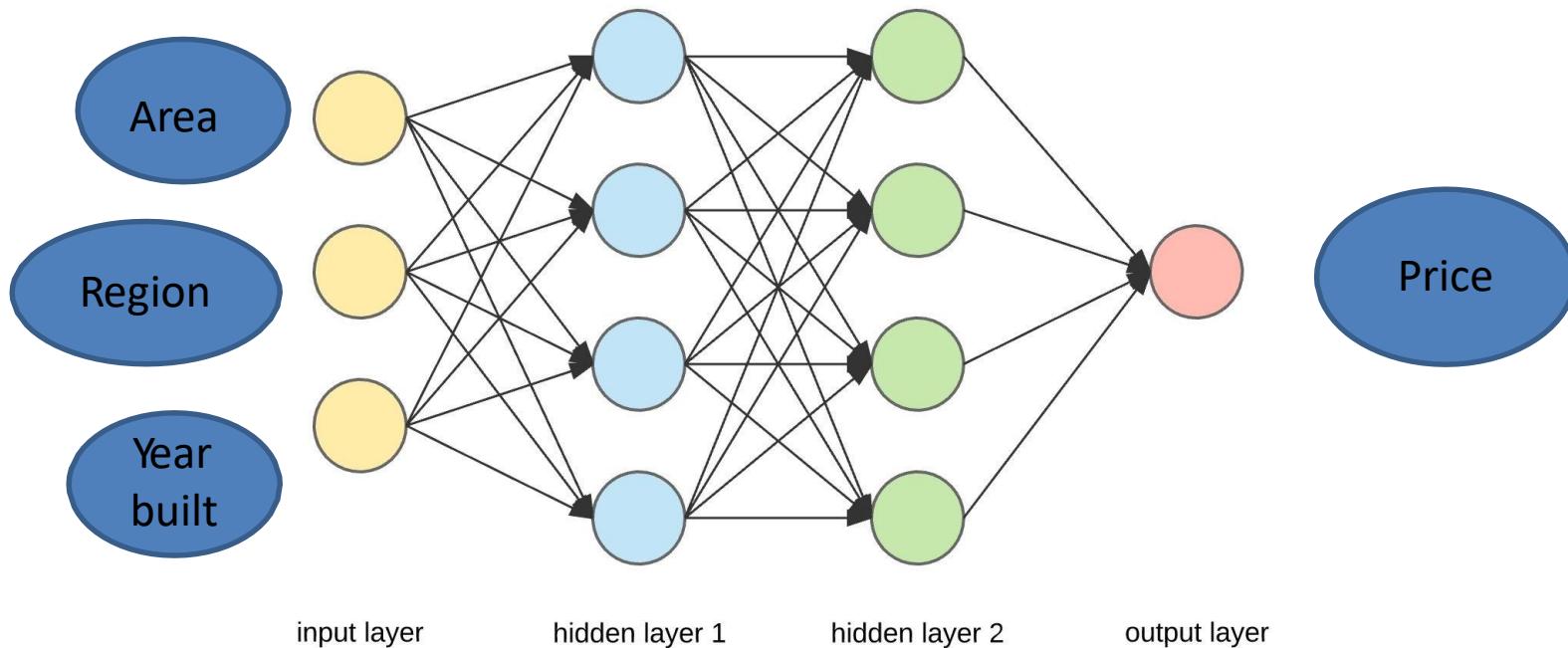


“Machine learning is the science of getting computers to act without being explicitly programmed”, Arthur Samuel 1959

Deep Learning vs. Machine Learning

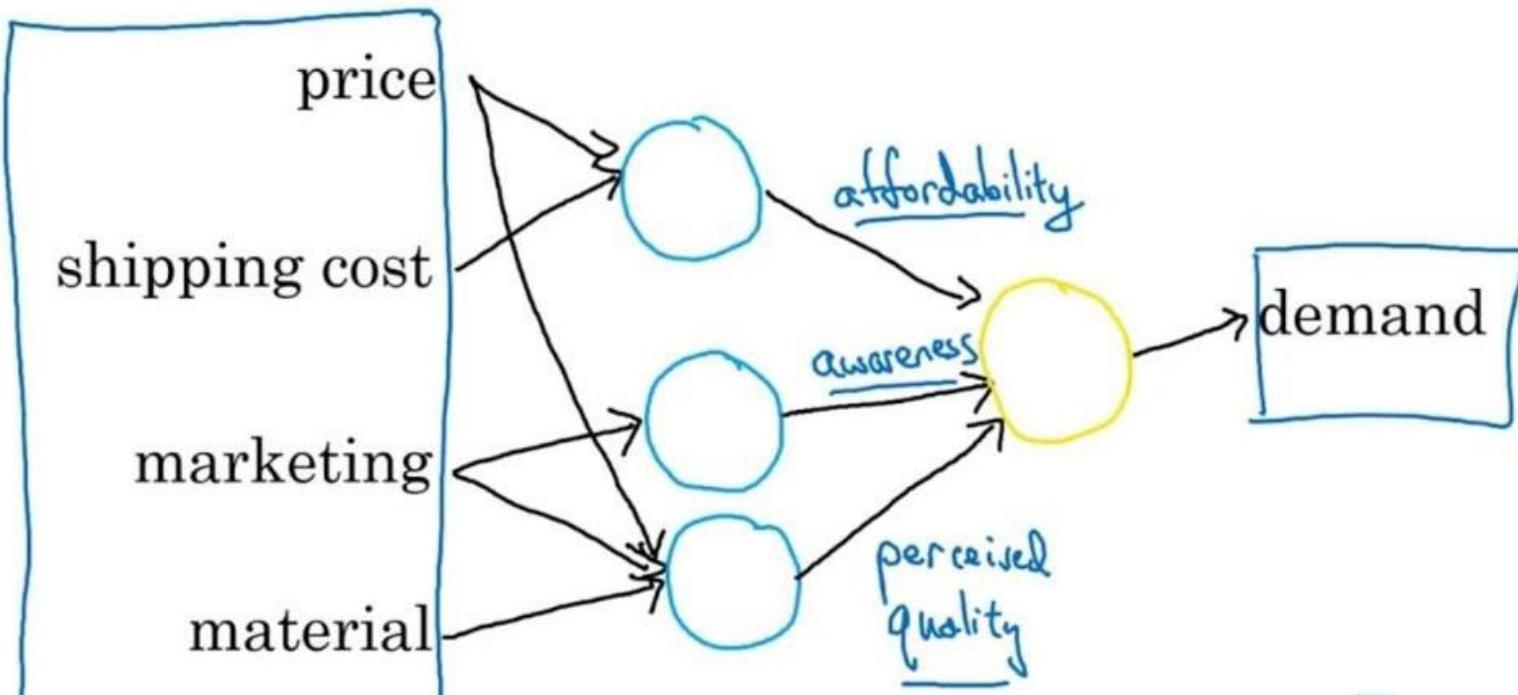


Deep Learning vs. Machine Learning



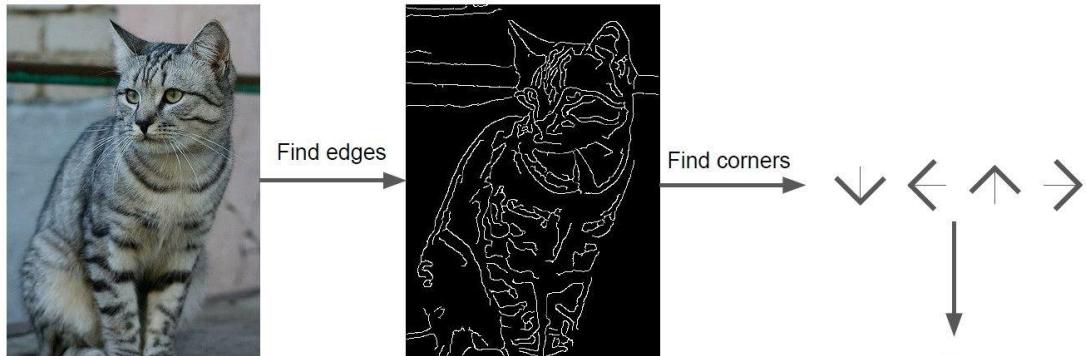
Price	Floor space	Rooms	Lot size	Appartment	Row house	Corner house	Detached
250000	71	4	92	0	1	0	0
209500	98	5	123	0	1	0	0
349500	128	6	114	0	1	0	0
250000	86	4	98	0	1	0	0
419000	173	6	99	0	1	0	0
225000	83	4	67	0	1	0	0
549500	165	6	110	0	1	0	0
240000	71	4	78	0	1	0	0
340000	116	6	115	0	1	0	0

What neurons represent?



ML = manually engineered features vector But learn the weights

Attempts have been made

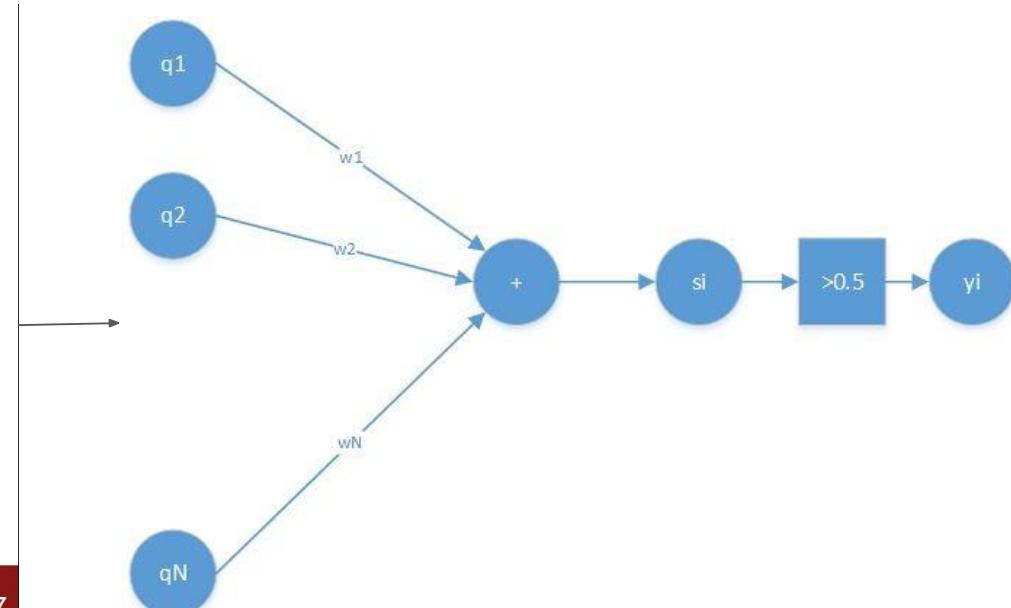


John Canny, "A Computational Approach to Edge Detection", IEEE TPAMI 1986

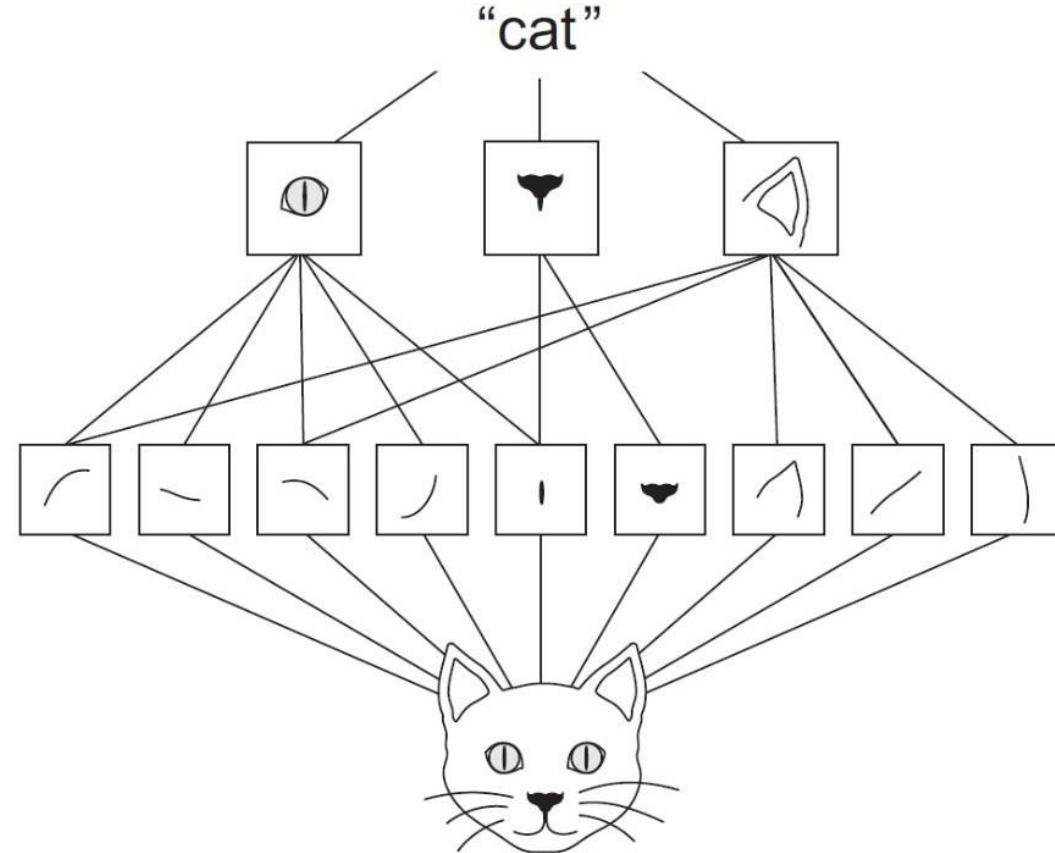
Fei-Fei Li & Justin Johnson & Serena Yeung

Lecture 2 - 15

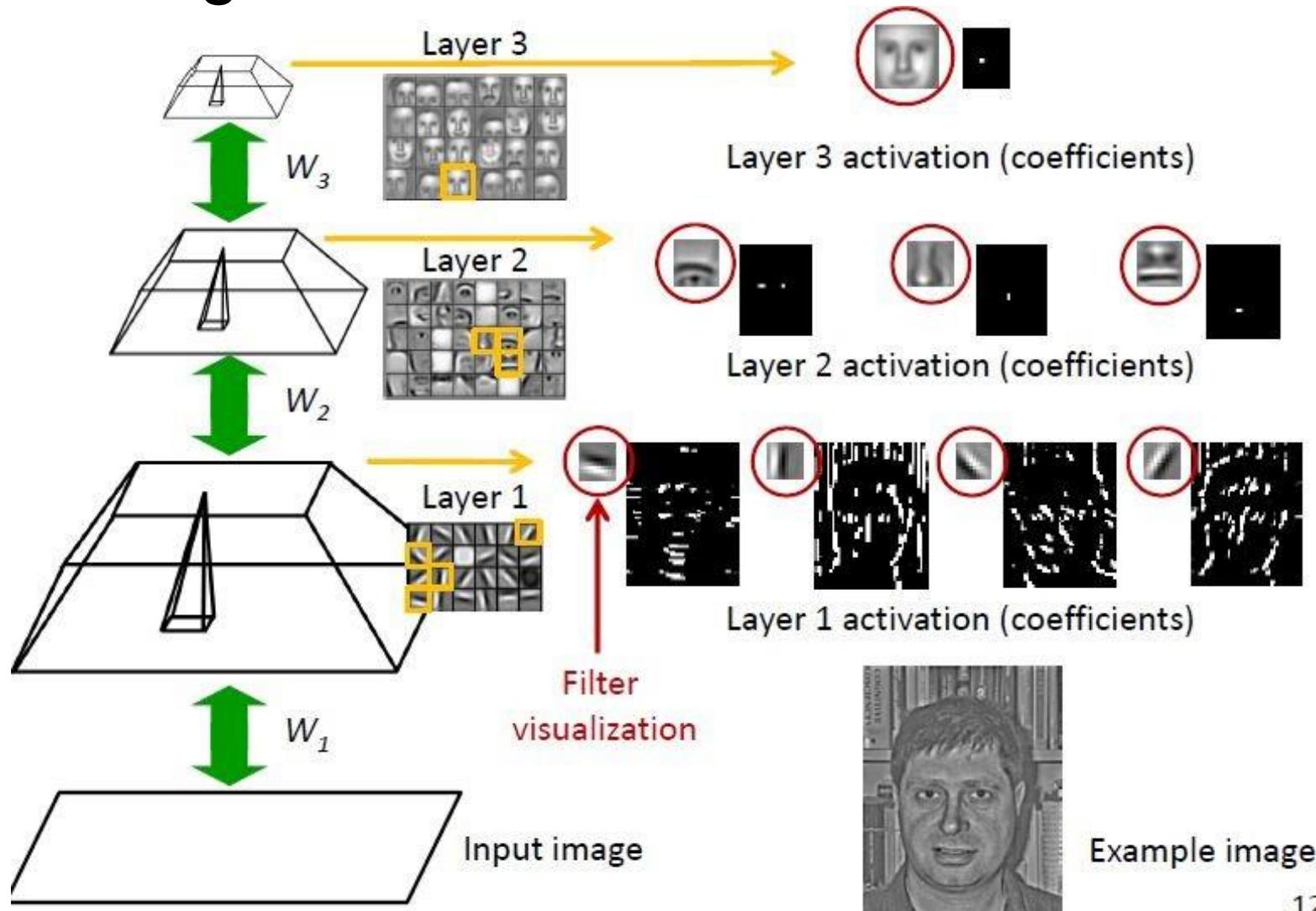
April 6, 2017



What could be good cat features?

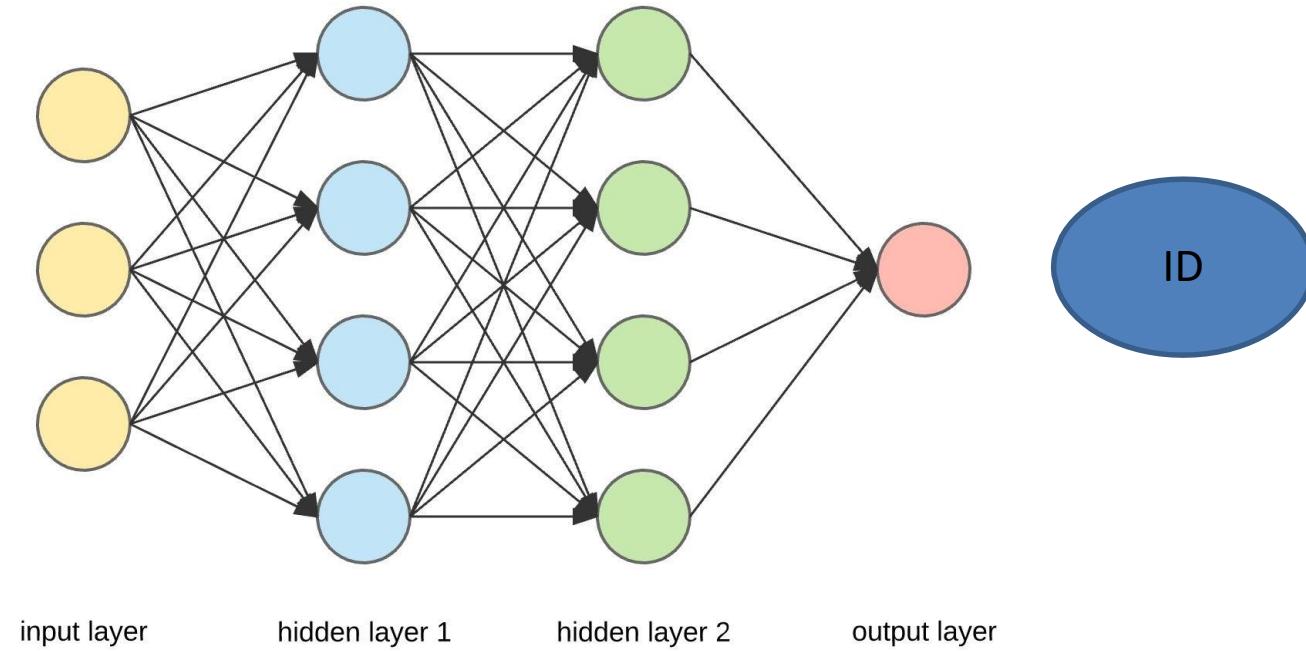


What could be good face features?



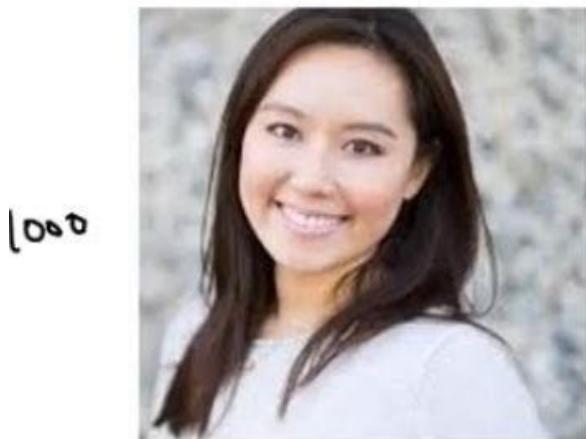
12

Deep Learning vs. Machine Learning



What neurons can see?

Face recognition



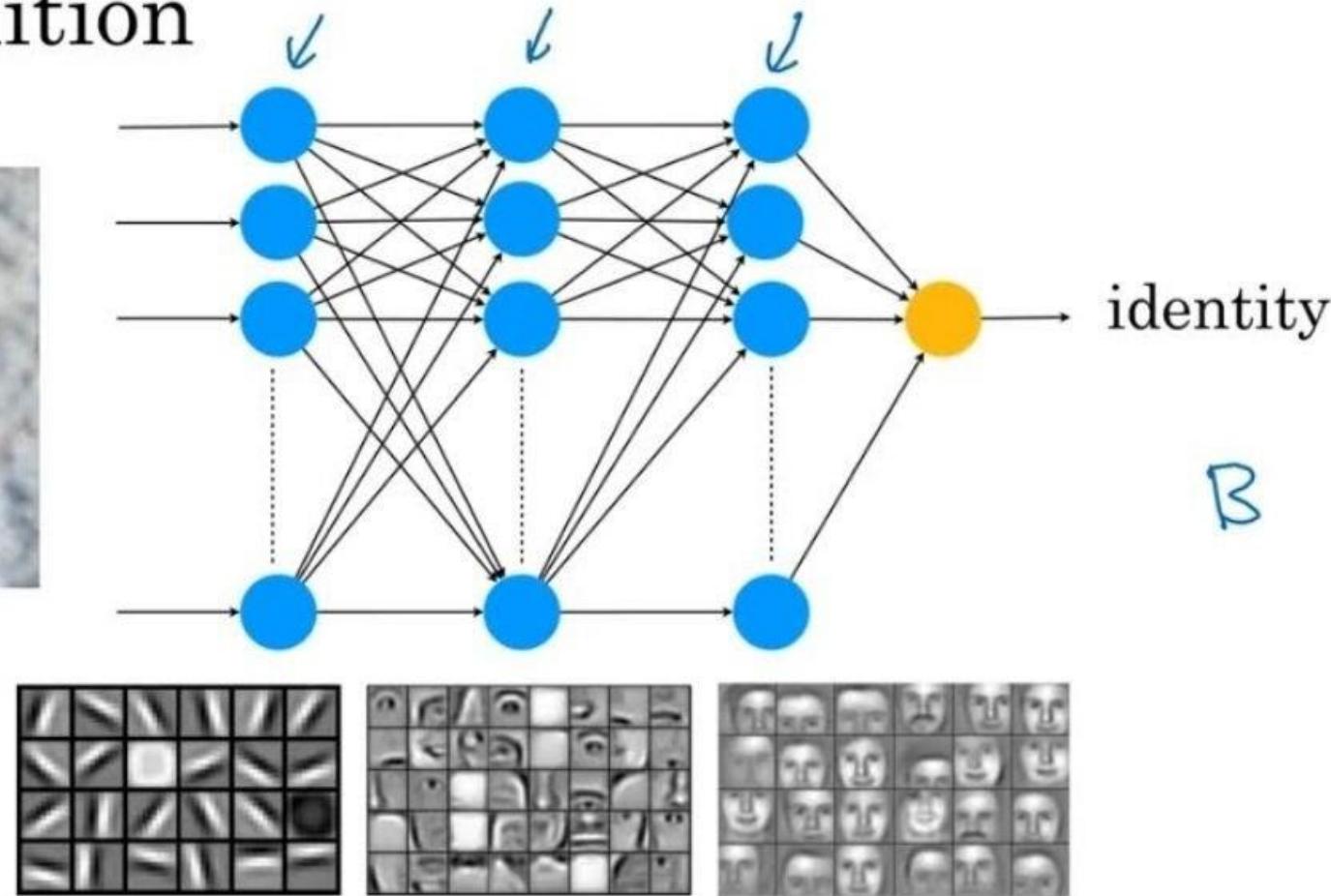
1000

1000

1,000,000

3,000,000

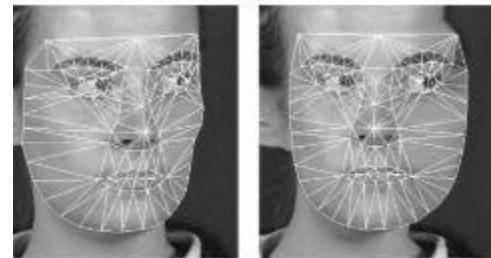
deeplearning.ai



Features

Raw data = pixels

Features = $f(\text{Raw data})$



Features = face points

Feature is a representation / transformation of raw data

Why representation matters so much?

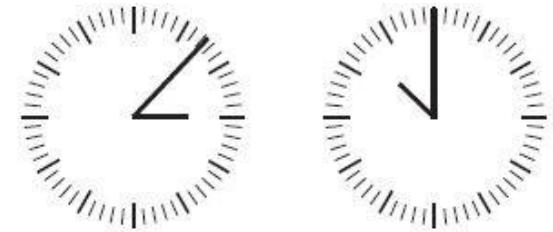
Which is easier to read the time?

-to make computer vision program to see the analog form

- just read x1,y1 and x2,y2 of the 2 pointers?

- just read theta 1 and theta 2 of the 2 pointers?

Raw data:
pixel grid



Better
features:
clock hands'
coordinates

{x1: 0.7,
y1: 0.7}
{x2: 0.5,
y2: 0.0}

{x1: 0.0,
y1: 1.0}
{x2: -0.38,
y2: 0.32}

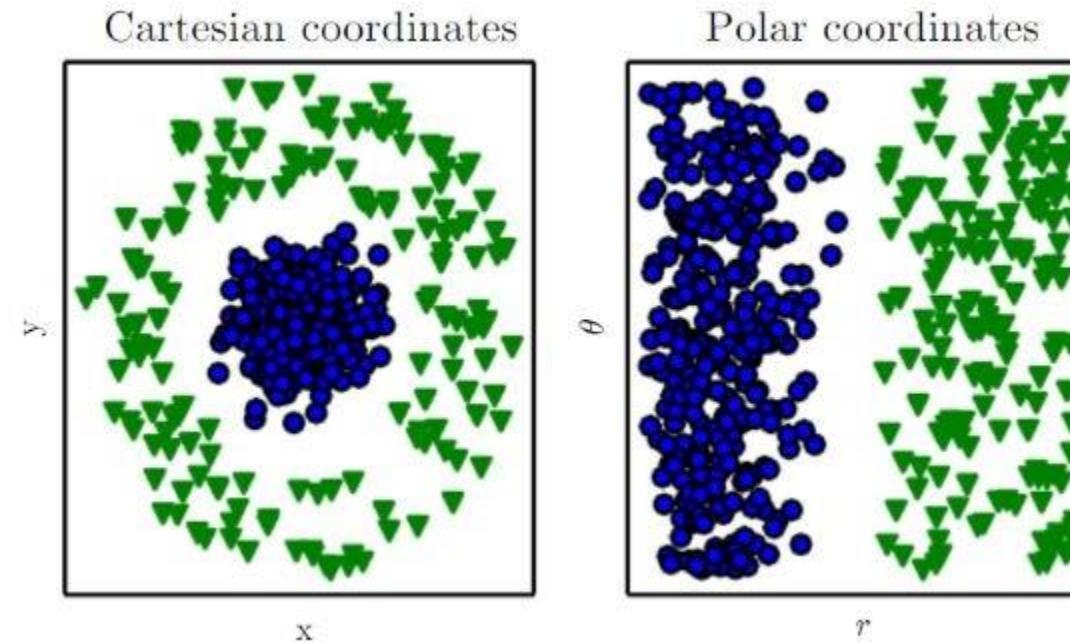
Even better
features:
angles of
clock hands

theta1: 45
theta2: 0

theta1: 90
theta2: 140

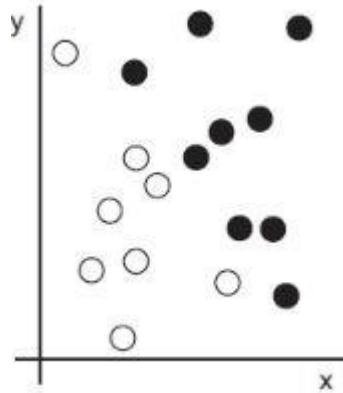
Why representation matters so much?

We have two categories (blue/green) which we want to cluster/separate/classify

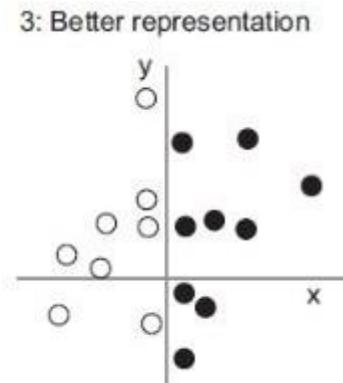
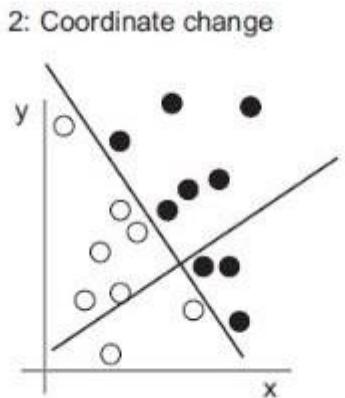
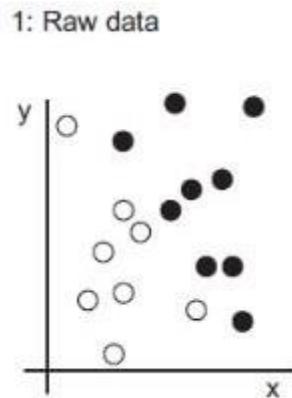


So separating the two classes in this case is much easier just by using polar representation

Why representation matters so much?

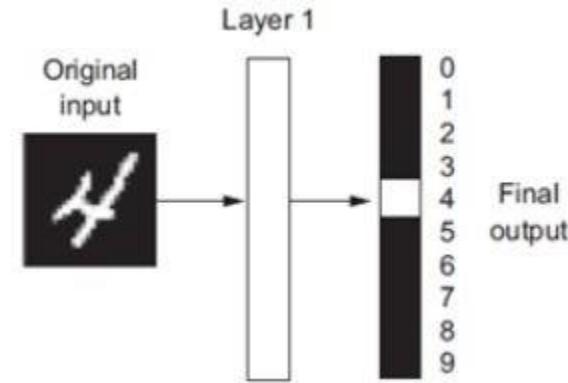


We want to develop an algorithm that can take the coordinates (x, y) of a point and output whether that point is likely to be black or to be white:
-Inputs: the coordinates of our points.
-Outputs: the colors of our points.

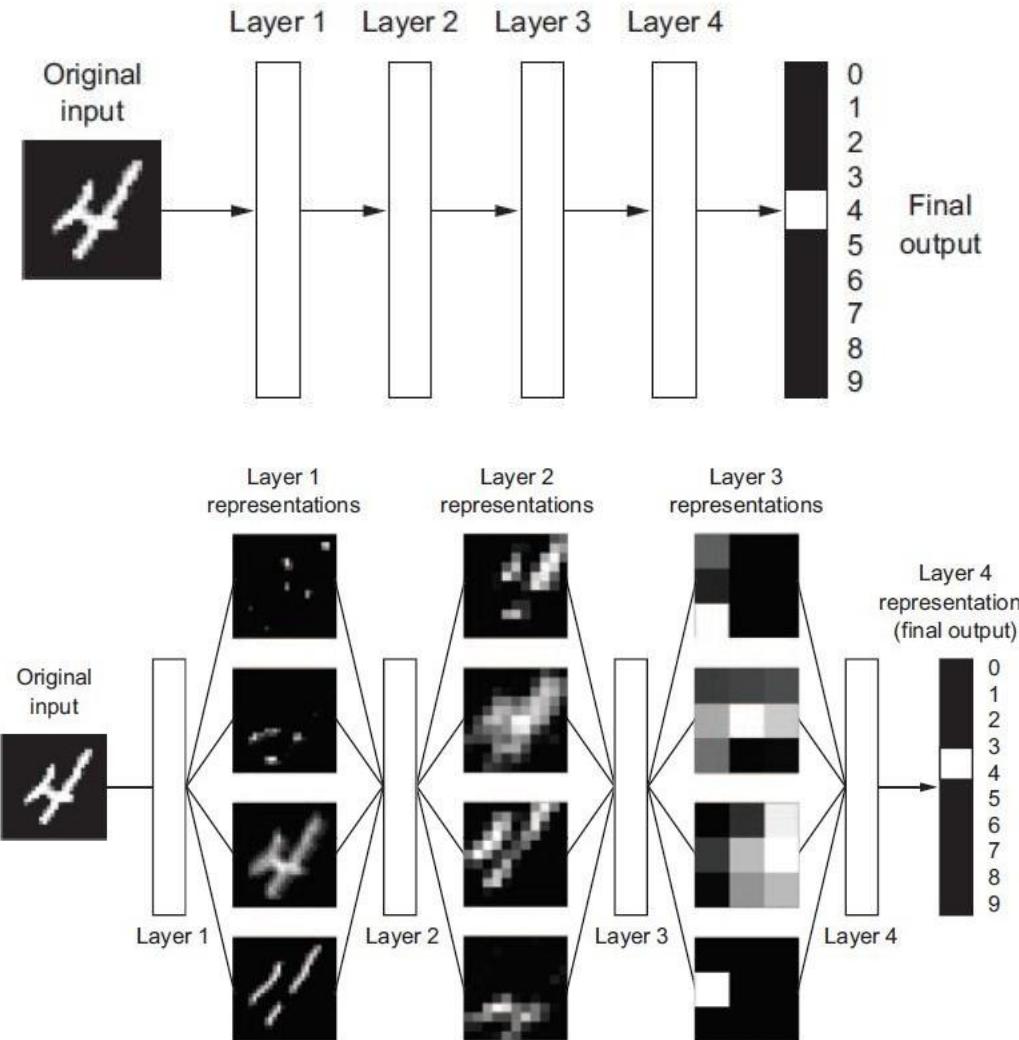


How to decide on which representation?

- Manual: ML
- Learning: DL
(Representation Learning)
- How ?
 - Hypothesis space: We have many possible transforms: f_1, f_2, \dots
 - Learning: Search in the *hypothesis space*



The “Deep” in Deep Learning Hierarchical Representation

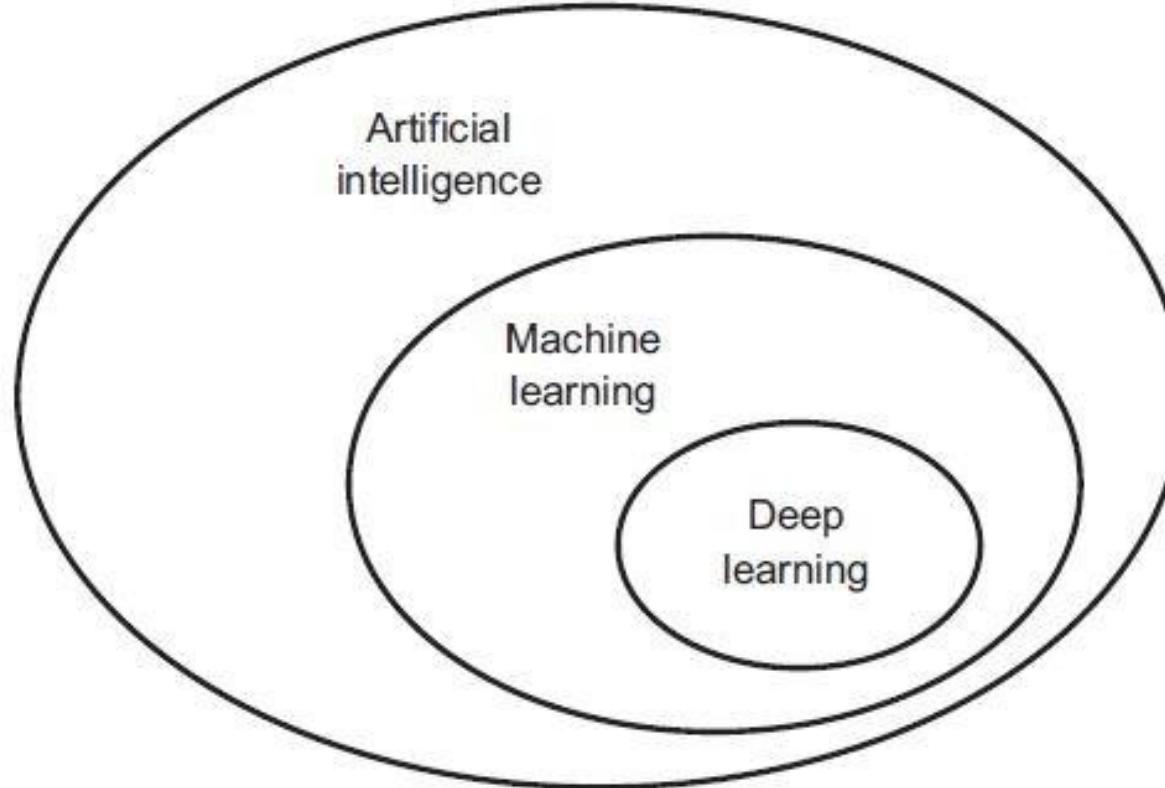


Now we converge to a definition of Deep Learning

Deep Learning = Deep Representation Learning

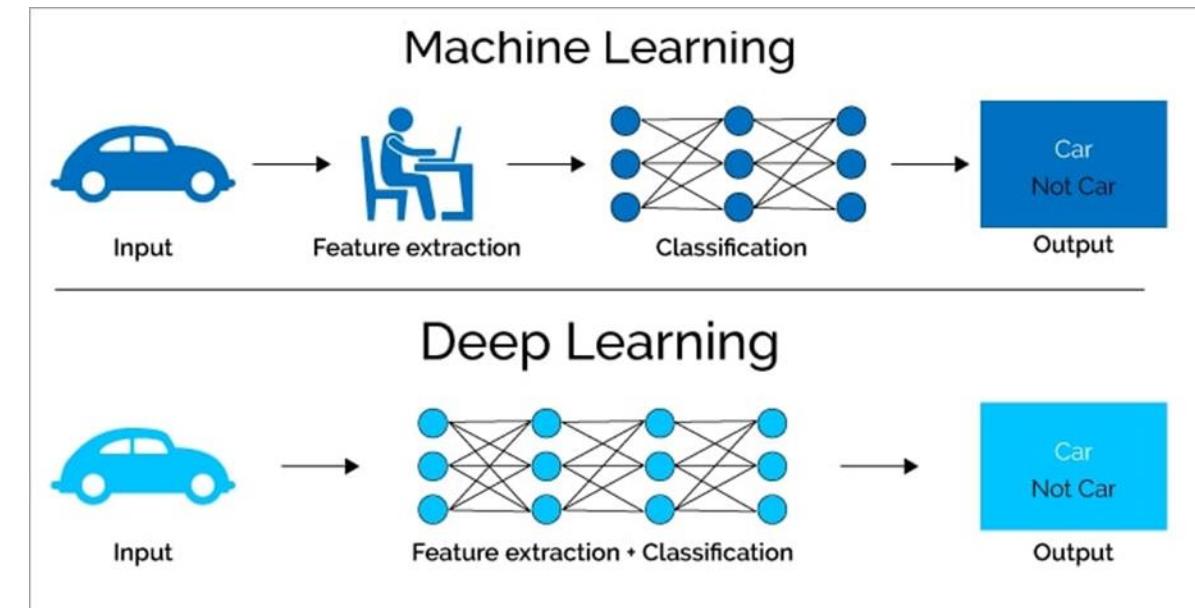
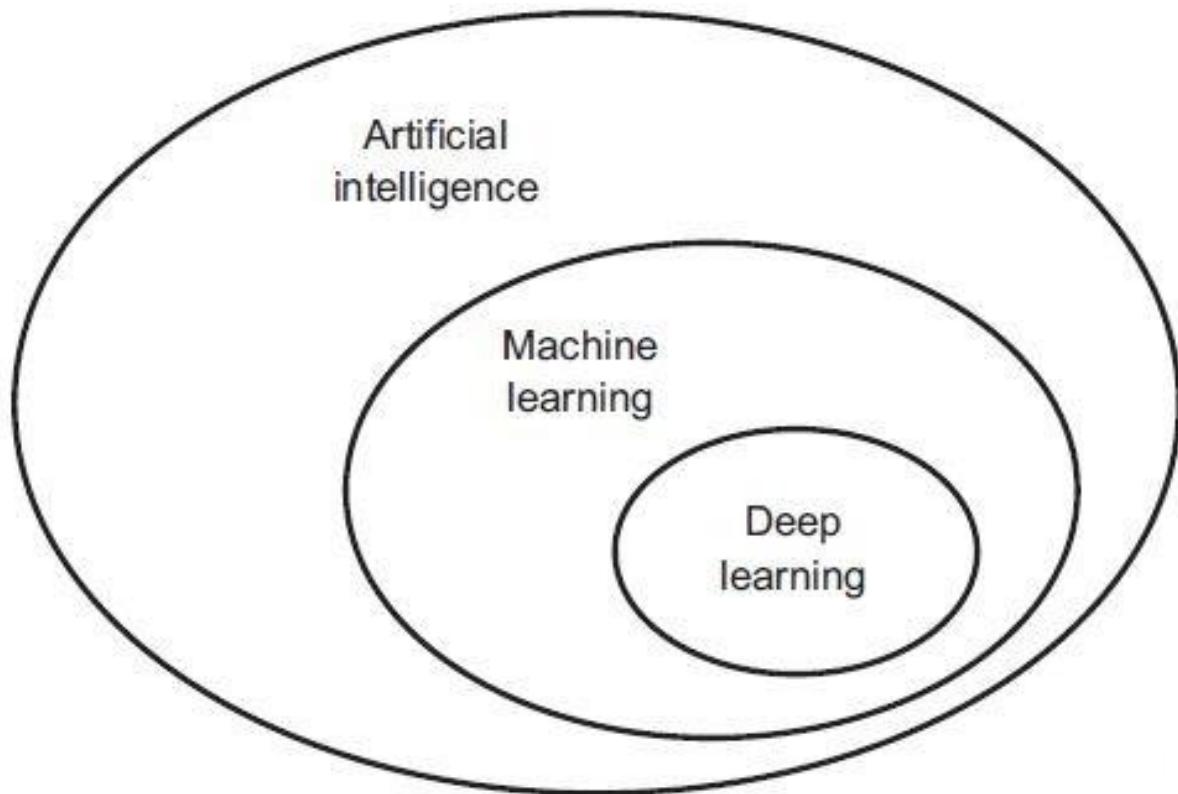
- “*Deep learning is a particular kind of machine learning that achieves great power and flexibility by learning to represent the world as a nested hierarchy of concepts, with each concept defined in relation to simpler concepts, and more abstract representations computed in terms of less abstract ones.*”
 - *Deep Learning Book, Ian Goodfellow, Yoshua Bengio, Aaron Courville*

between AI, ML, Representation Learning and DL “Deep Learning Book”



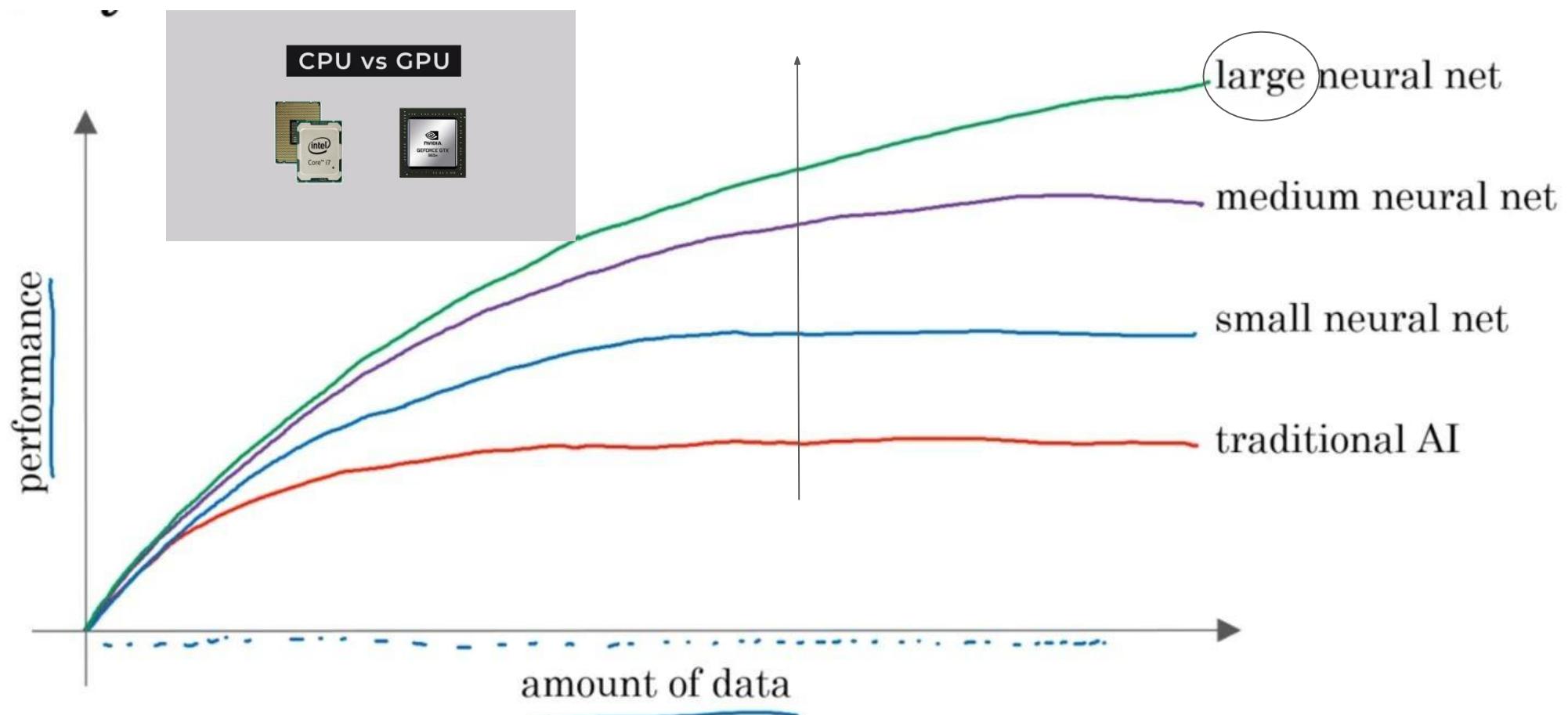
<https://www.deeplearningbook.org/>

between AI, ML, Representation Learning and DL “Deep Learning Book”

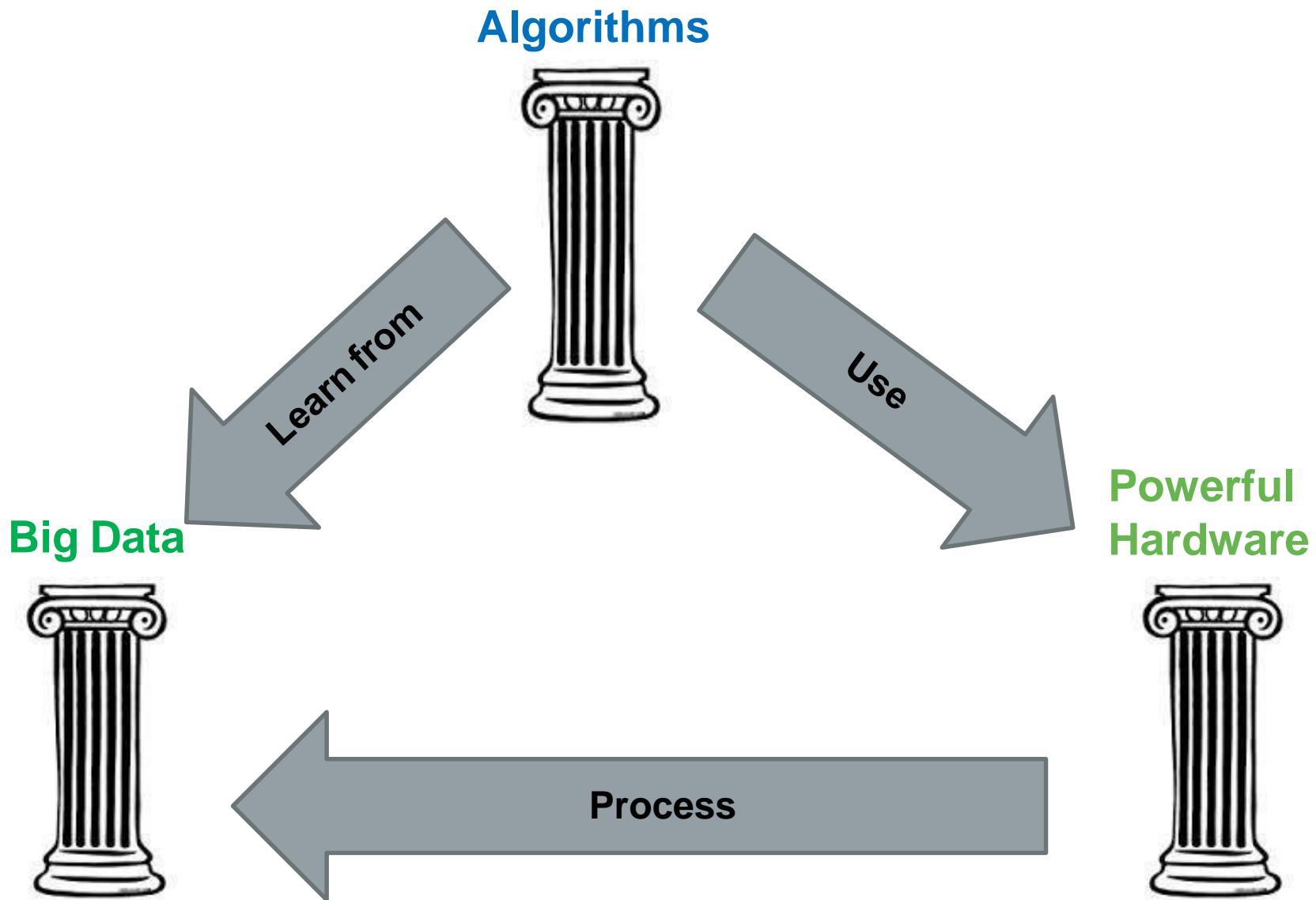


<https://www.deeplearningbook.org/>

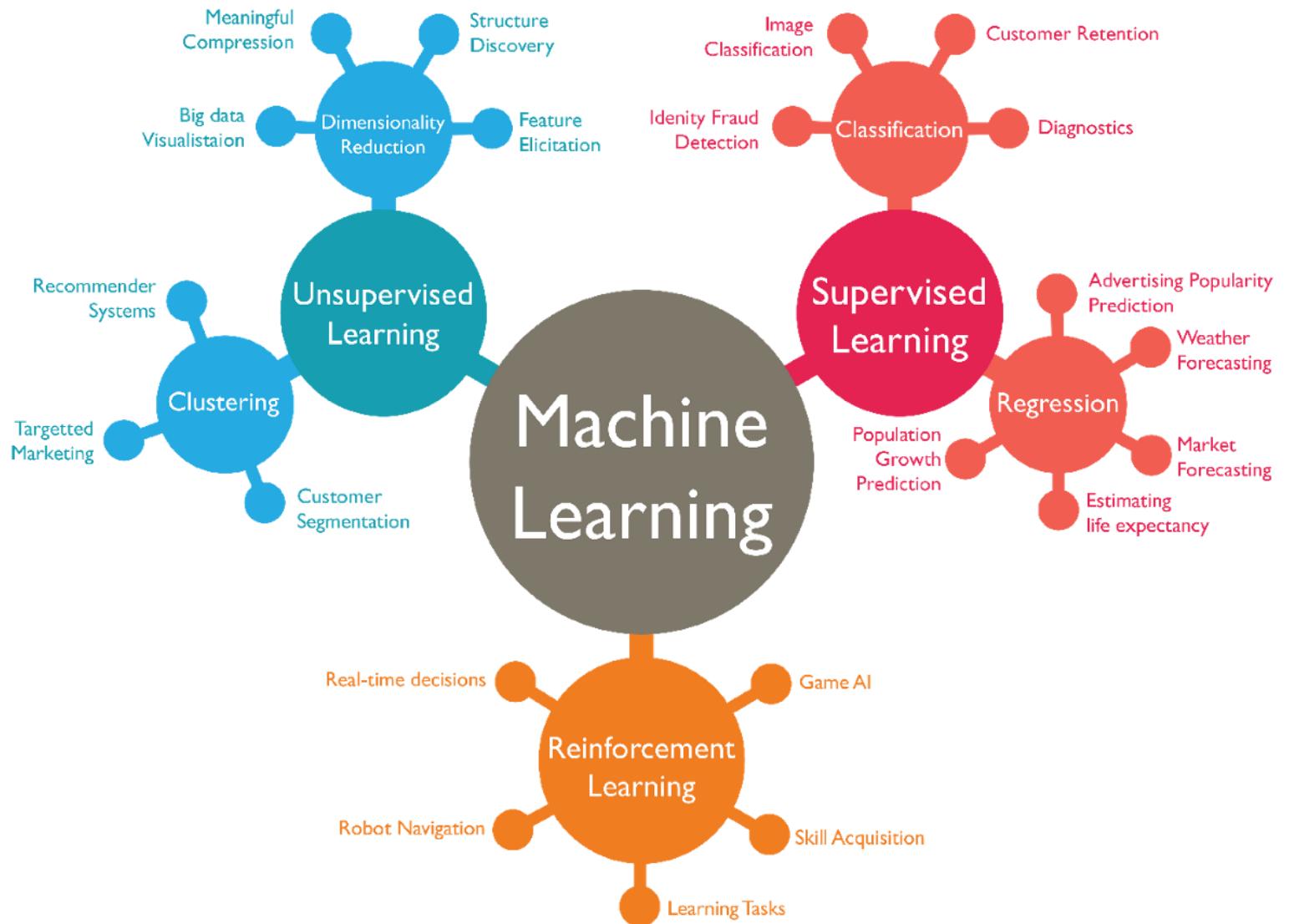
Why now?



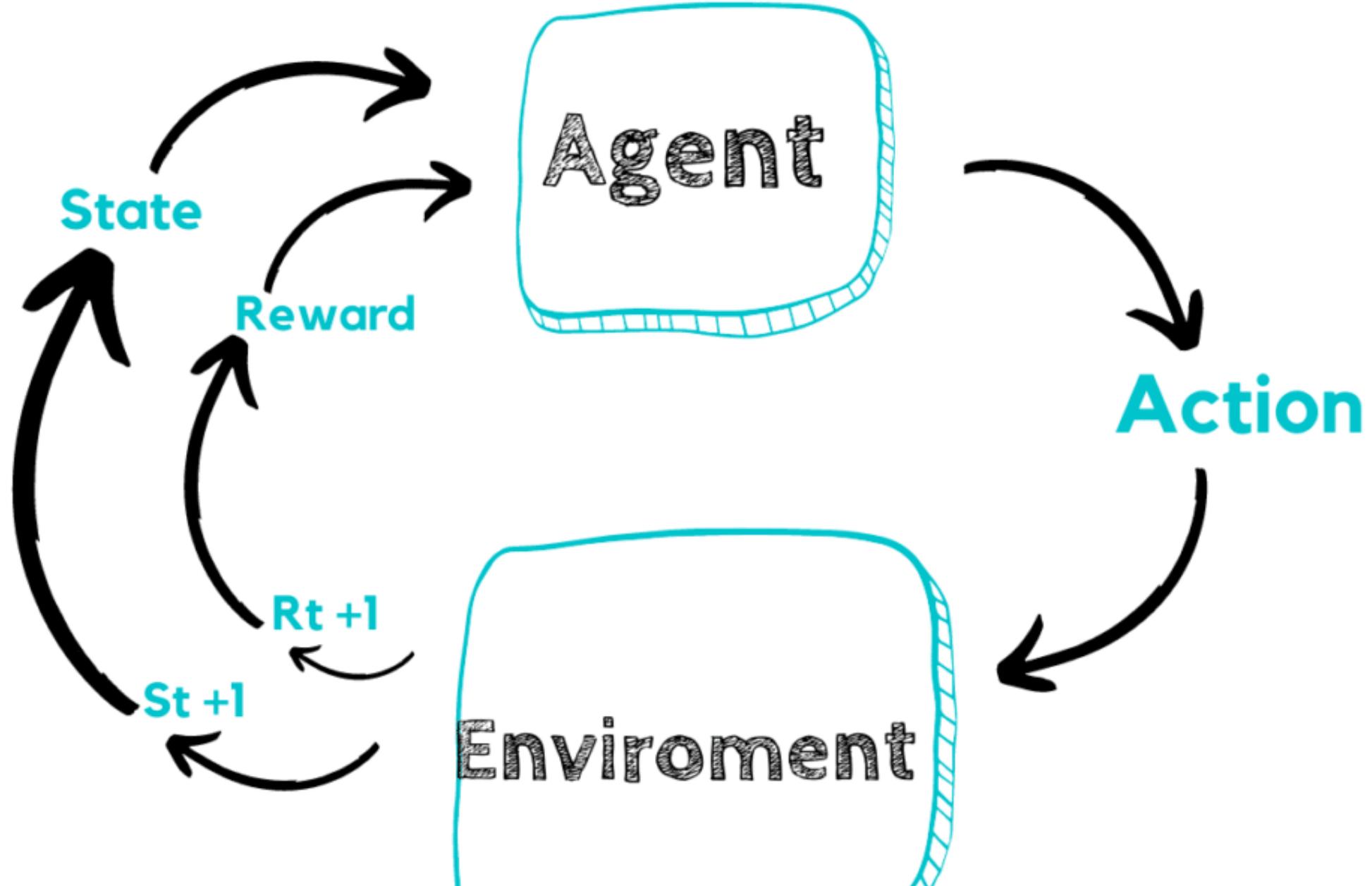
Why now?



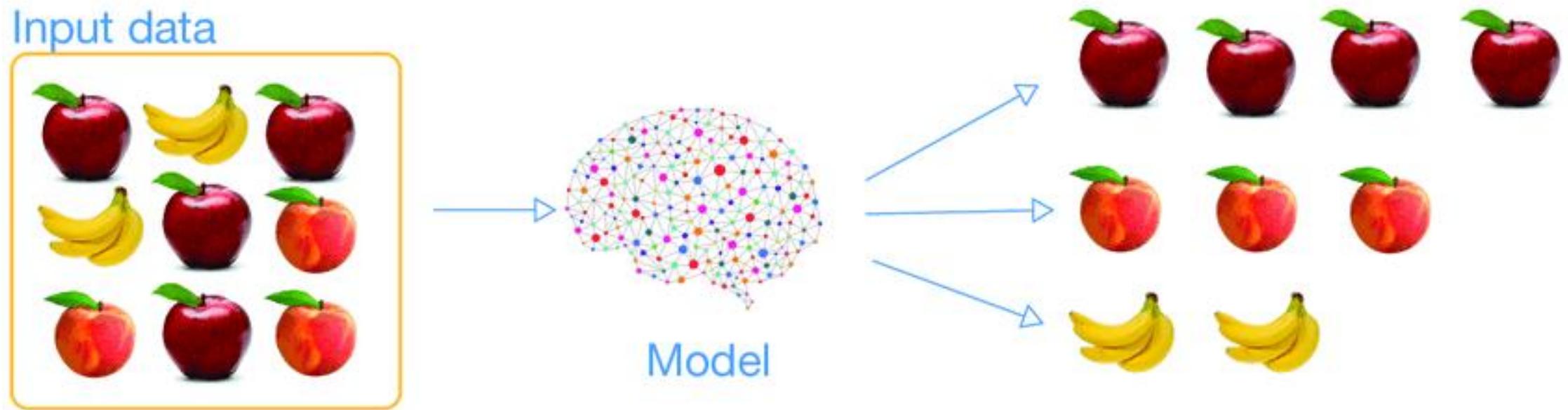
ML types



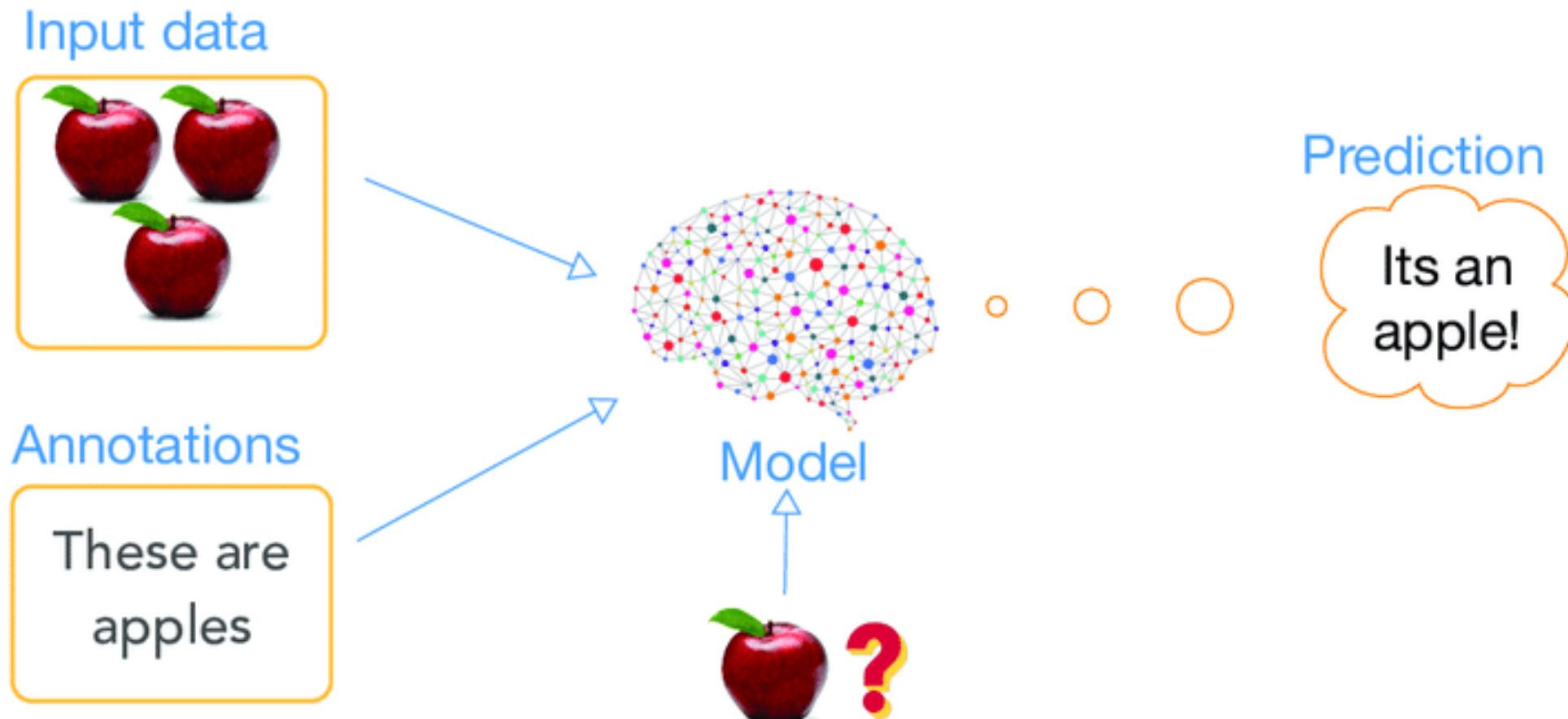
Reinforcement



Unsupervised



Supervised



Transfer Learning

Car detection



100,000 images

Golf cart detection



100 images

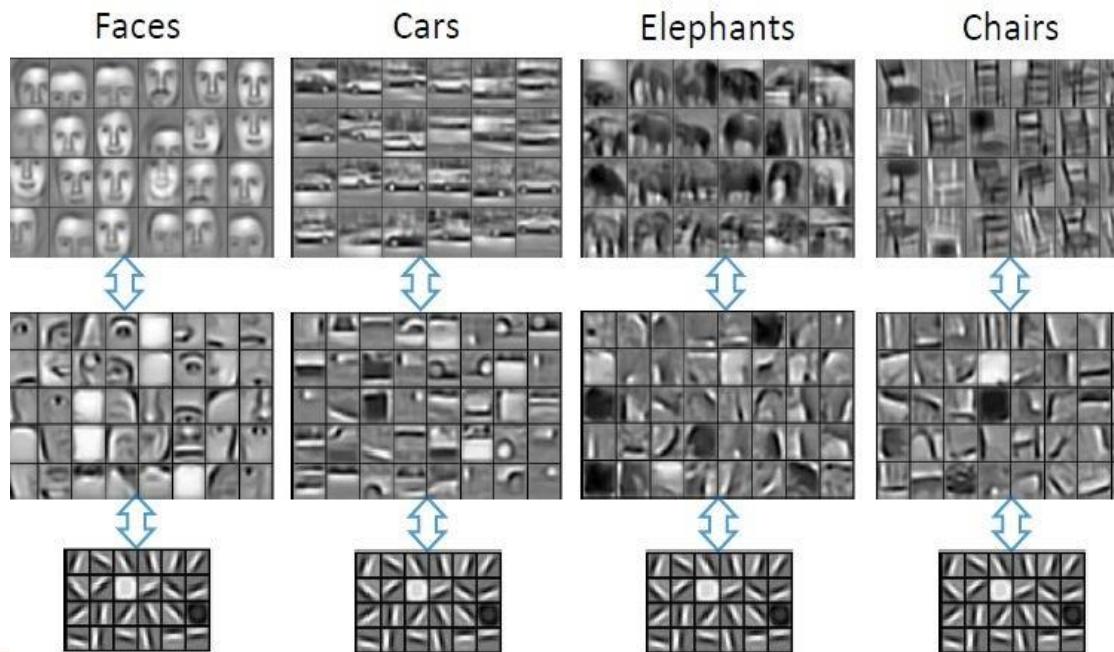
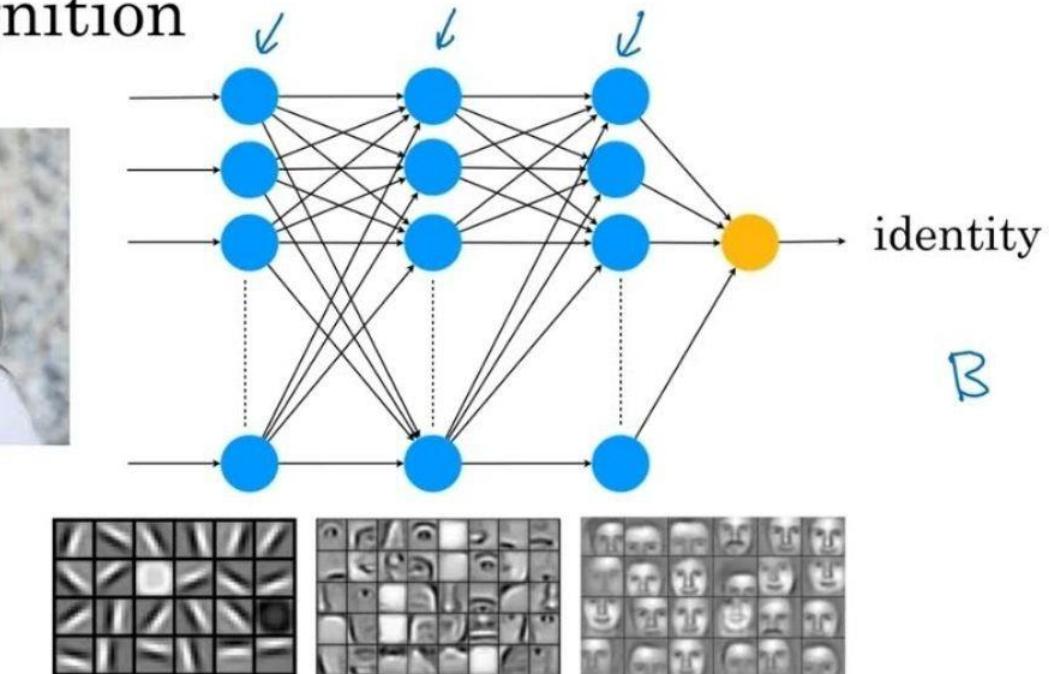
Learn from task A, and use knowledge to help on task B

Transfer Learning - How is it possible?

Face recognition

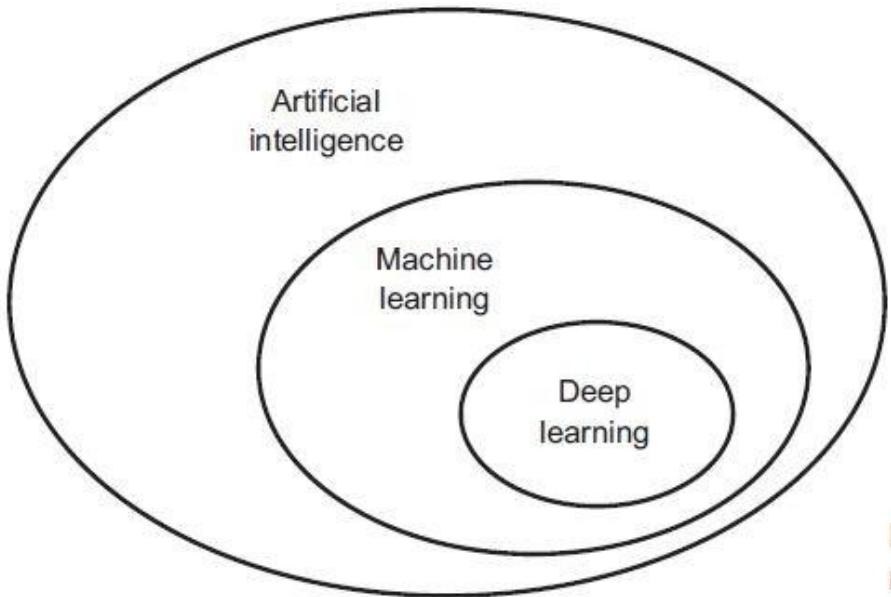


1000
1000
1,000,000
3,000,000



Data science vs. Machine Learning

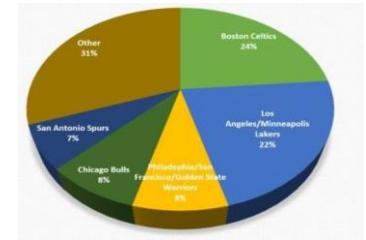
ML: “Make computers learn rather than explicitly programmed”, Arthur Samuel 1959



Output is software

DS: “Science of extracting Knowledge and insights from data”

Price	Floor space	Rooms	Lot size	Apartment	Row house	Corner house	Detached
250000	71	4	92	0	1	0	0
209500	98	5	123	0	1	0	0
349500	128	6	114	0	1	0	0
250000	86	4	98	0	1	0	0
419000	173	6	99	0	1	0	0
225000	83	4	67	0	1	0	0
549500	165	6	110	0	1	0	0
240000	71	4	78	0	1	0	0
340000	116	6	115	0	1	0	0



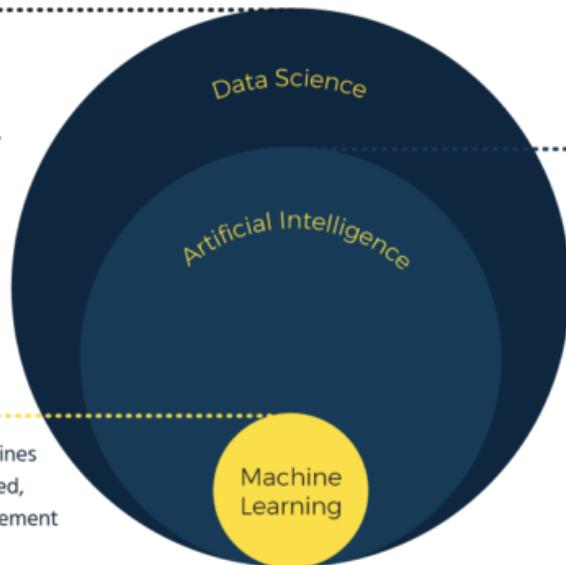
Looking at market data, what is the best sales strategy?
Output = slide deck supporting decision making
“Speak with Data”

Data science vs. Machine Learning

AI vs. Data Science vs. Machine Learning

Data Science

- Collection, preparation, and analysis of data
- Leverages AI/ML, research, industry expertise, and statistics to make business decisions

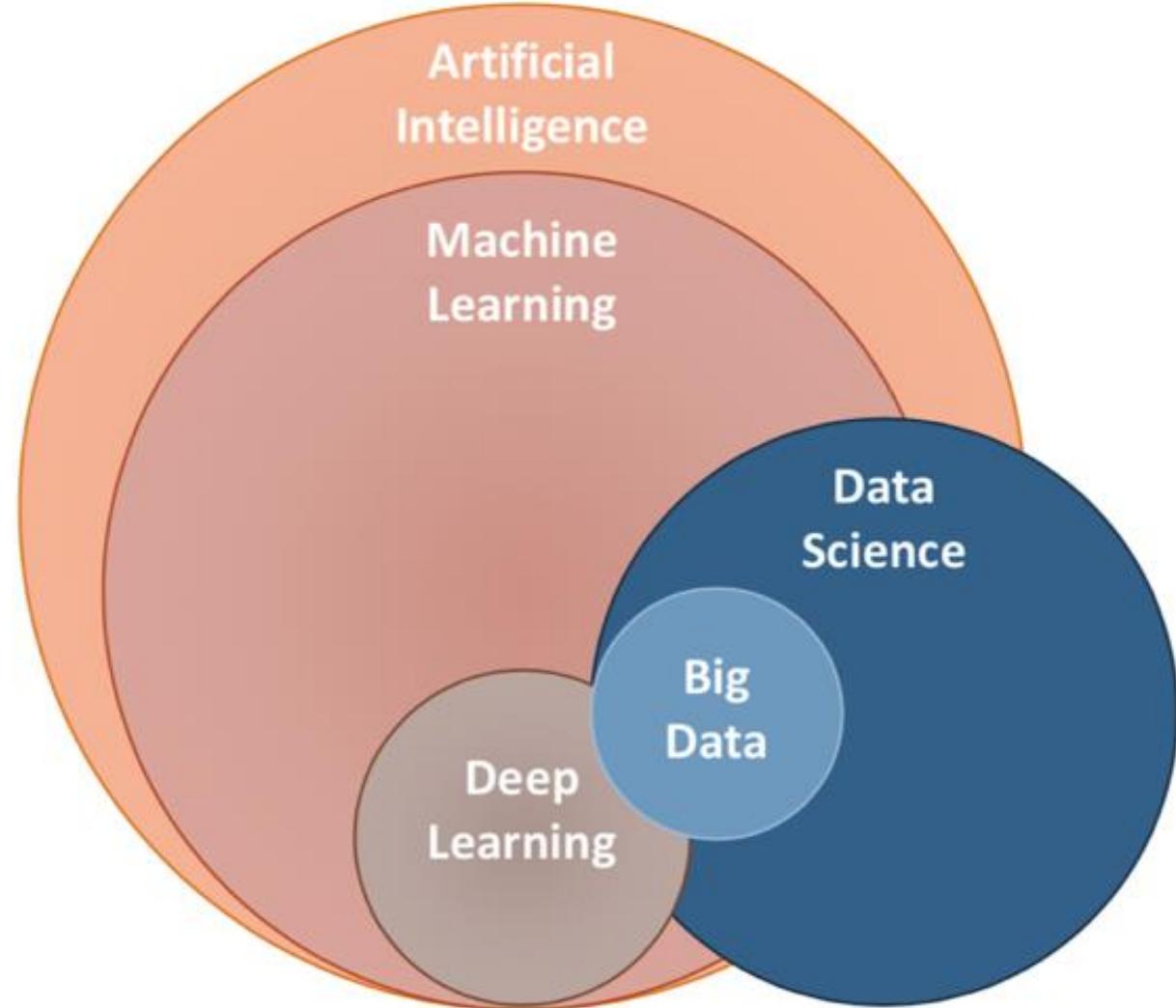


Machine Learning

- Algorithms that help machines improve through supervised, unsupervised, and reinforcement learning
- Subset of AI and Data Science tool

Artificial Intelligence

- Technology for machines to understand/interpret, learn, and make 'intelligent' decisions
- Includes Machine Learning among many other fields



Take away messages

ML is data driven

Global ML workflow

DL is more data driven + HW driven → Why now? GPU

AI is good at simple perception tasks and automation using analysis of tons of data

AI is not good at complex tasks. AGI is far.

Think statistical metrics. 100% is not possible.

AI Hype and Limitations: Bias, Attacks

These slides were created from the following sources

<https://www.coursera.org/learn/ai-for-everyone>

<https://www.coursera.org/learn/machine-learning-projects>

<https://course.fast.ai/>

<https://www.deeplearningbook.org/>

<https://www.coursat.ai>