Lecture 08
Introduction to
Data and Machine learning

Wan Fang

Southern University of Science and Technology

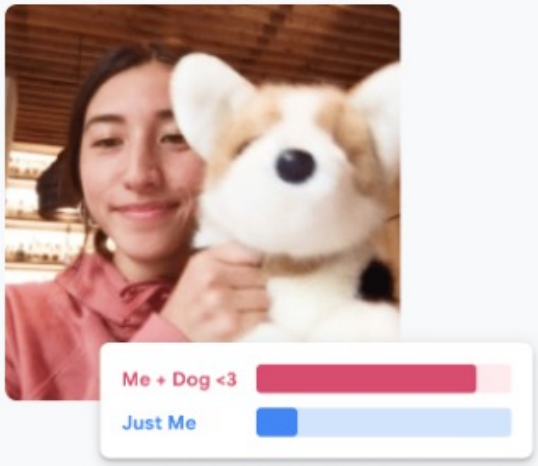
Agenda

- Introduction to Data and Machine Learning
  - Activity: Teachable Machine (50 mins)
  - What parts of ML can be Designed?

Activity: Teachable Machine

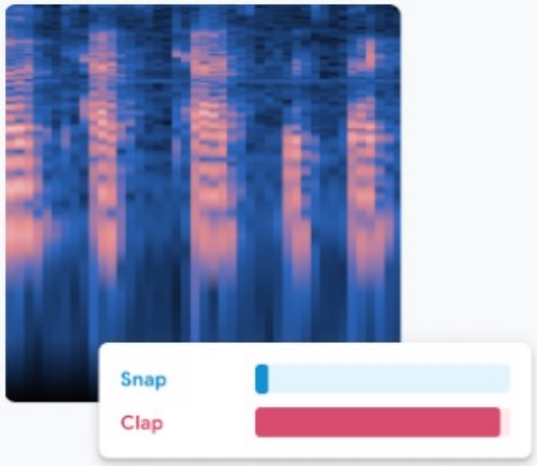
- https://train.aimaker.space/train

Teachable Machine is flexible – use files or capture examples live. It's respectful of the way you work. You can even choose to use it entirely on-device, without any webcam or microphone data leaving your computer.



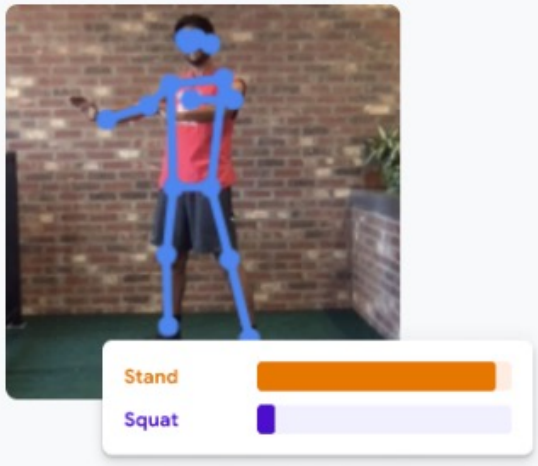
Images

Teach a model to classify images using files or your webcam.



Sounds

Teach a model to classify audio by recording short sound samples.



Poses

Teach a model to classify body positions using files or striking poses in your webcam.

Activity: Teachable Machine

- **Step 1. Let's train a model together!**
- Guiding questions:
  - What is going on in your own words?
  - What is the model "learning"?
  - What happens when we only use a few examples in each category (n=20)? A bunch of examples in each category (n=500)?

Activity: Teachable Machine

- **Step 2: What happens when we "break" it?**
- Guiding questions:
  - Does THIS machine have a concept of fingers? How about hands?
  - What would it take to have it recognize different fingers?
  - Even if we trained it on all 5 fingers, what might it still be missing?
  - When would we want to use a finger-classification? Who would want to use it? Why?

Activity: Teachable Machine

- **Step 3: Play around and apply to your domain! (30 mins)**
- Guiding questions:
  1. What happens when you **change the 'background**'? Go into another room or rotate your computer and try again. What's happening?
  2. What are some ways you think this could help your field? Who would want to use it? Why would they want to use it?
  3. What would they need as the 'training' data? How much of it would they need?
  4. What are some ways this might misclassify something? Would that be acceptable? How might you design ways to prevent that?
  5. What is the form factor you'd want it to take? Does a laptop and web camera work, or would you need something smaller that would fit into the space? Does it need to be a camera? What about another type of sensor (weight, photosensor)

What parts of ML can be Designed?

Partly Adapted from Designing Machine Learning at Stanford Design School

A crash course in AI + ML

A one-pager to get you up to speed on some core concepts including the difference between AI and ML, and the various types of machine learning.

artificial intelligence (AI)
= the science of getting machines to learn, think, act, and perform tasks in ways traditionally attributed to human intelligence

types of AI

narrow AI
= equals or exceeds human intelligence or efficiency at a very specific thing

general AI
= match human intelligence across domains + tasks

super AI
= exceeding human intelligence

not here (yet)

machine learning (ML)
= the ability for machines to learn and infer from large sets of examples and experience instead of explicitly programming the rules

deep learning
= artificial neural networks inspired by the human brain capable of learning from data that is unstructured

types of ML

types of learning

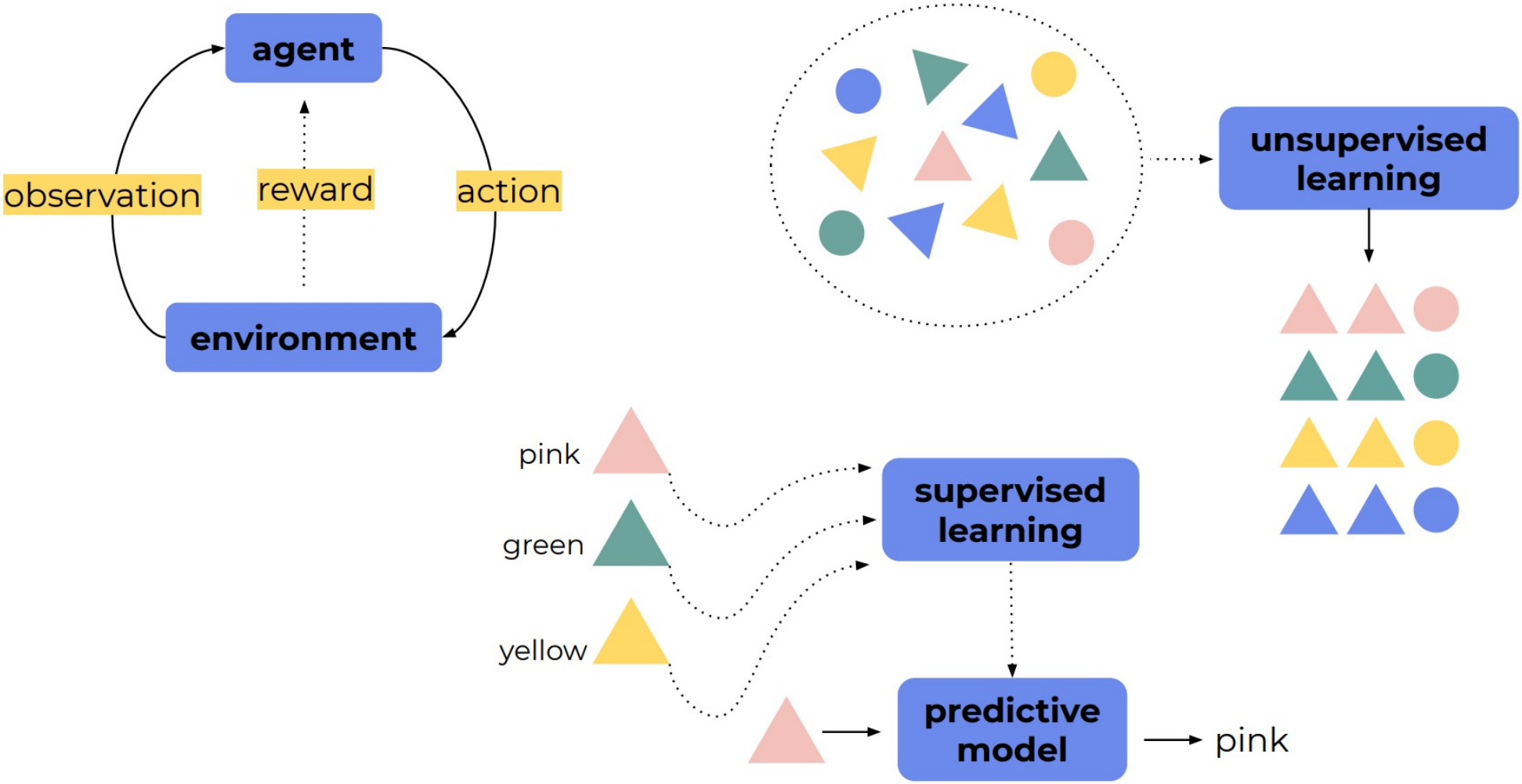
reinforcement learning
= collect data on the go and learn from trial and error to achieve an objective (below left)

supervised
= examples and data are labelled (below bottom)

unsupervised
= find patterns in large, non-labelled data sets (below right)

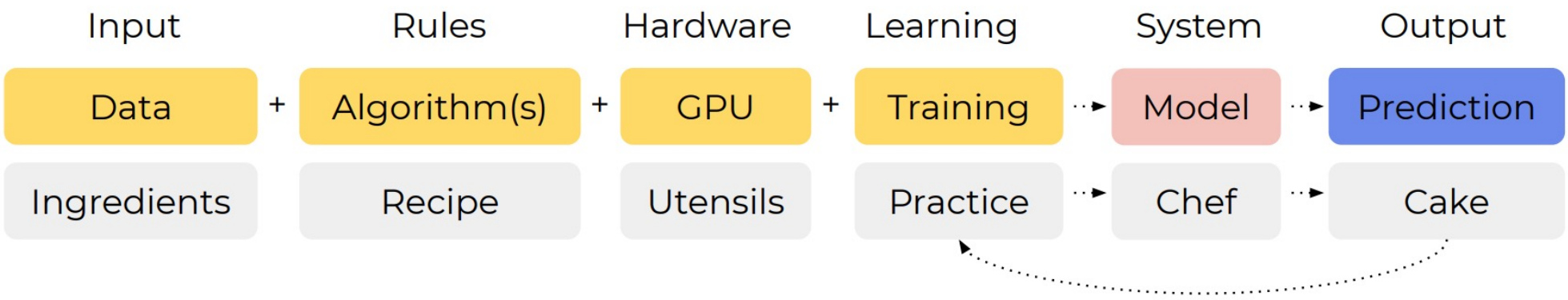
A crash course in AI + ML

A one-pager to get you up to speed on some core concepts including the difference between AI and ML, and the various types of machine learning.



The ML process

To get acquainted with terms and understand how a model arrives at a prediction, it can be helpful to draw an analogy with a process we're familiar with: baking a cake.



Data is the raw material you feed to the algorithm as input to produce a ML model.

An algorithm is a set of rules or step-by-step instructions to solve a problem.

The model requires GPU, and sometimes other resources, to run on.

Training process taking time and tweaking to learn, create and improve its model.

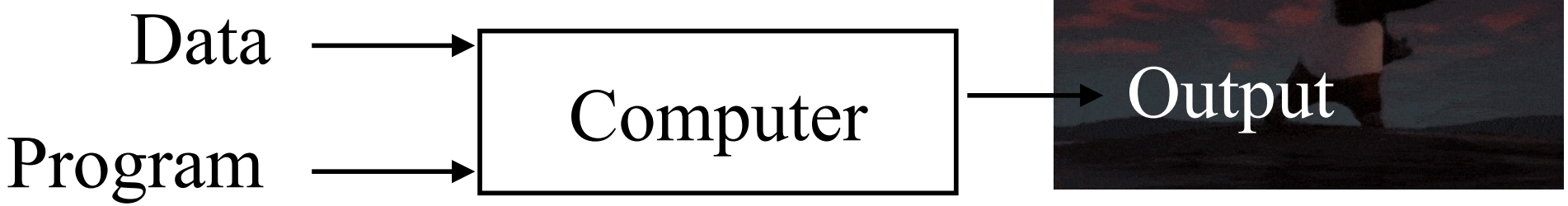
A model is a mathematical representation based on the algorithm(s) and data that is able to predict or produce an output and continues to learn over time.

Disclaimer: Please note this is a highly simplified representation of the real process which is a lot more complex and consists of plenty subtasks.

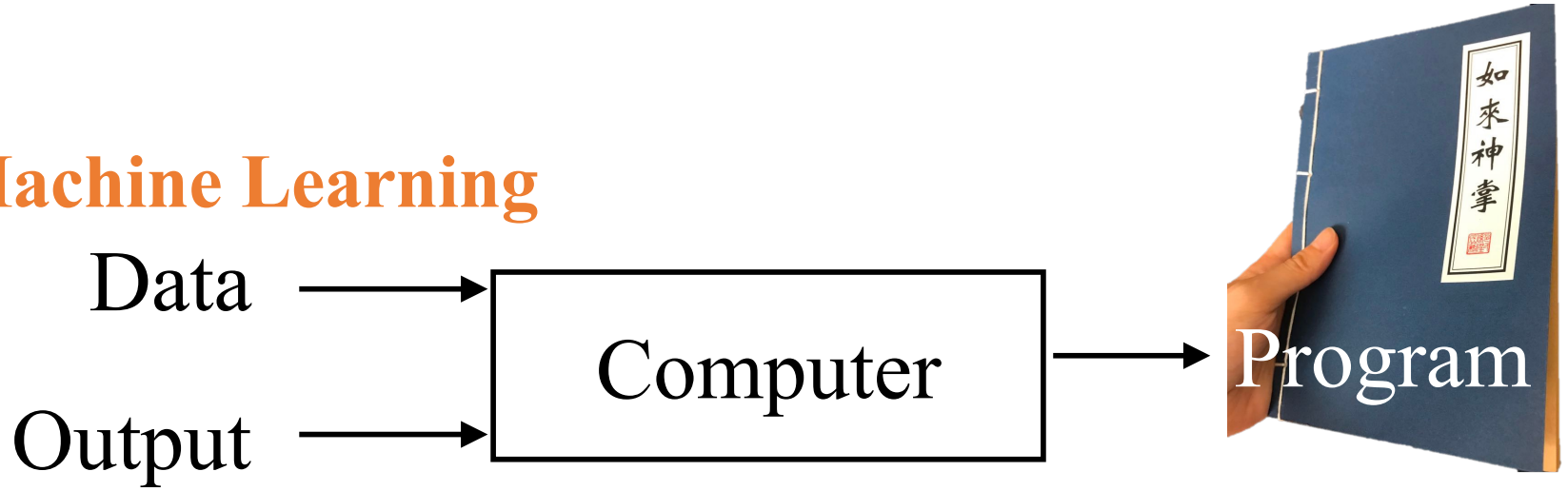
What is Machine Learning

- Machine Learning algorithms enable the computers to learn from data, and even improve themselves, without being explicitly programmed.

- Traditional Programming



- Machine Learning



AI –> Machine Learning–> Deep Learning

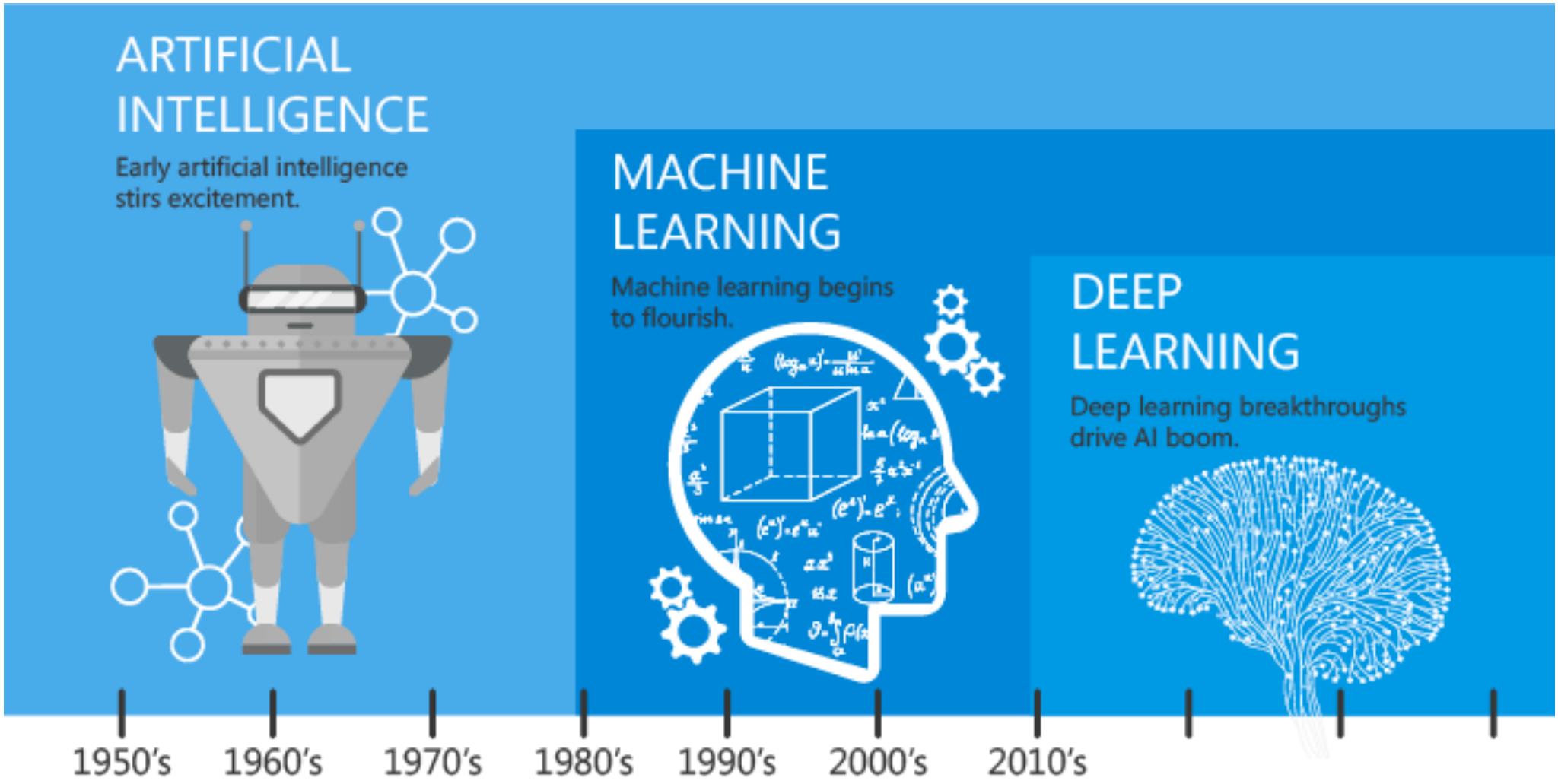
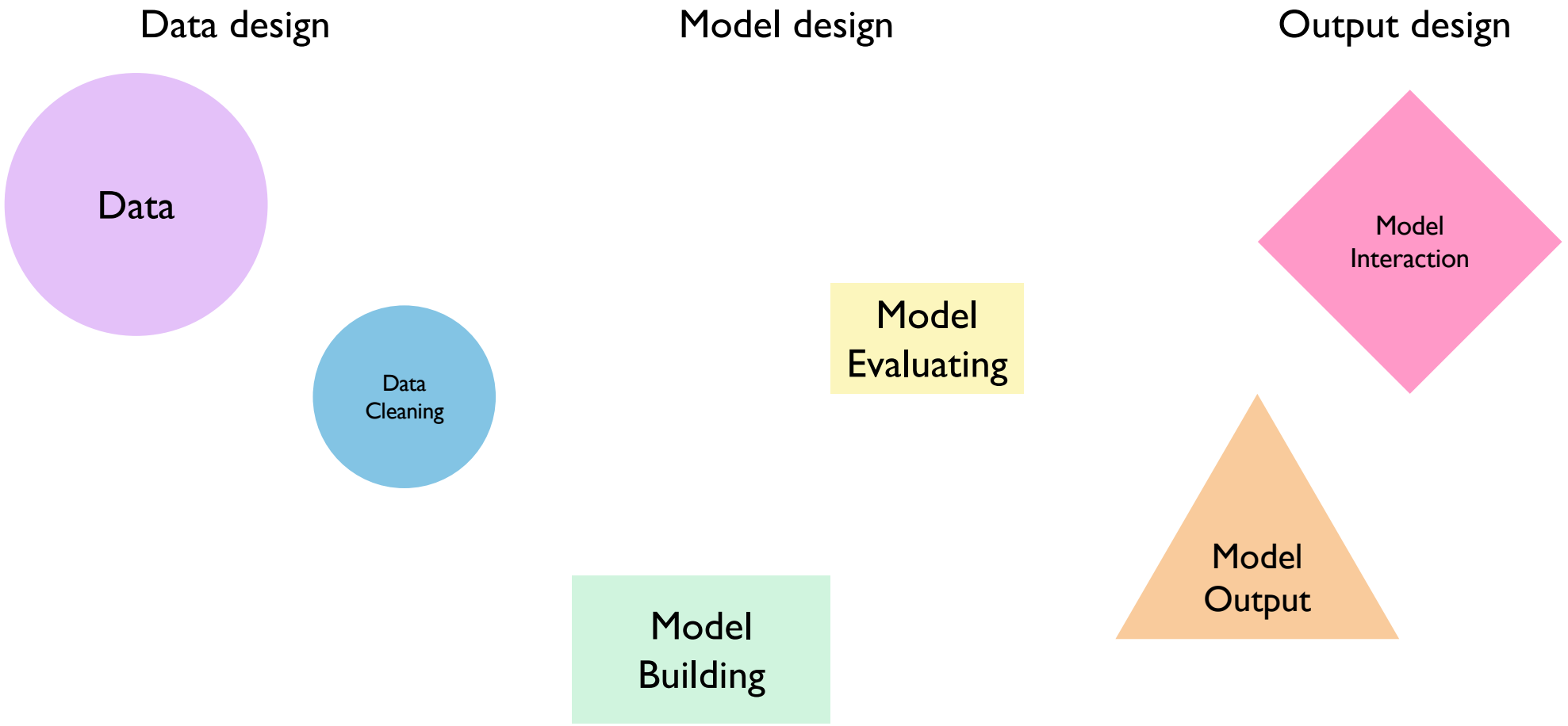
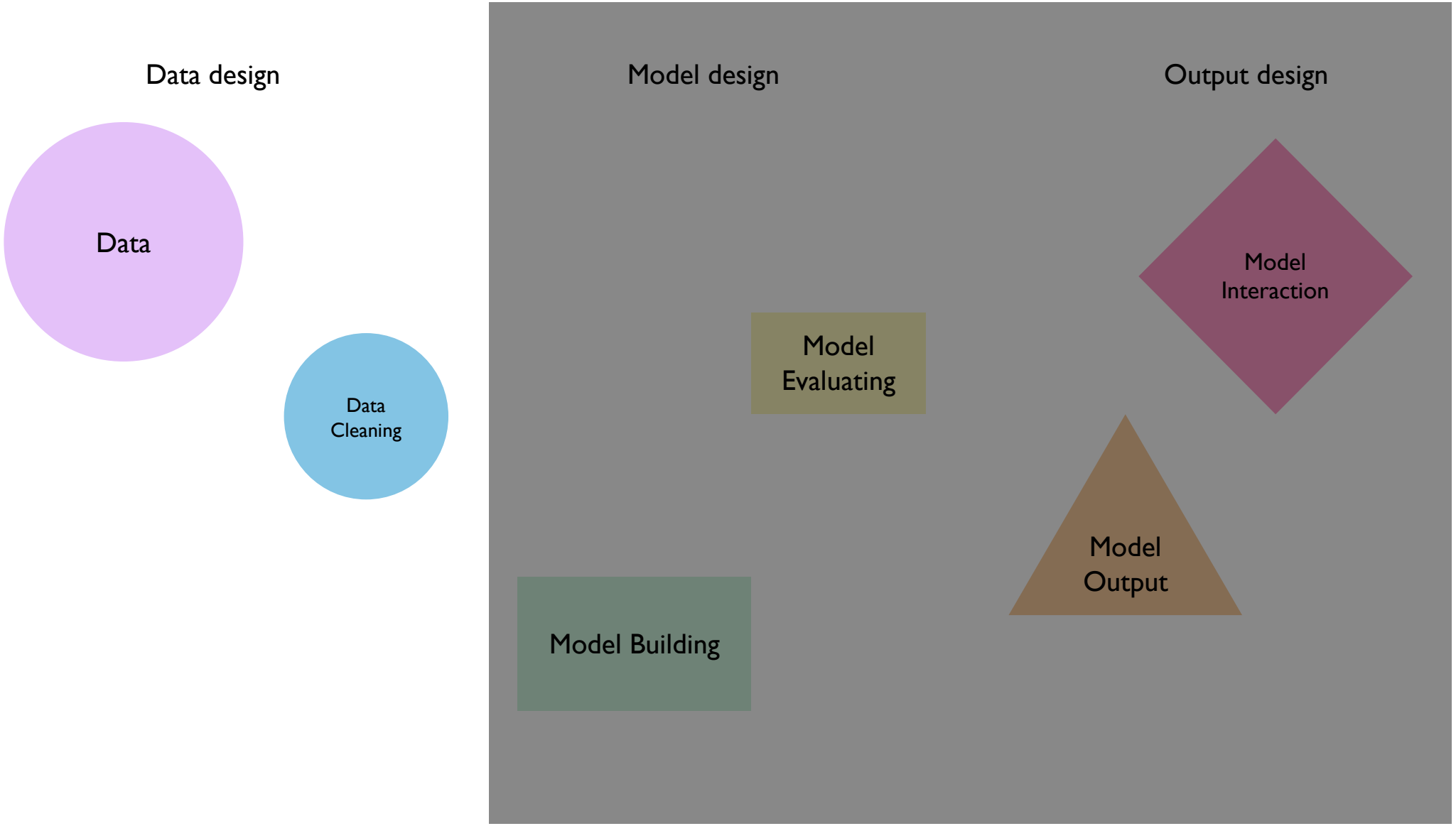


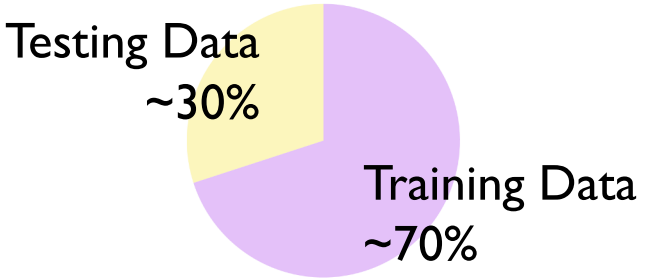
Image: Linked In | Machine Learning vs Deep learning


Data design
Model design
Output design
Data
Data Cleaning
Model Evaluating
Model Interaction
Model Building
Model Output


Data design
Data
Data
Cleaning
Model design
Model
Evaluating
Model Building
Output design
Model
Interaction
Model
Output

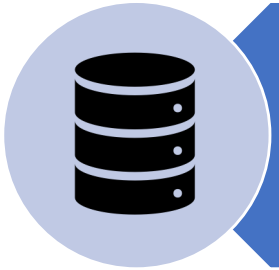
Data

- All machine learning models need data
  - What is Data?
  - Where does your data come from?
  - What is the type() of each of the features (columns)

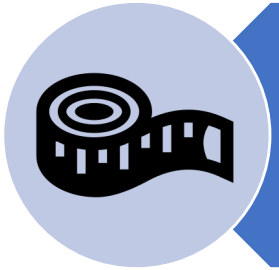


- Key vocab:
  - Training data - the data you use to train your ML model
  - Data type - the type/format of your data (string/integer)

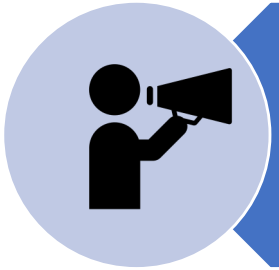
What is Data?



Data can be defined as a representation of facts, concepts, or instructions in a formalized manner,

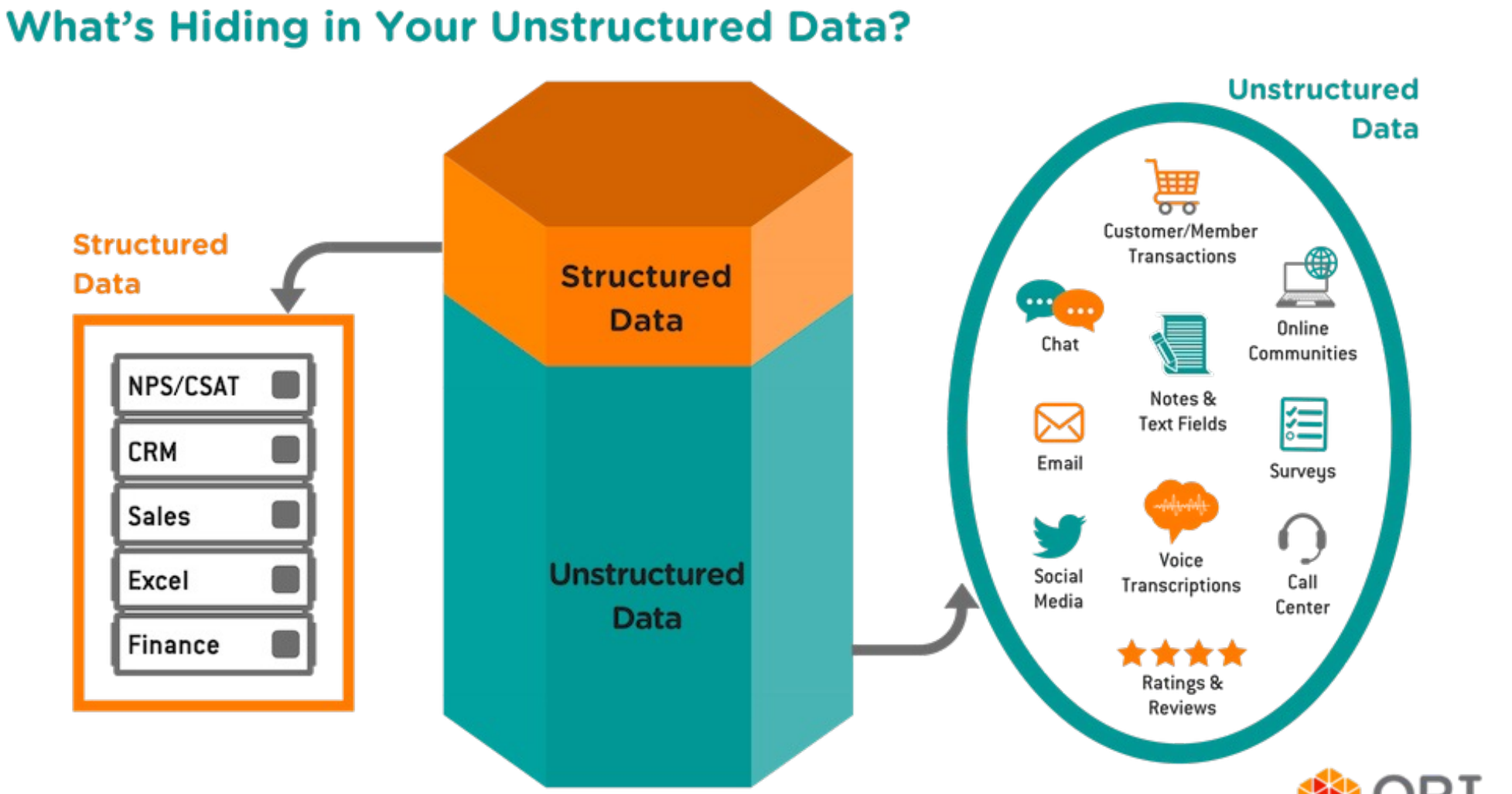


Collected through observations, measurements, or responses



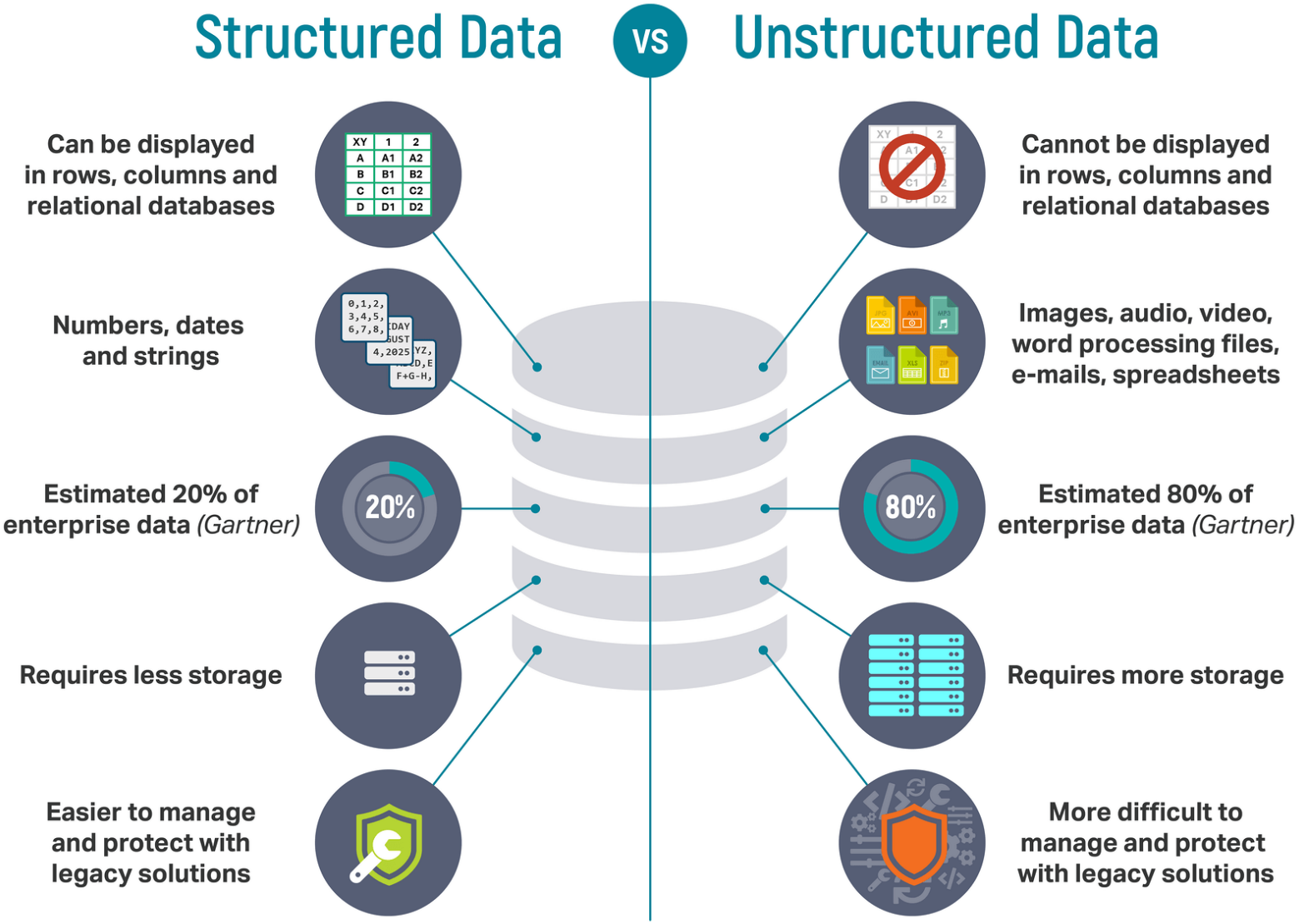
Suitable for communication, interpretation, or processing by human or electronic machines.

What is Data?



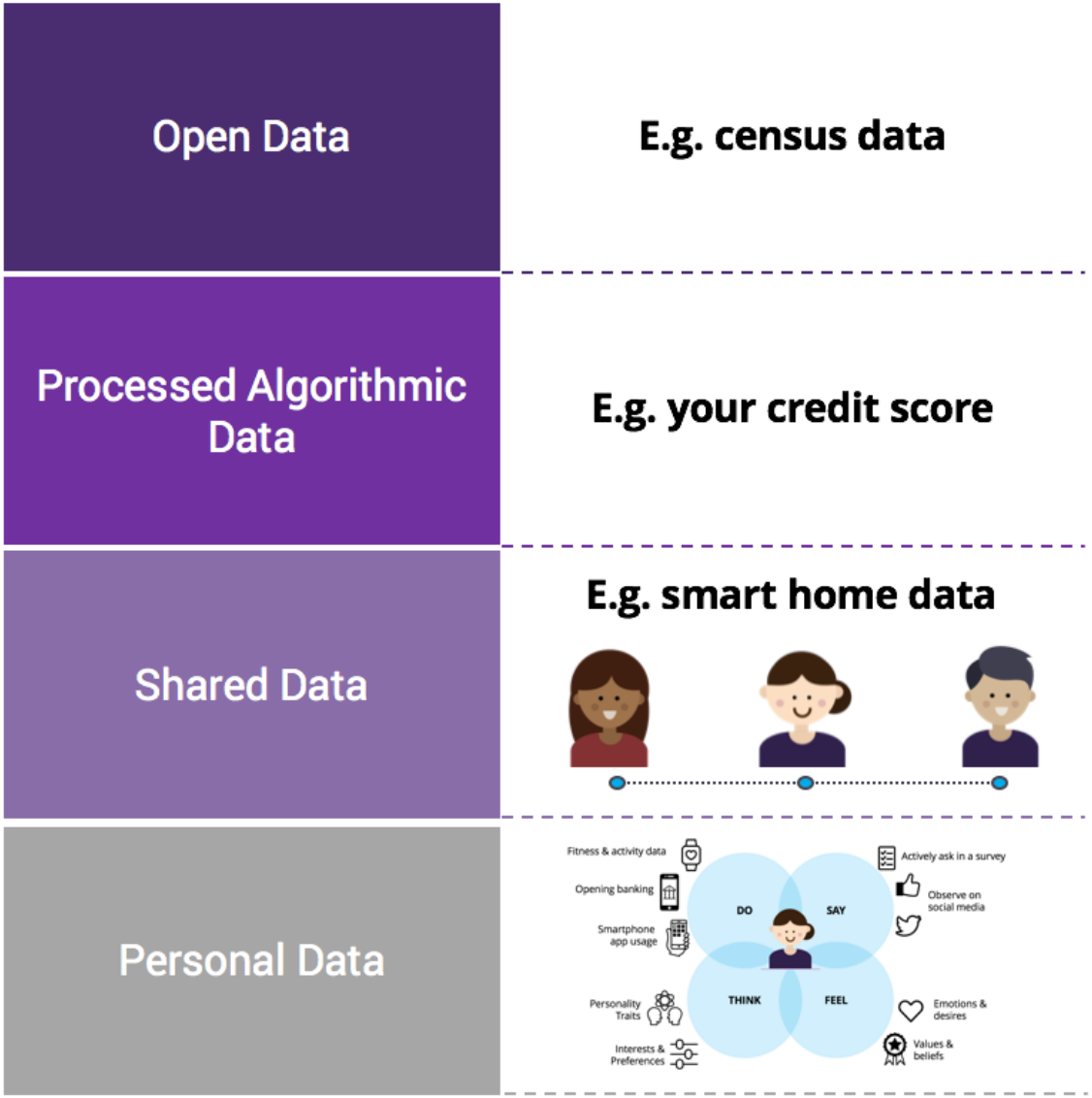
Source: Graphic adapted from January 2018 CXPA Presentation "The Why Behind the What," Jim Kitterman

What is Data?



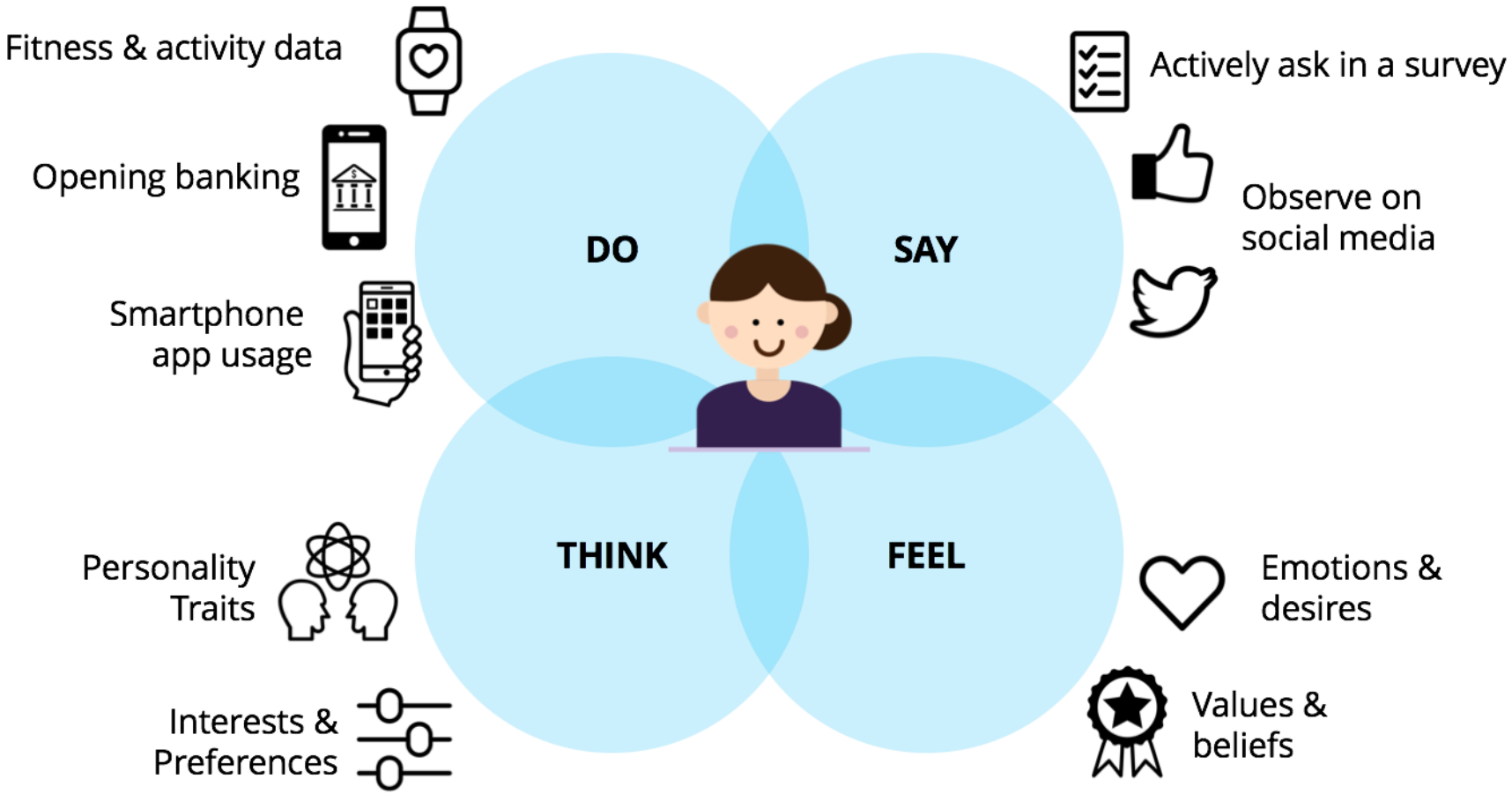
Source

The Full Human Data Stack



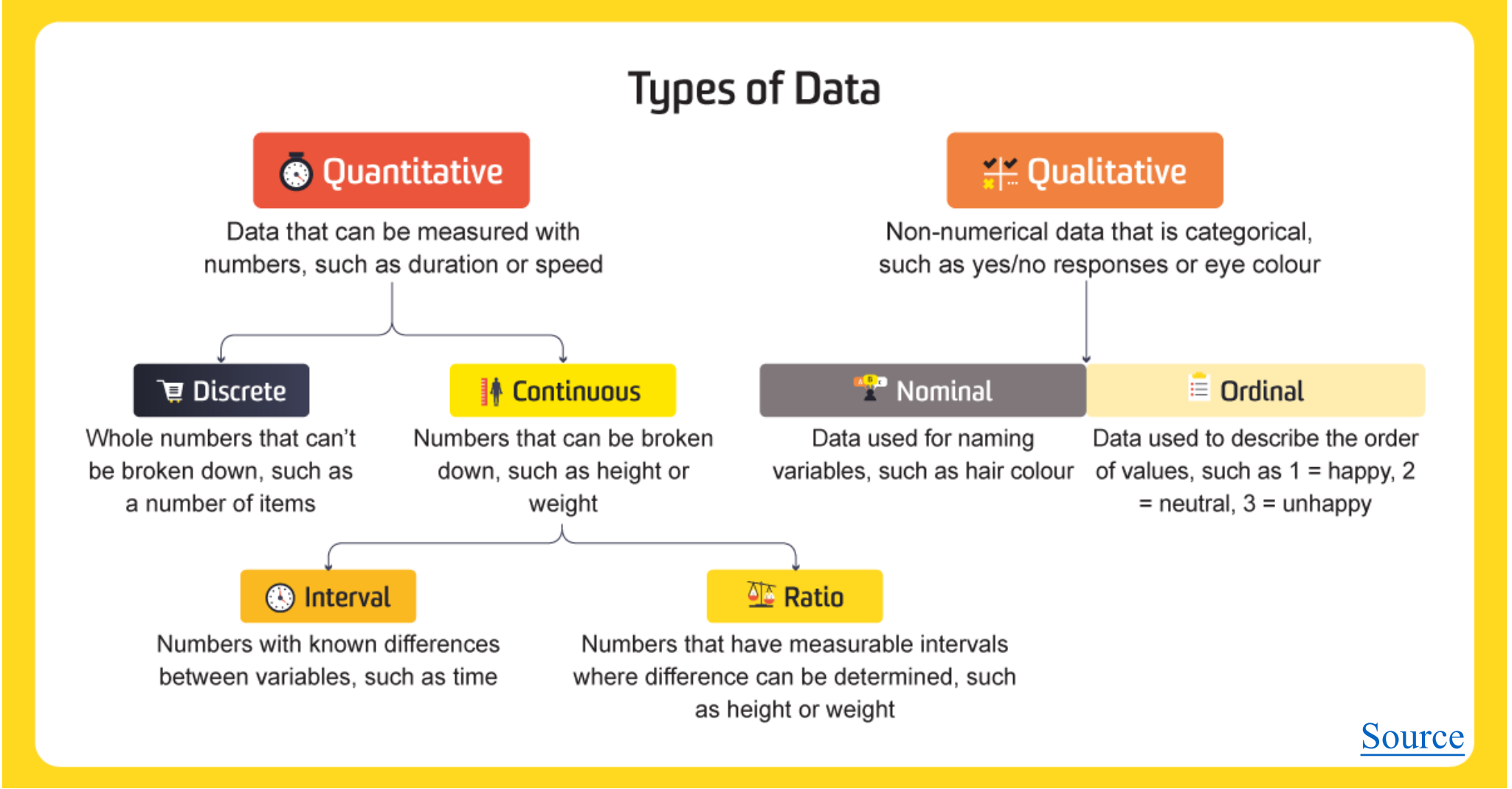
https://www.citizenme.com/human-data/

Personal Data

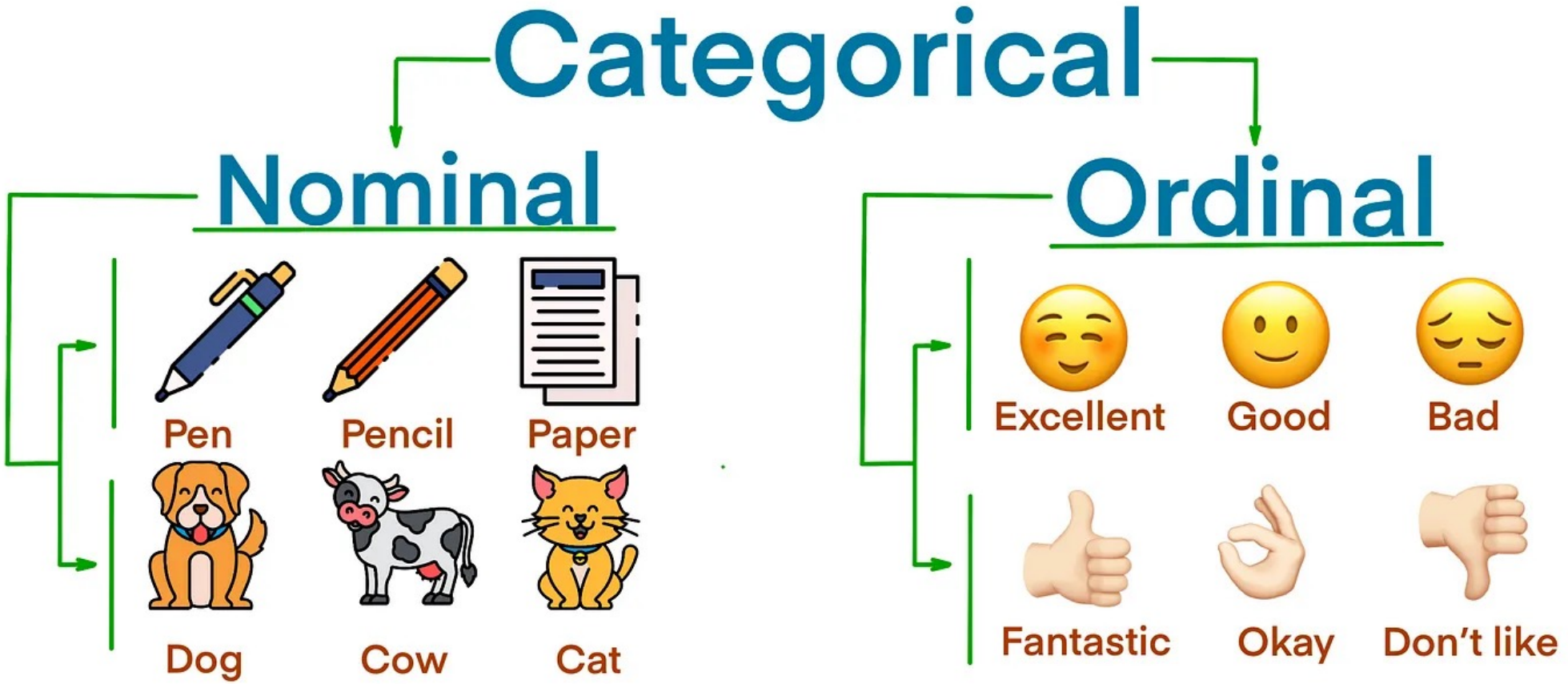


https://www.citizenme.com/human-data/

What type of data does machine learning need?



Source


Categorical
Nominal
Pen
Pencil
Paper
Dog
Cow
Cat
Ordinal
Excellent
Good
Bad
Fantastic
Okay
Don't like

Source

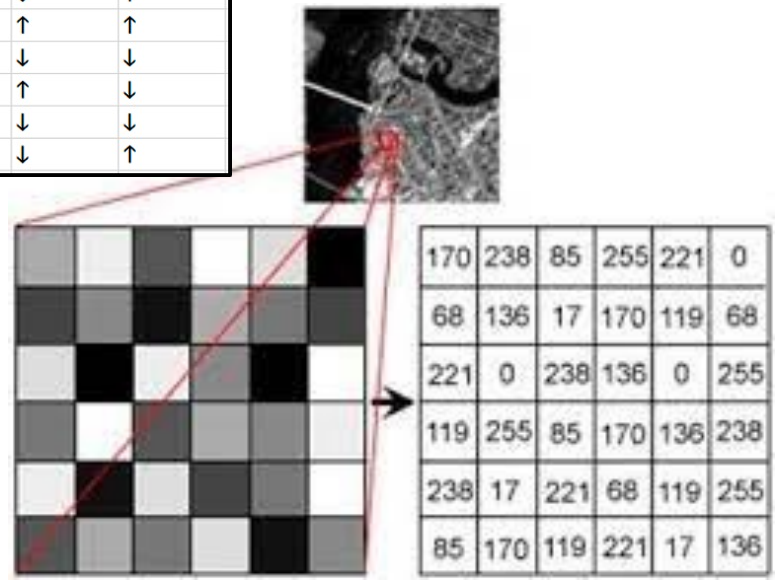
Text data

Source

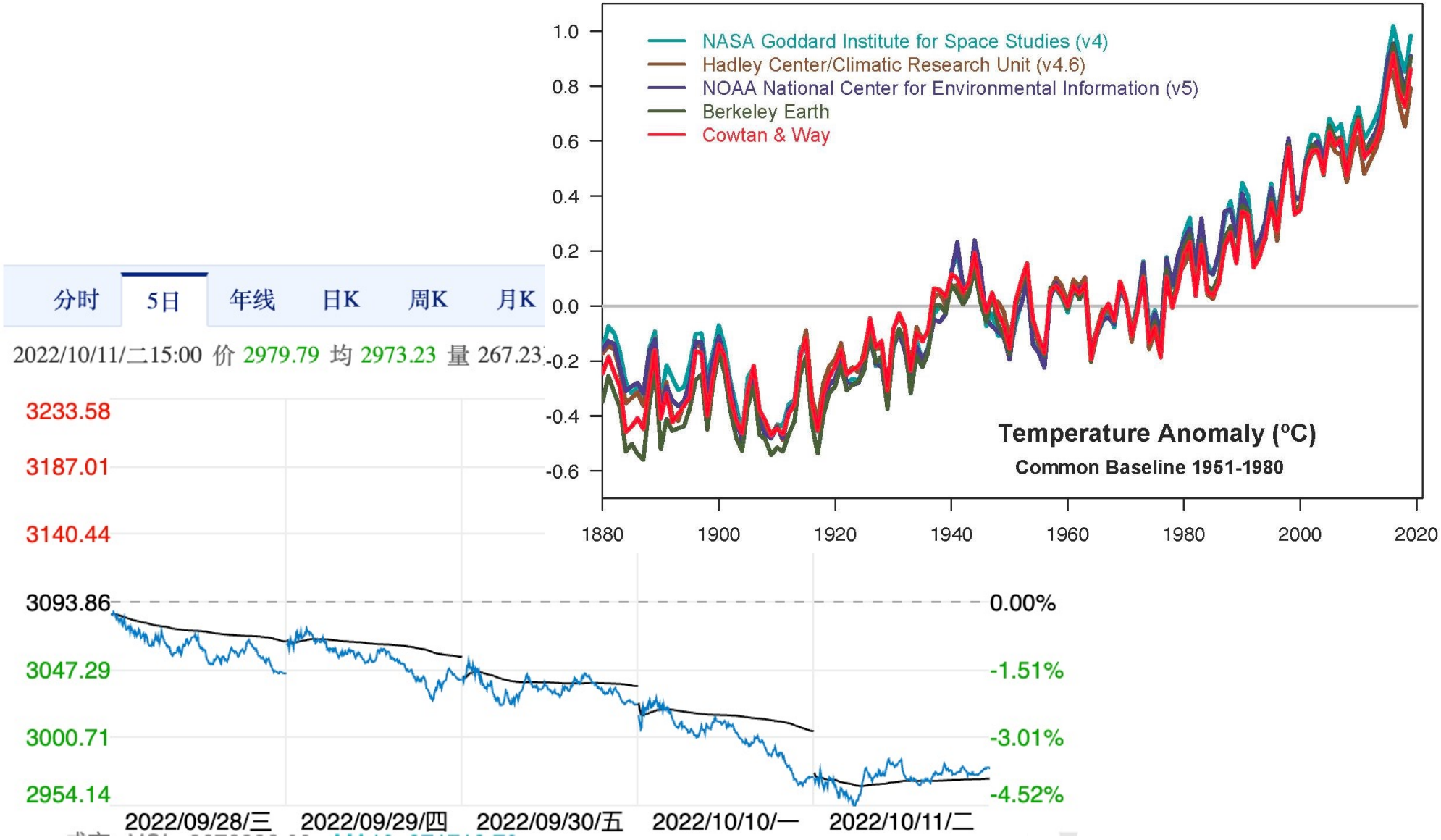
Numerical data

O9

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	DATA						RANK					TREND		
2	Sales Pers	May	June	July	Aug		May	June	July	Aug		June	July	Aug
3	1	84	138	72	45		35	2	42	60		↑	↓	↓
4	2	98	57	122	129		27	52	13	9		↓	↑	↑
5	3	83	108	107	107		36	19	20	20		↑	↓	→
6	4	91	135	120	56		30	3	14	54		↑	↓	↓
7	5	133	61	47	62		6	49	59	47		↓	↓	↑
8	6	73	80	86	113		41	37	33	18		↑	↑	↑
9	7	57	98	66	117		52	27	44	16		↑	↓	↑
10	8	86	52	134	132		33	56	5	8		↓	↑	↓
11	9	53	99	48	106		55	25	58	22		↑	↓	↑
12	10	96	80	59	69		29	37	50	43		↓	↓	↑
13	11	78	102	104	116		40	24	23	17		↑	↑	↑
14	12	133	119	90	89		6	15	31	32		↓	↓	↓
15	13	79	127	128	124		39	11	10	12		↑	↑	↓
16	14	49	66	64	62		57	44	46	47		↑	↓	↓
17	15	58	135	99	141		51	3	25	1		↑	↓	↑



Time series data



Where Do We Get Data for ML?

Five of the most popular ML dataset resources:

Google → Google's Dataset Search

Microsoft → Microsoft Research Open Data

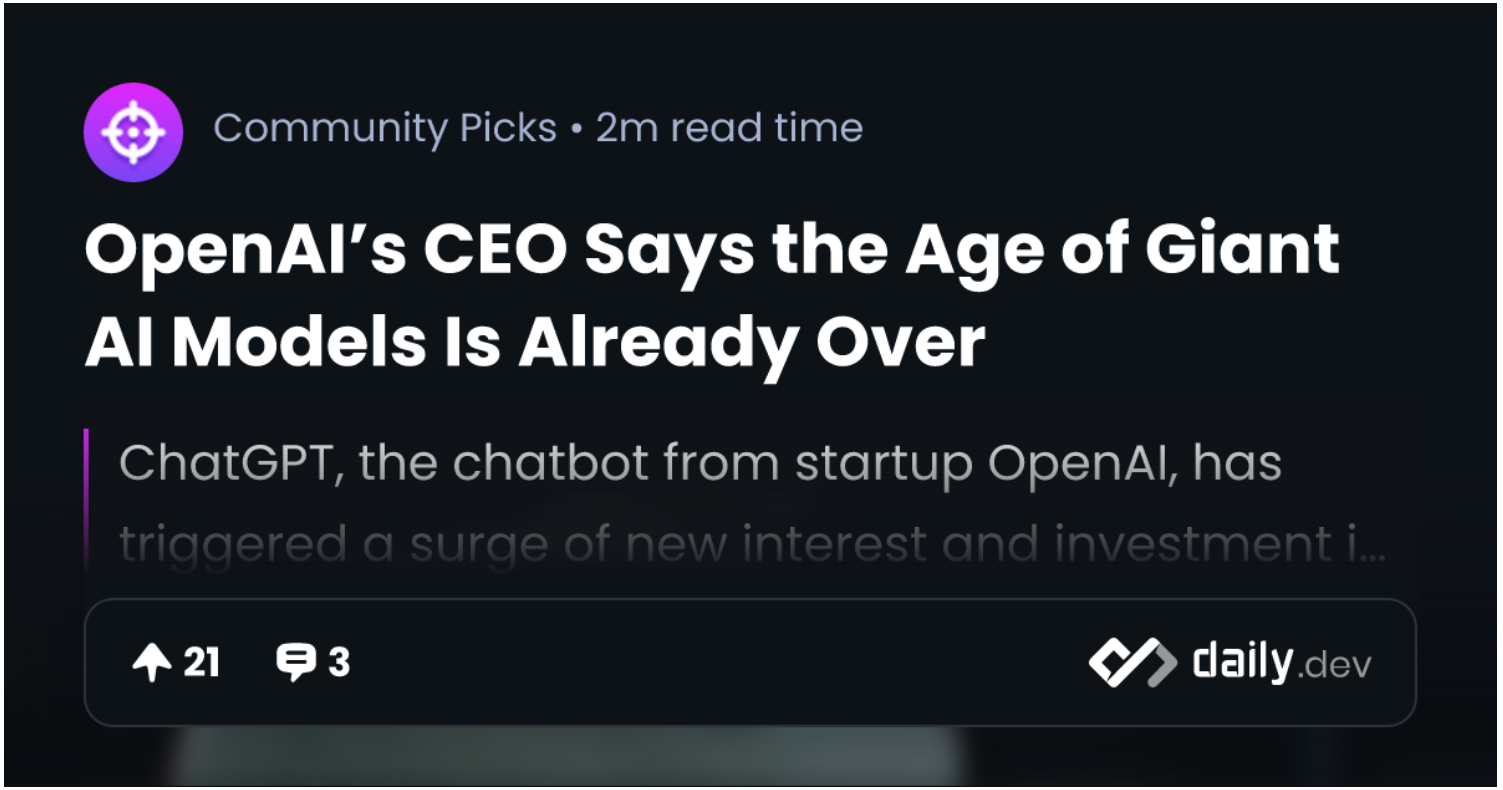
aws → Amazon Datasets

UCI → UCI Machine Learning Repository

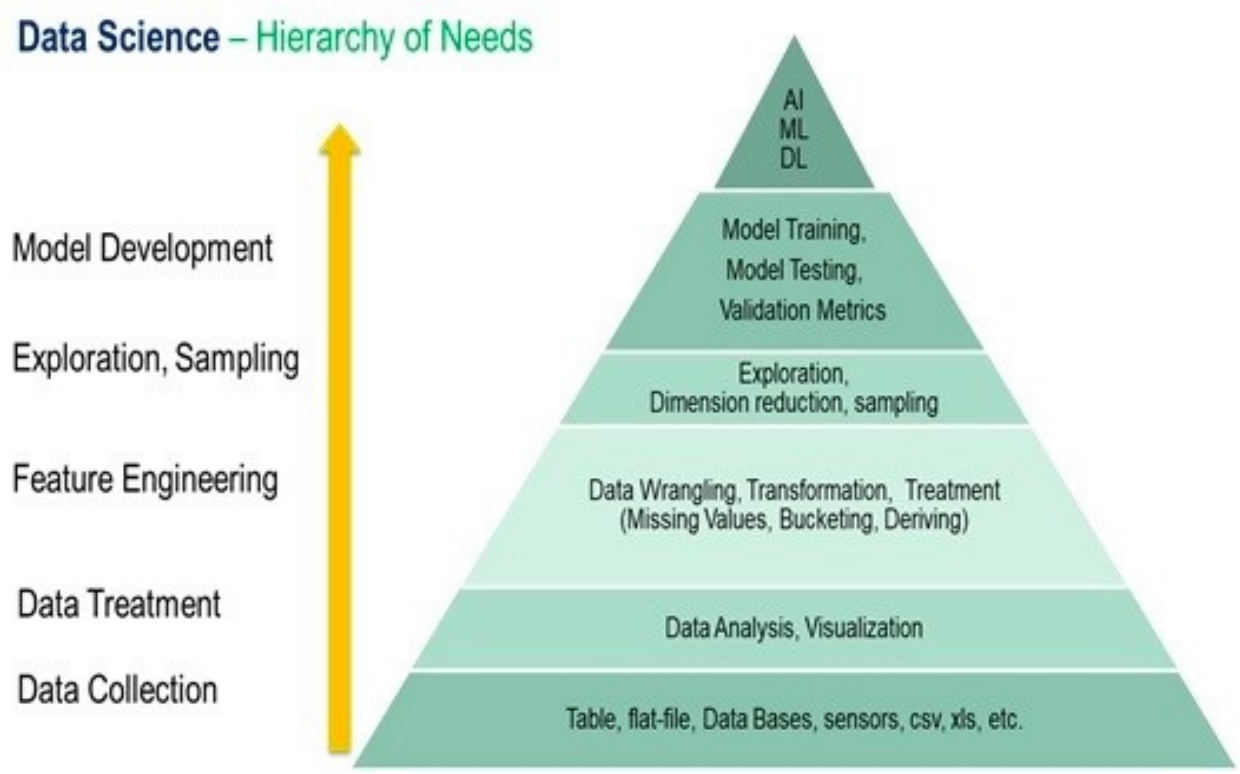
DATA.GOV → Government Datasets

Why is data important for ML?

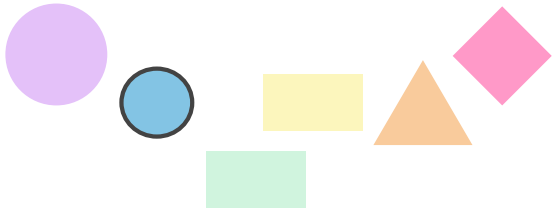
- Big data provides ample amounts of raw material from which machine learning systems can derive insights.
- A dumb algorithm with lots and lots of data beats a clever one with modest amounts of it.



Why is ML important to Data?

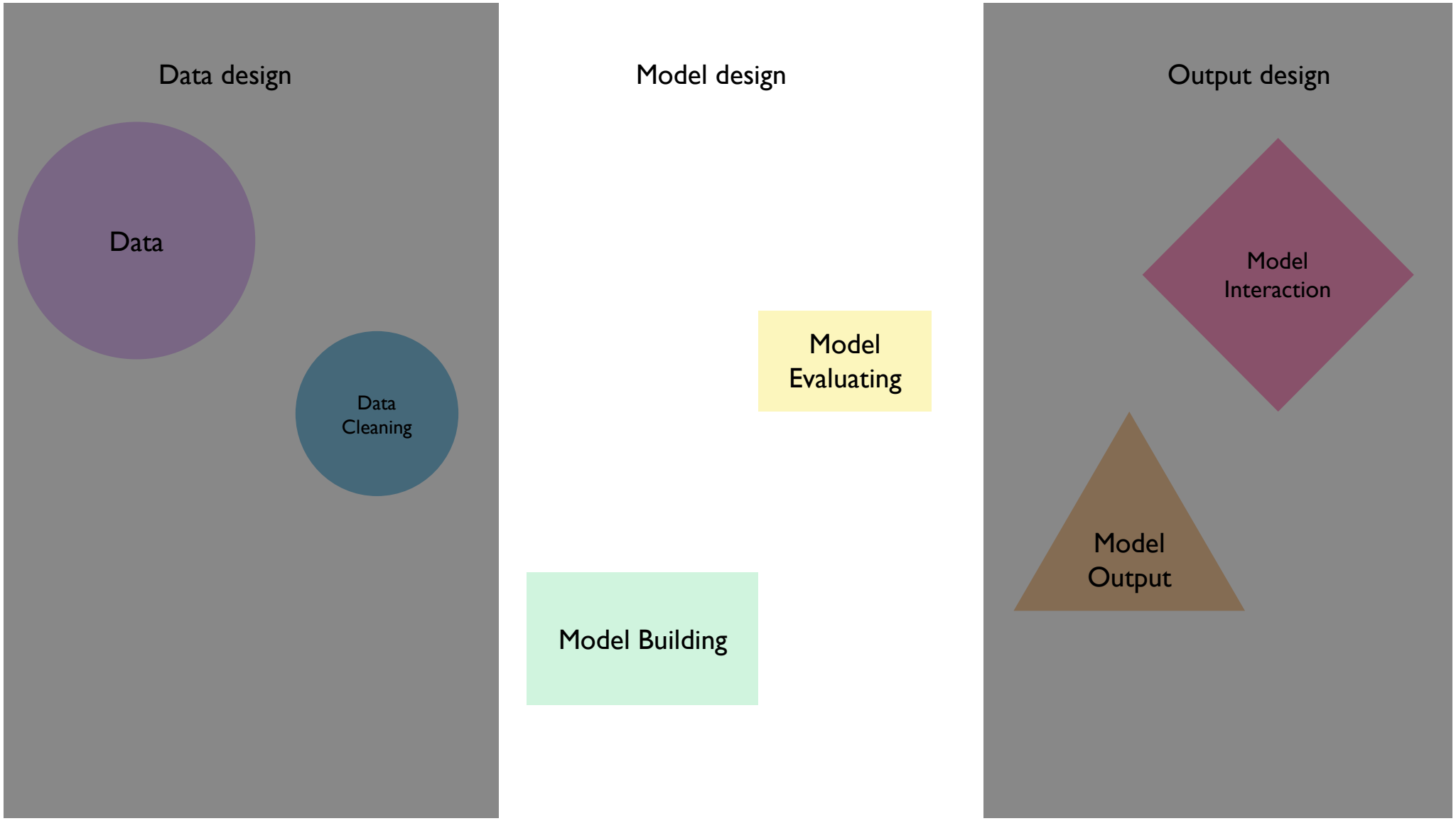


Data Cleaning



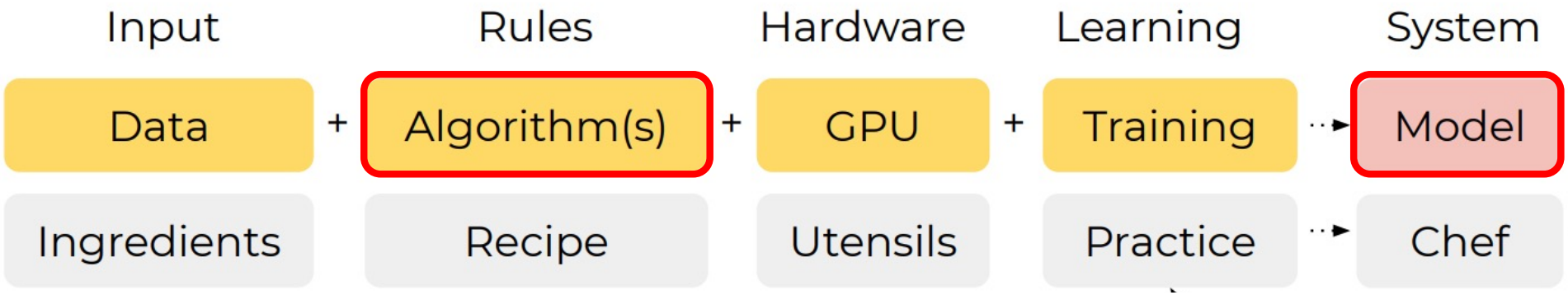
- Format the data in a way that the computer can read it
- Might choose to exclude missing values
- Explore your data - look for trends that might inform you
- Remember - how was your data collected?
  How is it going to be used?

- Key vocab:
  - Normalizing - adjusting your data to a common scale
  - Remove NA - removing “null” values

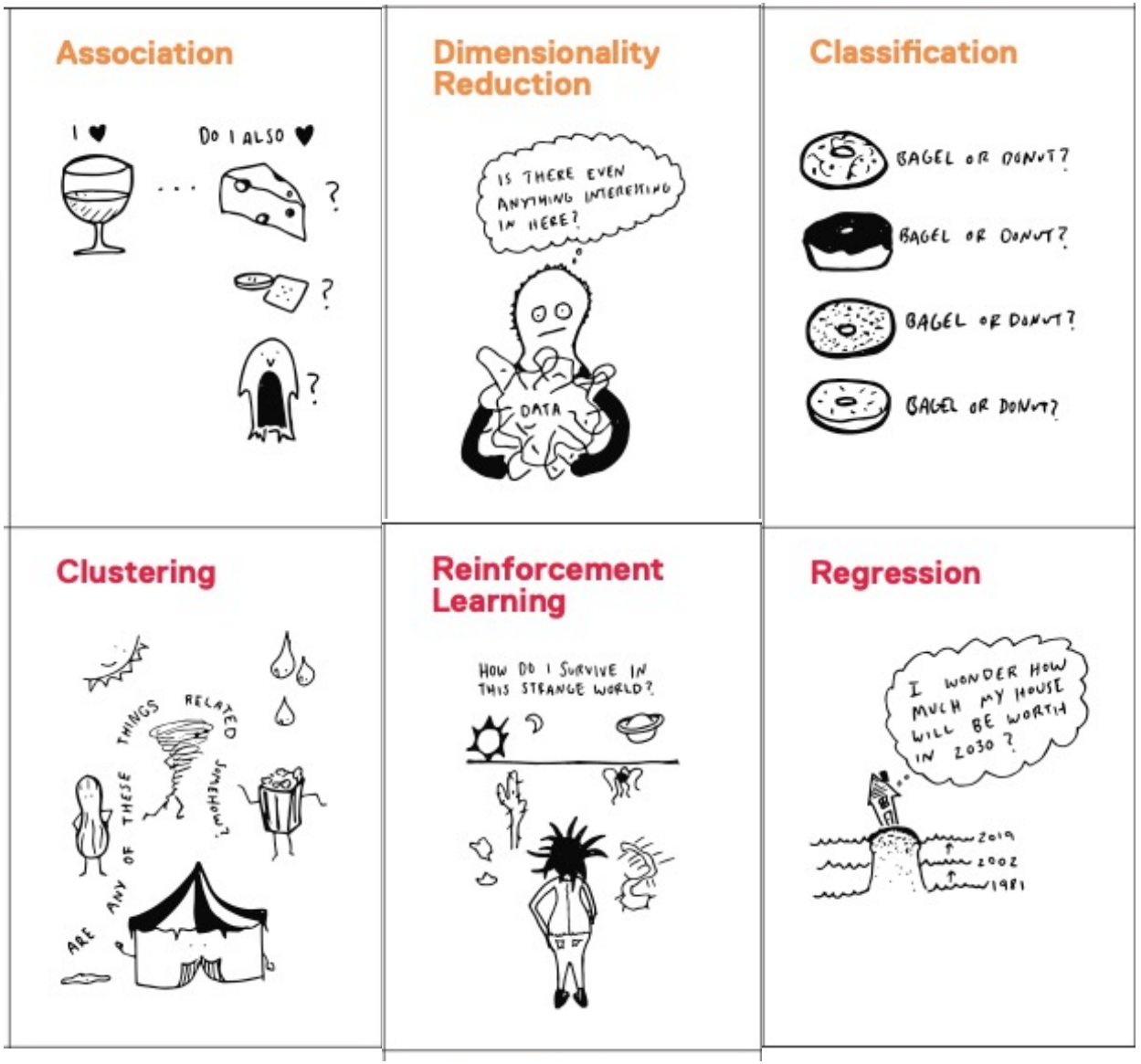

Data design
Data
Data Cleaning
Model design
Model Evaluating
Model Building
Output design
Model Interaction
Model Output

Model Building

- Algorithm vs Model
- Data + Algorithm = Model



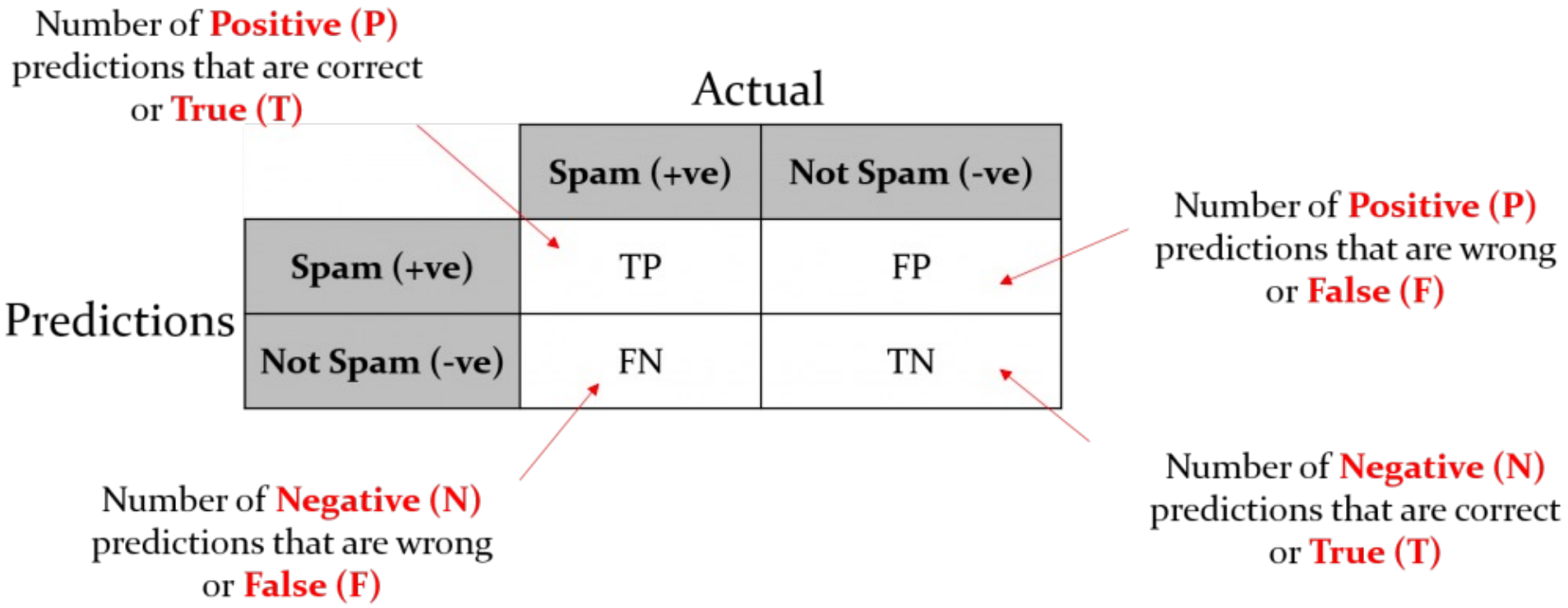
- Ask yourself: What type of problem are you trying to solve?

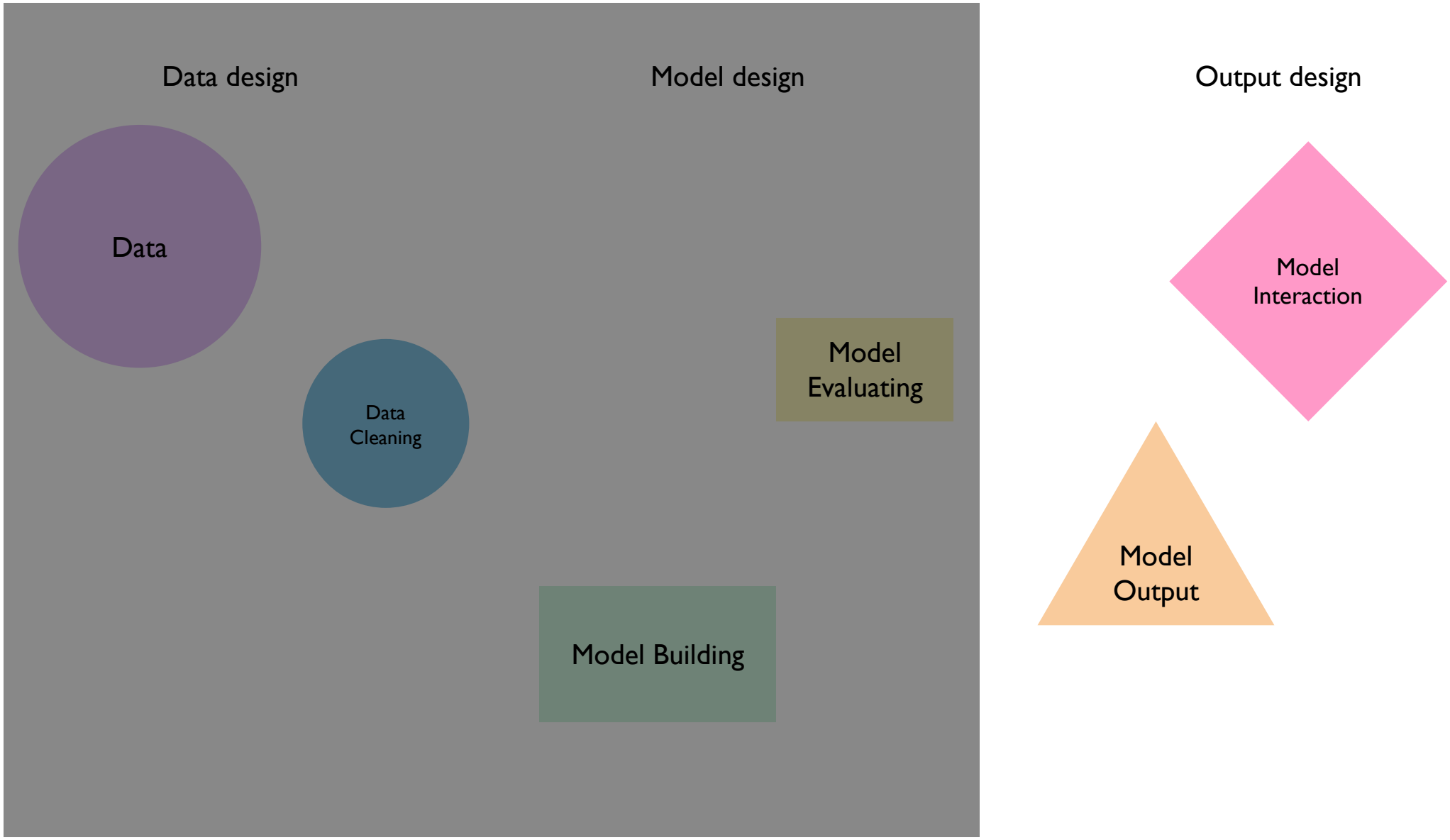


https://dschool.stanford.edu/resources/i-love-algorithms

Model Evaluating

- How well can your model [predict] unseen data?




Data design
Model design
Output design
Data
Data Cleaning
Model Evaluating
Model Building
Model Interaction
Model Output

Model Output

- What will the output of your model look like?

- Key vocab:
  - Confidence Interval - range of values we are fairly sure our true value lies in
  - Multi vs Single Classification - are you predicting membership in one cat, or multi
  - Generative Model - generates plausible values that look like values in data set

Model Interactions

- How do you give feedback to your model?
- How can you leverage the model capabilities to make it more impactful?
- What do you need to do to transform the model output to make it usable?

- Examples:
  - Interactions to improve model training
  - Human input and oversight
  - User Experience design to improve usability