An introduction to Machine Learning for Astronomy

Marc Huertas-Company

UNIVERSITY OF CALIFORNIA SANTA CRUZ

l'Observatoire de Paris LERMA

université PARIS DIDEROT

institut universitaire de France

INSTITUTO DE ASTROFISICA DE CANARIAS IAC

PROGRAM FOR THE WEEK

- INTRODUCTION TO SUPERVISED ML
- FOUNDATIONS TO NEURAL NETWORKS
- CONVOLUTIONAL NEURAL NETWORKS
- INTRODUCTION TO UNSUPERVISED ML

PRACTICAL INFO

FULL ONLINE COURSE VIA ZOOM:

https://rediris.zoom.us/j/99594075110?pwd=Y0dFcTVZWFFqb0xUQm41NWpweWs5dz09

Meeting ID: 995 9407 5110
Password: 842832

MORNING ~[9h30-12h30]: THEORY / LECTURES

AFTERNOON ~[14H30-17H30]: TUTORIALS

PRACTICAL INFO

FULL ONLINE COURSE VIA ZOOM:

https://rediris.zoom.us/j/99594075110?pwd=Y0dFcTVZWFFqb0xUQm41NWpweWs5dz09

Meeting ID: 995 9407 5110
Password: 842832

/E RULES DURING THE ZOOM MEETING:

Keep your Microphones Muted

If need to ask questions (which I recommend!) use the “raise hand” option

PRACTICAL INFO

OTHER AVAILABLE RESSOURCES:

GITHUB REPO WITH SLIDES AND TUTORIALS:

https://github.com/mhuertascompany/DL_ED127_2020

Slack Channel for interactive discussions / questions:

https://join.slack.com/t/iaed127/shared_invite/zt-f5sohajk-uP00KQtWdaPIC_Znw6H2EA

SOME PRELIMINARY NOTES

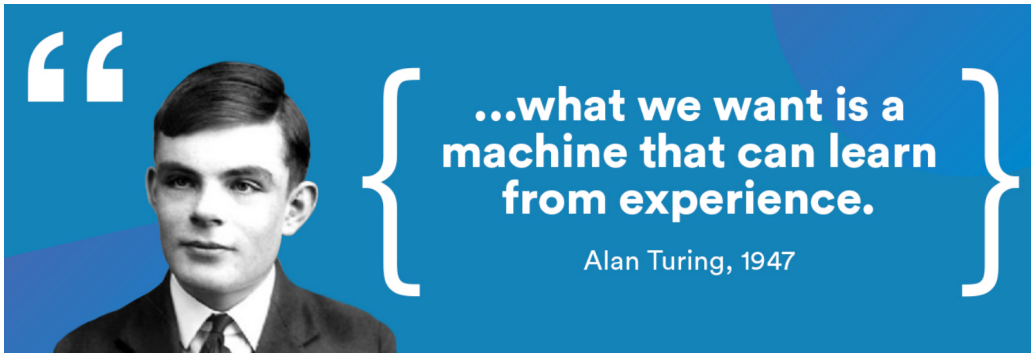
I AM NOT A MACHINE LEARNING RESEARCHER

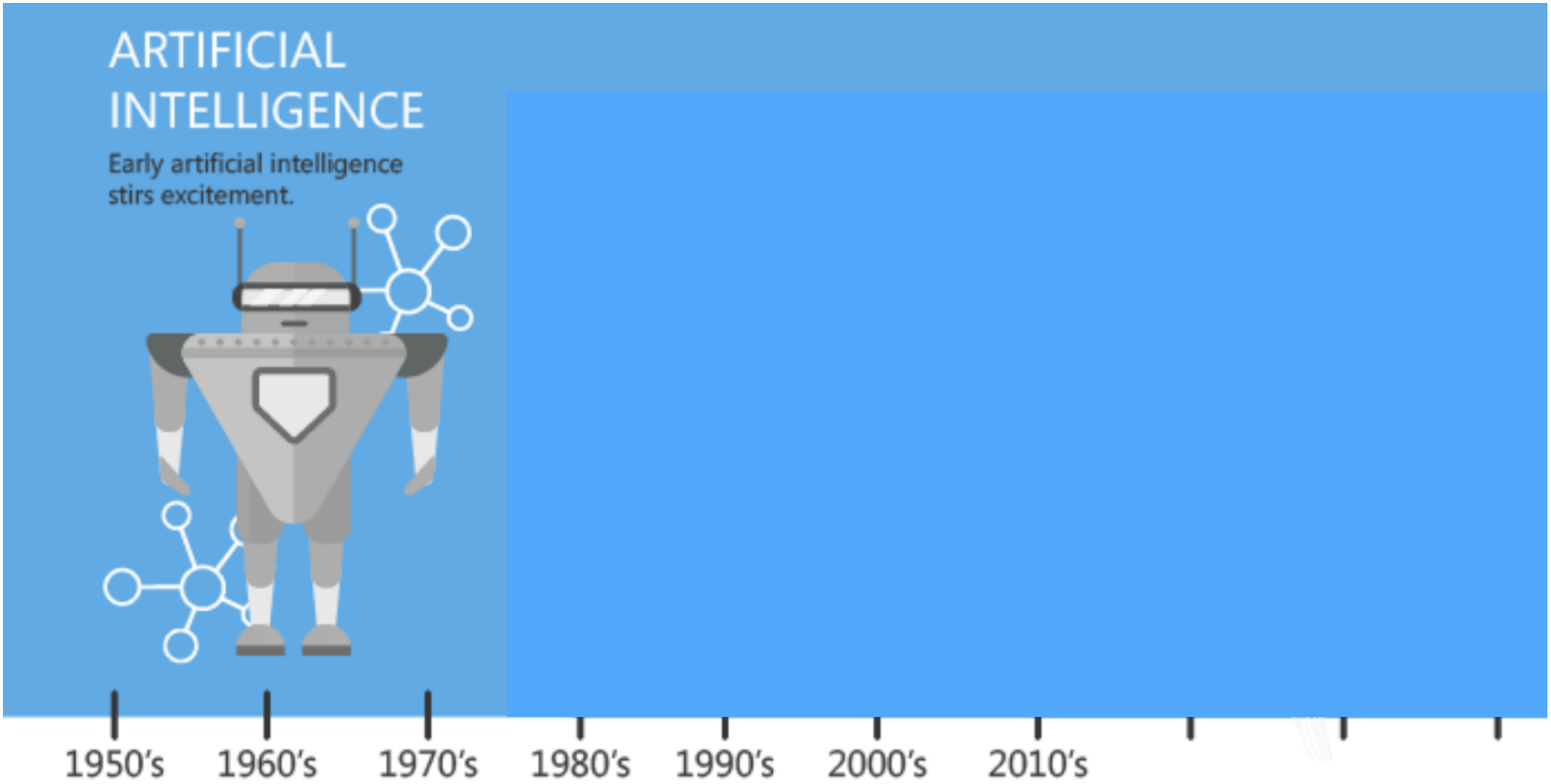
SOME PRELIMINARY NOTES

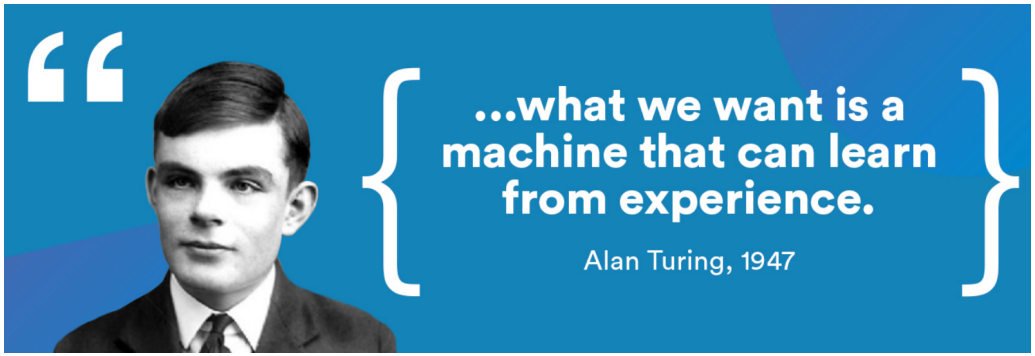
I AM NOT A MACHINE LEARNING RESEARCHER

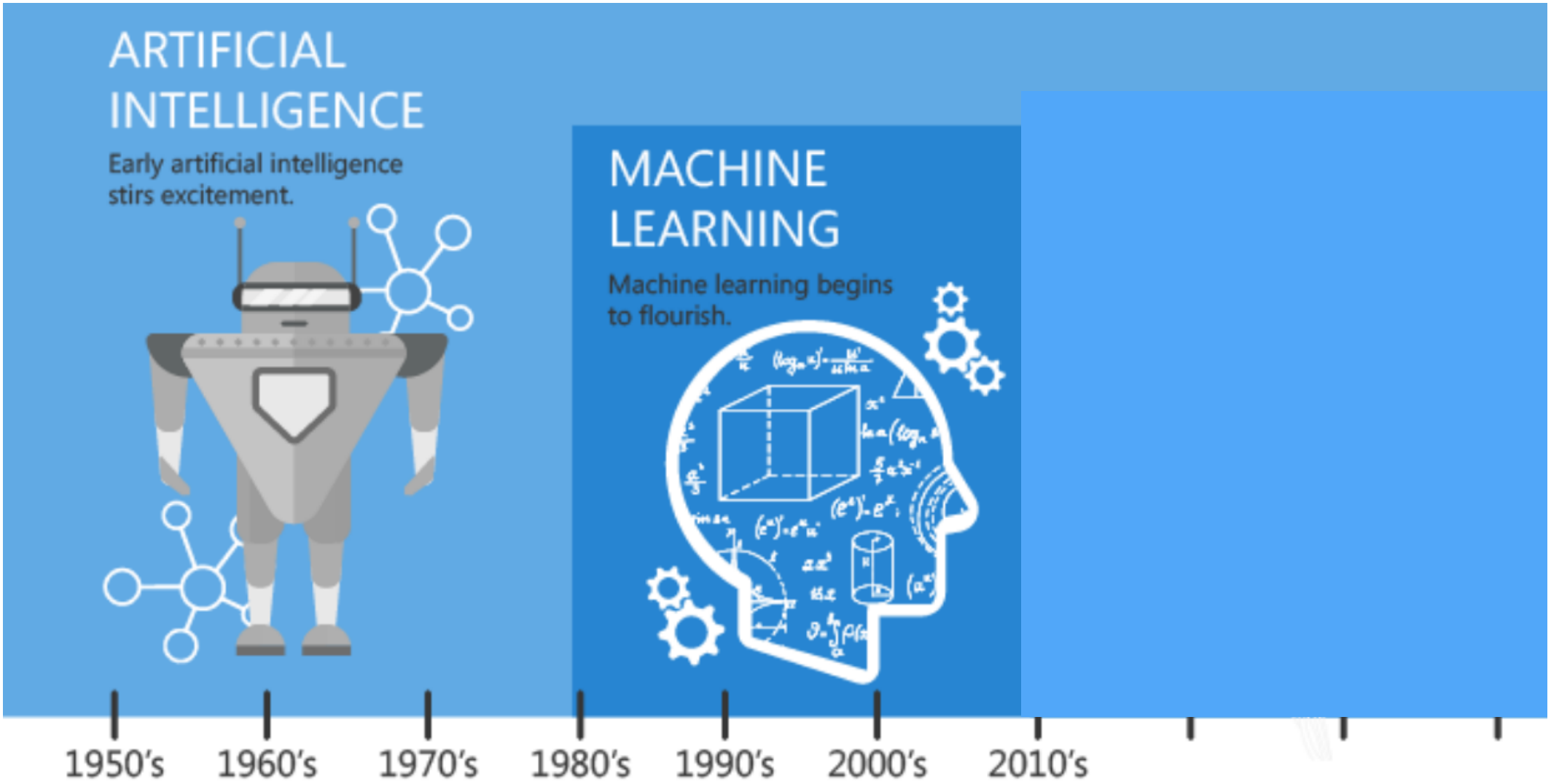
ONLY AN ASTRONOMER WHO HAS BEEN USING MACHINE LEARNING FOR THE LAST ~14 YEARS FOR MY RESEARCH

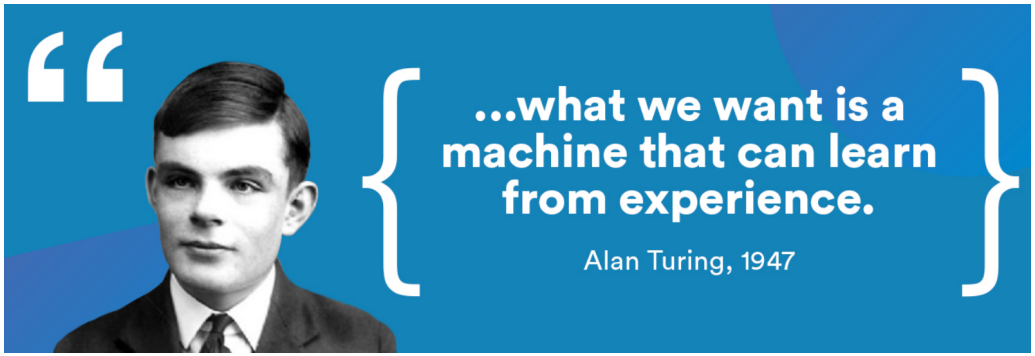
THESE LECTURES ARE INTENDED TO PROVIDE A **GLOBAL** UNDERSTANDING OF HOW AI TECHNIQUES WORK AND ESPECIALLY **HOW TO USE THEM FOR YOUR RESEARCH**

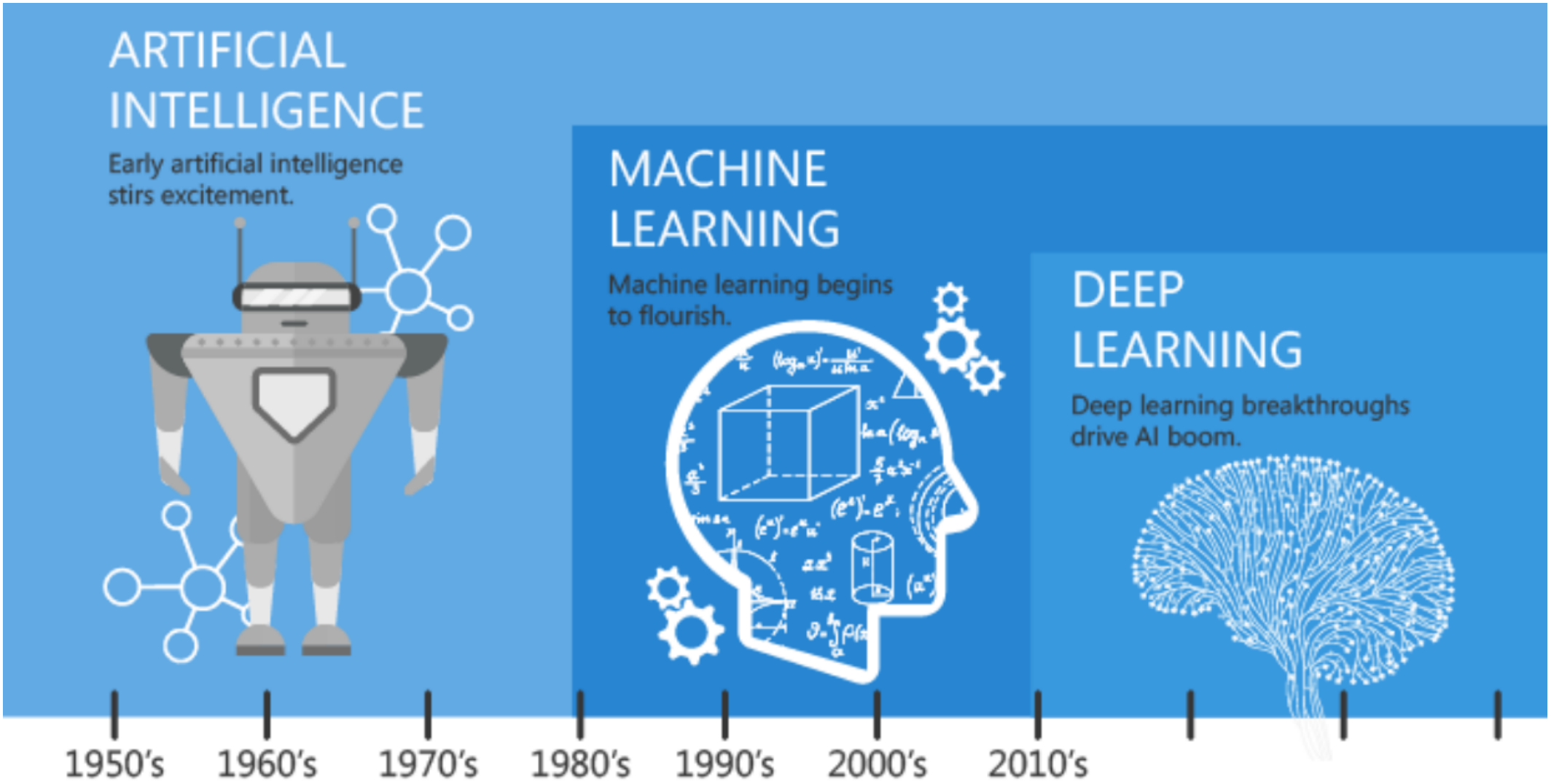

...what we want is a machine that can learn from experience.
Alan Turing, 1947


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

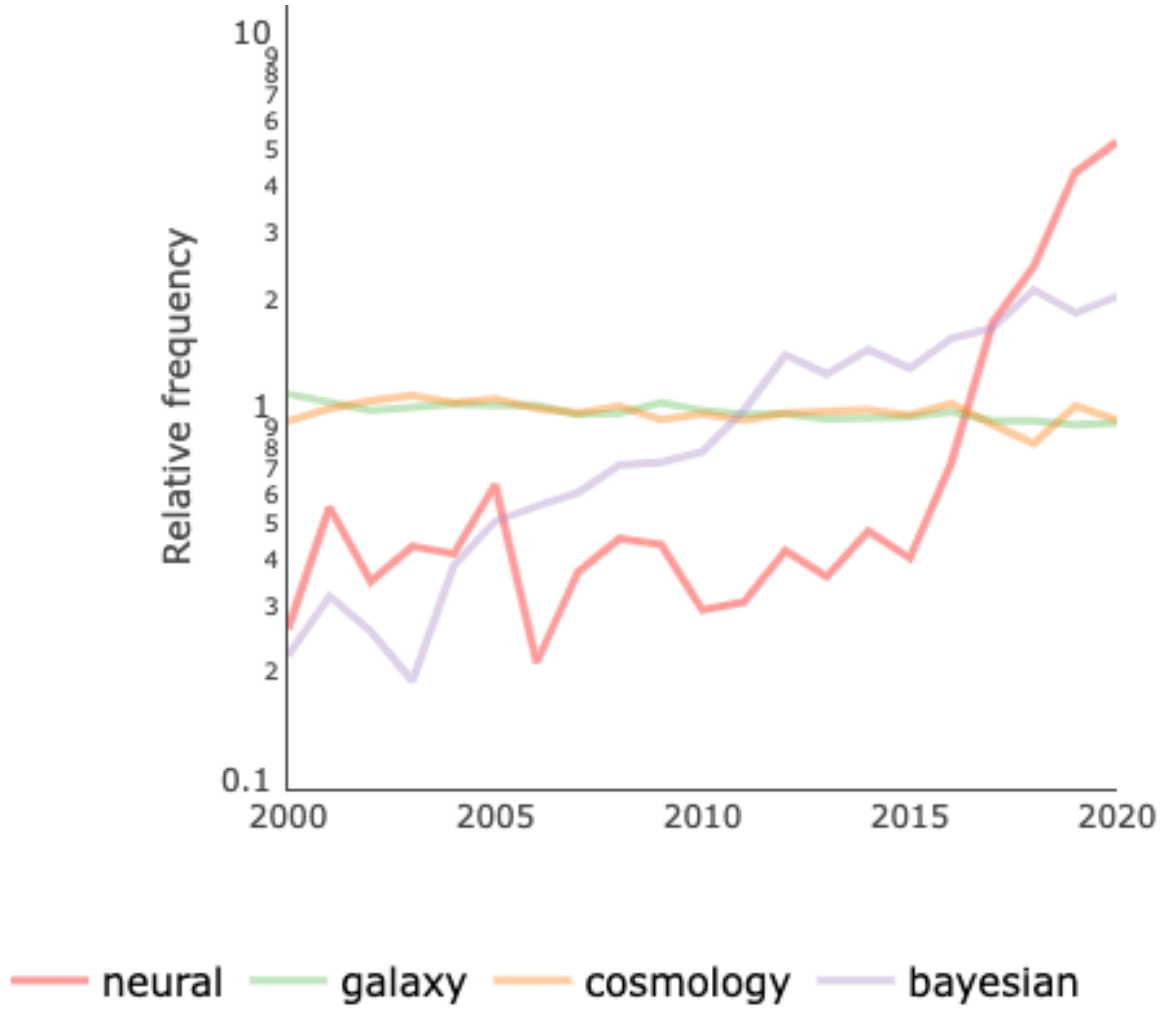

...what we want is a machine that can learn from experience.
Alan Turing, 1947


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

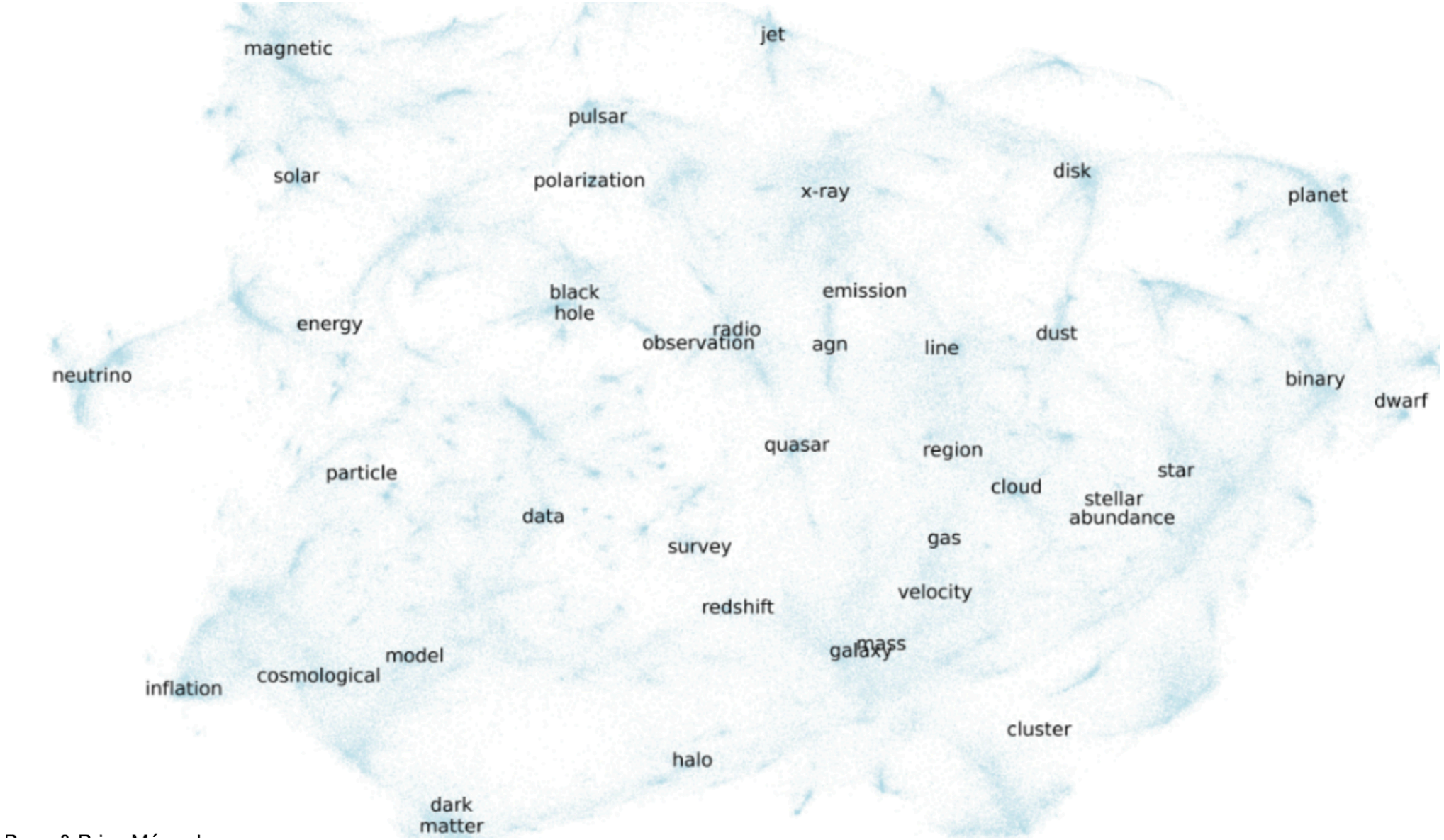

...what we want is a machine that can learn from experience.
Alan Turing, 1947

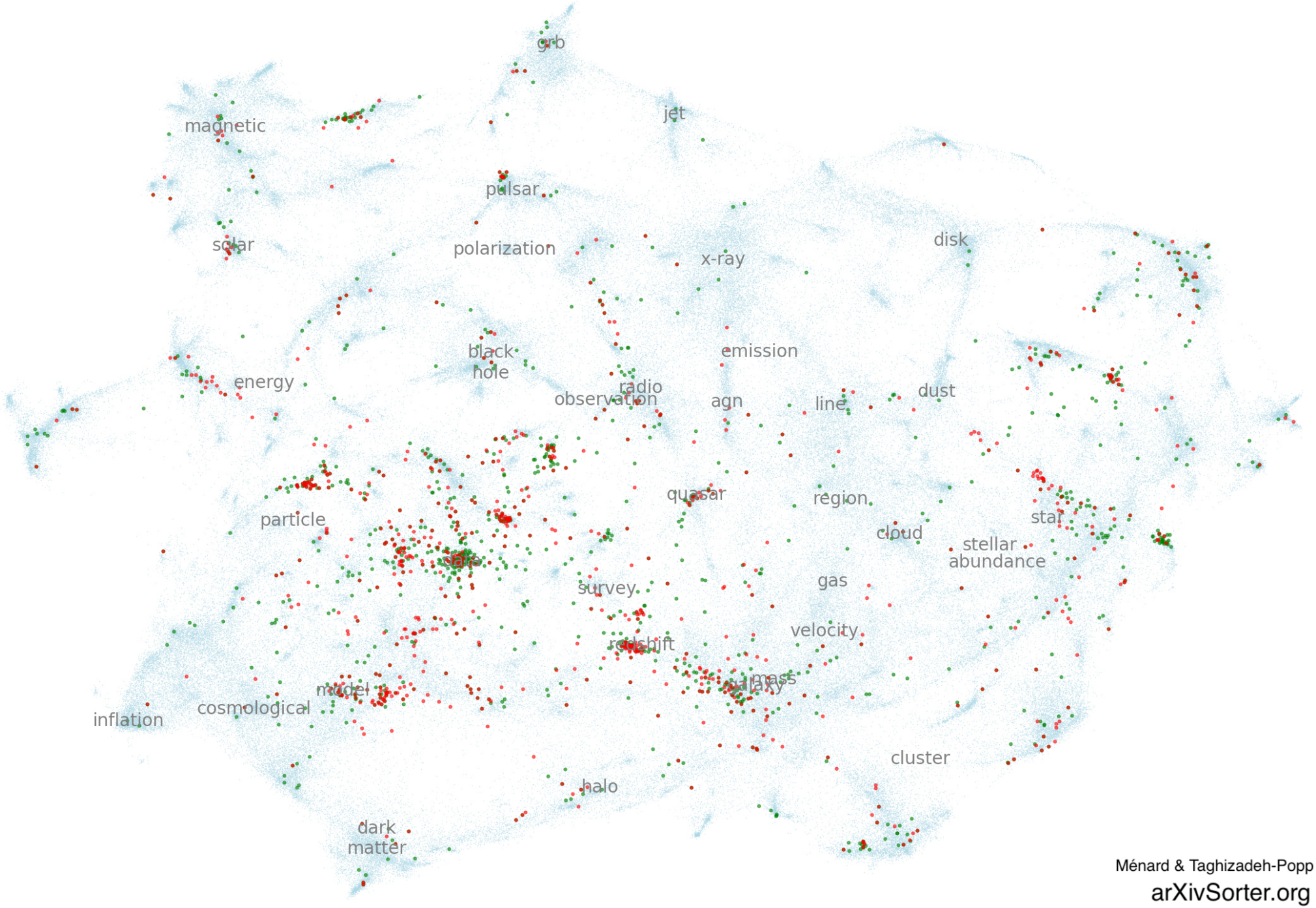

ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

Relative change of the number of papers on arXiv/astro-ph



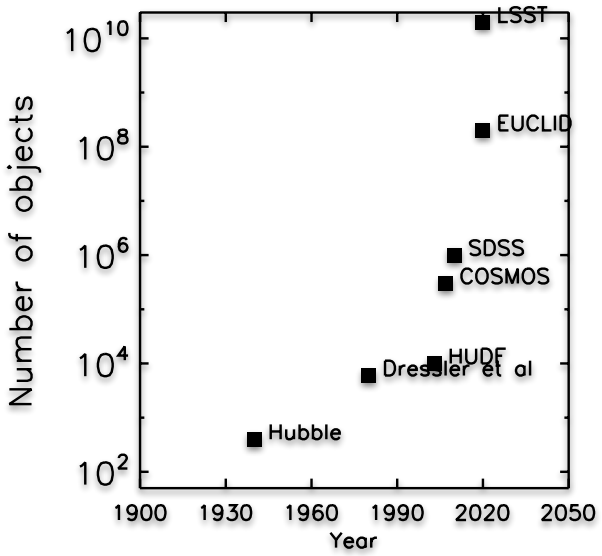
Huertas-Company+22




grb
jet
magnetic
pulsar
solar
polarization
x-ray
disk
black
hole
emission
energy
radio
observation
agn
line
dust
quasar
region
star
particle
cloud
stellar
abundance
gas
survey
velocity
mass
galaxy
model
cosmological
inflation
cluster
halo
dark
matter
Ménard & Taghizadeh-Popp
arXivSorter.org

WHY DO WE NEED THESE TOOLS IN ASTRONOMY?

AS IN MANY OTHER DISCIPLINES THE BIG-DATA REVOLUTION HAS ARRIVED TO ASTRONOMY TOO



← BIG-DATA REVOLUTION

← we are here

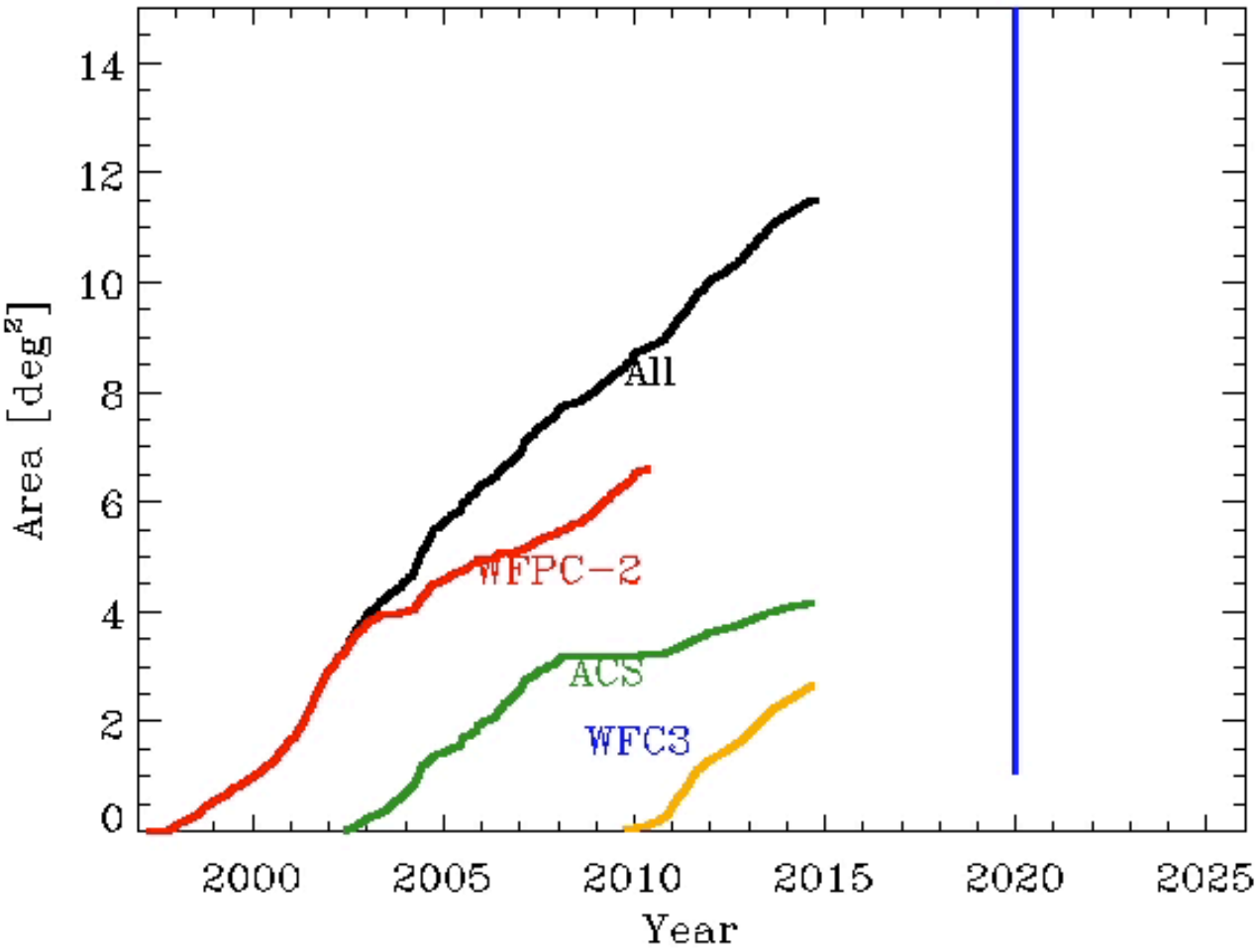
EXTREMELY LARGE IMAGING SURVEYS DELIVERING BILLIONS OF OBJECTS IN 2-5 YEARS

Large Synoptic Survey Telescope

euclid consortium

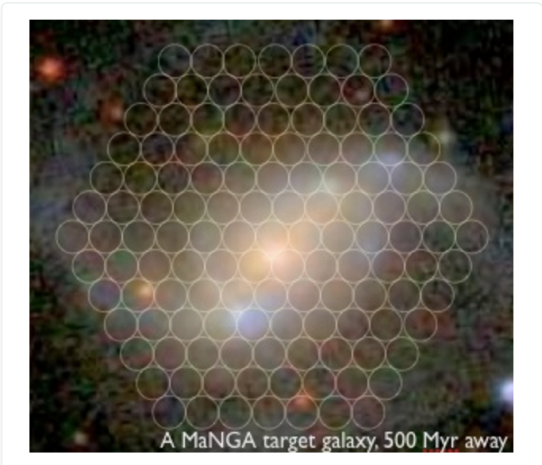
WIDE-FIELD INFRARED SURVEY TELESCOPE
WFIRST
ASTROPHYSICS • DARK ENERGY • EXOPLANETS

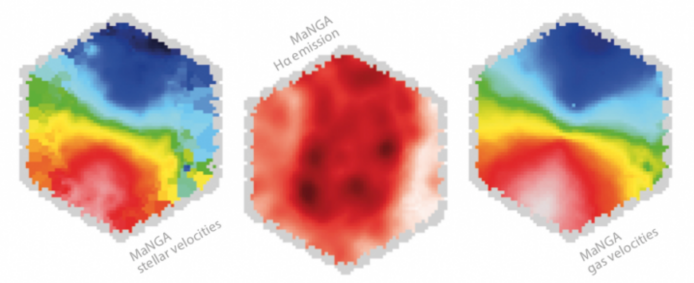
LSST simulation



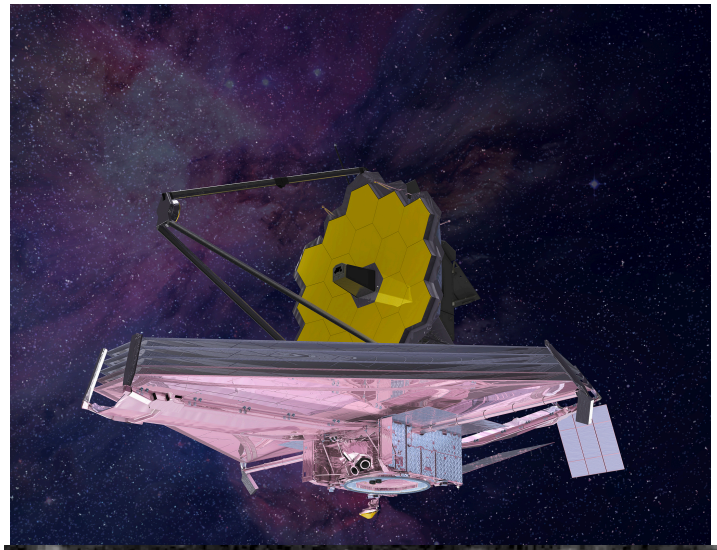
(Thanks to J. Brinchmann)

NOT ONLY VOLUME: AN INCREASING COMPLEXITY OF DATA

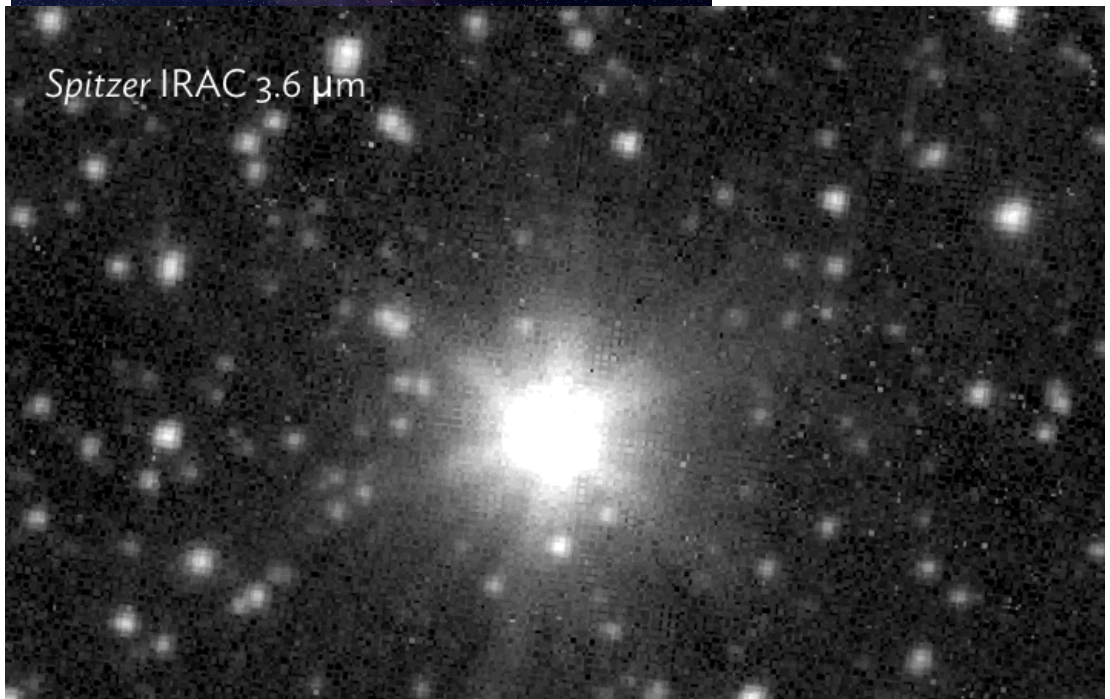


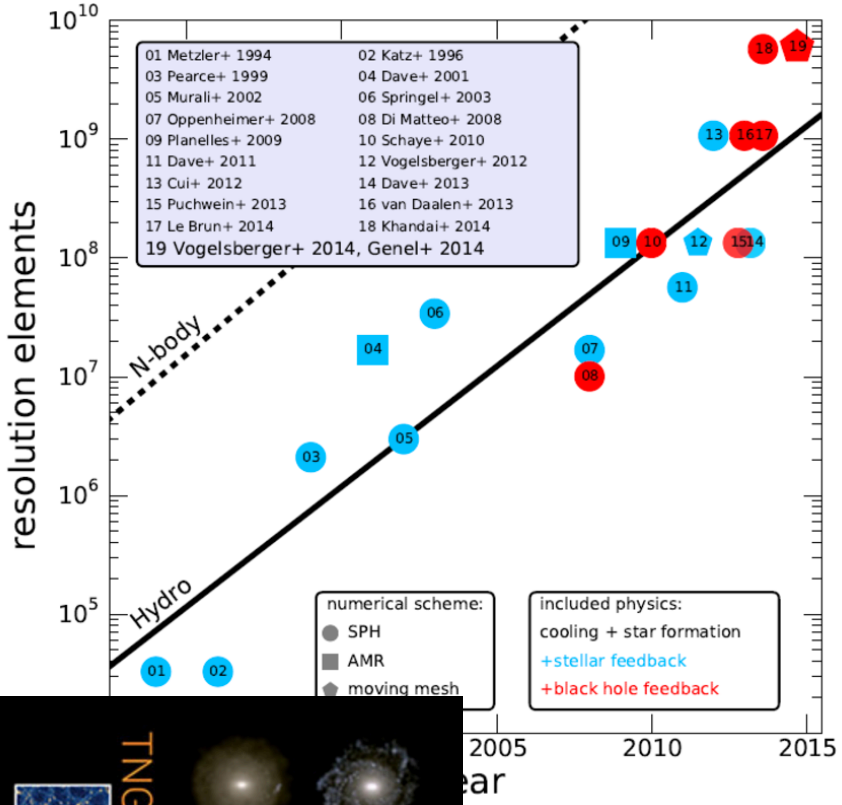


MANGA Survey



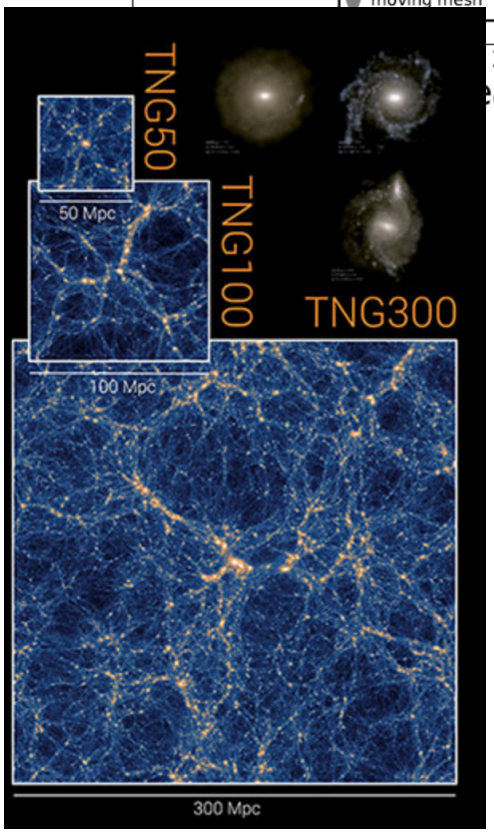
JWST

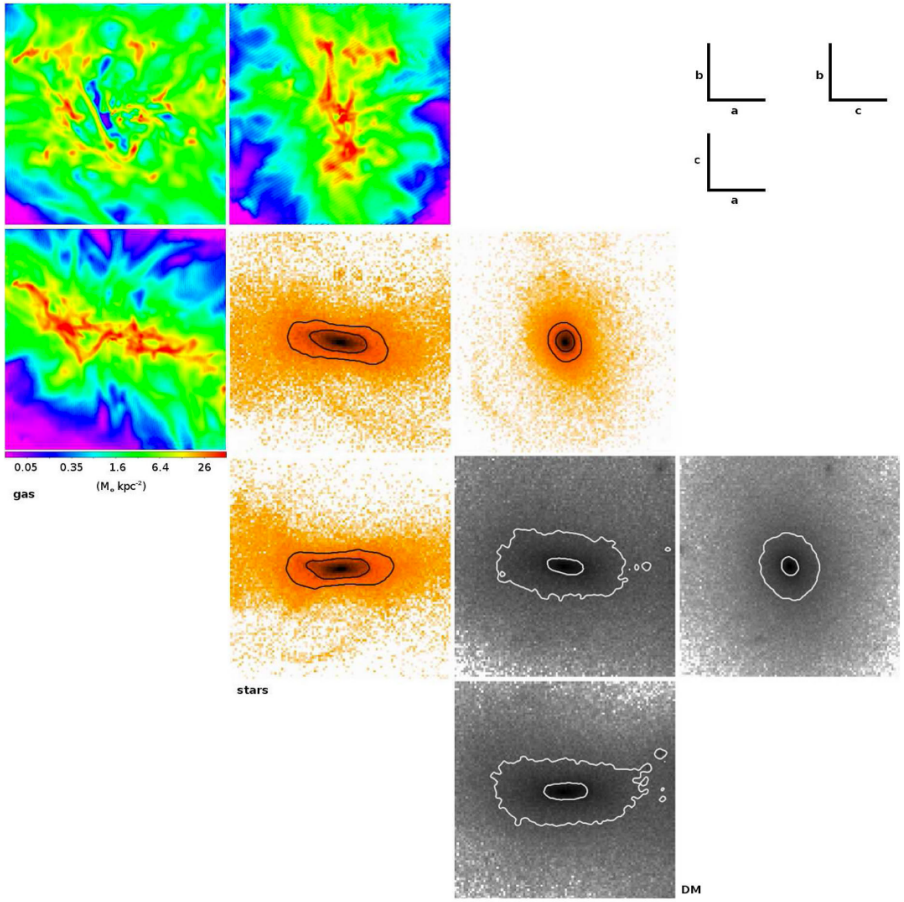




Genel+14

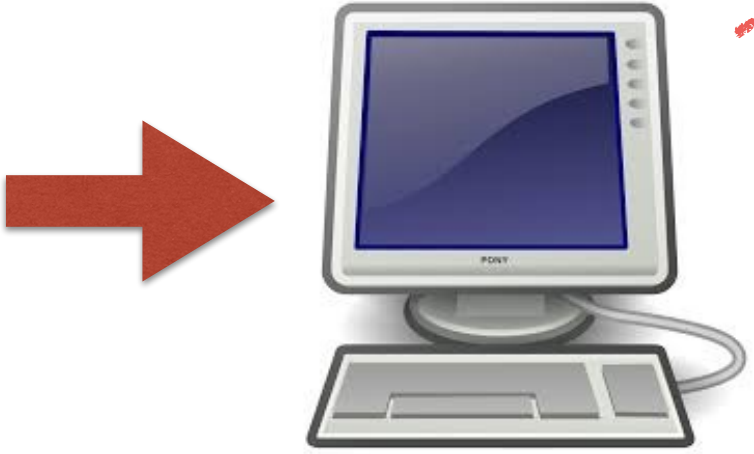
AND ALSO SIMULATIONS!





Ceverino+15

BEFORE 2012….

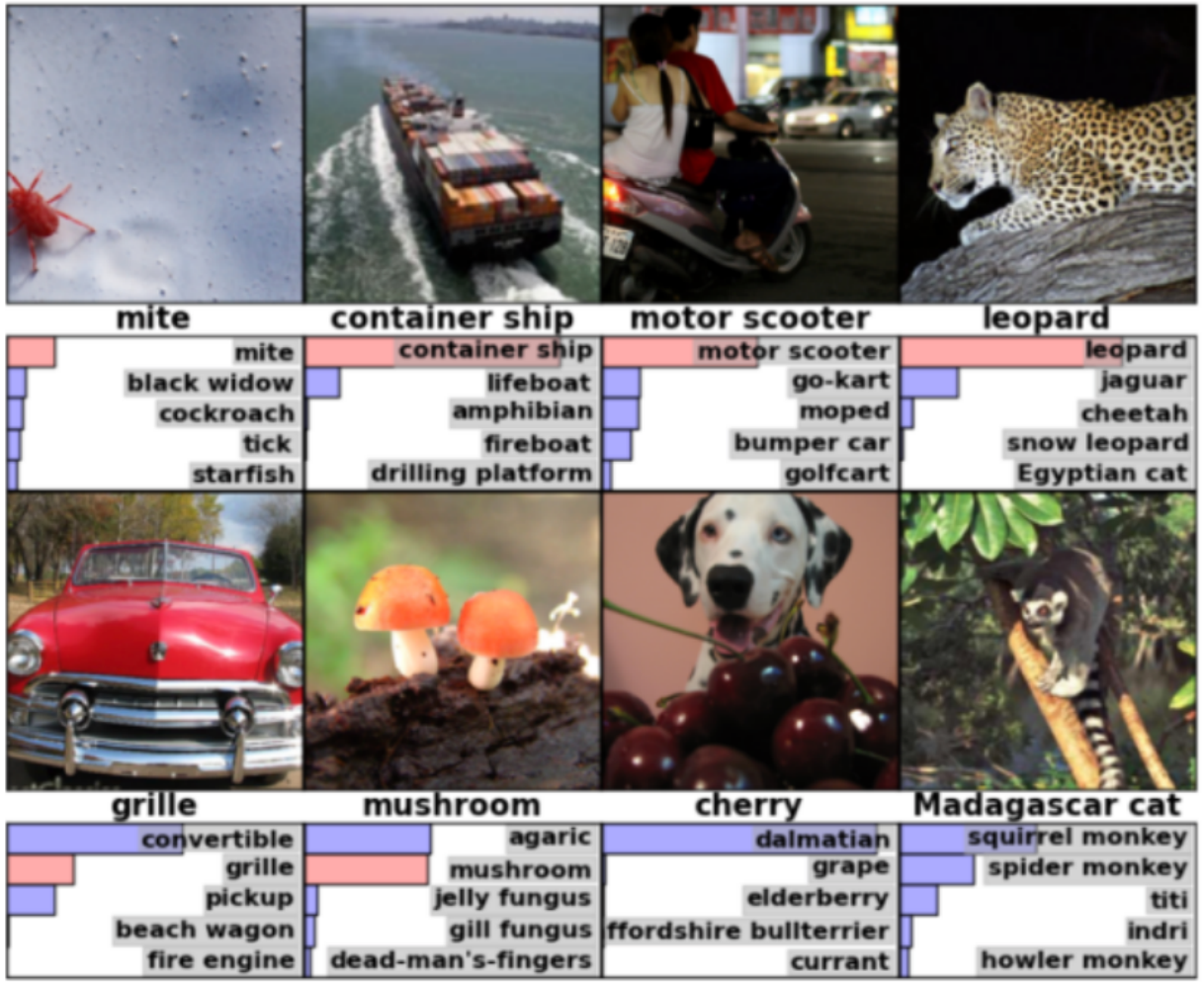


CAT?

DOG?

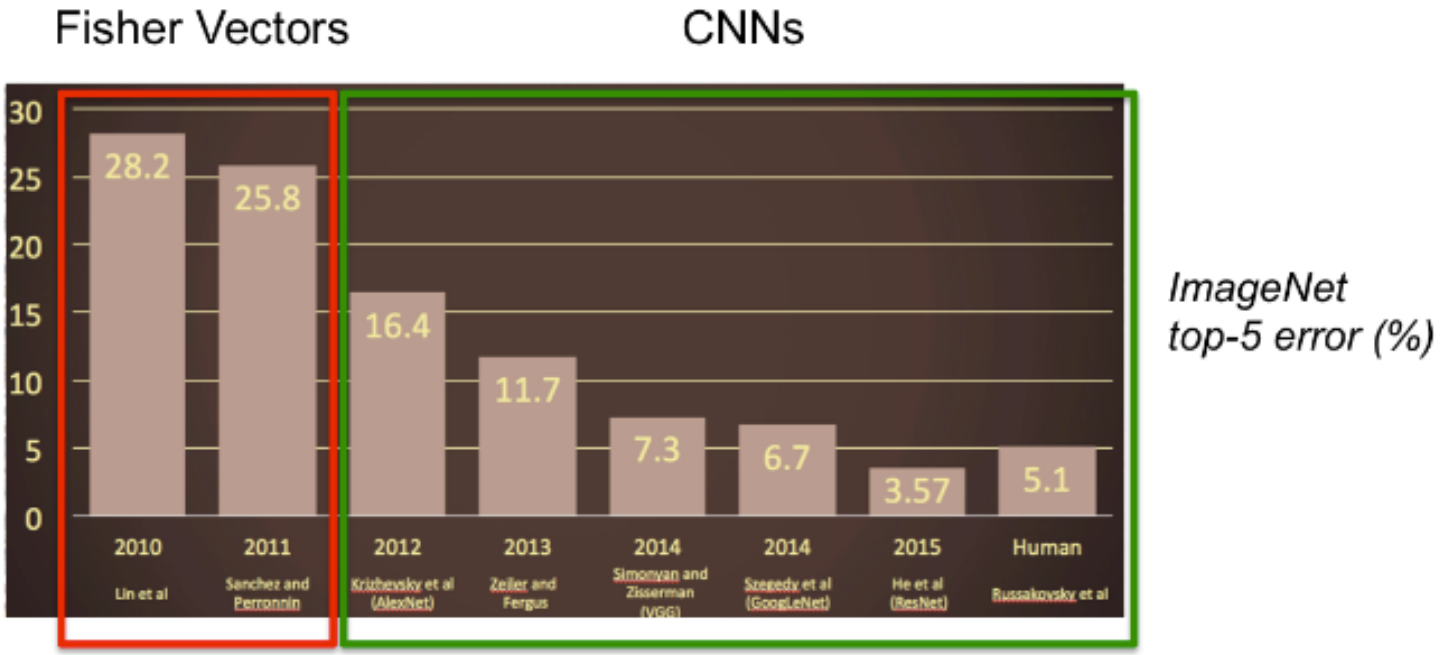
TRIVIAL HUMAN TASKS REMAINED CHALLENGING FOR COMPUTERS

AFTER 2012



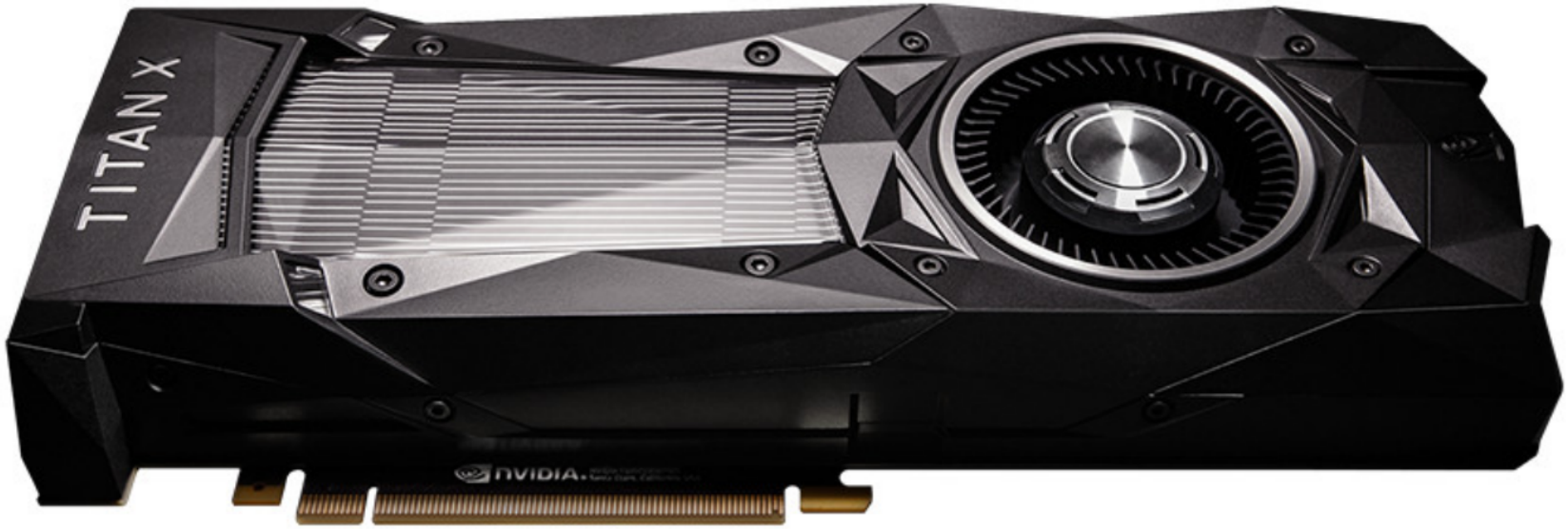
IT HAS BECOME TRIVIAL….

THIS IS A CHANGE OF PARADIGM!

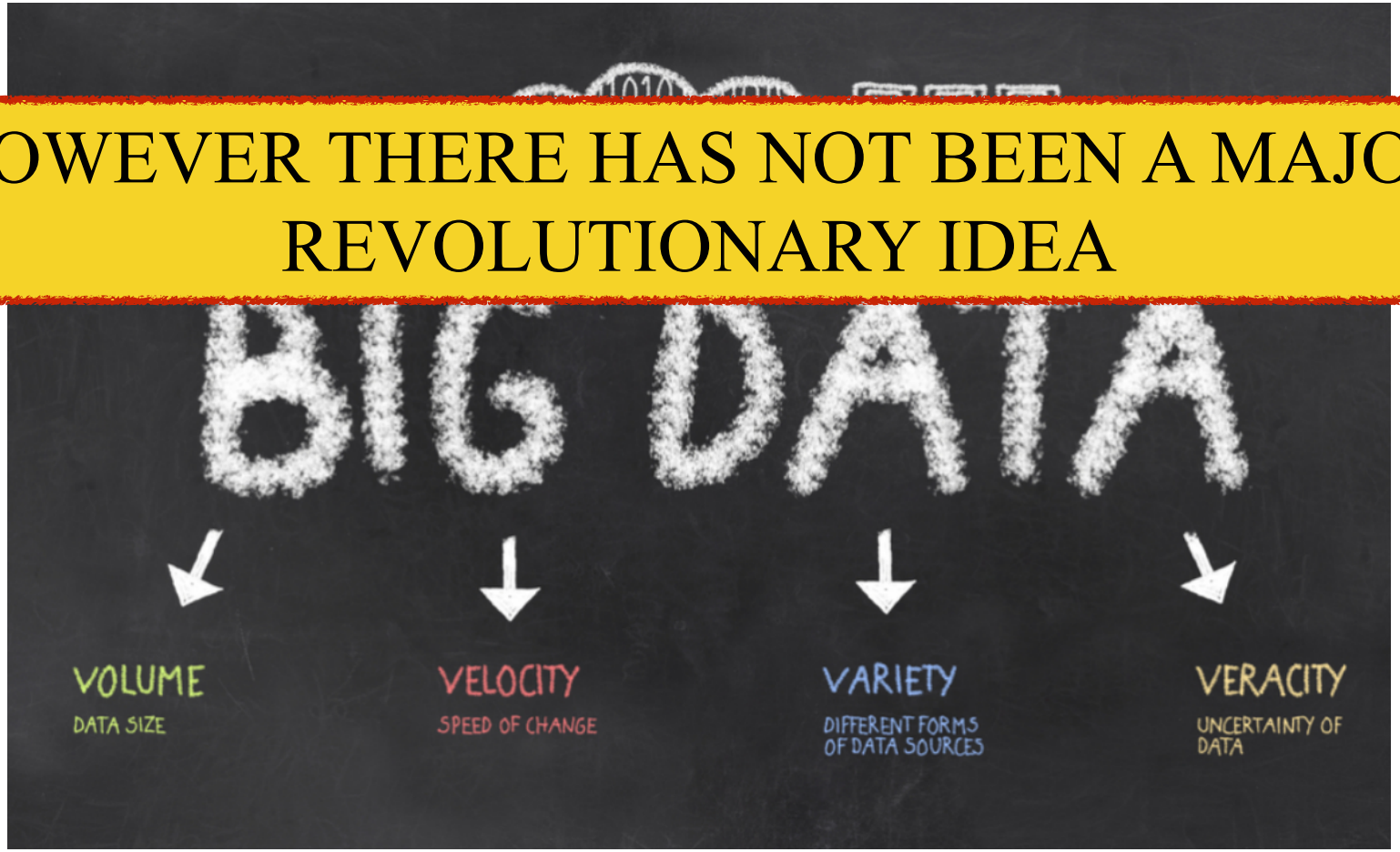


ONE OF THE MAIN REASONS OF THIS BREAKTHROUGH IS THE AVAILABILITY OF VERY LARGE DATASETS TO LEARN

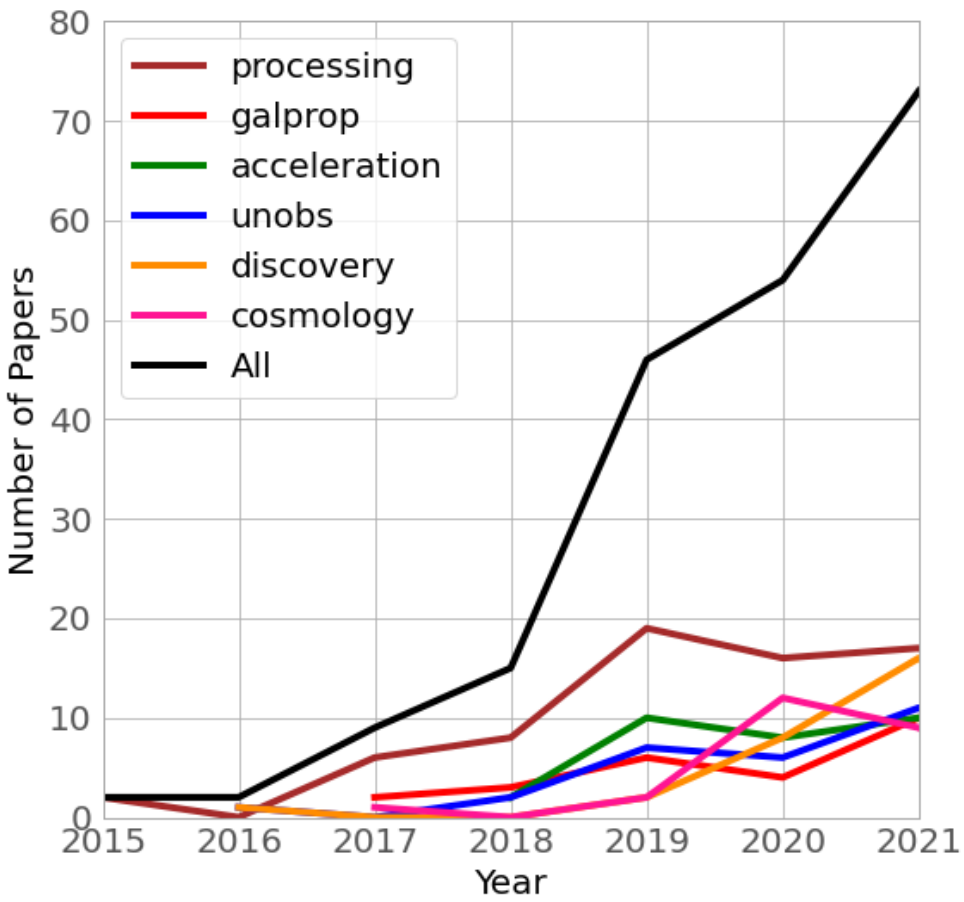
COMBINED WITH THE TECHNOLOGY TO PROCESS ALL THIS DATA

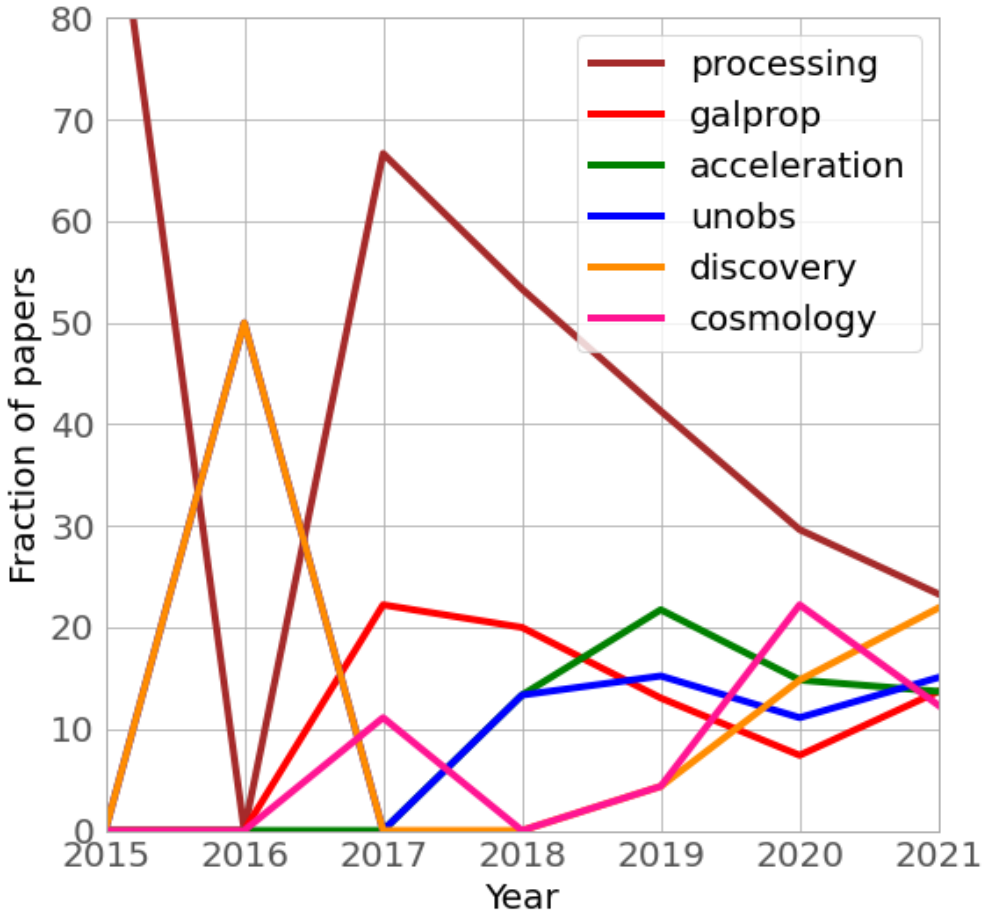


ONE OF THE MAIN REASONS OF THIS BREAKTHROUGH IS THE AVAILABILITY OF VERY LARGE DATASETS TO LEARN



HOWEVER THERE HAS NOT BEEN A MAJOR REVOLUTIONARY IDEA





Huertas-Company+22

WHAT IS THESE SERIES OF LECTURES ABOUT?

BASICS OF CLASSICAL MACHINE LEARNING

BASICS OF DEEP LEARNING
(BOTH SUPERVISED AND UNSUPERVISED)

HOPING THAT THIS WOULD BE USEFUL FOR YOUR RESEARCH!

(Apologies in advance for biases on Extra-Galactic Science + imaging)

WHY DO WE NEED THESE TOOLS IN ASTRONOMY?

PROGRAM FOR THE WEEK

- **PART I: A VERY QUICK INTRODUCTION TO 'CLASSICAL' SUPERVISED MACHINE LEARNING (FOR CLASSIFICATION)**
  - UNSUPERVISED / SUPERVISED
  - GENERAL STEPS TO "TEACH A MACHINE"
  - "CLASSICAL" CLASSIFIERS: RFs

PROGRAM FOR THE WEEK

- **PART II: FOUNDATIONS OF 'SHALLOW' NEURAL NETWORKS**
  - PERECPTRON, NEURON DEFINITION
  - LAYER OF NEURONS, HIDDEN LAYERS
  - ACTIVATION FUNCTIONS
  - OPTIMIZATION [GRADIENT DESCENT, LEARNING RATES]
  - BACKPROPAGATION
  - LOSS FUNCTIONS

PROGRAM FOR THE WEEK

- **PART III: CONVOLUTIONAL NEURAL NETWORKS**
  - CONVOLUTIONS AS NEURONS
  - CNNs [POOLING, DROPOUT]
  - VANISHING GRADIENT / BATCH NORMALIZATION
  - CNN VISUALIZATION

PROGRAM FOR THE WEEK

- **PART IV: BEYOND CLASSIFICATION WITH CNNs**
  - FCNNs (IMAGE2IMAGE NETWORKS)
  - OBJECT DETECTION
  - INSTANCE AND SEMANTIC SEGMENTATION
  - RECURSIVE NETWORKS

PROGRAM FOR THE WEEK

- **PART V: INTRODUCTION TO UNSUPERVISED MACHINE (DEEP) LEARNING**
  - DIMENSIONALITY REDUCTION
  - CLUSTERING
  - DEEP GENERATIVE MODELS
  - DEEP PROBABILISTIC MODELS

REFERENCES

SEVERAL SLIDES / INFOS SHOWN HERE ARE INSPIRED/ TAKEN FROM OTHER WORKS / COURSES FOUND ONLINE

- Deep Learning Book [Goodfellow, Bengio, Courville]
- Deep Learning: Do-It-Yourself! [Bursuc, Krzakala, Lelarge]
- DEEPLEARNING.AI [COURSERA, Ng, Bensouda, Katanforoosh]
- MACHINE LEARNING LECTURES [Keck]
- EPFL DEEP LEARNING COURSE [Fleuret]
- DEEPMIND / UCL Lecture series [DeepMind]

+ many others!

Thanks to all of them!

HANDS-ON SESSIONS

WE WILL TRY TO PRACTICALLY IMPLEMENT SOME OF THE NOTIONS LEARNED

TUTORIALS WILL USE **GOOGLE COLAB**. YOU WILL NEED A GOOGLE ACCOUNT AND SOME SPACE IN YOUR GDRIVE. WORKS ALSO BETTER ON CHROME.

NOTEBOOKS ARE GENERALLY SELF EXPLANATORY. YOU WILL BE ASKED TO FILL PIECES OF CODE AT DIFFERENT STAGES. SOLUTIONS ARE ALSO PROVIDED.

LET'S TRY TO DISCUSS AS MUCH AS POSSIBLE!

INTERACTION

AS MY COLLEAGUES, I WILL USE POLLING TOOLS

PLEASE GO TO pollev.com/marchuertasc257

ASK QUESTIONS ON SLACK

I WILL BE ON WONDER.ME
AFTER THE LECTURES AS WELL

How would you define Machine Learning?

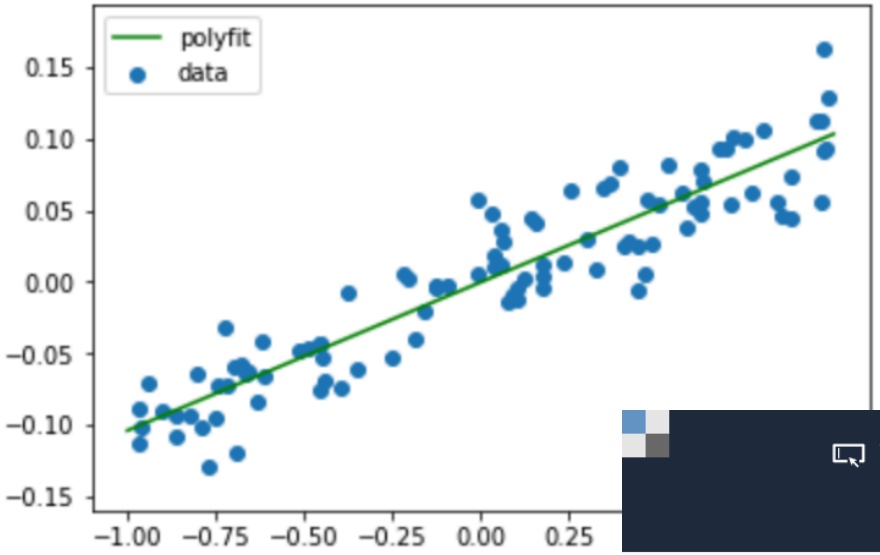
Have you ever used ML for your work?

Yes

No

I don't know

PART I: AN INTRODUCTION TO "CLASSICAL" SUPERVISED MACHINE LEARNING (FOR CLASSIFICATION)



When poll is active, respond at **pollev.com/marchuertasc257**

Text **MARCHUERTASC257** to **22333** once to join

Is this linear fit machine learning?

Yes

No

It depends

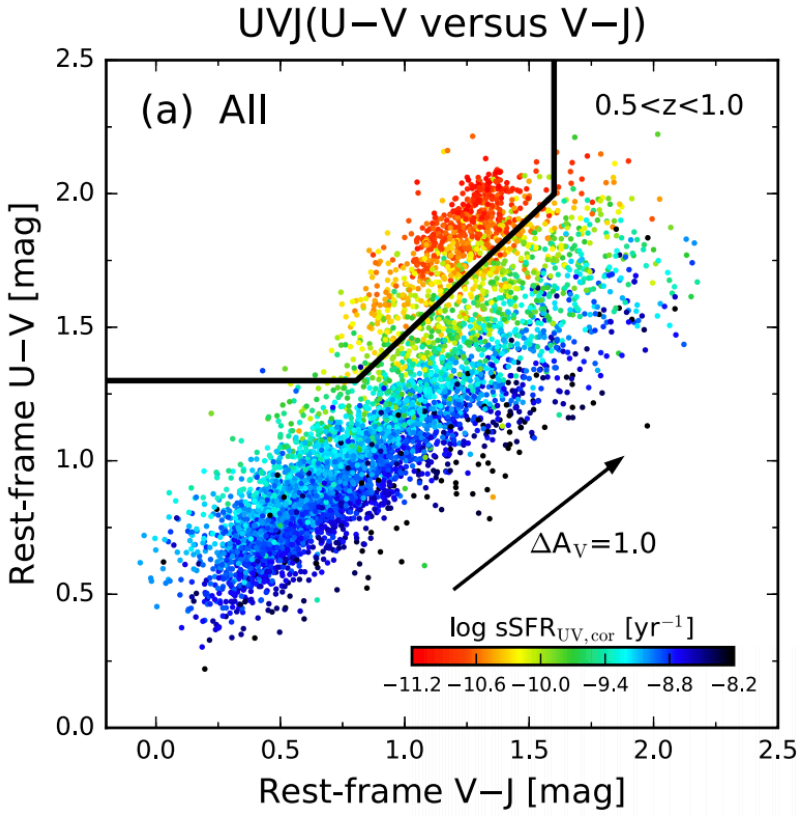
I don't know

Powered by Poll Everywhere

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

GO HERE:

https://github.com/mhuertascompany/Saas-Fee/blob/main/hands-on/session1/hello_ANN.ipynb



When poll is active, respond at **pollev.com/marchuertasc257**

Text **MARCHUERTASC257** to **22333** once to join

Is this color-color plot machine learning?

Yes

No

It depends

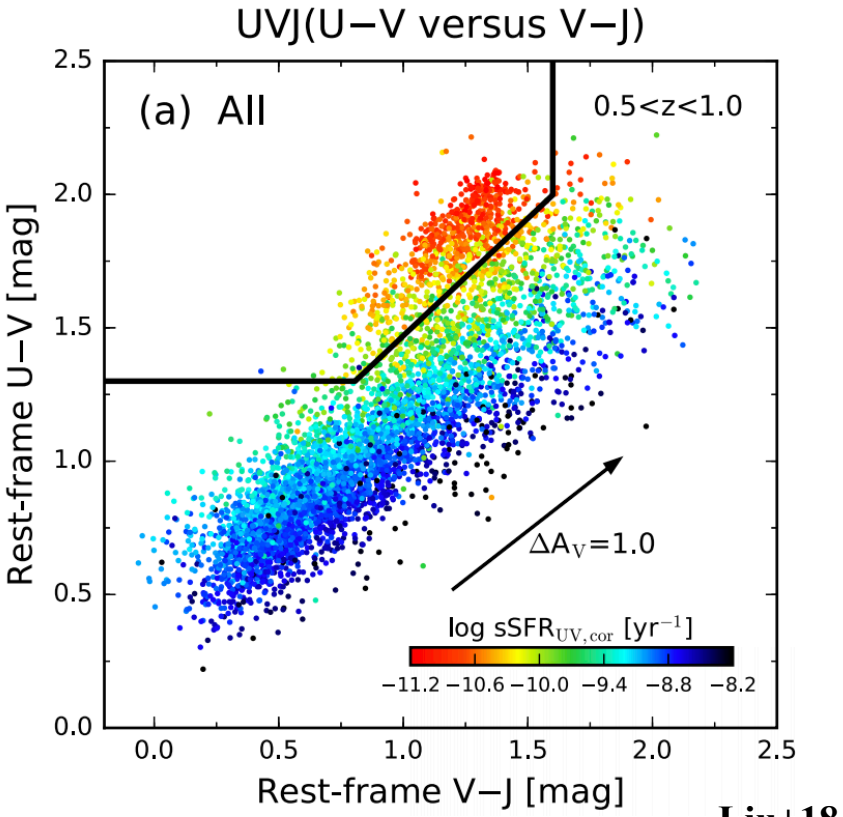
I don't know

Powered by Poll Everywhere

Start the presentation to see live content. For screen share software, share the entire screen. Get help at **pollev.com/app**

Liu+18

$$f_W(\vec{x}) = \vec{y}$$

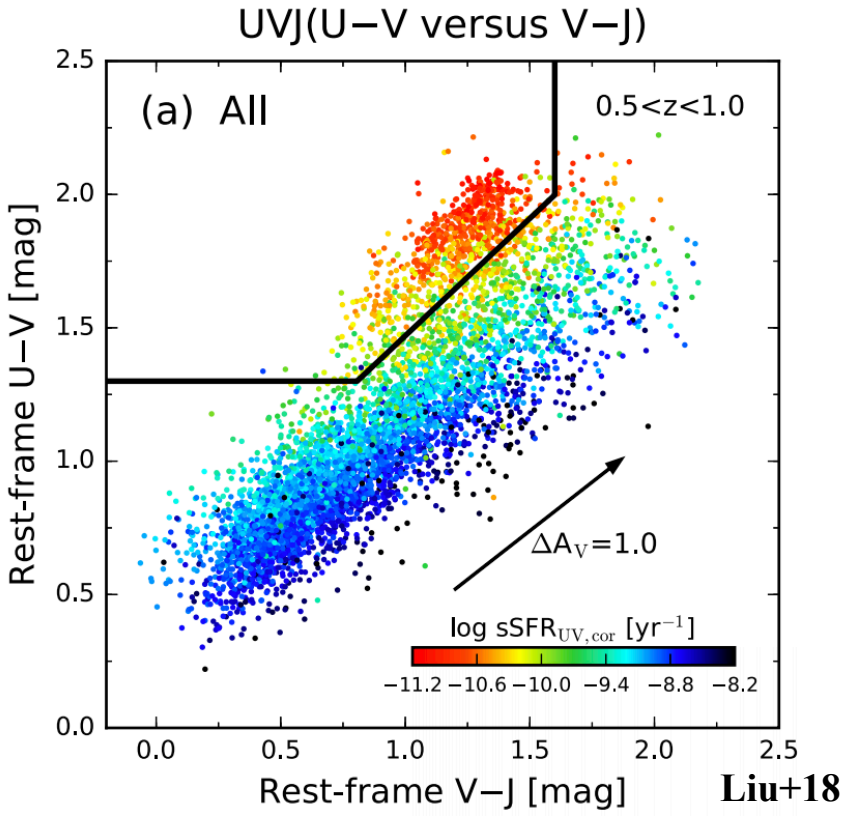


Liu+18

$$f_W(\vec{x}) = \vec{y}$$

LABEL

Q , SF



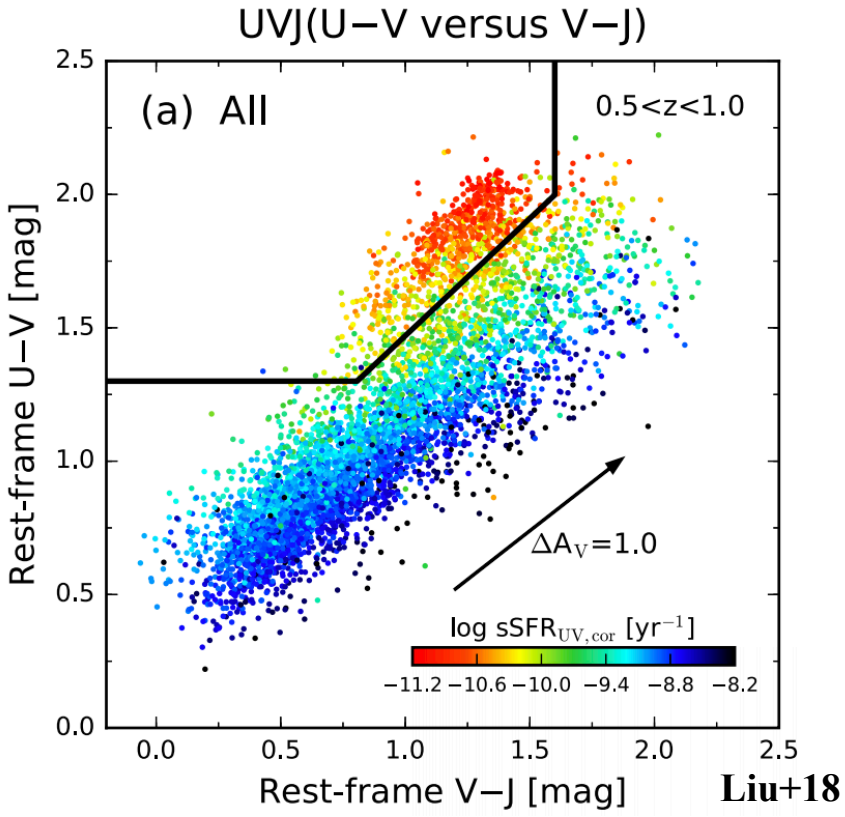
Liu+18

$$f_{\vec{W}}(\vec{x}) = \vec{y}$$

LABEL

Q(0) , SF(1)

(U-V, V-J) FEATURES



Liu+18

$$f_{\vec{W}}(\vec{x}) = \vec{y}$$

LABEL

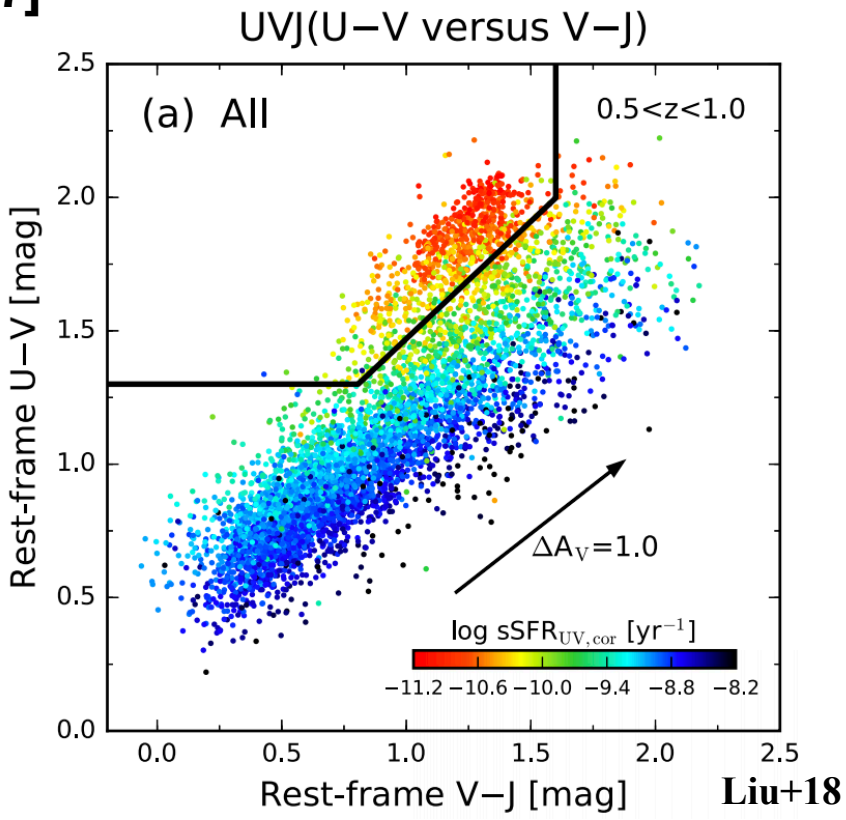
Q(0) , SF(1)

NETWORK FUNCTION

(U-V, V-J) FEATURES

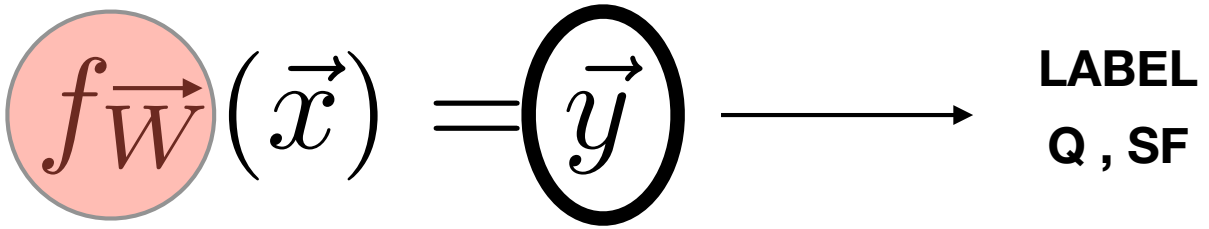
sgn[(u-v)-0.8*(v-j)-0.7]

WEIGHTS



Liu+18

“CLASSICAL”
MACHINE LEARNING



sgn[(u-v)-W1*(v-j)-W2]

REPLACE THIS BY A GENERAL
NON LINEAR FUNCTION WITH SOME PARAMETERS W

WHAT DOES MACHINE LEARNING DO?

the machine is told what to look for

SUPERVISED

the machine is NOT told what to look for

UN-SUPERVISED

TWO VERY BROAD TYPES OF MACHINE LEARNING ALGORITHMS

SUPERVISED LEARNING

CAT

CAT

the machine is told what to look for

DOG

HUMAN LABELLING

DOG

SUPERVISED LEARNING

CAT

CAT

the machine is told what to look for

DOG

HUMAN LABELLING

TRAINING SET OF LABELED EXAMPLES

DOG

CAT

CAT

DOG

DOG

SUPERVISED LEARNING

ML

CAT

CAT

CAT

DOG

DOG

SUPERVISED LEARNING

ML

DOG

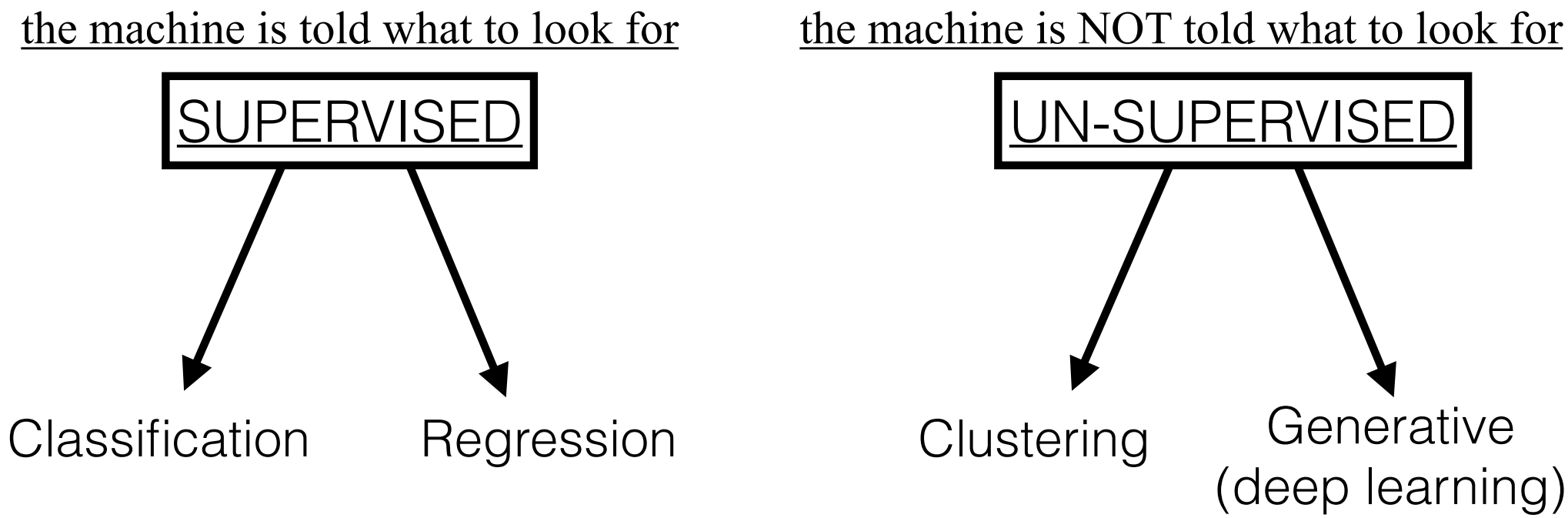
UNSUPERVISED LEARNING

UNSUPERVISED LEARNING

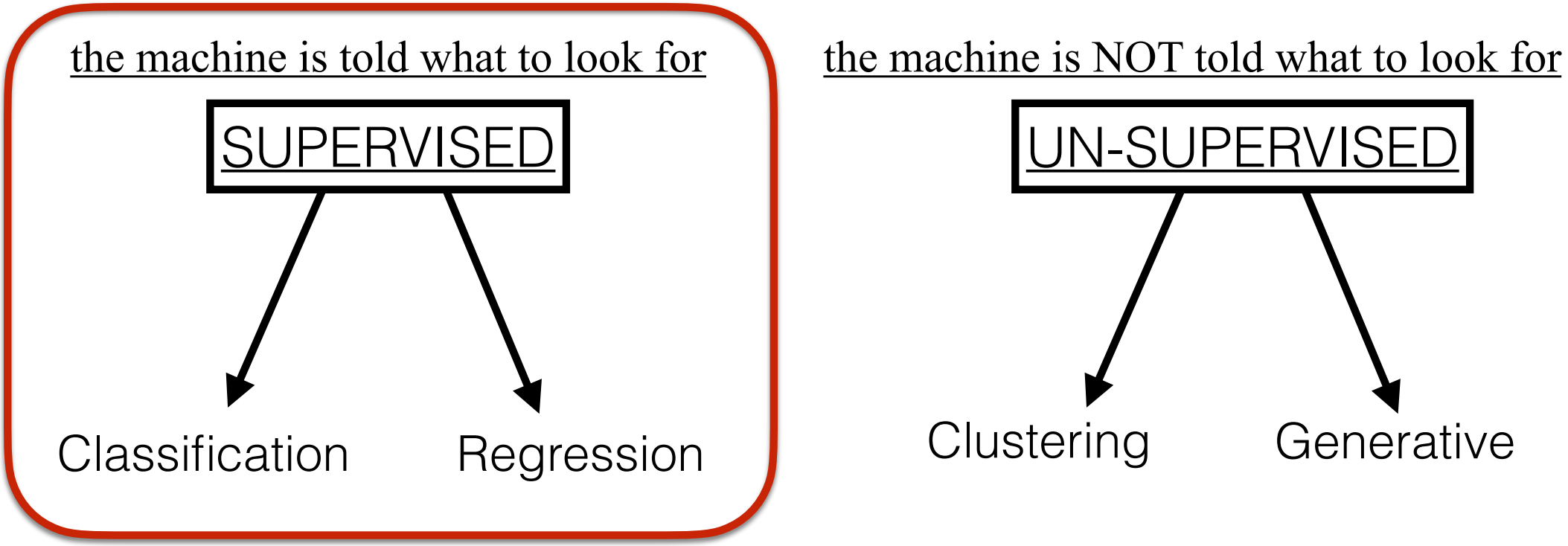
UNSUPERVISED LEARNING

THE DEFINITION OF CLASSES IS SOMETIMES NOT OBVIOUS

WHAT DOES MACHINE LEARNING DO?

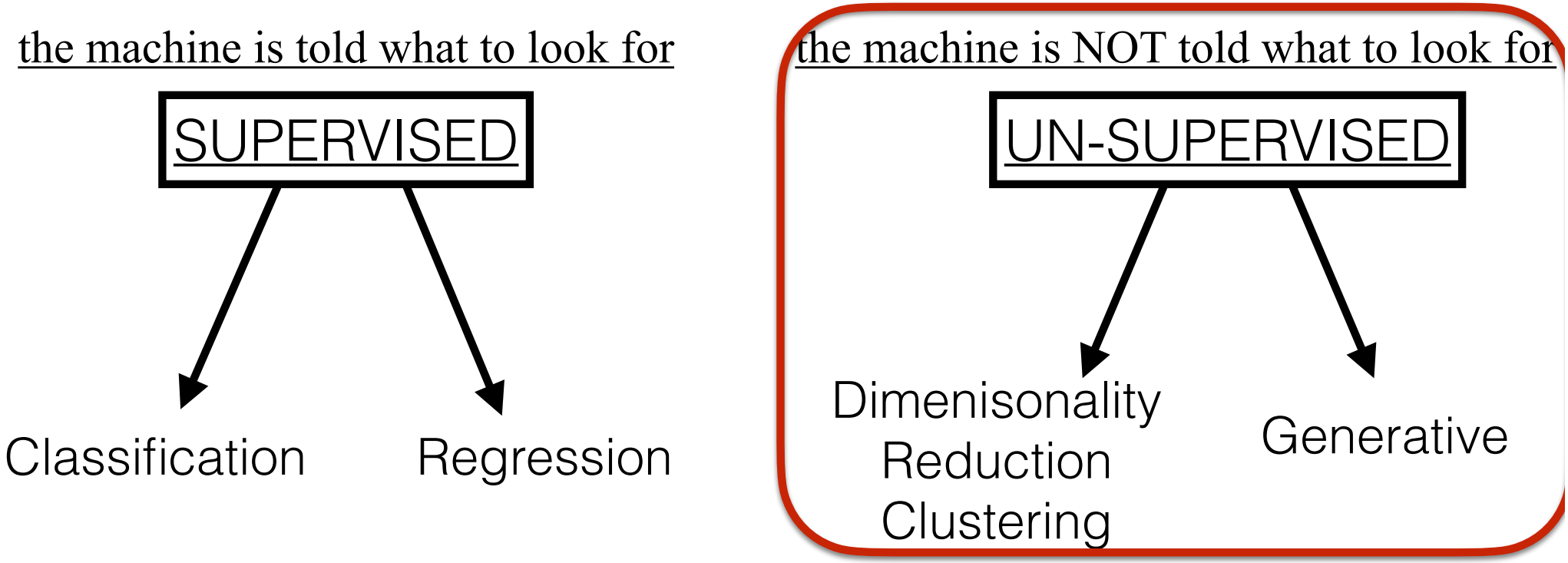


WHAT DOES MACHINE LEARNING DO?



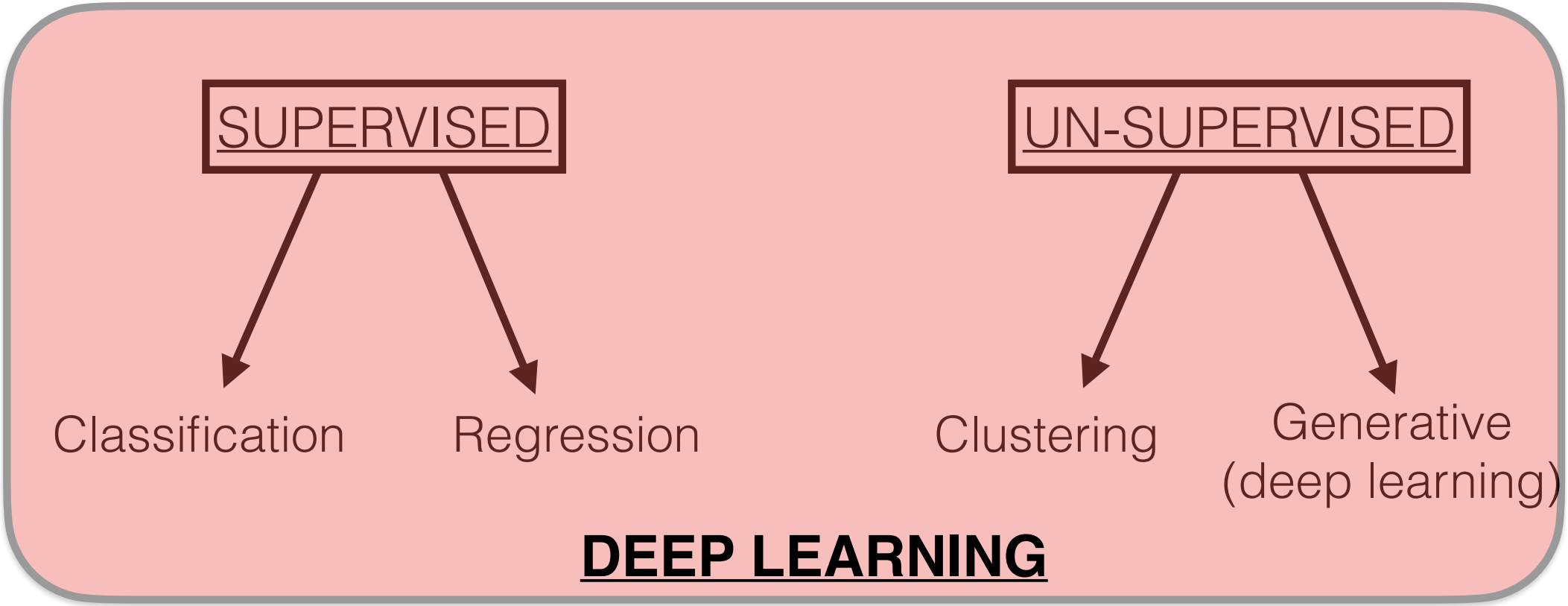
WE WILL START WITH THIS

WHAT DOES MACHINE LEARNING DO?



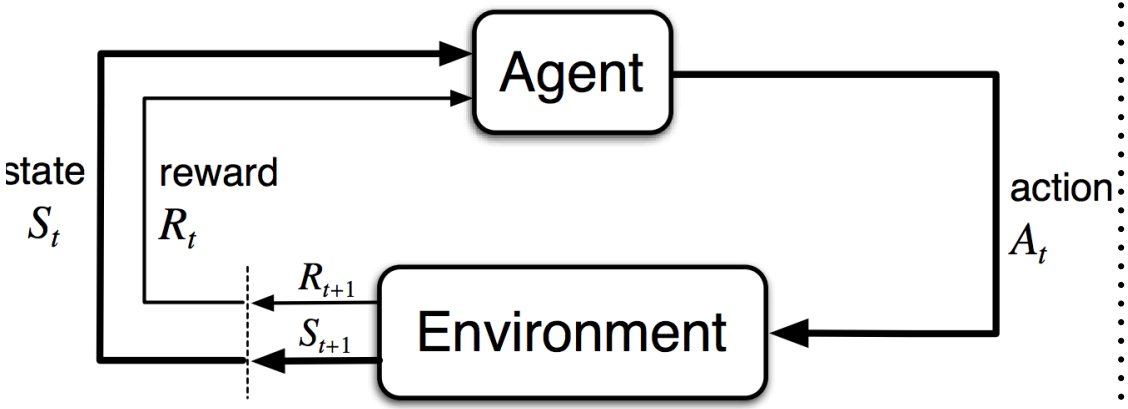
LAST LECTURES

WHAT DOES MACHINE LEARNING DO?



ACTUALLY, THERE IS A THIRD AND FORTH TYPE THAT WE WILL NOT HAVE TIME TO COVER

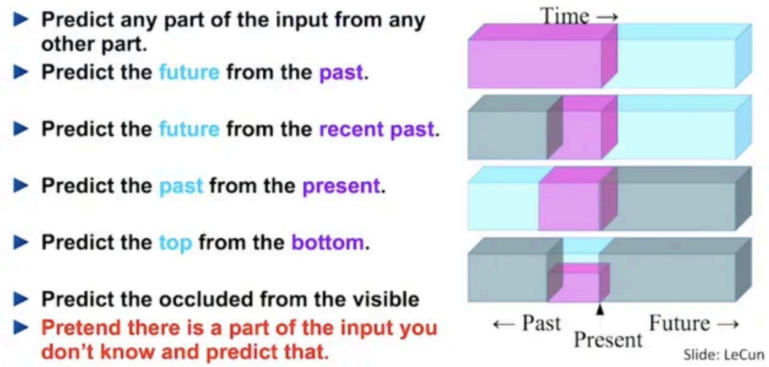
EINFORCMENT LEARNING:



source: Sutton & Barto

ITERATIVE LEARNING THROUGH TRIAL/ERROR: USED FOR ALPHAGO FOR EXAMPLE

SELF-SUPERVISED LEARNING:



> That's also why more knowledge about the structure of the world can be learned through self-supervised learning than from the other two paradigms: the data is unlimited, and amount of feedback provided by each example is huge.
>
> Y. LeCun

a self-supervised learning system attempts to predict parts of its inputs based on the other parts of its inputs

SUPERVISED LEARNING

Given a dataset with known labels (measurements) - find a function that can assign (predict) measurements for an unlabeled dataset

Training set

$(\vec{x_1}, \vec{x_2}, \vec{x_3}, ..., \vec{x_n})$ → Measurements (colors, fluxes, spectra indices…)

$(\vec{y_1}, \vec{y_2}, \vec{y_3}, ..., \vec{y_n})$ → Label (morphology, object type, redshift …)

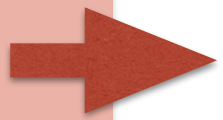
SUPERVISED LEARNING

Given a dataset with known labels (measurements) - find a function that can assign (predict) measurements for an unlabeled dataset

Training set

$(\vec{x_1}, \vec{x_2}, \vec{x_3}, ..., \vec{x_n})$

$(\vec{y_1}, \vec{y_2}, \vec{y_3}, ..., \vec{y_n})$



$f_W(\vec{x}) = \vec{y}$

?

SUPERVISED LEARNING

Unlabeled set

$(\vec{x_1}', \vec{x_2}', \vec{x_3}', ..., \vec{x_n}')$

$(\vec{y_1}', \vec{y_2}', \vec{y_3}', ..., \vec{y_n}')$

Training set

$(\vec{x_1}, \vec{x_2}, \vec{x_3}, ..., \vec{x_n})$

$(\vec{y_1}, \vec{y_2}, \vec{y_3}, ..., \vec{y_n})$

$f_W(\vec{x}) = \vec{y}$

?

$$(\vec{x_1}, \vec{x_2}, \vec{x_3}, \ldots, \vec{x_n}) \qquad \vec{x} \in \mathbb{R}^d$$

$$(\vec{y_1}, \vec{y_2}, \vec{y_3}, \ldots, \vec{y_n}) \qquad \vec{y} \in \mathbb{R} \qquad \vec{y} \in \mathbb{N}$$

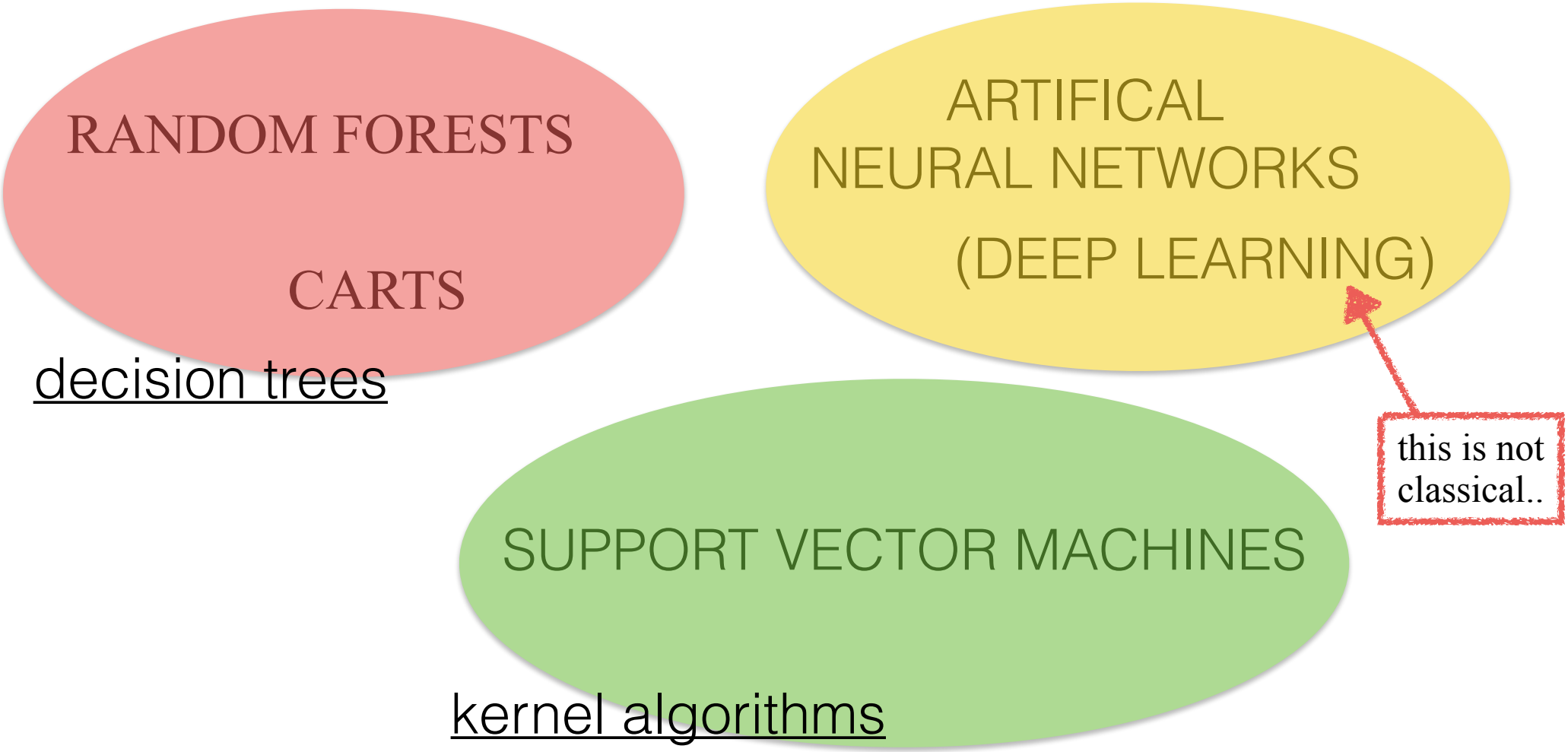
GENERAL GOAL: Find a (non-linear) function that outputs the correct class / measurement for a given input object:

$$f_W(\vec{x})$$

Number of parameters - can be large

It is translated into a minimization problem : find W such as the prediction error is minimal <u>over all unseen vectors</u>

Different "classical" supervised machine learning methods



The differences are in the function that is used

$$f_W(\vec{x})$$