



**UNIVERSITY
OF WARWICK**

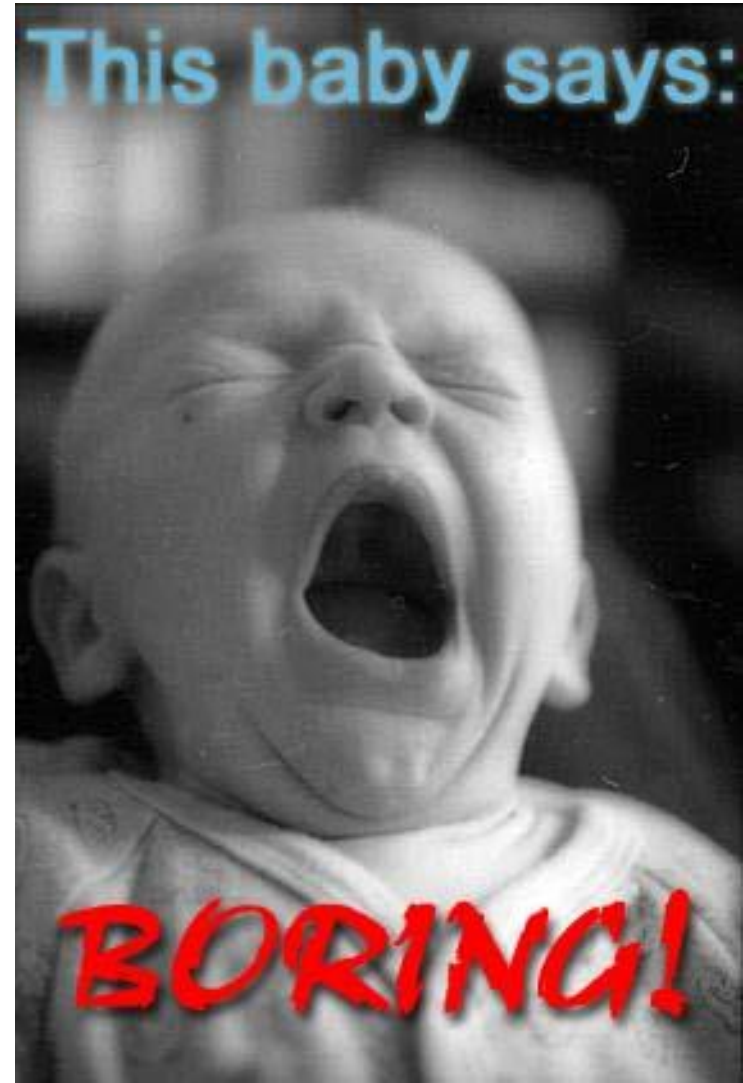
Lecture 1

Introduction to Machine Learning

Gabriele Pergola and Amina Asif
Machine Learning Algorithms and Practice
CS429

October 2026

Module Overview



Information

Lecturers Gabriele Pergola -- (First half of the module)

Gabriele.Pergola.1@warwick.ac.uk

Amina Asif – (Second half of the module)

Amina.Asif@warwick.ac.uk

Teaching Assistants



Boyang Fu



Piotr Keller

Module Website

<https://warwick.ac.uk/fac/sci/dcs/teaching/material/cs909/>

About the Module

- **Lectures:** Monday at 1pm & Tuesday at 9am & Friday at 12pm
Lecture recordings are on [Moodle](#)
- **Seminars:** Up to 1-hour seminar each week, depending on guests availability
- **Labs:** 1-hour lab session each week

More details at <https://warwick.ac.uk/fac/sci/dcs/teaching/material/cs909/>

Module Aims

- Foundation in Machine Learning Concepts: To deepen the **understanding of foundational concepts in Machine Learning**, encompassing both traditional data mining algorithms and advanced machine learning methodologies.
- Grasping the Role of **Machine Learning in Real-World Applications**: To provide a comprehensive understanding of how machine learning techniques can be leveraged to solve complex real-world problems.
- Mastery of Machine Learning Algorithms: To gain a broad-scope and thorough **knowledge of a variety of algorithms** used in machine learning.
- Practical Application of Machine Learning Tools: To enhance the ability to **apply advanced machine learning tools** and techniques to real-world data.

Learning Outcome

- Demonstrate a comprehensive understanding of different data mining tasks and the algorithms most appropriate for addressing them.
- Evaluate models/algorithms with respect to their accuracy.
- Demonstrate capacity to perform a self-directed piece of practical work that requires the application of data mining techniques.
- Critique the results of a data mining exercise.
- Develop hypotheses based on the analysis of the results obtained and test them.
- Conceptualise a data mining solution to a practical problem.

Topics

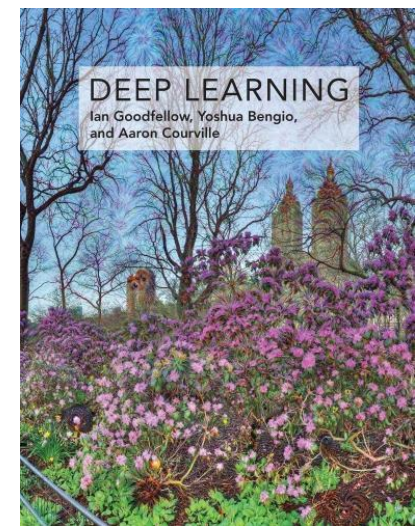
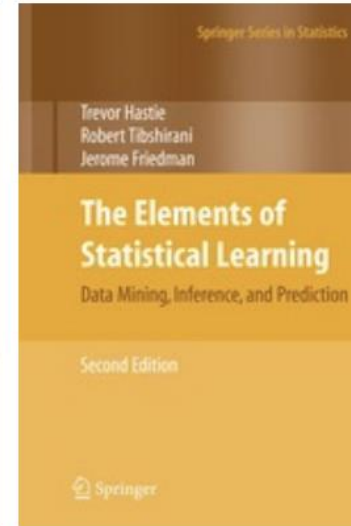
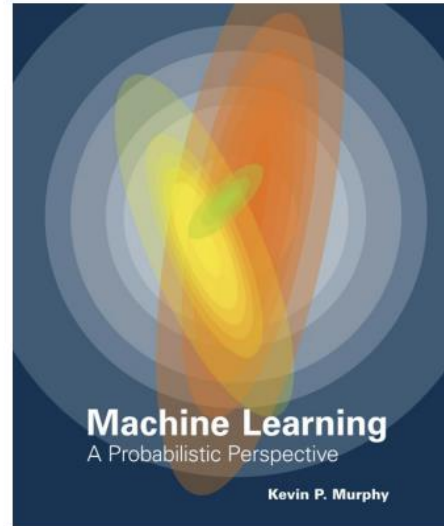
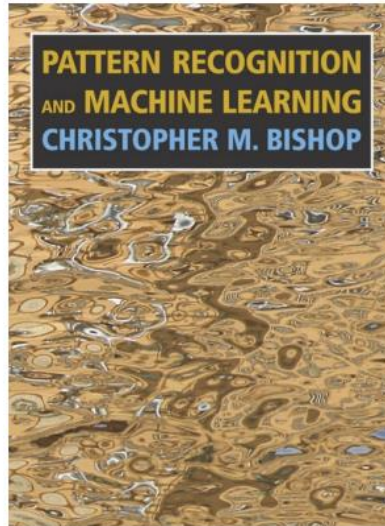
- Introduction to machine learning, basic concepts and motivation;
 - Data pre-processing and basic data transformations;
 - Regression models;
 - Classification: linear discriminants, SVMs, other classifiers such as decision trees and forests;
 - Model evaluation, bias-variance trade-off;
 - Ensemble methods: boosting, bagging & random forests;
 - Dimensionality reduction: Principal Component Analysis (PCA), T-distributed Stochastic Neighbour Embedding (t-SNE/UMAP);
-
- Introduction to deep learning, backpropagation, gradient descent;
 - Convolutional neural networks;
 - Word embeddings;
 - Sequence-to-sequence models, Residual Neural Networks;
 - Attention mechanisms and memory networks;
 - Language processing;
 - Unsupervised deep learning and generative models;
 - Learning Strategies such as Transfer learning, etc.

Assessment

- 50%: Exam in May/June
- 25%: 1st Assignment
 - Deadline of Submission 1: Week 6
- 25%: 2nd Assignment
 - Deadline of Submission 2: Week 11

Textbooks

**No single textbook is required, but
reading from them will help immensely!**



- Mitchell, T. M. (1997). ***Machine Learning***. McGraw-Hill.
- Bishop, C. M. (2006). ***Pattern Recognition and Machine Learning***. Springer.
- Murphy, K. P. (2012). ***Machine Learning: A Probabilistic Perspective***. MIT Press.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). ***The Elements of Statistical Learning: Data Mining, Inference, and Prediction*** (2nd ed.). Springer.

Programming in the Module

- We will be using **Python!**
 - Many packages available for text processing



spaCy



 PyTorch



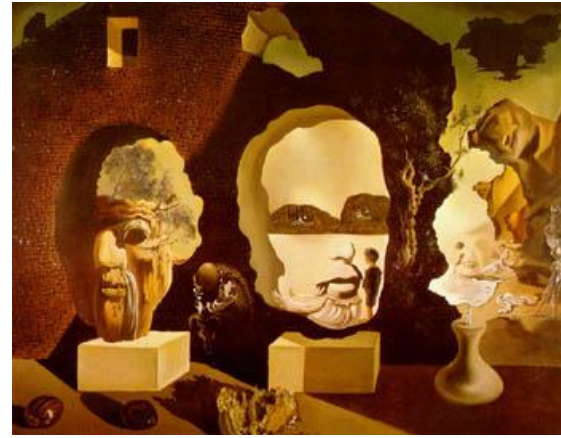
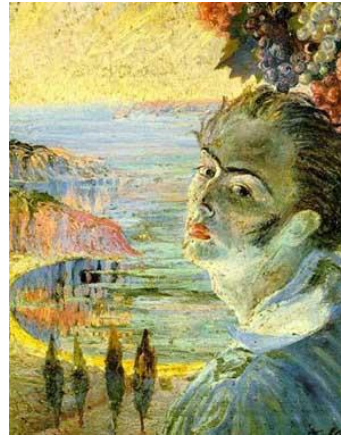
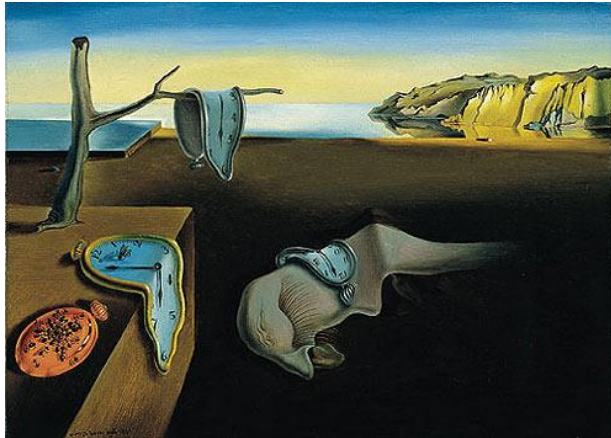
Hugging Face

Outline

- What is Machine Learning?
- Why Machine Learning?
- Machine Learning Principles

What is Machine Learning?

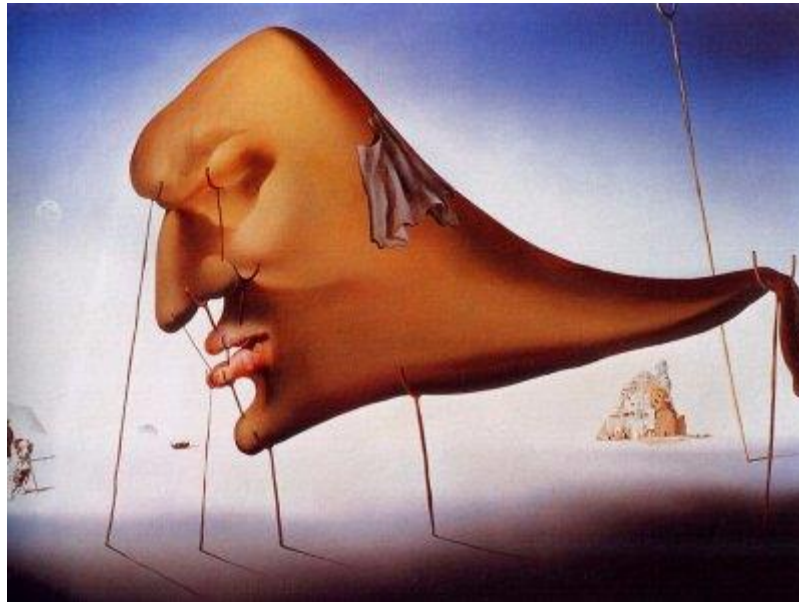
Paintings by two different painters



Who's painting is this?



And this?

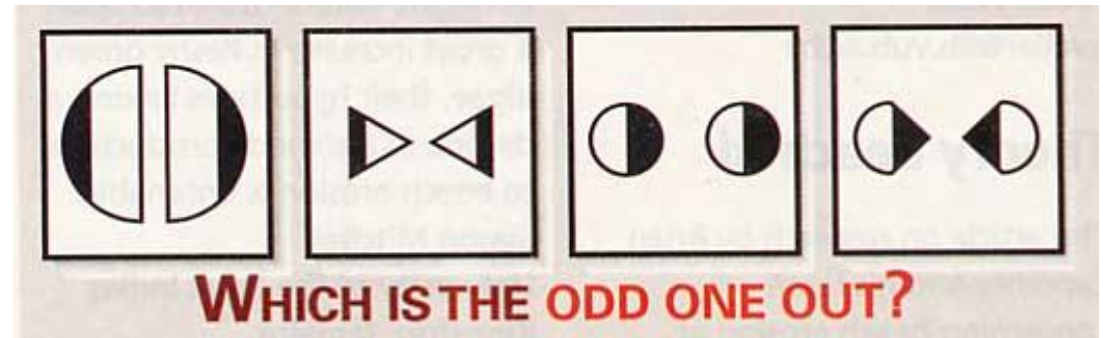


learning from data for generalization to unseen cases

How many categories (clusters) are there?



Find the odd one out!



Predict the next element

- 1,1,2,3,5,8,13,____
 - 21
- A yellow car that transforms into a transformer is called____
 - bumble bee
-  can transform into _____



Question?

- Consider the vectors
 - $X_1 = [1 \ 2 \ 1 \ 4]^T$
 - $X_2 = [2 \ 4 \ 2 \ 4]^T$
 - $X_3 = [0 \ 0 \ 0 \ 4]^T$
 - $X_4 = [3 \ 6 \ 3 \ 4]^T$
 - $X_5 = [4 \ 8 \ 4 \ 4]^T$
- To store each vector, how many dimensions (or variables) do we need?

Learning to drive



Questions

• How were you able to recognize that the object shown was indeed an apple?	Classification
• How were you able to discriminate between the paintings from two different painters?	Classification
• How were you able to find out the different types of apples in the picture?	Clustering
• How did you manage to find the next number in the series?	Regression
• How did you manage to predict the next element?	Next Element Prediction
• How were you able to find which dimension was redundant?	Dimensionality Reduction
• How were you able to find the odd one out?	Anomaly Detection
• Learning to drive / write?	Reinforcement learning

Definitions

- Computers are _____.
 - **Dumb**
- Making a machine (computer) perform the same tasks which you have just done is called _____.
 - **Artificial Intelligence**
- If you make a computer learn to do these tasks using existing data, then this is called _____.
 - **Machine Learning (learning from data)**
- Mining for patterns in data (for understanding it better and making predictions) is
 - **Data Mining**
- Solving a machine learning task using a deep layered network of artificial neurons is called _____.
 - **Deep Learning (Learning from data with neural networks)**

Why do we need Machine Learning?

Why Machine Learning?



Traditional CS:



Vincent van Gogh

Why Machine Learning?



Traditional CS:



Vincent van Gogh

The bottleneck is no longer access to information; now it's our ability to keep up.

AI can be trained on a variety of different types of texts and summary lengths.

A model that can generate long, coherent, and meaningful summaries remains an open research problem.

The last few decades have witnessed a fundamental change in the challenge of taking in new information. The bottleneck is no longer access to information; now it's our ability to keep up. We all have to read more and more to keep up-to-date with our jobs, the news, and social media. We've looked at how AI can improve people's work by helping with this information deluge and one potential answer is to have algorithms automatically summarize longer texts. Training a model that can generate long, coherent, and meaningful summaries remains an open research problem. In fact, generating any kind of longer text is hard for even the most advanced deep learning algorithms. In order to make summarization successful, we introduce two separate improvements: a more contextual word generation model and a new way of training summarization models via reinforcement learning (RL). The combination of the two training methods enables the system to create relevant and highly readable multi-sentence summaries of long text, such as news articles, significantly improving on previous results. Our algorithm can be trained on a variety of different types of texts and summary lengths. In this blog post, we present the main contributions of our model and an overview of the natural language challenges specific to text summarization.

...ATTCGAGGATTACACCGTAAGAAATTT...

...ATCGCCTGATTACATATATACCGTTGG...

....AGATTAAATCGTTTCGATTACATTGAC

...ATTCGAGGATTACACCGTAAGAAATTT...

...ATCGCCTGATTACATATATACCGTTGG...

....AGATTACATCGTTTCGATTACATTGAC

Why Machine Learning?



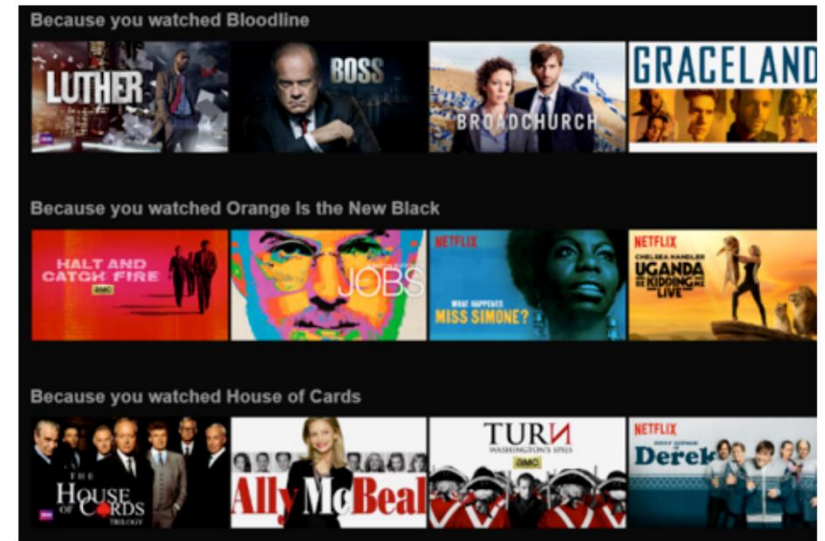
Traditional CS:



Vincent van Gogh



"Because you watched this TV show..."



Why Machine Learning?



Traditional CS:

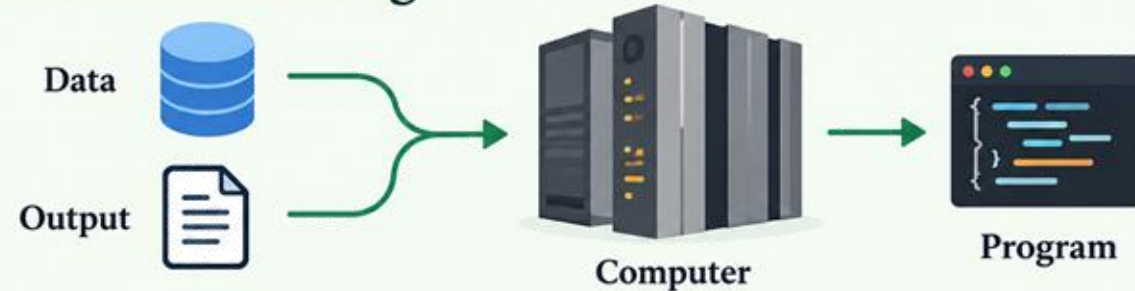


Vincent van Gogh



Vincent van Gogh

Machine Learning:



Why Machine Learning?



Traditional CS:

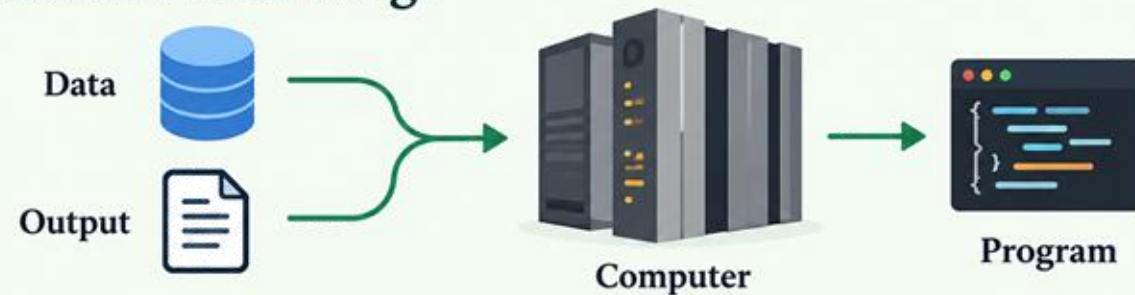


Vincent van Gogh



Vincent van Gogh

Machine Learning:

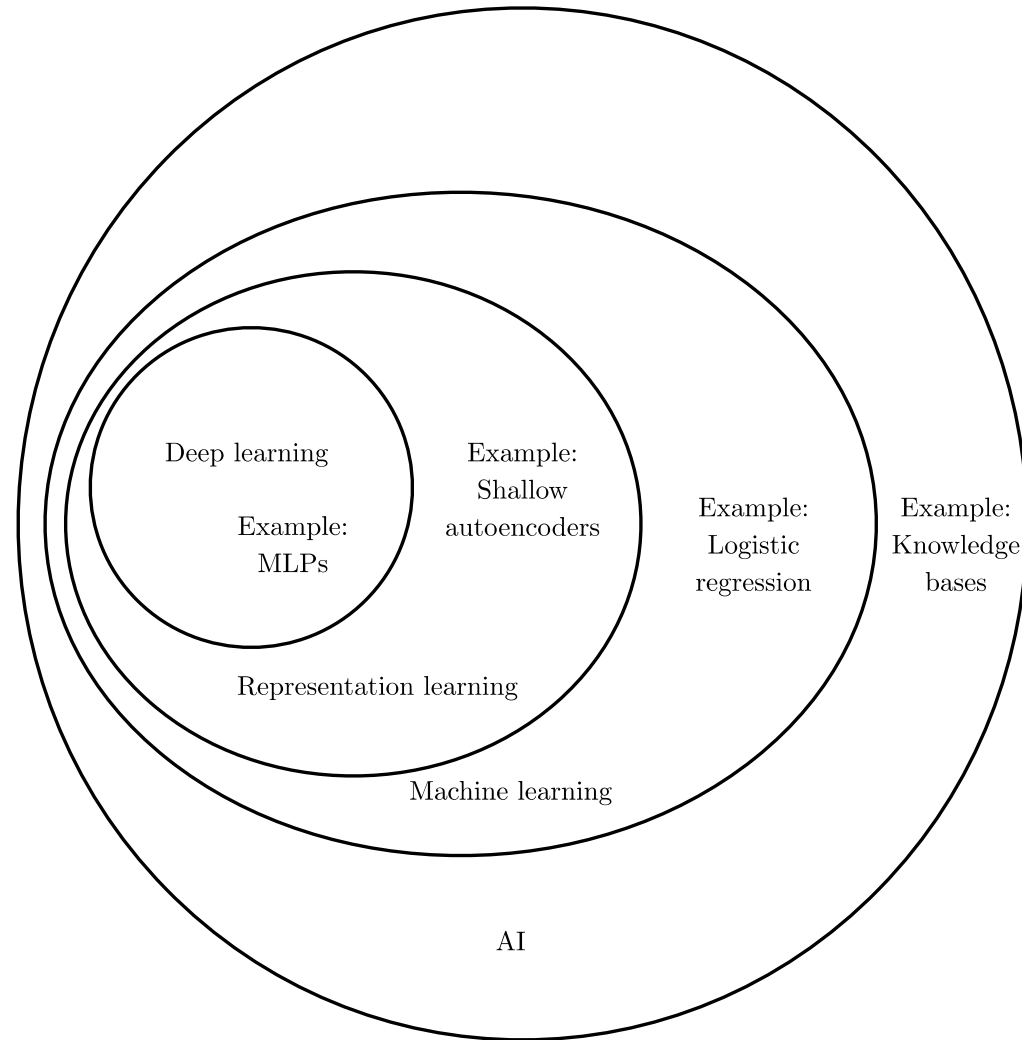


Deduction (top-down) vs. Induction (bottom-up) Reasoning

- **Required Reading**

- Halevy, Alon, Peter Norvig, and Fernando Pereira.
“The Unreasonable Effectiveness of Data.” *IEEE Intelligent Systems*, 2009.

The Hierarchy of Definitions



Formal Setup

- Input Space $\mathcal{X}$



- Output (label) space $\mathcal{Y}$ Vincent van Gogh

- We have a training set $\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ with $x_i \in \mathcal{X}$, $y_i \in \mathcal{Y}$.

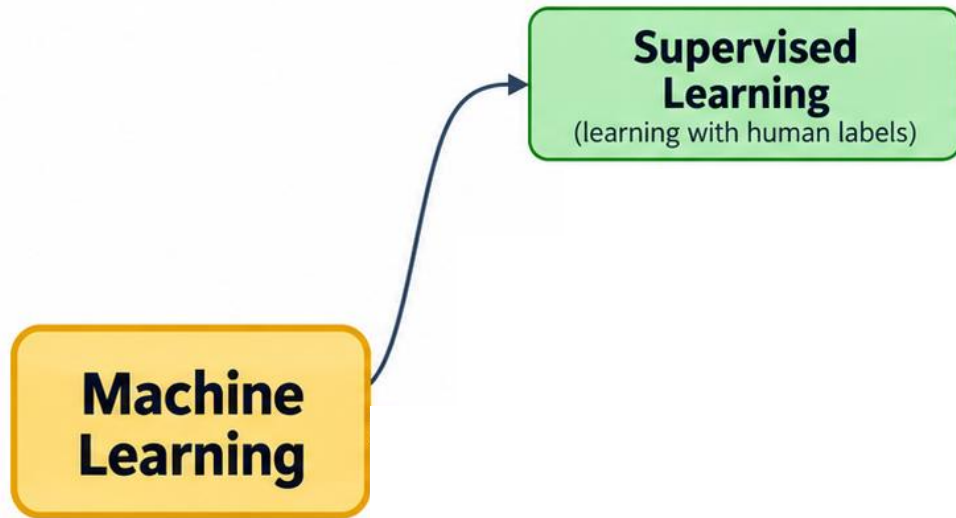
- Aim: Learn a function $f: \mathcal{X} \rightarrow \mathcal{Y}$



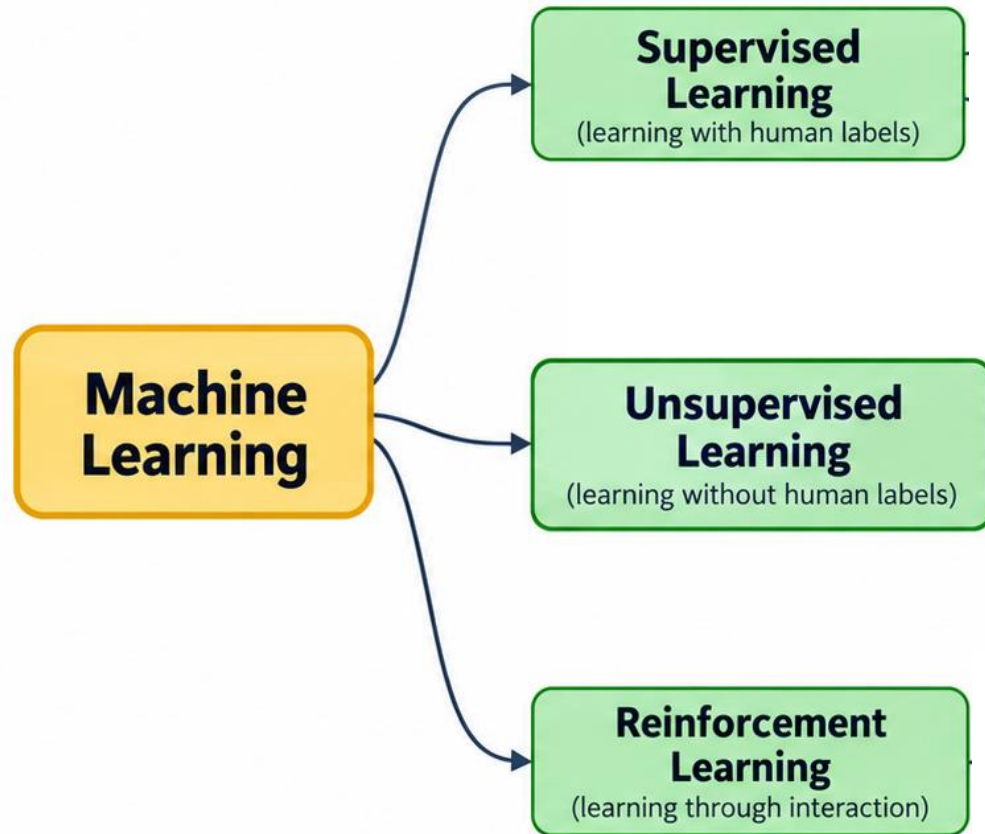
Painter
names

- If $\mathcal{Y}$ is discrete $\rightarrow$ **Classification**
- If $\mathcal{Y}$ is continuous $\rightarrow$ **Regression**

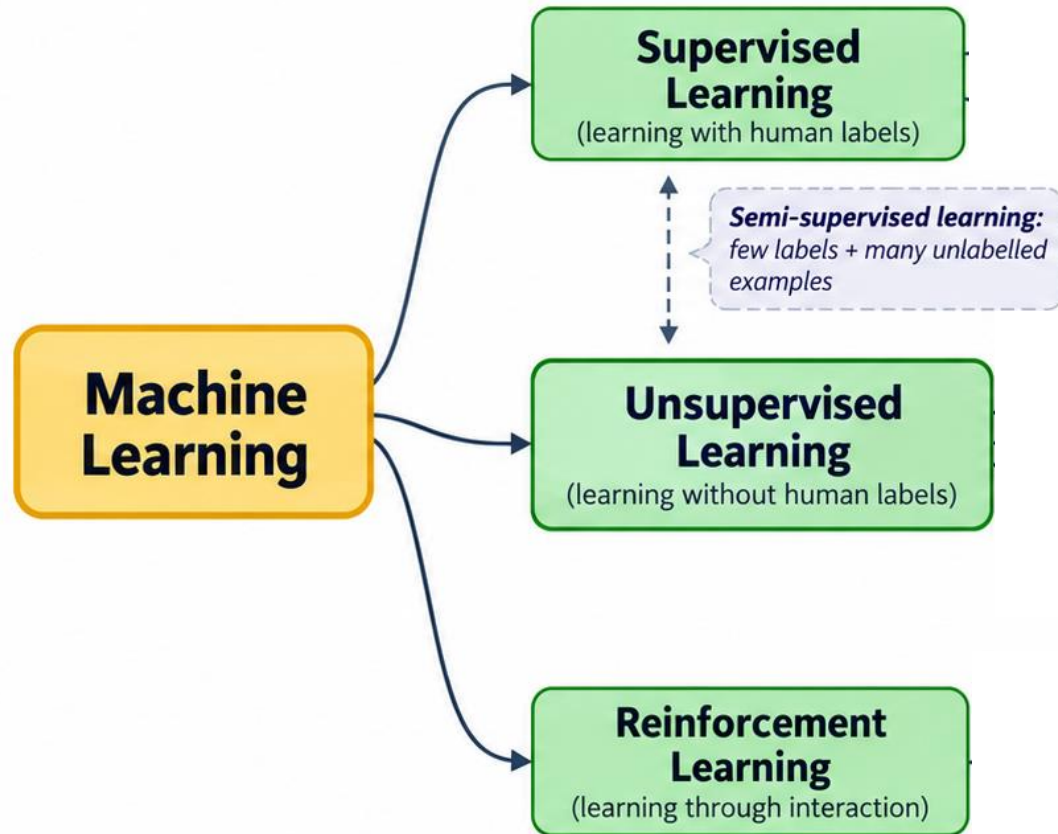
(Machine) Learning Paradigms



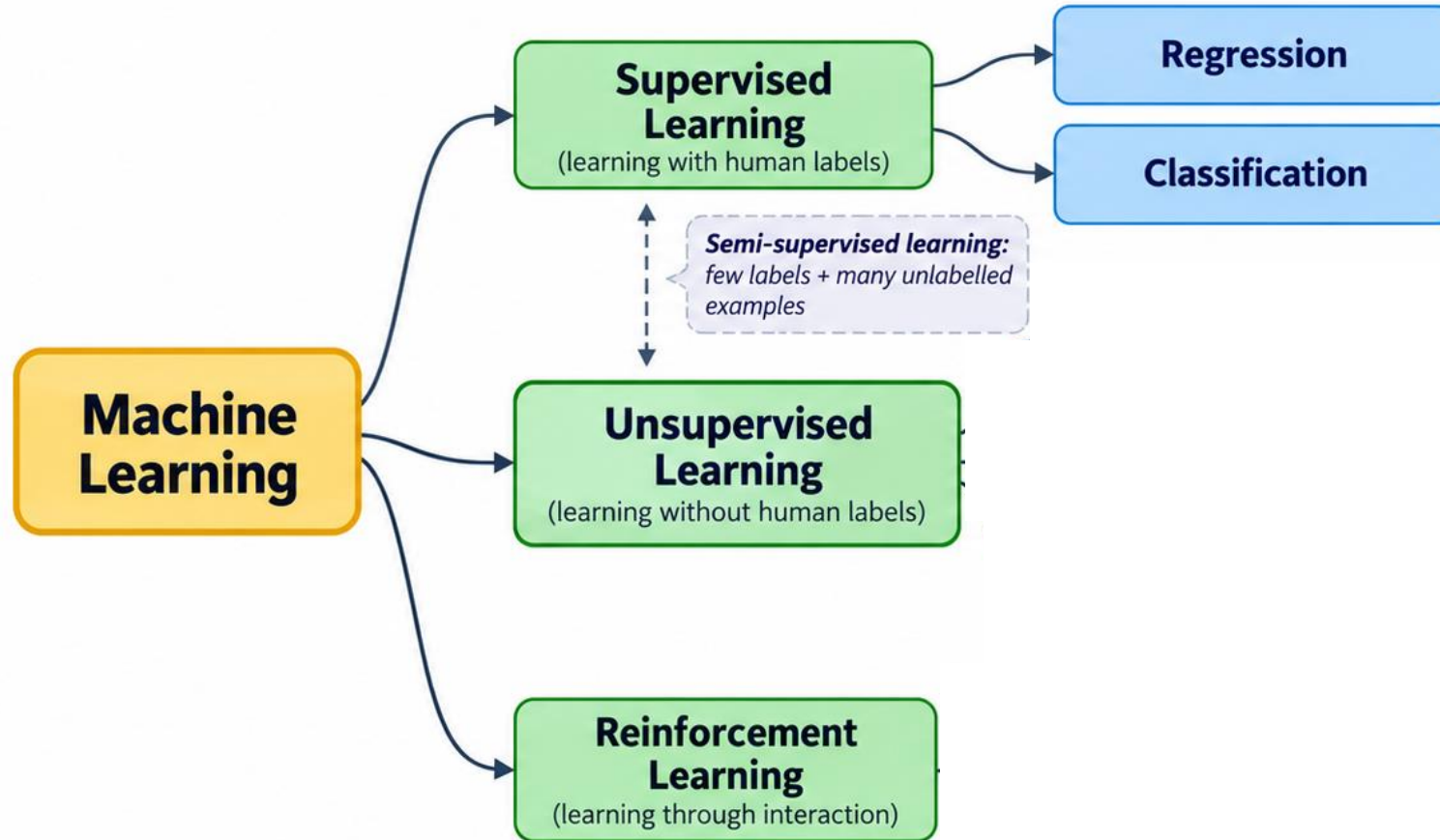
(Machine) Learning Paradigms



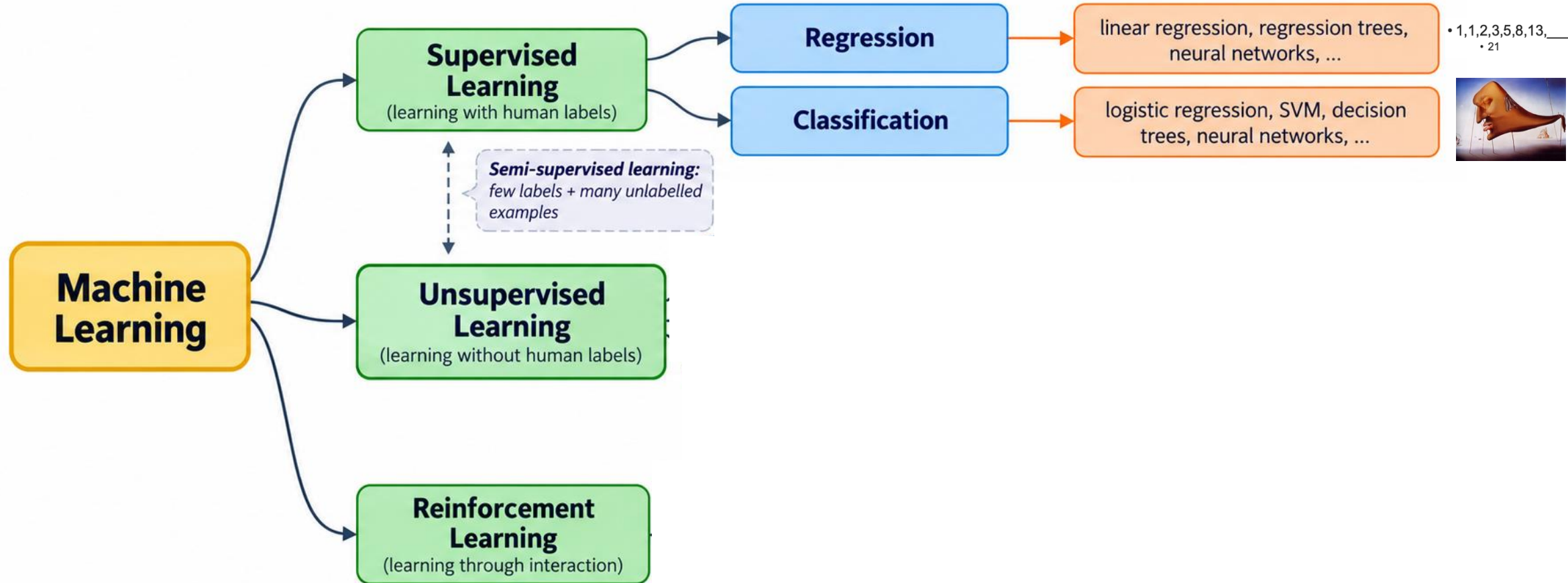
(Machine) Learning Paradigms



(Machine) Learning Paradigms



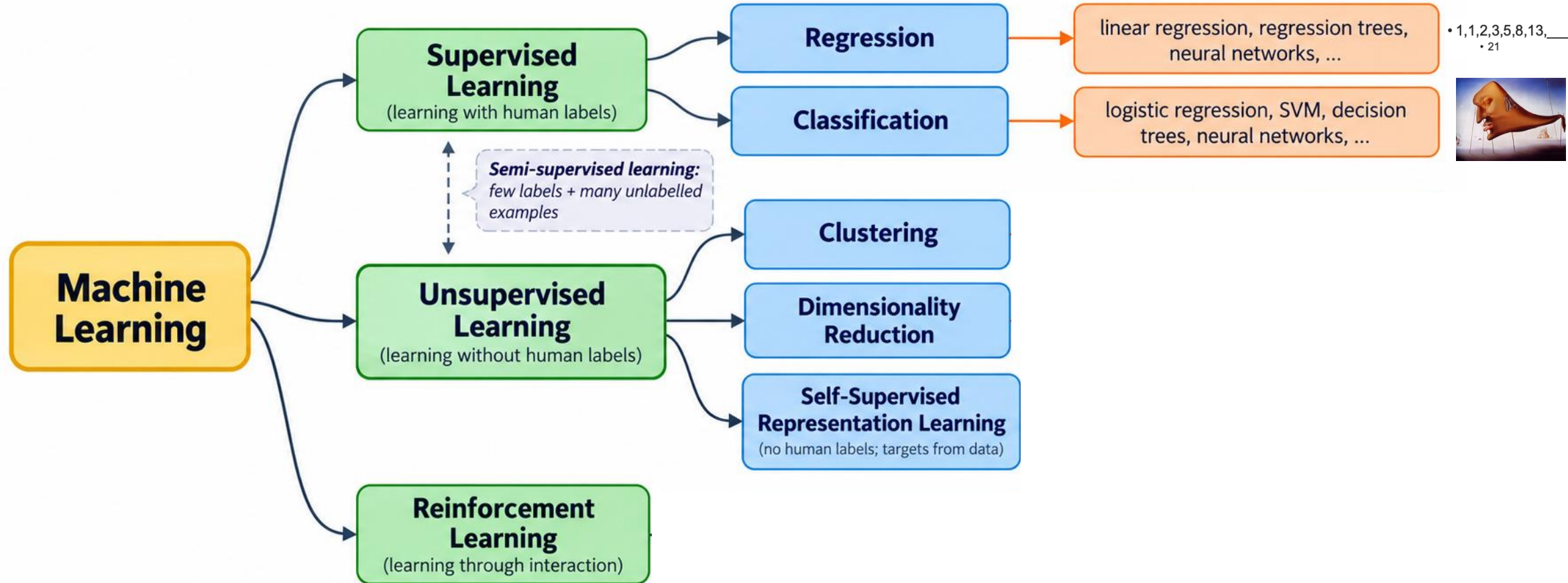
(Machine) Learning Paradigms



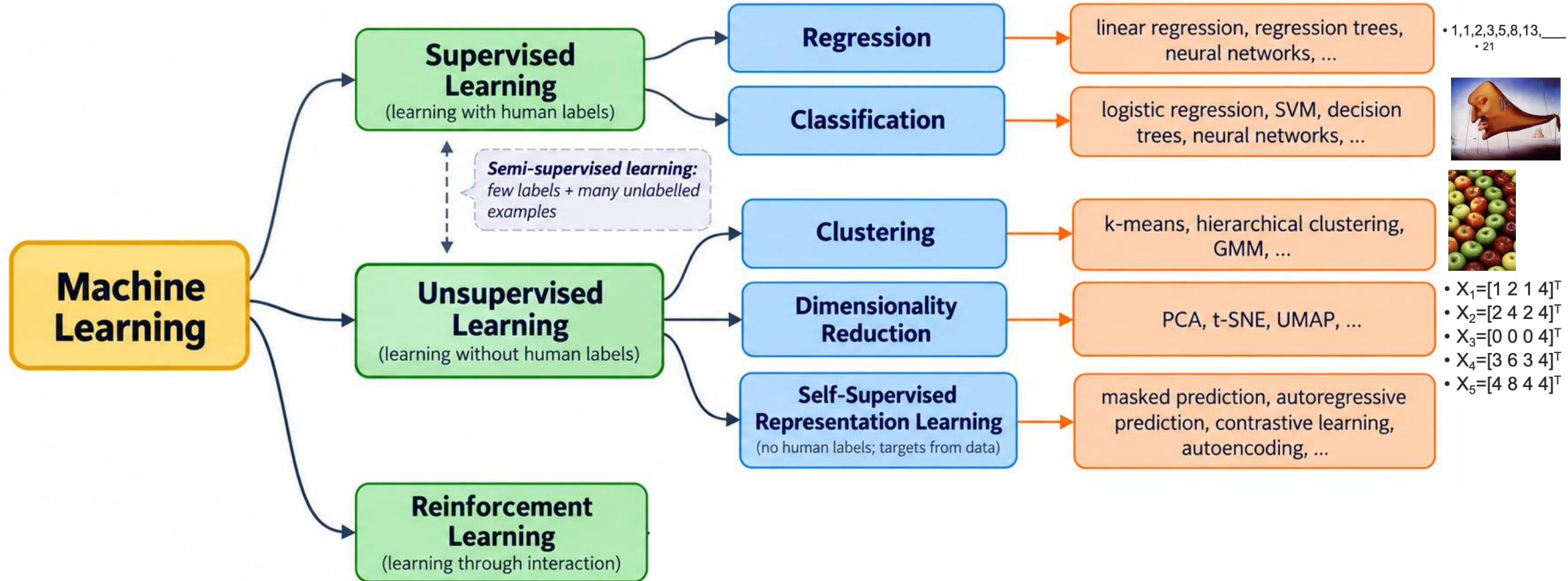
• 1,1,2,3,5,8,13,____
• 21



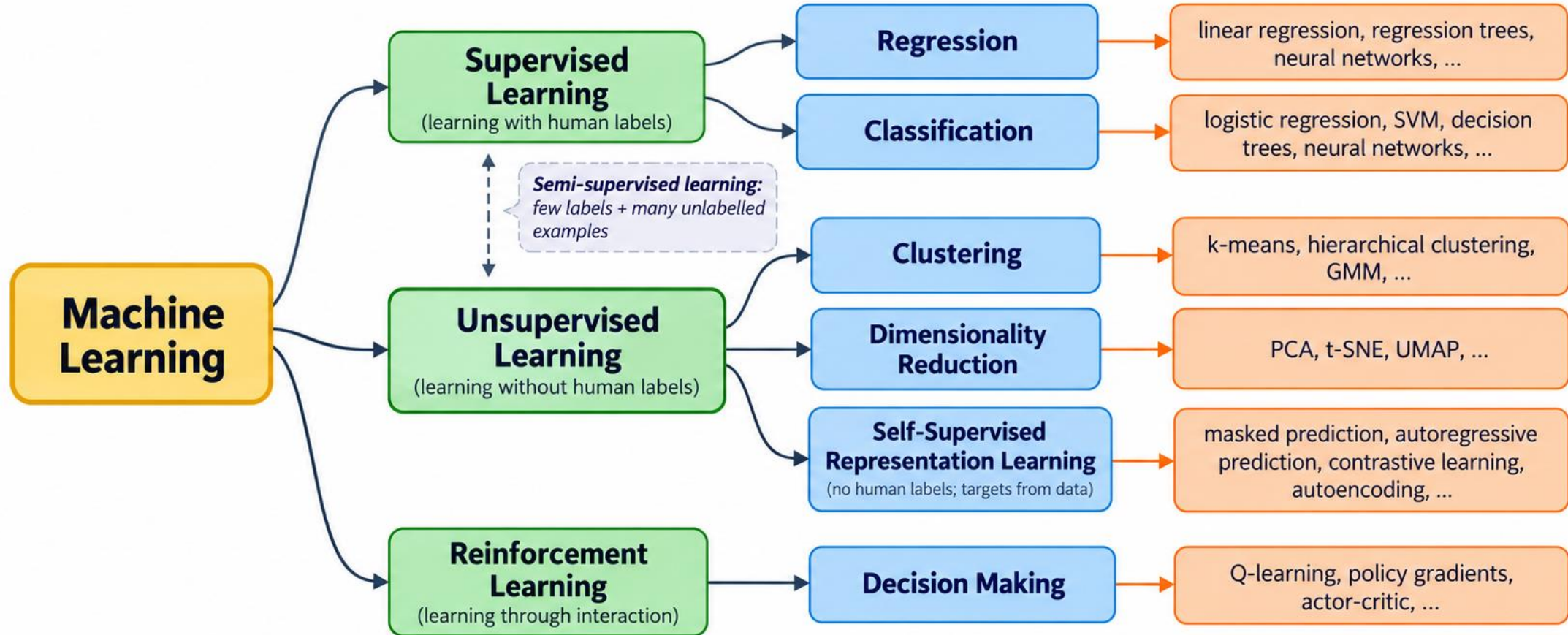
(Machine) Learning Paradigms



(Machine) Learning Paradigms



(Machine) Learning Paradigms



Welcome to “Big” AI!

- Since (2021 ~)
- HUGE neural networks
- VAST amounts of training data
- MASSIVE compute power for training

On the Opportunities and Risks of Foundation Models

Rishi Bommasani* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji
Annie Chen Kathleen Creel Jared Quincy Davis Dorothea Demszky Chris Donahue
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kavin Ethayarajh
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avani Narayan
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou
Percy Liang*¹

Center for Research on Foundation Models (CRFM)
Stanford Institute for Human-Centered Artificial Intelligence (HAI)
Stanford University

AI is undergoing a paradigm shift with the rise of models (e.g., BERT, DALL-E, GPT-3) trained on broad data (generally using self-supervision at scale) that can be adapted to a wide range of downstream tasks. We call these models foundation models to underscore their critically central yet incomplete character. This report provides a thorough account of the opportunities and risks of foundation models, ranging from their capabilities (e.g., language, vision, robotic manipulation, reasoning, human interaction) and technical principles (e.g., model architectures, training procedures, data, systems, security, evaluation, theory) to their applications (e.g., law, healthcare, education) and societal impact (e.g., inequity, misuse, economic and environmental impact, legal and ethical considerations). Though foundation models are based on standard deep learning and transfer learning, their scale results in new emergent capabilities, and their effectiveness across so many tasks incentivizes homogenization. Homogenization provides powerful leverage but demands caution, as the defects of the foundation model are inherited by all the adapted models downstream. Despite the impending widespread deployment of foundation models, we currently lack a clear understanding of how they work, when they fail, and what they are even capable of due to their emergent properties. To tackle these questions, we believe much of the critical research on foundation models will require deep interdisciplinary collaboration commensurate with their fundamentally sociotechnical nature.

¹Corresponding author: pliang@cs.stanford.edu

*Equal contribution.

arXiv:2108.07258v3 [cs.LG] 12 Jul 2022

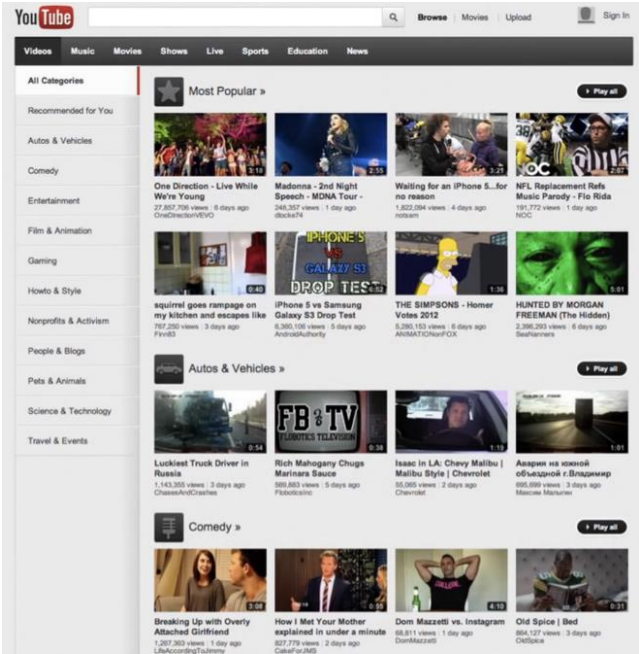
Key Idea

Intelligence
Is primarily a problem of
data

A series of factors

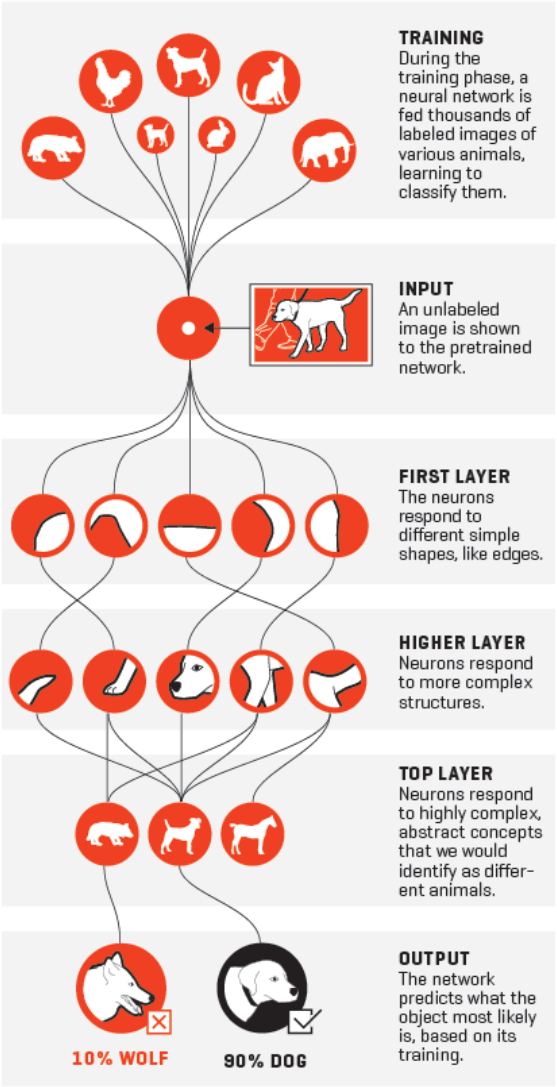


GPUs



Data

HOW NEURAL NETWORKS RECOGNIZE A DOG IN A PHOTO

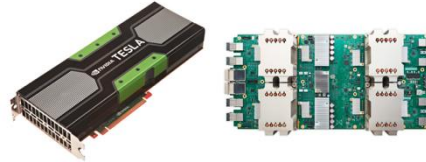


Models

Why Deep Learning Now?

- Better algorithms and understanding

- Computing power (GPUs, TPUs, ...)



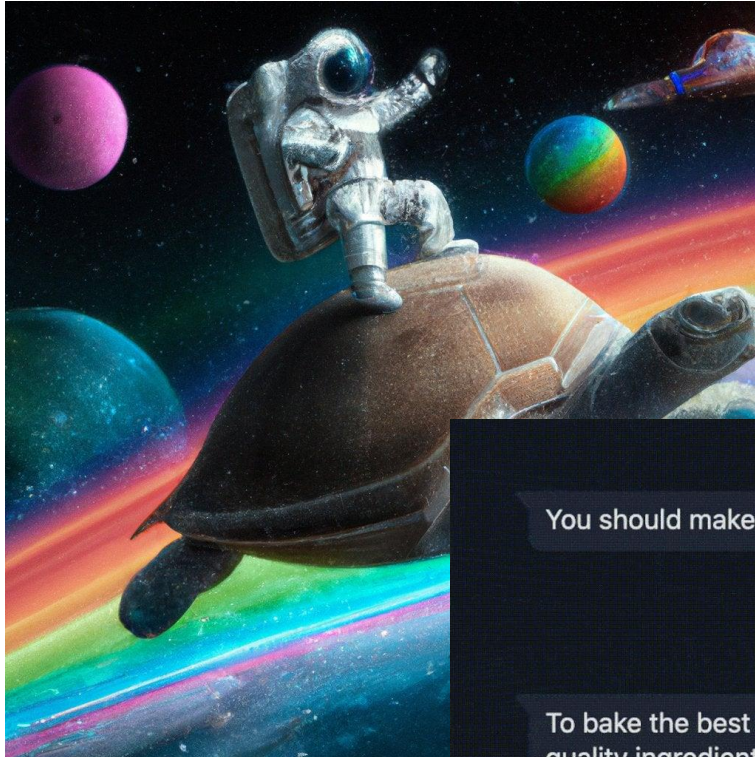
- Abundant data

- Open source tools and mod



Large (Language) Models

- On going story..



with it?

You should make a chocolate mousse!

what's the secret for baking the best chocolate chip cookies?

5:07 PM ✓✓

To bake the best chocolate chip cookies, you should use high quality ingredients. Also, you should not over bake the cookies.

5:07 PM



Type a message

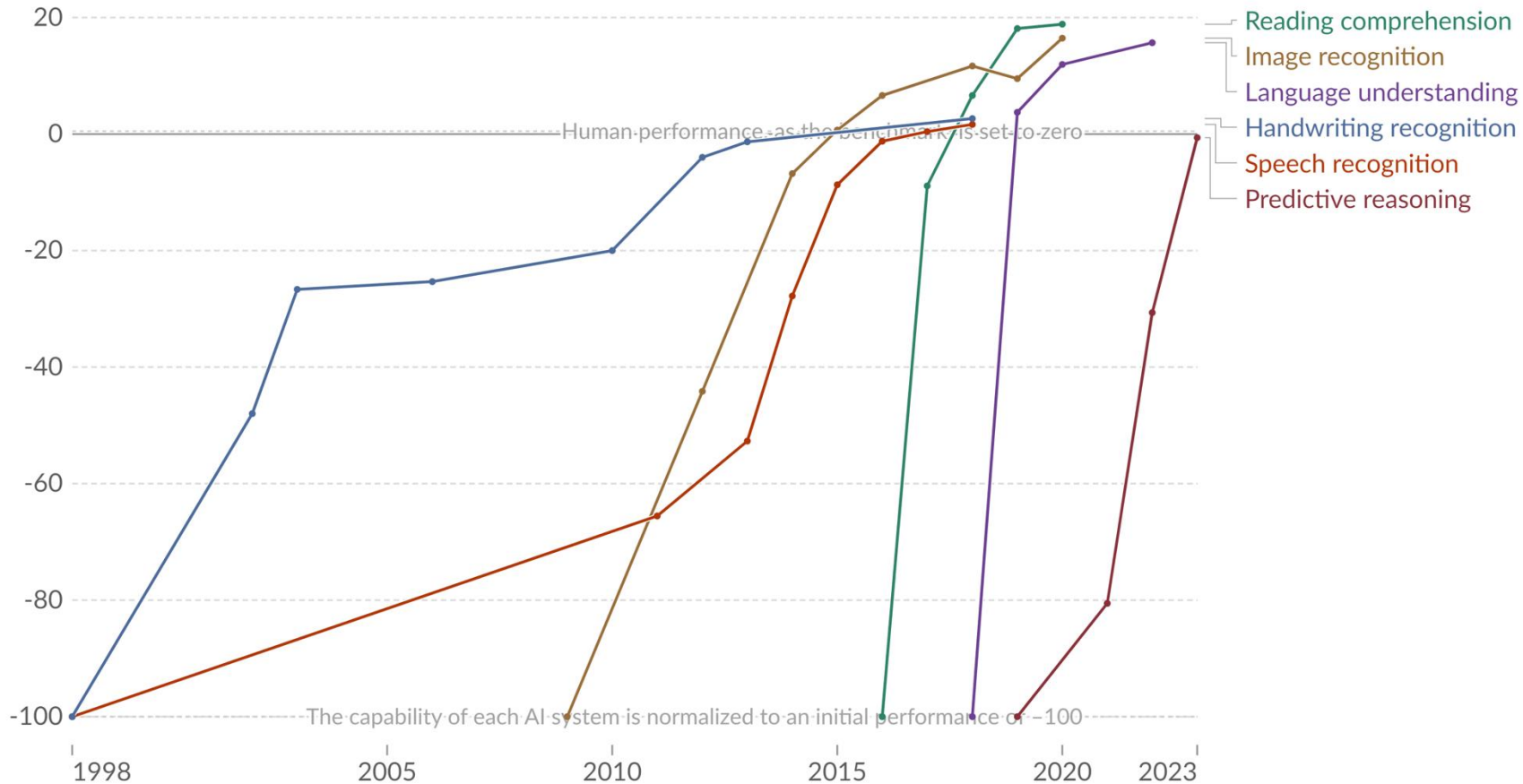


send_tweet.py

```
10 |  
11  
12  
13  
14  
15
```

Test scores of AI systems on various capabilities relative to human performance

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.



Data source: Kiela et al. (2023)

OurWorldinData.org/artificial-intelligence | CC BY

Note: For each capability, the first year always shows a baseline of -100, even if better performance was recorded later that year.

The Philosophy of ML Algorithms

The Philosophical foundations of Machine Learning



Train PhD students to be thinkers not just specialists

Many doctoral curricula aim to produce narrowly focused researchers rather than critical thinkers. That can and must change, says Gundula Bosch.

Under pressure to turn out productive lab members quickly, many PhD programmes in the biomedical sciences have shortened their courses, squeezing out opportunities for putting research into its wider context. Consequently, most PhD curricula are unlikely to nurture the big thinkers and creative problem-solvers that society needs.

That means students are taught every detail of a microbe's life cycle but little about the life scientific. They need to be taught to recognize how errors can occur. Trainees should evaluate case studies derived from flawed real research, or use interdisciplinary detective games to find logical fallacies in the literature. Above all, students must be shown the scientific process as it is — with its limitations and potential pitfalls as well as its fun side, such as serendipitous discoveries and hilarious blunders.

This is exactly the gap that I am trying to fill at Johns Hopkins University in Baltimore, Maryland, where a new graduate science programme is entering its second year. Microbiologist Arturo Casadevall and I began pushing for reform in early 2015, citing the need to put the philosophy back into the doctorate of philosophy: that is, the 'Ph' back into the PhD. We call our programme R3, which means that our students learn to apply rigour to their design and conduct of experiments; view their work through the lens of social responsibility; and to think critically, communicate better, and thus improve reproducibility. Although we are aware of many innovative individual courses developed along these lines, we are striving for more-comprehensive reform.

Our offerings are different from others at the graduate level. We have critical-thinking assignments in which students analyse errors in reasoning in a *New York Times* opinion piece about 'big sugar', and the ethical implications of the arguments made in a *New Yorker* piece by surgeon Atul Gawande entitled 'The Mistrust of Science'. Our courses on rigorous research, scientific integrity, logic, and mathematical and programming skills are integrated into students' laboratory and field-work. Those studying the influenza virus, for example, work with real-life patient data sets and wrestle with the challenges of applied statistics.

A new curriculum starts by winning allies. Both students and faculty members must see value in moving off the standard track. We used informal interviews and focus groups to identify areas in which students and faculty members saw gaps in their training. Recurring themes included the inability to apply theoretical knowledge in statistical tests in the laboratory, frequent mistakes in choosing an appropriate set of experimental controls, and significant difficulty in explaining work to non-experts.

Introducing our programme to colleagues in the Johns Hopkins life-sciences departments was even more sensitive. I was startled by the oft-expressed opinion that scientific productivity depended more

on rote knowledge than on competence in critical thinking. Several principal investigators were uneasy about students committing more time to less conventional forms of education. The best way to gain their support was coffee: we repeatedly met lab heads to understand their concerns.

With the pilot so new, we could not provide data on students' performance, but we could address faculty members' scepticism. Some colleagues were apprehensive that students would take fewer courses in specialized content to make room for interdisciplinary courses on ethics, epistemology and quantitative skills. In particular, they worried that the R3 programme could lengthen the time required for students to complete their degree, leave them insufficiently knowledgeable in their subject areas and make them less productive in the lab.

We made the case that better critical thinking and fewer mandatory discipline-specific classes might actually position students to be more productive. We convinced several professors to try the new system and participate in structured evaluations on whether R3 courses contributed to students' performance.

So far, we have built 5 new courses from scratch and have enrolled 85 students from nearly a dozen departments and divisions. The courses cover the anatomy of errors and misconduct in scientific practice and teach students how to dissect the scientific literature. An interdisciplinary discussion series encourages broad and critical thinking about science. Our students learn to consider societal consequences of research advances, such

as the ability to genetically alter sperm and eggs.

Discussions about the bigger-picture problems of the scientific enterprise get students to reflect on the limits of science, and where science's ability to do something competes with what scientists should do from a moral point of view. In addition, we have seminars and workshops on professional skills, particularly leadership skills through effective communication, teaching and mentoring.

It is still early days for assessment. So far, however, trainees have repeatedly emphasized that gaining a broader perspective has been helpful. In future, we will collect information about the impact that the R3 approach has on graduates' career choices and achievements.

We believe that researchers who are educated more broadly will do science more thoughtfully, with the result that other scientists, and society at large, will be able to rely on this work for a better, more rational world. Science should strive to be self-improving, not just self-correcting. ■

Gundula Bosch directs the R3 Graduate Science Initiative at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. e-mail: gbosch@jhu.edu

PUT THE
PHILOSOPHY
BACK
INTO THE
DOCTORATE
OF
PHILOSOPHY.

The Philosophical foundations of Machine Learning



Train PhD students to be thinkers not just specialists

Many doctoral curricula aim to produce narrowly focused researchers rather than critical thinkers. That can and must change, says Gundula Bosch.

Under pressure to turn out productive lab members quickly, many PhD programmes in the biomedical sciences have shortened their courses, squeezing out opportunities for putting research into its wider context. Consequently, most PhD curricula are unlikely to nurture the big thinkers and creative problem-solvers that society needs.

That means students are taught every detail of a microbe's life cycle but little about the life scientific. They need to be taught to recognize how errors can occur. Trainees should evaluate case studies derived from flawed real research, or use interdisciplinary detective games to find logical fallacies in the literature. Above all, students must be shown the scientific process as it is — with its limitations and potential pitfalls as well as its fun side, such as serendipitous discoveries and hilarious blunders.

This is exactly the gap that I am trying to fill at Johns Hopkins University in Baltimore, Maryland, where a new graduate science programme is entering its second year. Microbiologist Arturo Casadevall and I began pushing for reform in early 2015, citing the need to put the philosophy back into the doctorate of philosophy: that is, the 'Ph' back into the PhD. We call our programme R3, which means that our students learn to apply rigour to their design and conduct of experiments; view their work through the lens of social responsibility; and to think critically, communicate better, and thus improve reproducibility. Although we are aware of many innovative individual courses developed along these lines, we are striving for more-comprehensive reform.

Our offerings are different from others at the graduate level. We have critical-thinking assignments in which students analyse errors in reasoning in a *New York Times* opinion piece about 'big sugar', and the ethical implications of the arguments made in a *New Yorker* piece by surgeon Atul Gawande entitled 'The Mistrust of Science'. Our courses on rigorous research, scientific integrity, logic, and mathematical and programming skills are integrated into students' laboratory and field-work. Those studying the influenza virus, for example, work with real-life patient data sets and wrestle with the challenges of applied statistics.

A new curriculum starts by winning allies. Both students and faculty members must see value in moving off the standard track. We used informal interviews and focus groups to identify areas in which students and faculty members saw gaps in their training. Recurring themes included the inability to apply theoretical knowledge in statistical tests in the laboratory, frequent mistakes in choosing an appropriate set of experimental controls, and significant difficulty in explaining work to non-experts.

Introducing our programme to colleagues in the Johns Hopkins life-sciences departments was even more sensitive. I was startled by the oft-expressed opinion that scientific productivity depended more

on rote knowledge than on competence in critical thinking. Several principal investigators were uneasy about students committing more time to less conventional forms of education. The best way to gain their support was coffee: we repeatedly met lab heads to understand their concerns.

With the pilot so new, we could not provide data on students' performance, but we could address faculty members' scepticism. Some colleagues were apprehensive that students would take fewer courses in specialized content to make room for interdisciplinary courses on ethics, epistemology and quantitative skills. In particular, they worried that the R3 programme could lengthen the time required for students to complete their degree, leave them insufficiently knowledgeable in their subject areas and make them less productive in the lab.

We made the case that better critical thinking and fewer mandatory discipline-specific classes might actually position students to be more productive. We convinced several professors to try the new system and participate in structured evaluations on whether R3 courses contributed to students' performance.

So far, we have built 5 new courses from scratch and have enrolled 85 students from nearly a dozen departments and divisions. The courses cover the anatomy of errors and misconduct in scientific practice and teach students how to dissect the scientific literature. An interdisciplinary discussion series encourages broad and critical thinking about science. Our students learn to consider societal consequences of research advances, such

as the ability to genetically alter sperm and eggs.

Discussions about the bigger-picture problems of the scientific enterprise get students to reflect on the limits of science, and where science's ability to do something competes with what scientists should do from a moral point of view. In addition, we have seminars and workshops on professional skills, particularly leadership skills through effective communication, teaching and mentoring.

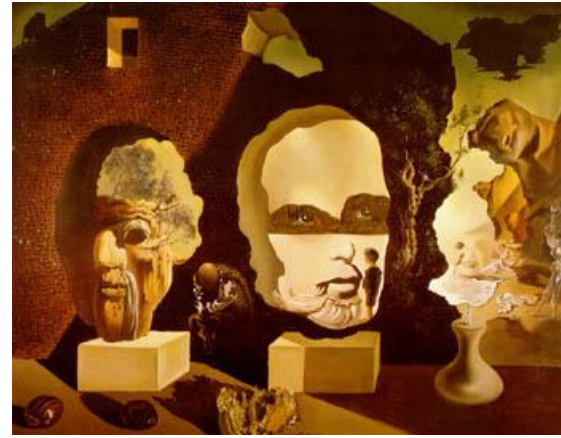
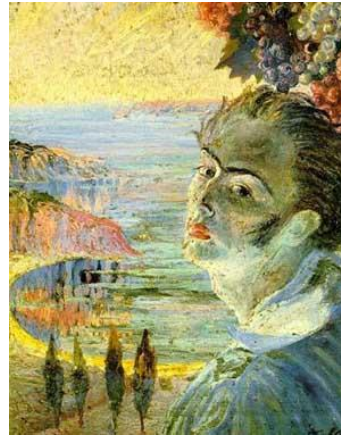
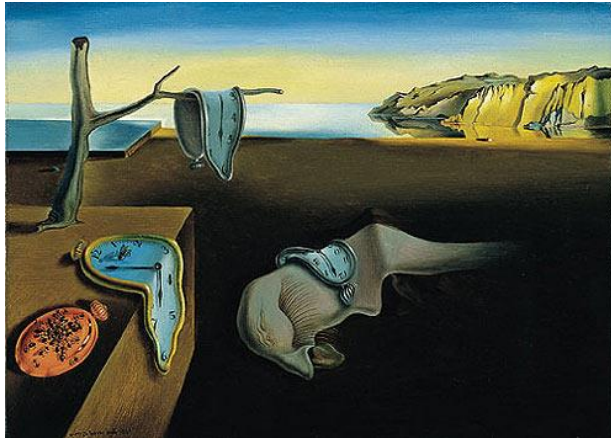
It is still early days for assessment. So far, however, trainees have repeatedly emphasized that gaining a broader perspective has been helpful. In future, we will collect information about the impact that the R3 approach has on graduates' career choices and achievements.

We believe that researchers who are educated more broadly will do science more thoughtfully, with the result that other scientists, and society at large, will be able to rely on this work for a better, more rational world. Science should strive to be self-improving, not just self-correcting. ■

Gundula Bosch directs the R3 Graduate Science Initiative at Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. e-mail: gboesch@jhu.edu

PUT THE
PHILOSOPHY
BACK
INTO THE
DOCTORATE
OF
PHILOSOPHY.

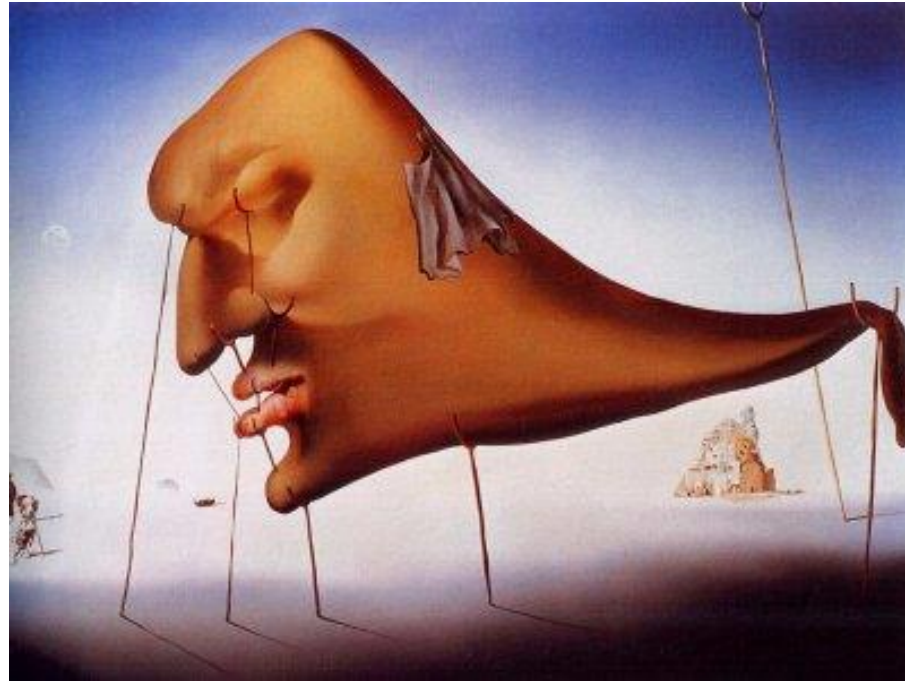
Paintings by two different painters



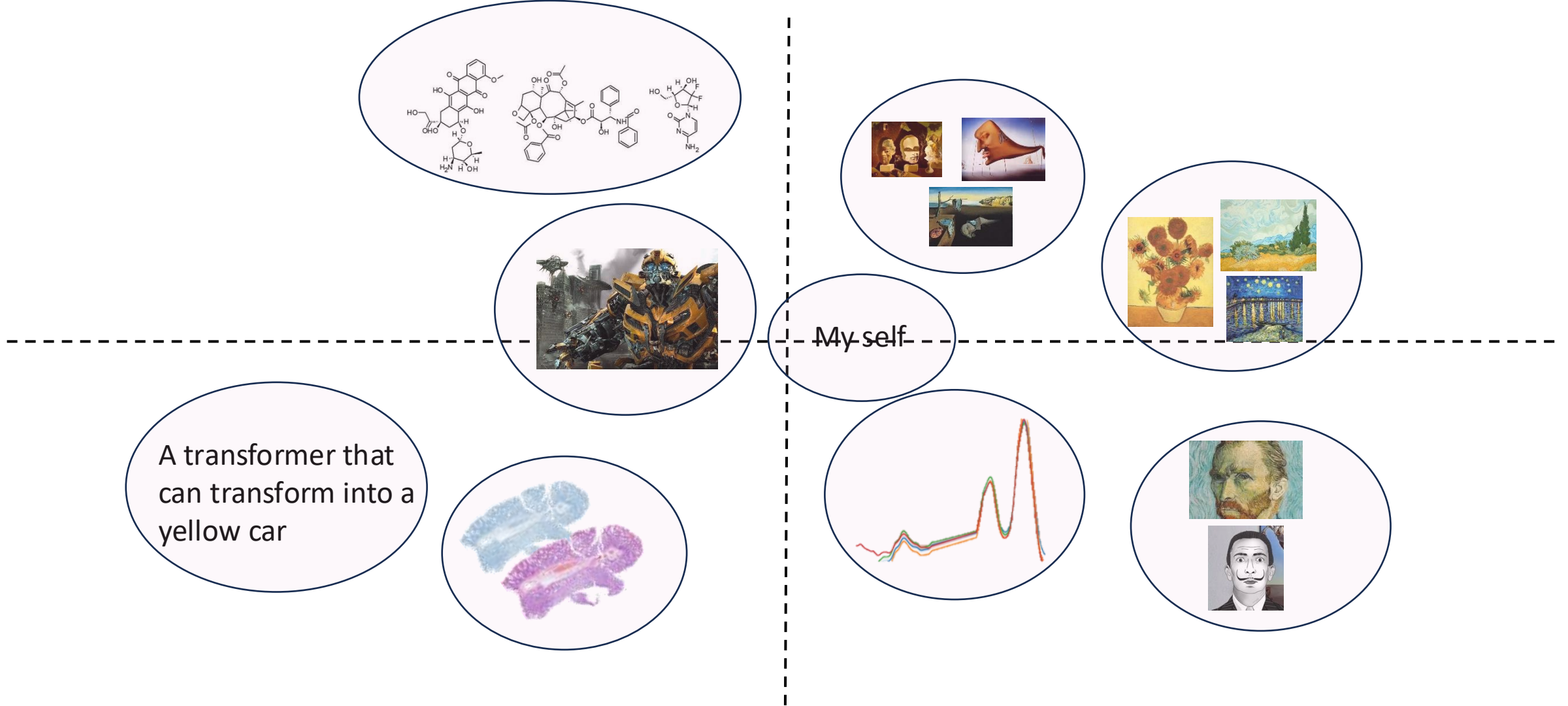
Who's painting is this?



And this?



learning from data for generalization to unseen cases
inductive inference



I. Entities have (explicit or implicit) representations



“Bank” in which statement is more semantically related to the picture?

- **A:** As he walked by the **bank**, he saw some tillers
- **B:** As he walked by the **bank**, he saw some tellers

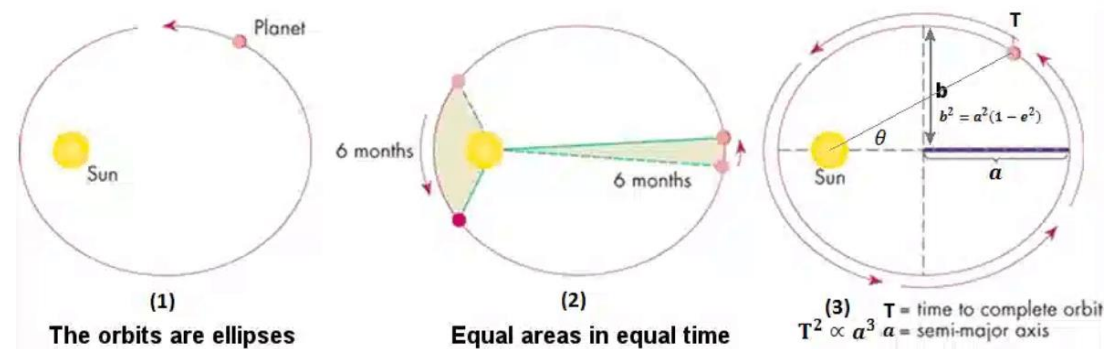


As he walked by the **bank**, he saw some tillers



As he walked by the **bank**, he saw some tellers

II. Semantic relatedness of entities is context dependent
and thus their **representations are contextual**



III. It is possible to develop representations in an **inductive** manner (i.e., through empirical observations)

https://en.wikipedia.org/wiki/Johannes_Kepler

<https://earthobservatory.nasa.gov/features/OrbitsHistory>

<https://plato.stanford.edu/entries/induction-problem/>



IV. Representation of an entity can allow us to reconstruct or “generate” it



US Airways Flight 1549



- V.** Generalise beyond the training observations &
- VI.** Learning through interactions with actions and consequences.

Modelling Principles

Algorithms

I. Entities have (explicit or implicit) representations

Feature analysis / Representation learning

II. Inductive bias and structure: Relations between entities depend on the context, and thus their representations are contextual.

Using Convolutions, Transformers, Graph Layers, or any other architectural priors.

III. Learning from data: It is possible to develop representations in an inductive manner (through empirical observations)

Loss functions, (gradient) optimisation

IV. Learning signal: Representation of any entity can allow us to reconstruct or “generate” it to a “sufficient” degree

Supervised, self-supervised, generative objectives

V. Generalise beyond the training observations

Regularisation, validation, augmentation, etc.

VI. Learning through interactions with actions and consequences

Reinforcement learning, policies, world models, etc.

Conclusion

- What is meant by a feature representation?
- What is a feature space?
- What are the philosophical foundations of machine learning?
- What is linear separability?
- Lab assignments
- What applications is ML particularly suited to?
- What are some interesting applications of machine learning?
- What interesting application would you like to develop once you learn the module?
- What do you think is lacking in the philosophical foundations of machine learning we discussed?