UNIVERSIDADE CATÓLICA PORTUGUESA

BRAGA

Machine Learning

Session 1 - T

Introduction to Machine Learning

Degree in Applied Data Science

2024/2025

Objectives

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

- Learn the theoretical foundations of advanced machine learning concepts;

- Implement advanced machine learning algorithms using Python;

Classes

- Typically, classes will be divided in two parts:
  - Foundations of Data Science and Machine Learning (1 hour).
  - Hands-on Implementation in Python (1 hour).

Evaluation

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

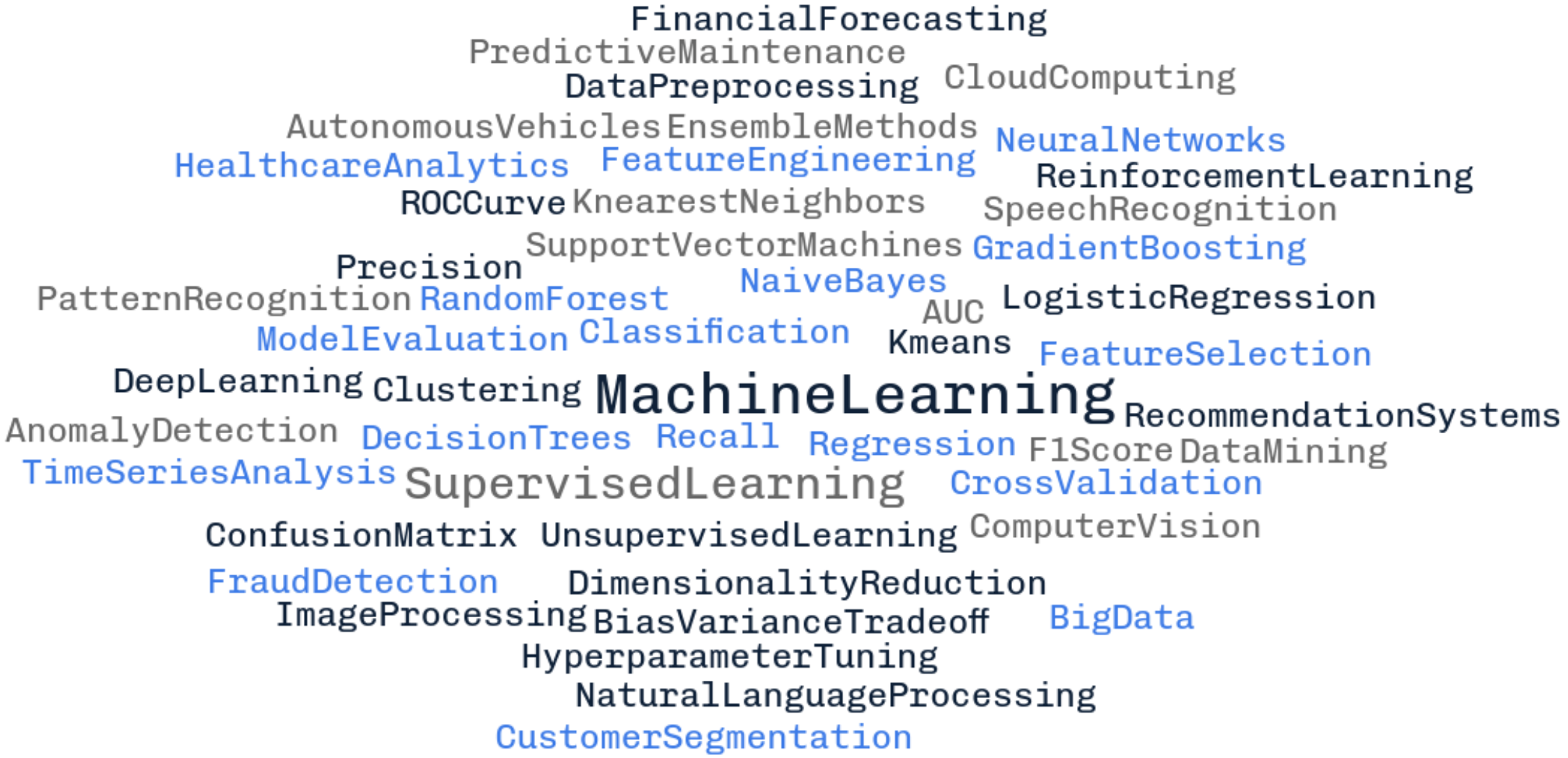
- Exam (35%)
  - Date: 9th of May 2025
- Group project (35%)
  - Groups of 3 students;
  - Presentation (16th of May 2025);
  - Jupyter Notebook (Submission: 30th of May 2025);
- ML package (30%)
  - Group development of an ML package from scratch on GitHub.
  - Oral defense (30th of May 2025);

UNIVERSIDADE
CATÓLICA
PORTUGUESA

BRAGA

Introduction to Machine Learning

Terminology



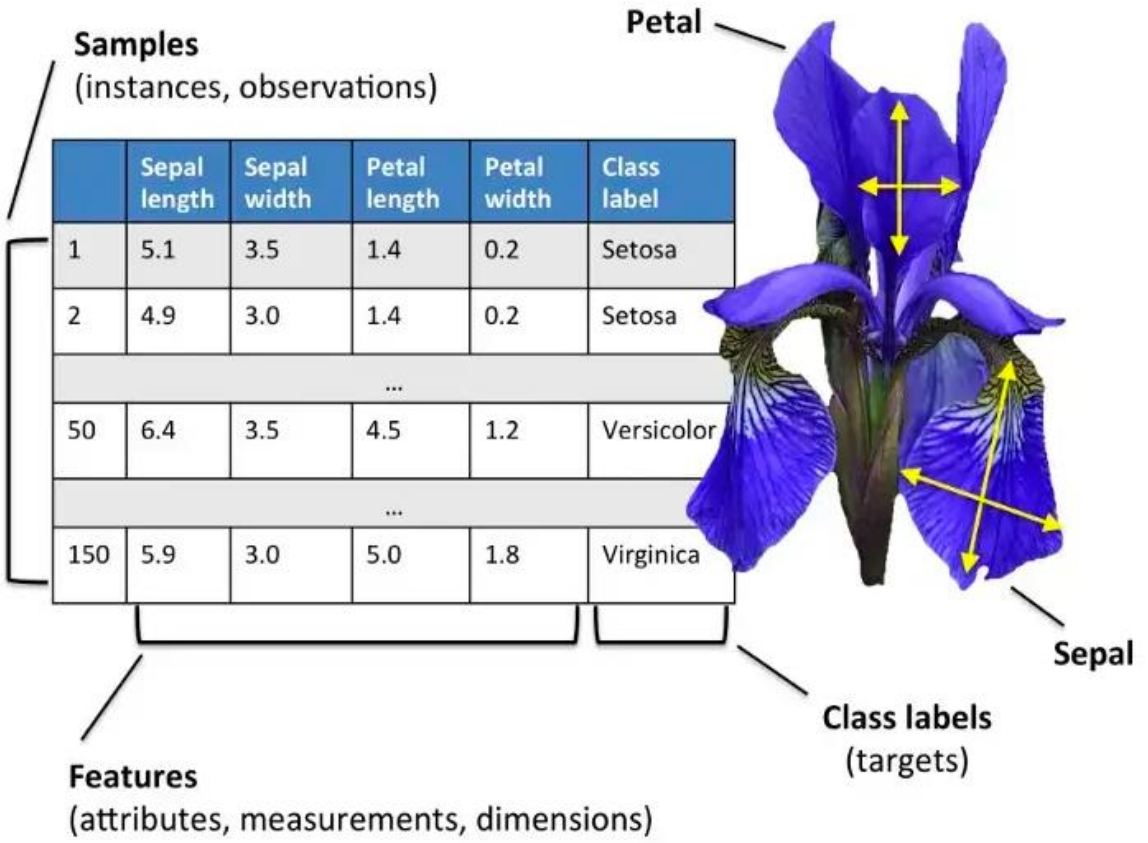
Data everywhere!

- 328.77 million terabytes of data are created each day.*
- By 2025, it is expected to skyrocket to a staggering 181 zettabytes.*
- Last two years alone, account for 90% of the world's data.*

* https://news.wildintelligence.xyz/p/328-million-terabytes-of-data

Data Types

- Tabular data:

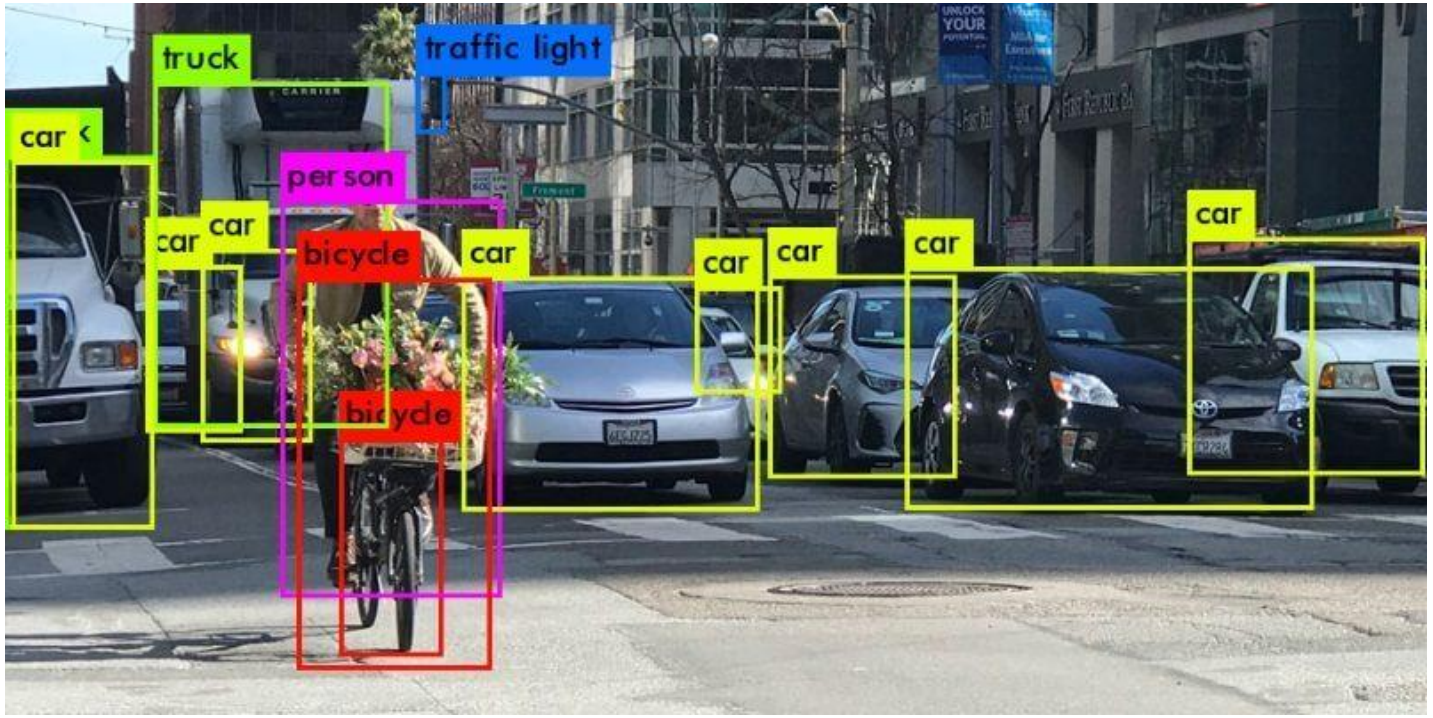


	Sepal length	Sepal width	Petal length	Petal width	Class label
1	5.1	3.5	1.4	0.2	Setosa
2	4.9	3.0	1.4	0.2	Setosa
...					
50	6.4	3.5	4.5	1.2	Versicolor
...					
150	5.9	3.0	5.0	1.8	Virginica

https://eminebozkus.medium.com/exploring-the-iris-flower-dataset-4e000bcc266c

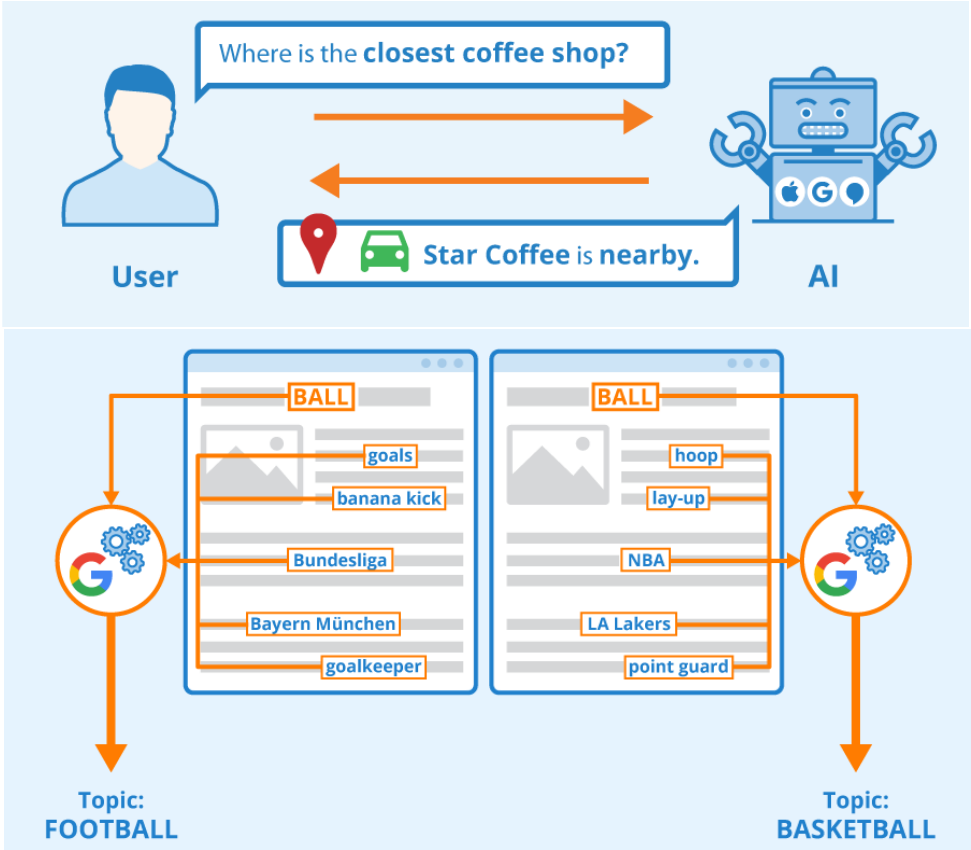
Data Types

- Image data:



Data Types

- Text data:

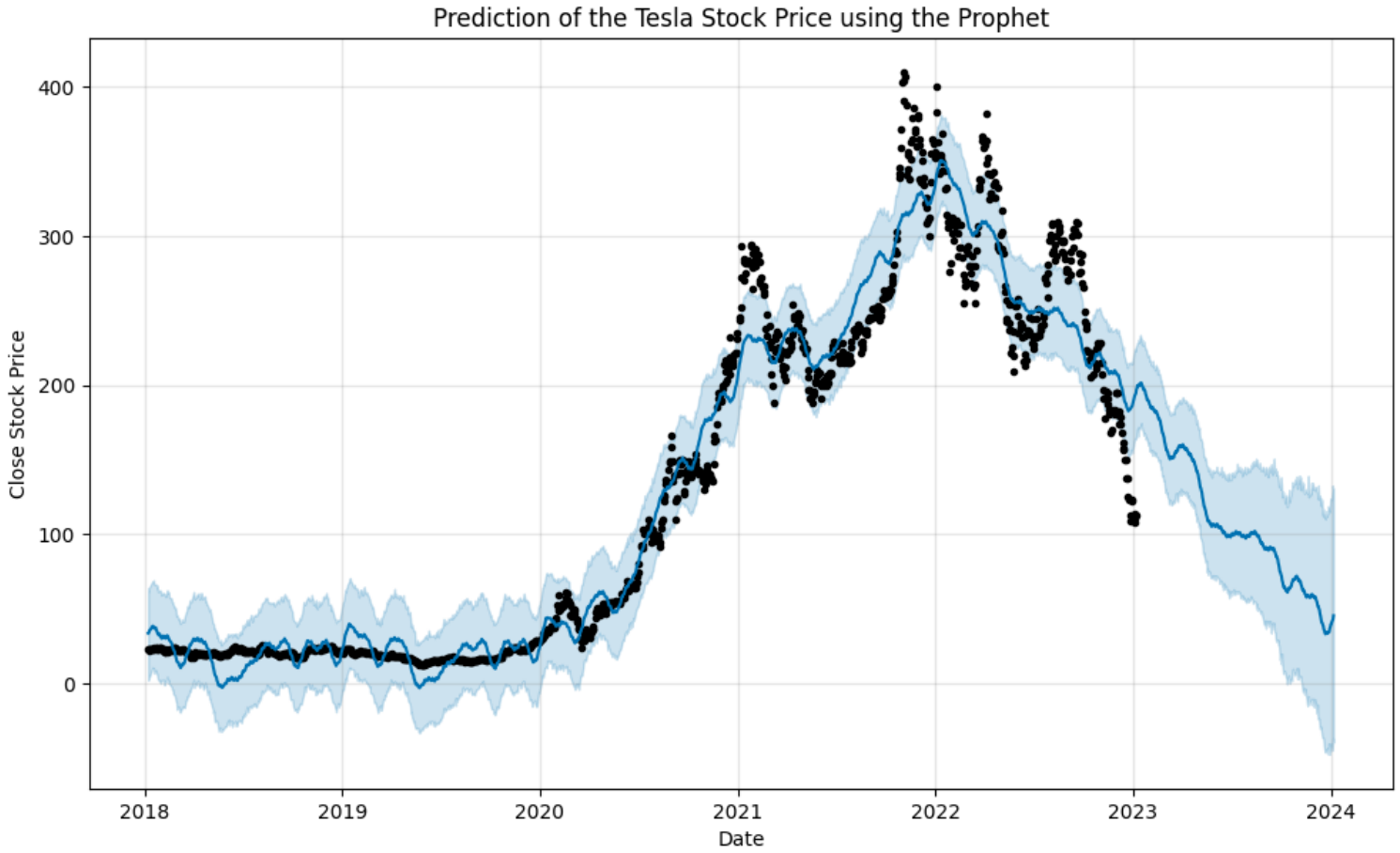


Data Types

- Audio data:

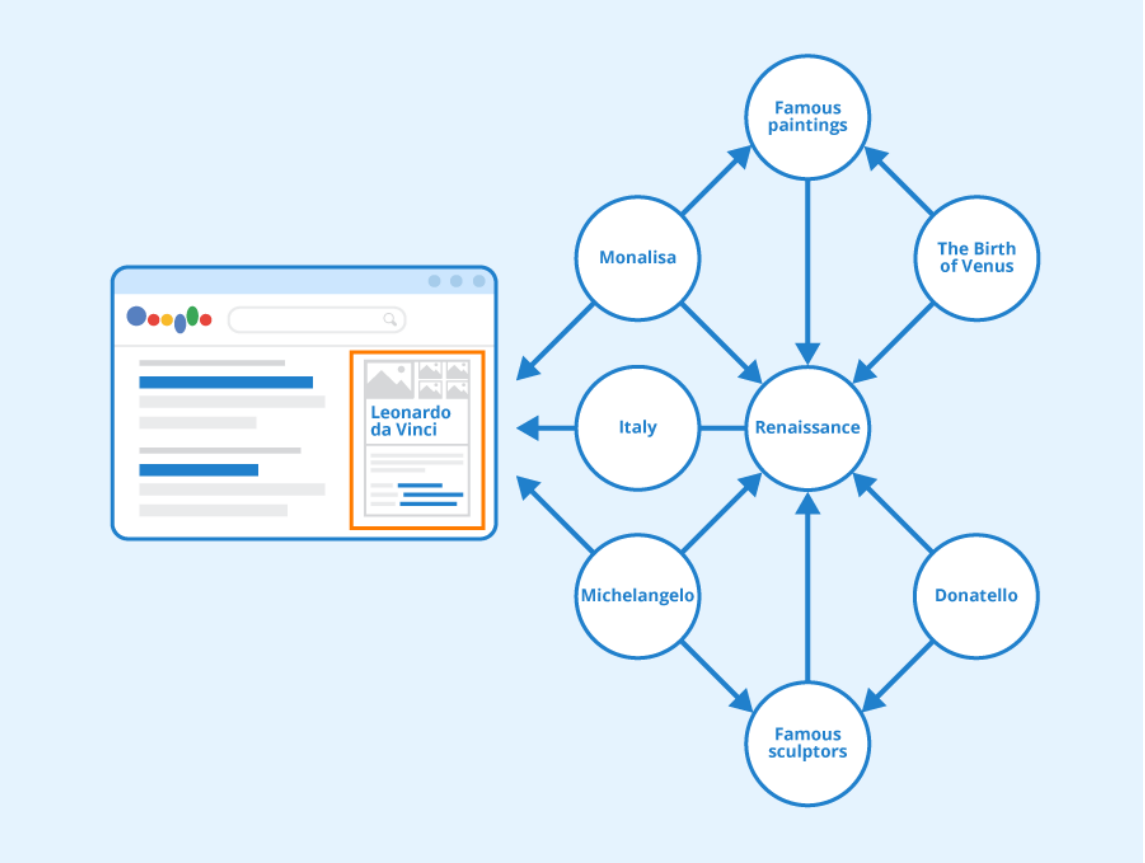
Data Types

- Time series data:

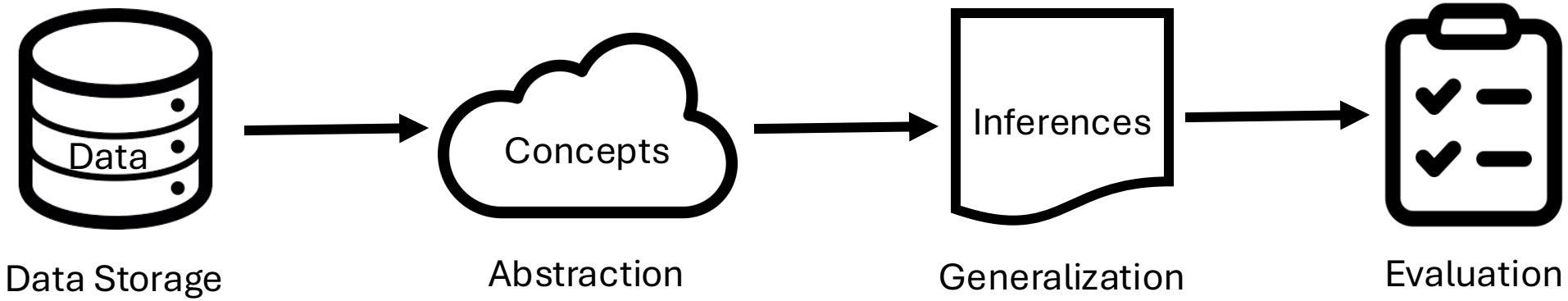


Data Types

- Graph data:



The Components of Learning



- **Data storage:** every learning process starts with data. Both humans and computers rely on data storage as a fundamental basis for more advanced reasoning..
- **Abstraction:** involves extracting knowledge from stored data by forming general concepts that represent the data as a whole.
- **Generalization** involves leveraging abstracted data to generate knowledge and inferences, which subsequently inform actions in novel situations.
- **Evaluation:** serves as a feedback mechanism, measuring the effectiveness of acquired knowledge and offering insights for possible enhancements.

Machine Learning

UNIVERSIDADE CATÓLICA PORTUGUESA
BRAGA

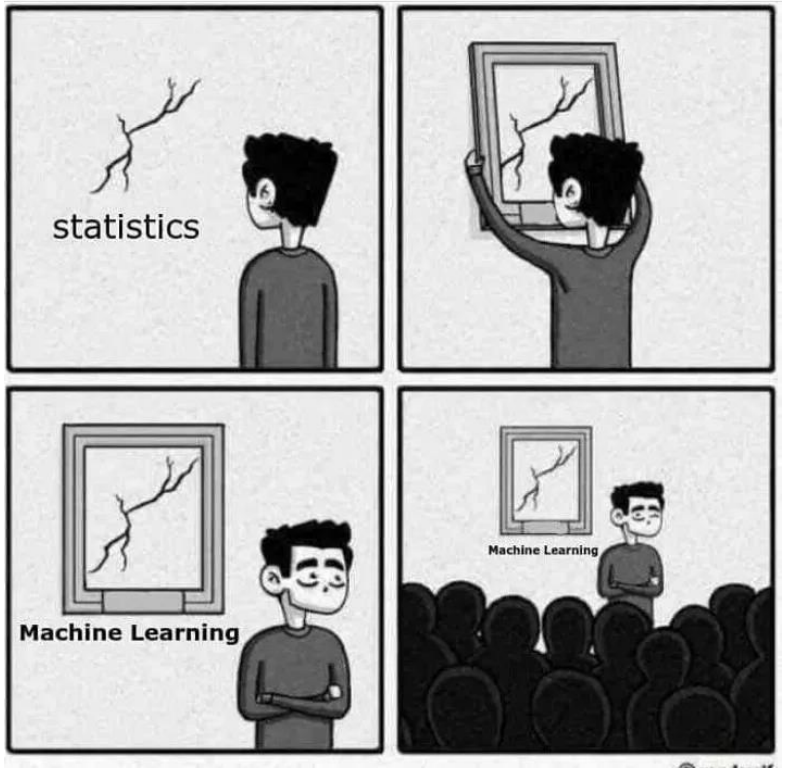
"A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. "

Tom Mitchell. Machine Learning 1997.

What is Machine Learning?

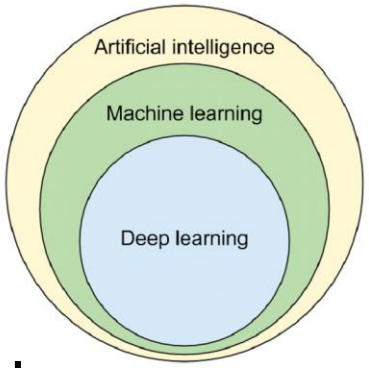
- It’s similar to statistics…
  - Both aim to **uncover patterns within data**.
  - Both heavily rely on **calculus, probability, and linear algebra**, and share many **common algorithms**.

- But it's not statistics!
  - Statistics primarily helps drawing **reliable conclusions**; ML focuses on **building autonomous agents**.
  - Statistics gives greater importance on **interpretability** and **mathematical rigor**; ML prioritizes **predictive performance, scalability, and autonomy**.



https://medium.com/nybles/understanding-machine-learning-through-memes-4580b67527bf

Machine Learning and Artificial Intelligence

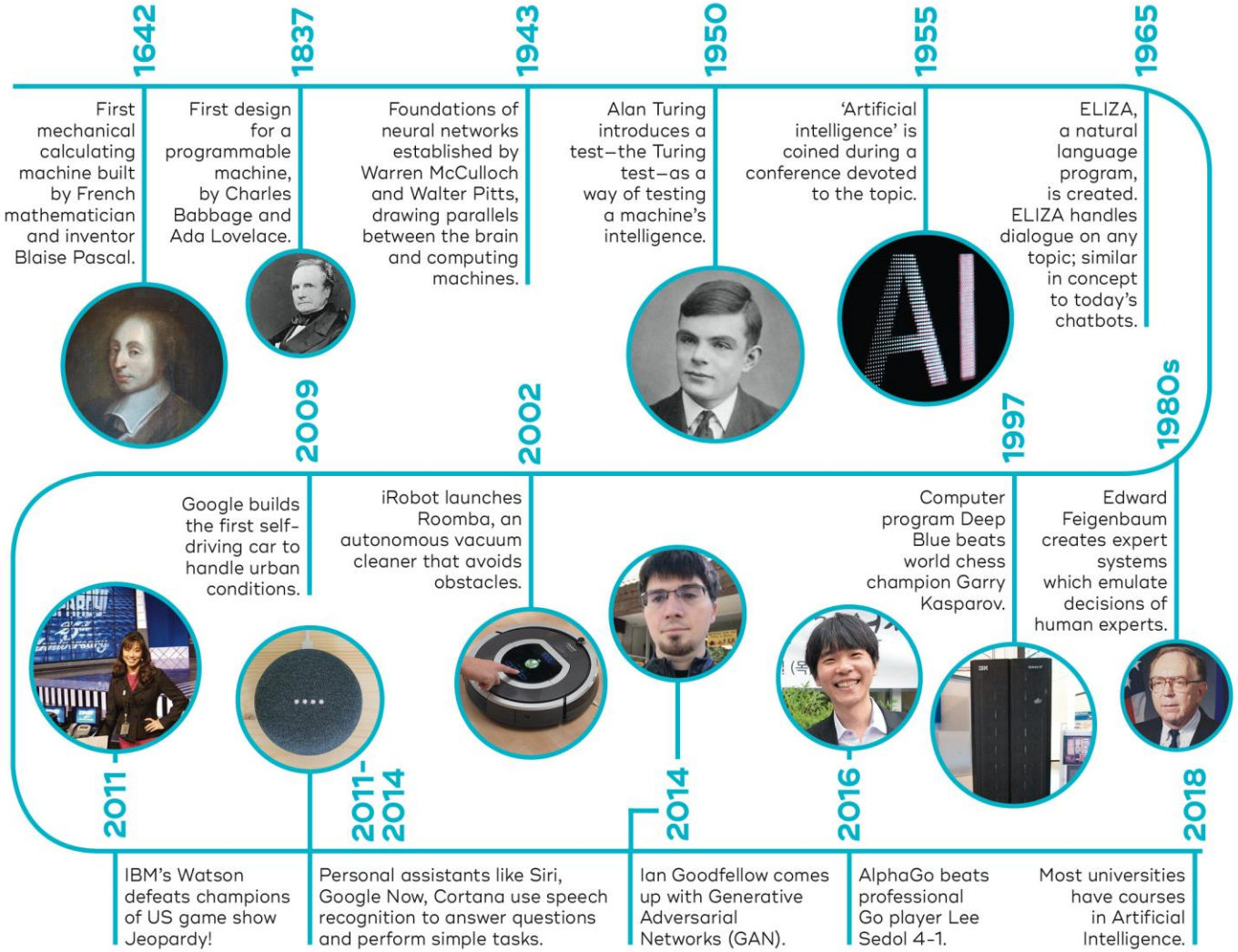


- Machine Learning is a **subset of Artificial Intelligence** (AI).

- AI encompasses various approaches beyond learning-based systems:
  - Symbolic reasoning;
  - Rule-based systems;
  - Tree search;

- Learning-based systems, such as Machine Learning, are more flexible and proficient in solving problems through **data-driven** learning processes.

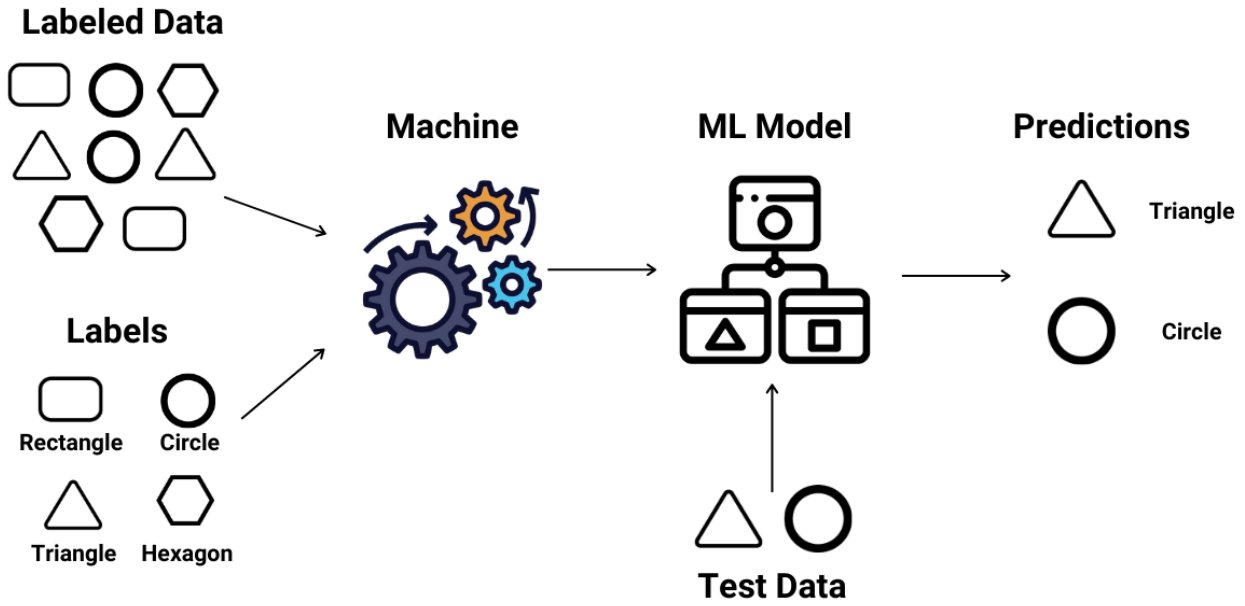
Brief History of Machine Learning



https://qbi.uq.edu.au/brain/intelligent-machines/history-artificial-intelligence

Machine Learning Paradigms

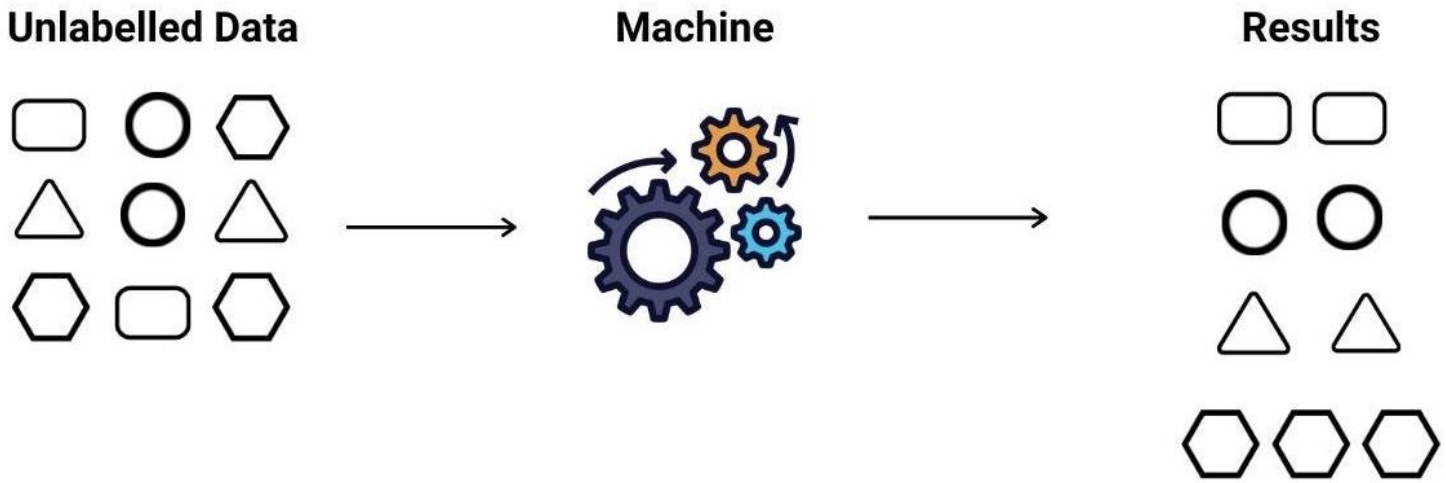
- **Supervised:** the algorithm is trained on a labeled dataset, where the input data is paired with corresponding output labels. The goal is to learn a mapping from inputs to outputs, allowing the algorithm to make predictions on new, unseen data.



https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning

Machine Learning Paradigms

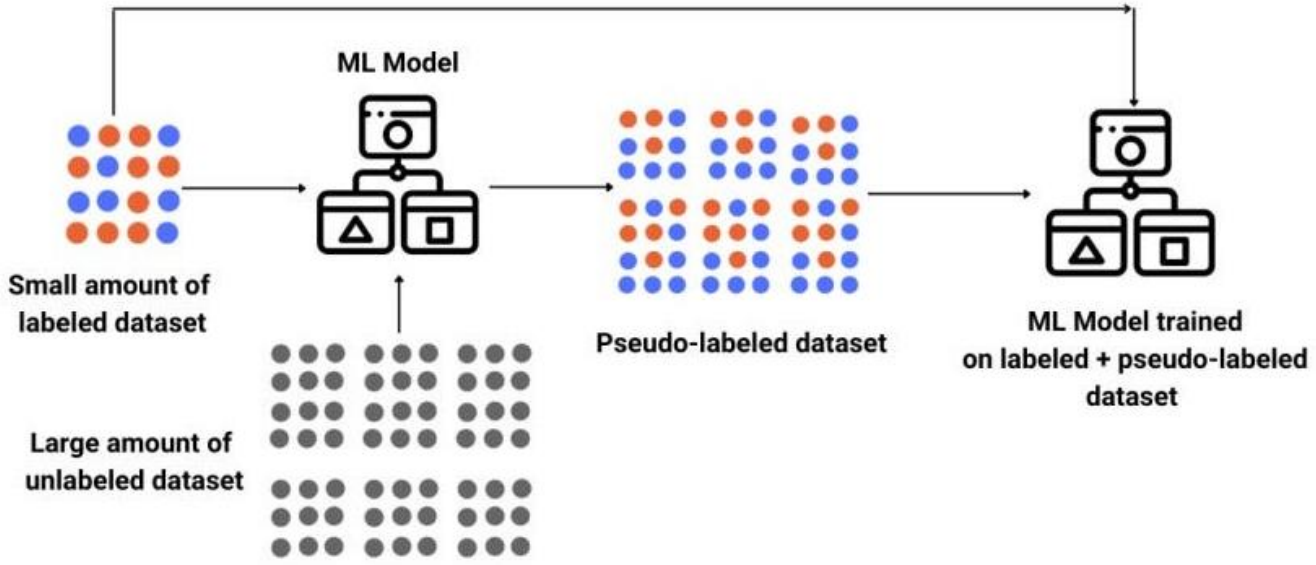
- **Unsupervised:** involves working with unlabeled data, where the algorithm explores the inherent structure and patterns within the input without explicit output guidance.



https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning

Machine Learning Paradigms

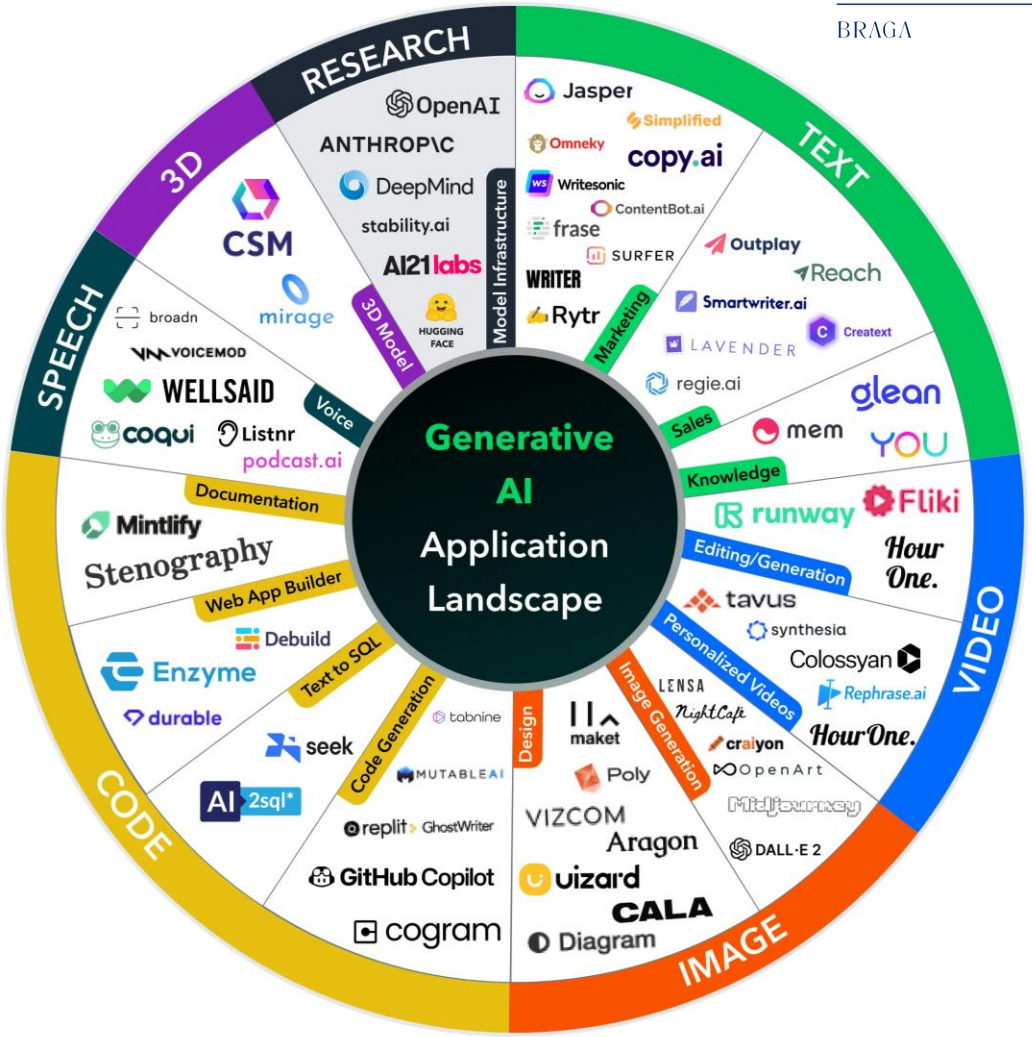
- **Semi-supervised:** combines labeled and unlabeled data for training. It aims to leverage the benefits of both supervised and unsupervised learning, often useful when obtaining labeled data is costly or time-consuming.



https://www.enjoyalgorithms.com/blogs/supervised-unsupervised-and-semisupervised-learning

Machine Learning Paradigms

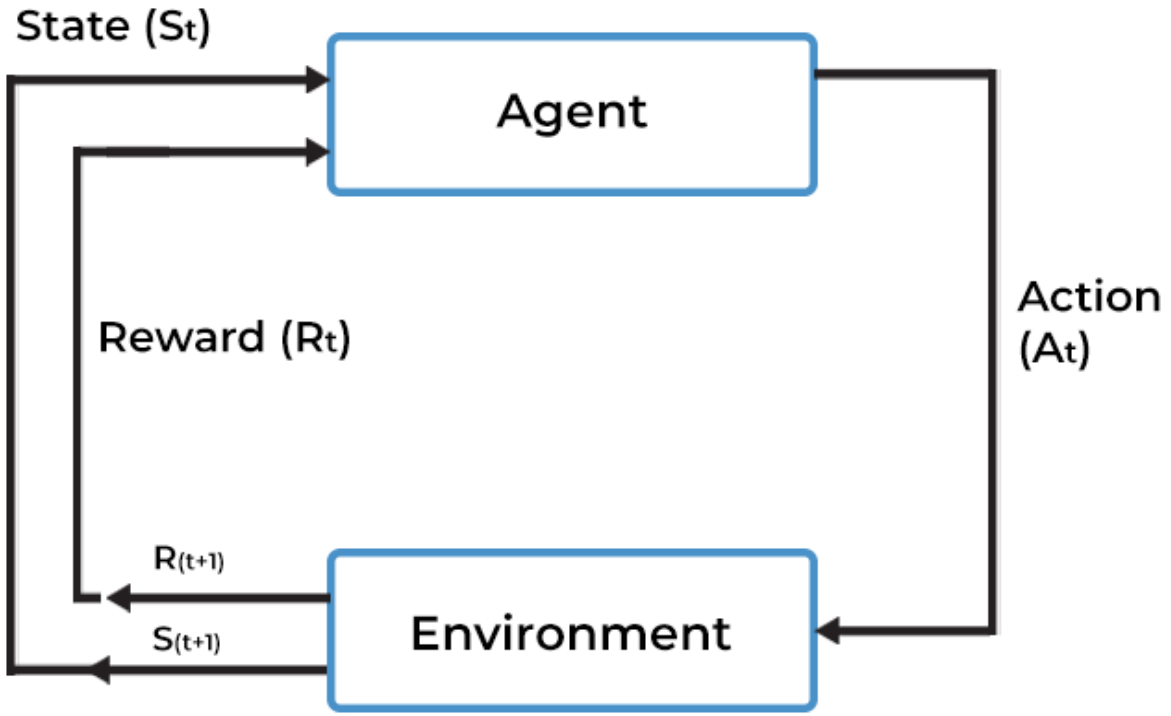
- **Generative:** the focus is on learning the underlying probability distribution of the input data. It can generate new, synthetic samples that resemble the training data.



https://www.rapidops.com/blog/generative-ai-tools/

Machine Learning Paradigms

- **Reinforcement Learning:** an agent learns to make decisions by interacting with an environment. The agent receives feedback in the form of rewards or penalties, enabling it to learn optimal strategies over time.



Machine Learning Applications

- Spam filtering;
- Fraud detection;
- Disease detection;
- Drug discovery;
- Recommendation systems;
- Forecasting;
- Autonomous Vehicles;
- Chatbots;
- Object detection;
- Anomaly detection;
- …

Resources

- Kelleher, J. D., Namee, B. M., & D'Arcy, A. (2015). Fundamentals of machine learning for predictive data analytics. London, England: MIT Press.

- Guido, S., & Mueller, A. C. (2016). Introduction to machine learning with python. Sebastopol, CA: O'Reilly Media.