ESCP EUROPE
BUSINESS SCHOOL

BERLIN
LONDON
MADRID
PARIS
TURIN
WARSAW

Big Data & Analytics

Lecture 4: M.L.&a.i. – Data visualization

DESIGNING TOMORROW

affiliated to
CCI PARIS ILE-DE-FRANCE

What's ML?

According to Google Trends, interest in the term machine learning (ML) has increased over 790% since 2013.

ML has gone from the realm of a relatively small number of data scientists to the mainstream of analysis and business.

The aim of this lecture is to try to answer to the question "What is machine learning?" while using a minimum of technical terms because basic concepts of machine learning are actually not very difficult to grasp while applications and opportunities are enormous.

ESCP EUROPE BUSINESS SCHOOL

What's an algorithm?

An algorithm is a strategy used to solve a problem and consists of a finite sequence of operations.

The term derives from the Latin transcription of the name of the Persian mathematician al-Khwarizmi, who lived in the 9th century AD and is considered one of the first authors to have referred to this concept when he wrote the book "Rules of Restoration and Reduction."

An algorithm can be:	
GENERAL	When the solution is the same for all problems of the same class
FINITE	When it consists of a finite number of instructions, it has an end
NON-AMBIGUOUS	The operations must be interpretable in the same way by everyone, even if the executor is different
DETERMINISTIC	When starting from the same input data, you obtain the same results as output

ESCP EUROPE BUSINESS SCHOOL

Why is relevant for you?

There are several companies in the food and beverage industry that have ALREADY adopted machine learning for various purposes

- **Coca-Cola:** Coca-Cola uses machine learning for predictive analytics to forecast demand for their products, which helps them make better decisions about inventory management, production planning, and pricing.
- **PepsiCo:** PepsiCo uses machine learning to analyze customer behavior and preferences, which helps them tailor their marketing and product offerings to specific customer segments.
- **Nestle:** Nestle uses machine learning to improve supply chain efficiency and reduce waste by predicting demand for their products and optimizing production schedules.
- **Kraft Heinz:** Kraft Heinz uses machine learning for clustering analysis to segment their customers based on demographics, purchase history, and other data. This helps them target their marketing and product offerings more effectively.
- **Starbucks:** Starbucks uses machine learning for predictive analytics to forecast demand for their products and optimize staffing at their stores, which helps them improve efficiency and reduce costs.

What's ML?

Machine learning is, more or less, a way for computers to learn things without being specifically programmed.
But
Machines learn through commands provided by humans.

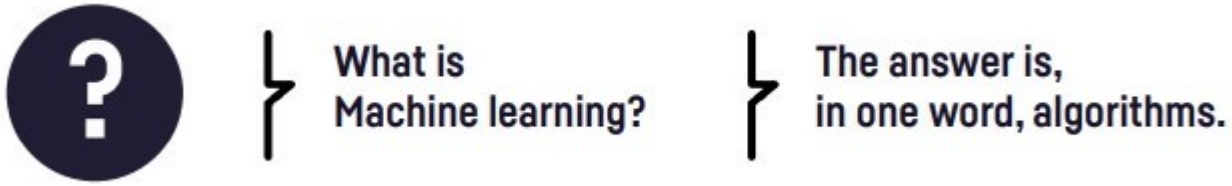
These sets of rules are known as algorithms

Algorithms are sets of rules (written in a formal language) that a computer is able to follow.

ESCP EUROPE BUSINESS SCHOOL

Apparently this is a **contradiction** but it is not .. Because it's all about DATA used by the algorithms and methodology **for deriving syntetic data from "raw" data**

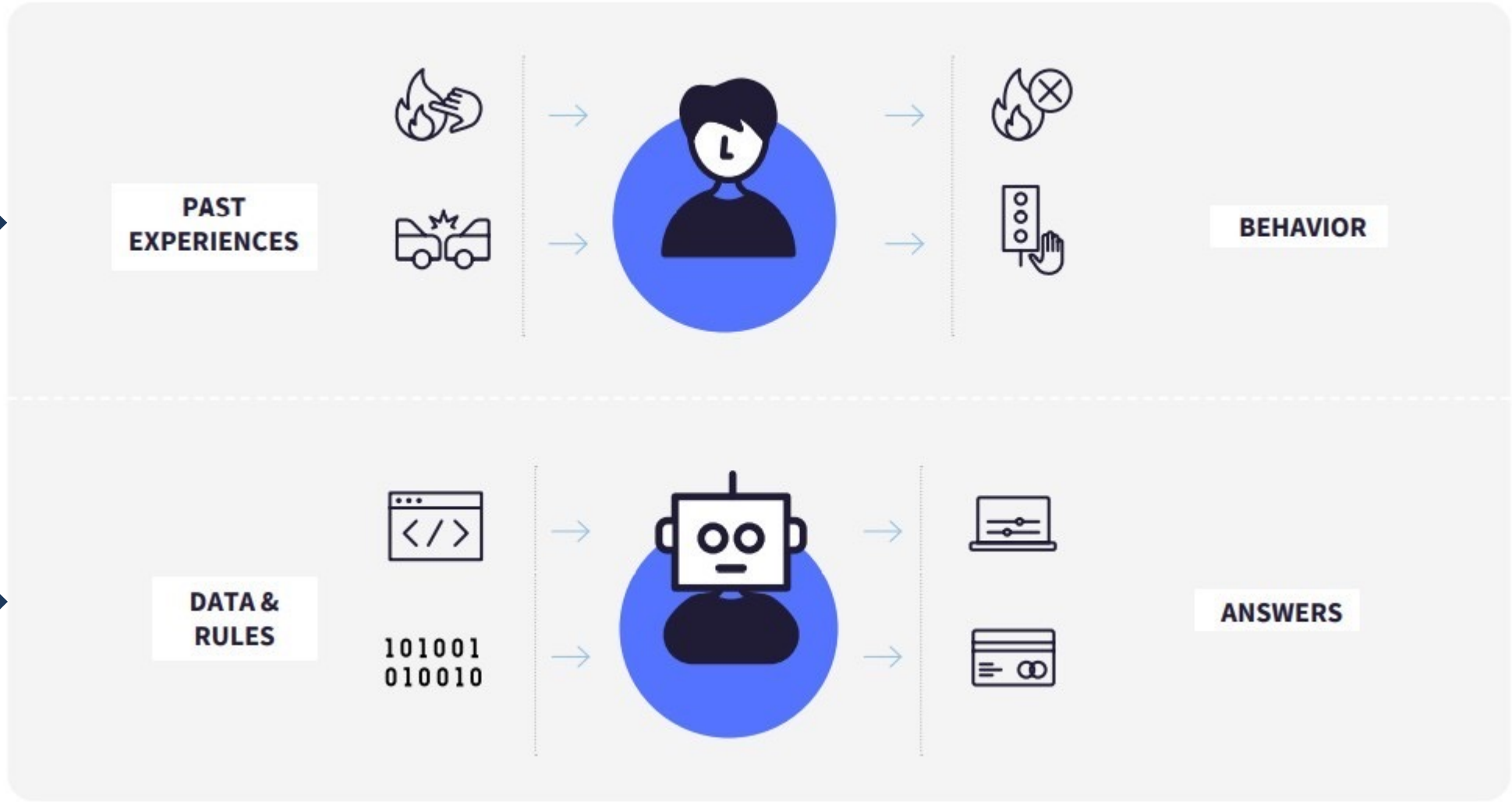
What's ML?



More precisely …
Machine learning is a branch of artificial intelligence that deals with the design and development of algorithms and statistical models that allow computers to learn from data, without being explicitly programmed. The goal of machine learning is to enable computers to identify patterns in data, make predictions, and make decisions, and to continually improve their performance on these tasks as they are exposed to more data.

There are many different types of machine learning algorithms, ranging from simple linear regression models to complex deep neural networks. The choice of which algorithm to use depends on the nature of the problem being solved and the type of data being used.

Learning experience

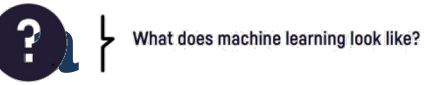


ESCP EUROPE BUSINESS SCHOOL

Prediction or clustering? This is the dilemma

- In machine learning, our goal is either **prediction** or **clustering**.
  - **Prediction** is a process where, from a set of input variables, we estimate the value of an output variable.
  - This technique is used for data that has a precise mapping between input and output, referred to as labeled data.
  - This is known as supervised learning.
  - **Clustering** is the process to discover hidden patterns within our data that will let us identify groups, or clusters, within that data
  - This is known as unsupervised learning

Prediction or clustering? This is the dilemma

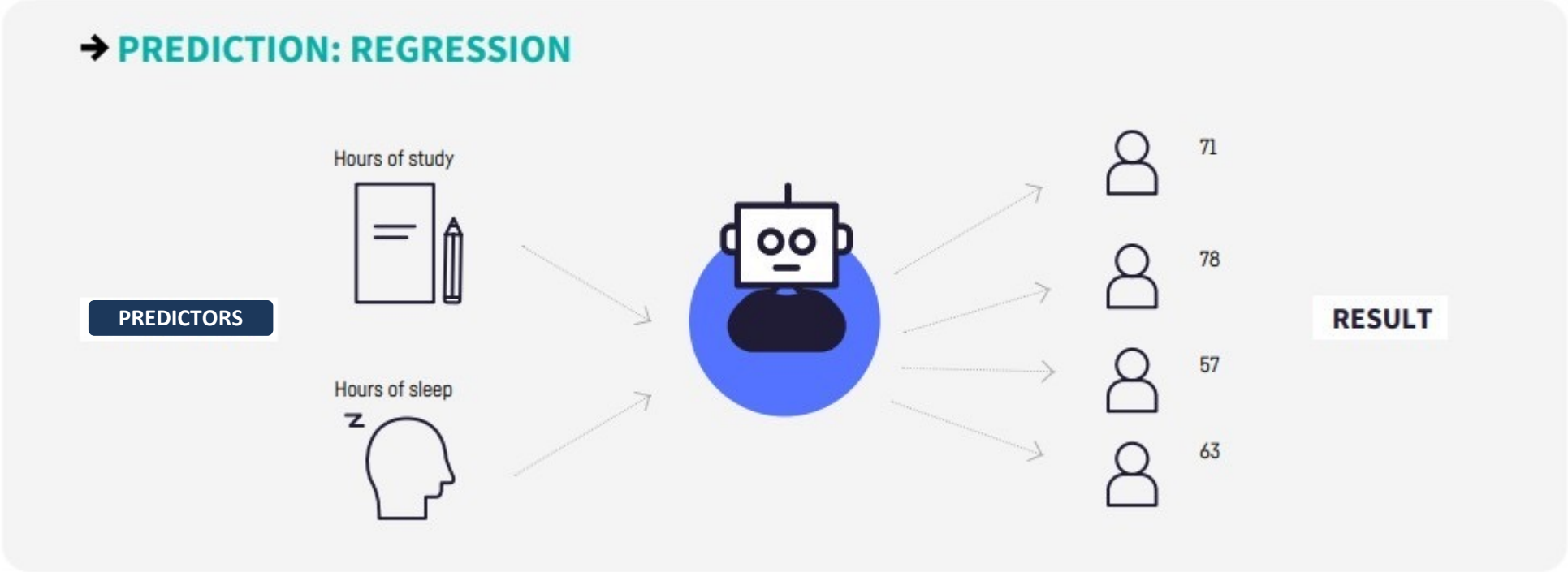


- **Predictive Method**: One application of a predictive method in a food and beverage company could be demand forecasting. By analyzing historical sales data and other relevant information, such as weather patterns and marketing campaigns, a predictive model can be built to forecast future demand for the company's products. This can help the company optimize production schedules, reduce waste, and make more informed decisions about inventory management and pricing.

- **Clustering Method**: One application of a clustering method in a food and beverage company could be customer segmentation. By analyzing customer data, such as demographics, purchase history, and preferences, a clustering algorithm can be used to group customers into similar segments. This information can then be used to target marketing and product offerings more effectively, and to better understand the needs and preferences of different customer segments

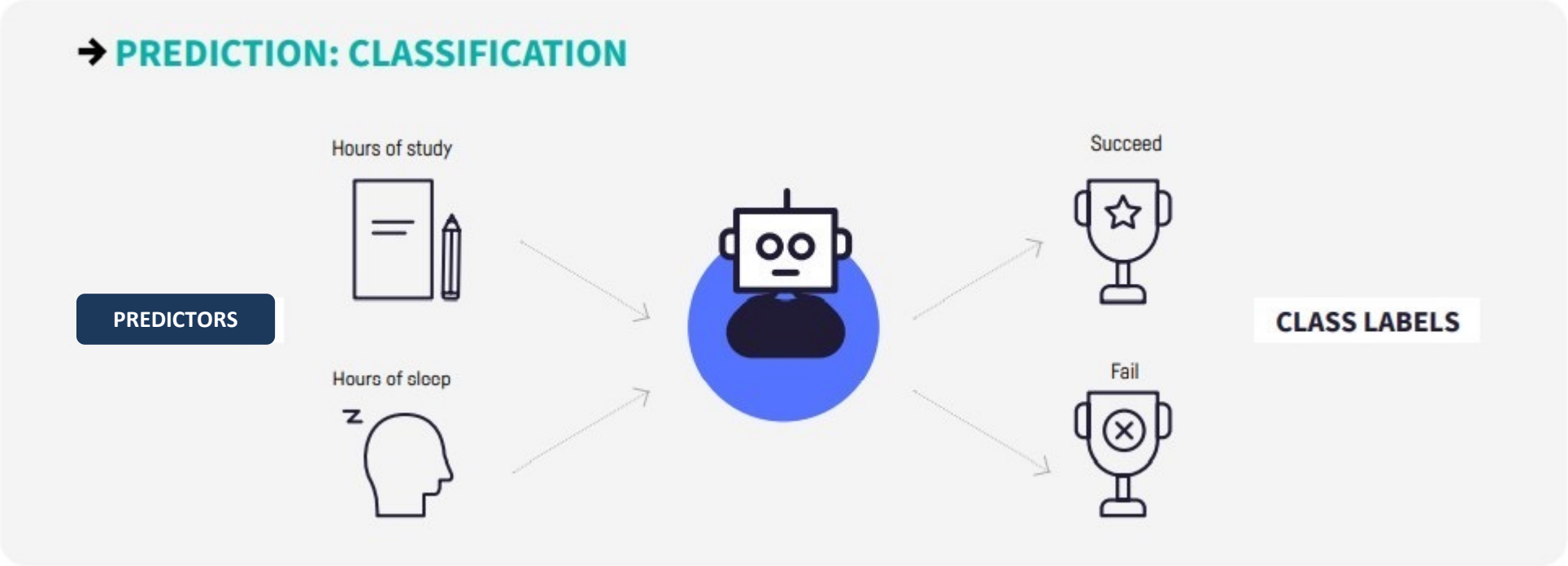
Prediction problems: Regression and classification

Regression problems;the variable to predict is numerical. For example: Understand how the number of sleep hours and study hours (the predictors) determine a students' test scores.



Prediction problems: Regression and classification

Classification problems: the variable to predict is part of one of some number of predefined categories, which can be as simple as "yes" or "no." For example: Understand how the number of sleep hours and study hours (the predictors) determine if a student will Succeed or Fail, where “Succeed” and “Fail” are two class labels



Prediction Algorithms: most common, pro & cons

The most prominent and common algorithms used in machine learning historically and today come in three groups: **linear models, tree-based models**, and **neural networks**.

- A linear model uses a simple formula to find the “best fit” line through a set of data points.
- Subtype: logistic Regression
- Old School Statistics.
- Simple: the variable you want to predict (the dependent variable) is represented as an equation of variables you know (independent variables)
- Can be simple or multiple. In multiple linear regression, the value of the target variable changes based on the value of more than one independent variable, or x.

- For example, you might want to know how long it will take to bake a cake, and your regression analysis might yield an equation t = 0.5x + 0.25y, where t is the baking time in hours, x is the weight of the cake batter in kg, and y is a variable which is 1 if it is chocolate and 0 if it is not. I

ESCP EUROPE BUSINESS SCHOOL

Prediction Algorithms: most common, pro & cons

The most prominent and common algorithms used in machine learning historically and today come in three groups: **linear models, tree-based models**, and **neural networks**.

- Tree-based = decision trees, i.e., a sequence of branching operations.
- Subtype: Random Forest, gradient Boosting
- A decision tree is a graph that uses a branching method to show each possible outcome of a decision.
- In machine learning, the branches used are binary yes/no answers
- Like if you're ordering a salad, you first decide the type of lettuce, then the toppings, then the dressing. We can represent all possible outcomes in a decision tree.

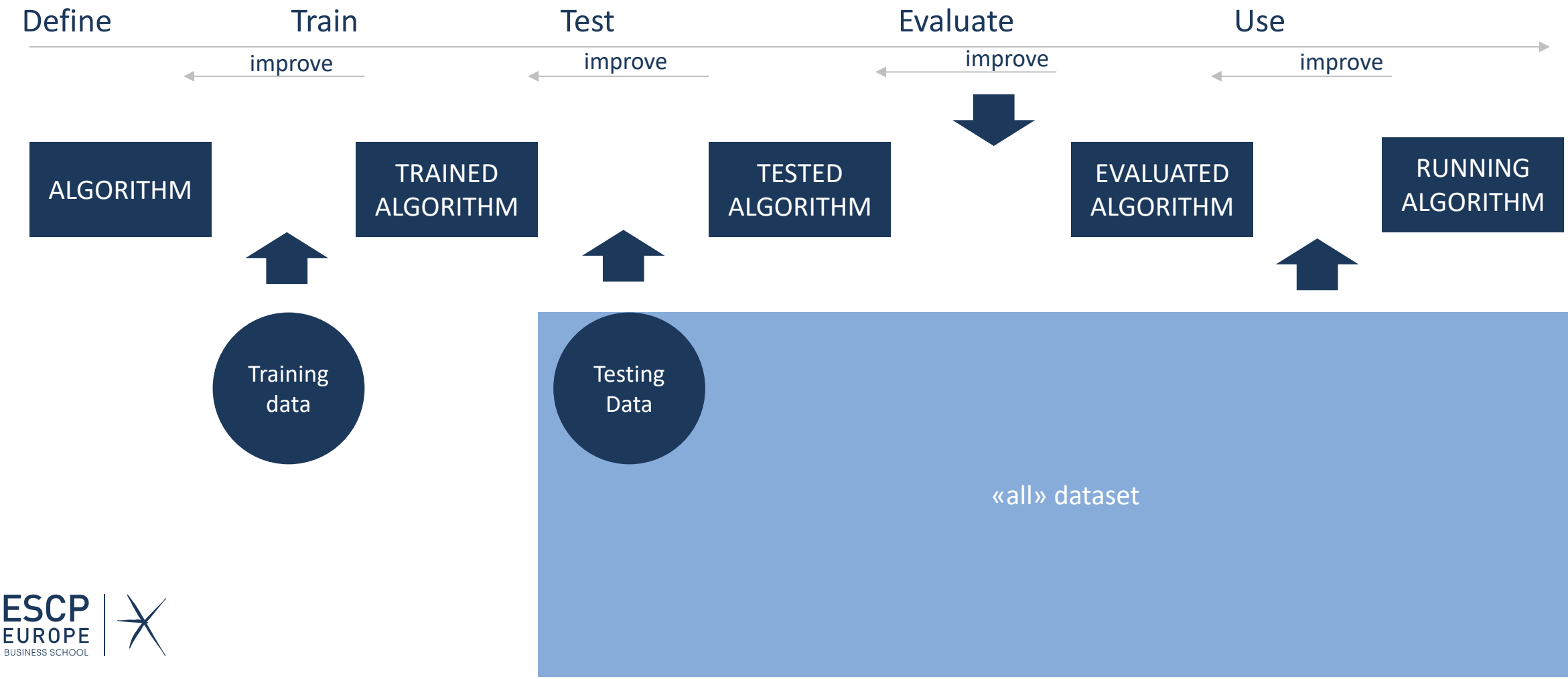
ESCP EUROPE BUSINESS SCHOOL

Prediction Algorithms: most common, pro & cons

The most prominent and common algorithms used in machine learning historically and today come in three groups: **linear models, tree-based models**, and **neural networks**.

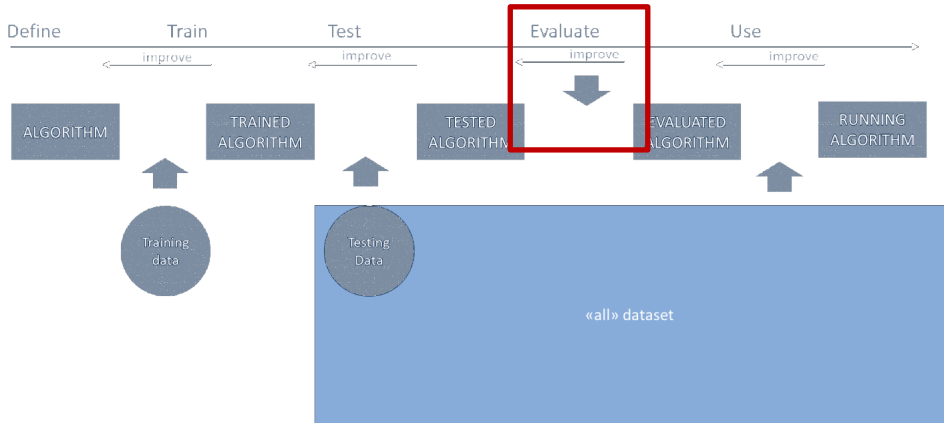
- Neural networks refer to a biological phenomenon comprised of interconnected neurons that exchange messages with each other.
- This idea has now been adapted to the world of machine learning and is called ANN (Artificial Neural Networks).
- Deep learning can be done with several layers of neural networks put one after the other. ANNs are a family of models that are taught to adopt cognitive skills

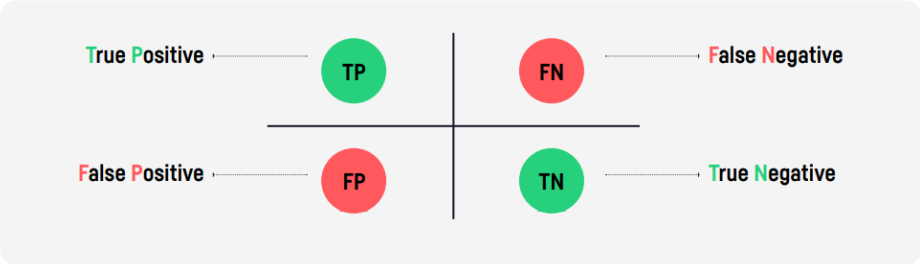
Train, test and evaluate an algorithm



ESCP EUROPE BUSINESS SCHOOL

Train, test and evaluate a prediction algorithm



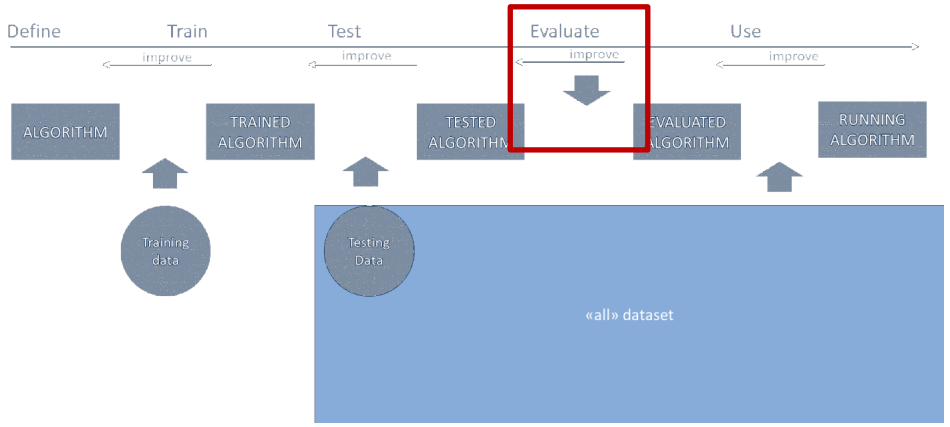


THE CONFUSION MATRIX

- Simple table known as a confusion matrix.
- The number of columns and rows in the table depends on the number of possible outcomes.
- The labels that make up a confusion matrix are TP, or true positive, FN, or false negative, FP, or false positive, and TN, or true negative.
- Alghorithms are never “perfects”
  - Type I error – FP
  - Type II error – FN
- Depending on our use case, we have to decide if we are more willing to accept higher numbers of type I or type II errors.

ESCP EUROPE BUSINESS SCHOOL

Train, test and evaluate a prediction algorithm

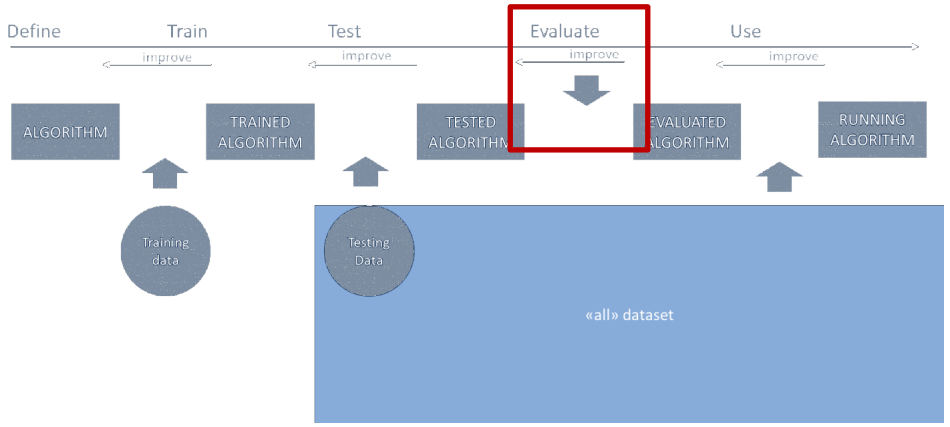


OVERFITTING AND REGULARIZATION

- During learning phase model also learns peculiarities of that data that don't have any predictive value.
- Those peculiarities could influence the prediction (Bias)
- The problem: overfitting → the remedy: regularization

ESCP EUROPE BUSINESS SCHOOL

Train, test and evaluate a prediction algorithm



REGULARIZATION

For decision tree models, regularization can be achieved through setting tree depth. A deep tree — that is, one with a lot of decision nodes — will be complex, and the deeper it is, the more complex it is. By limiting the depth of a tree, making it more shallow, we accept losing some accuracy, but it will be more general

For linear regression, regularization takes the form of L2 and L1 regularization.
y = C1a + C2b + C3c…, where the Cs are coefficients and a, b, and c are variables.
L2 regularization do reduce the magnitude of the coefficients, so that the impact of individual variables is somewhat dulled.
L1 regularization just eliminates a lot of these variables, working under the assumption that much of what they're capturing is just noise

ESCP EUROPE BUSINESS SCHOOL

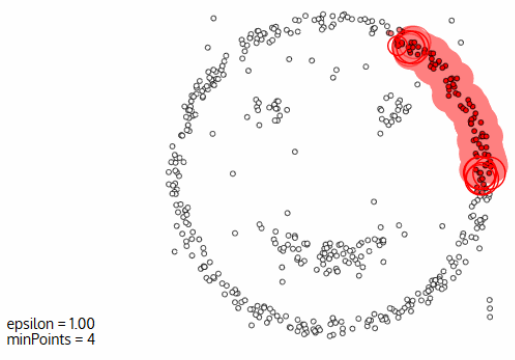
Clustering as a way to learn

Clustering is a type of unsupervised learning in machine learning, where the goal is to group similar data points together into clusters. The idea is to find a structure in the data, such that data points within the same cluster are more similar to each other than to data points in other clusters.

There are many different algorithms for clustering, each with its own strengths and weaknesses. Some popular clustering algorithms include:

- **DBSCAN** (Density-Based Spatial Clustering of Applications with Noise): This is a density-based clustering algorithm that groups data points into clusters based on their proximity to each other. It is particularly useful for finding clusters of arbitrary shapes.
- **K-Means**: This is a simple and fast algorithm that works by dividing the data into a specified number of clusters. The algorithm iteratively updates the centroid (representative point) of each cluster until the clusters are stable.
- **Hierarchical Clustering**: This is a type of clustering that builds a hierarchy of clusters, starting with individual data points and merging clusters until a single cluster remains. There are two main types of hierarchical clustering: Agglomerative Clustering and Divisive Clustering.

Clustering as a way to learn (DBSCAN)

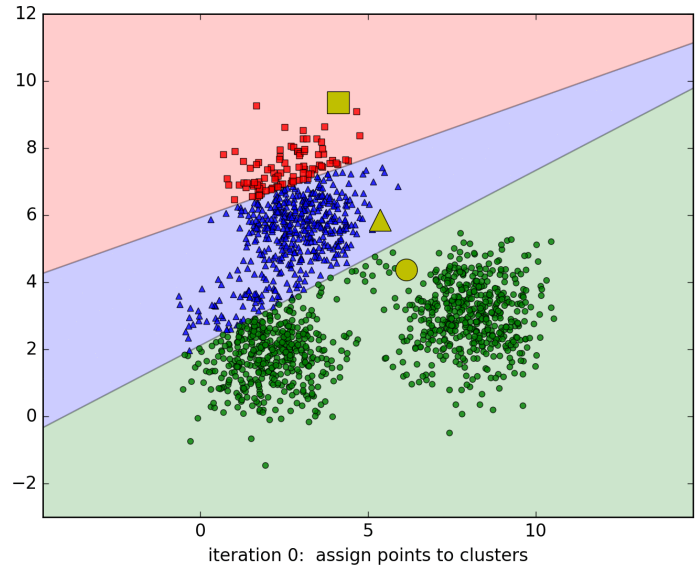


The DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm is a method for clustering data points in a dataset. The basic idea of DBSCAN is to identify groups of data points that are close together and separate from other groups.

- Choose a starting point and a distance threshold.
- Find all the data points within the distance threshold of the starting point. These points are considered to be in the same cluster.
- Repeat step 2 for all the data points in the cluster, expanding the cluster to include any points that are within the distance threshold of any point in the cluster.
- Repeat the process for each cluster until all data points have been assigned to a cluster.
- The main advantage of DBSCAN is that it can automatically identify clusters of different shapes and sizes, without having to specify the number of clusters in advance. This makes it useful for exploring datasets where the number and shape of the clusters is unknown.

In simple words, DBSCAN is a method for grouping data points into clusters based on their proximity to each other. The algorithm starts with a single data point and finds all other data points that are close by. These points are then grouped together into a cluster, and the process is repeated for each cluster until all data points have been assigned to a cluster.

Clustering as a way to learn (K-Mea



K-Means is a popular and widely used algorithm for clustering data in machine learning. The basic idea behind K-Means is to divide a dataset into a specified number of clusters (k) in a way that minimizes the sum of squared distances between data points and the centroid (representative point) of each cluster.

- Choose the number of clusters (k).
- Randomly select k initial centroids from the data.
- Assign each data point to the closest centroid.
- Compute the mean of the data points assigned to each centroid and use this as the new centroid for each cluster.
- Repeat steps 3 and 4 until the centroids no longer change or a maximum number of iterations is reached.

The K-Means algorithm is simple and easy to implement, but it has some limitations. For example, it assumes that clusters are spherical in shape and of similar size, which may not always be the case. Additionally, the initial choice of centroids can have a significant impact on the final clusters, so multiple runs of the algorithm with different initial centroids may be necessary to ensure that the best clusters are found.

ESCP EUROPE BUSINESS SCHOOL

Clustering as a way to learn (Hiearchical Clustering)

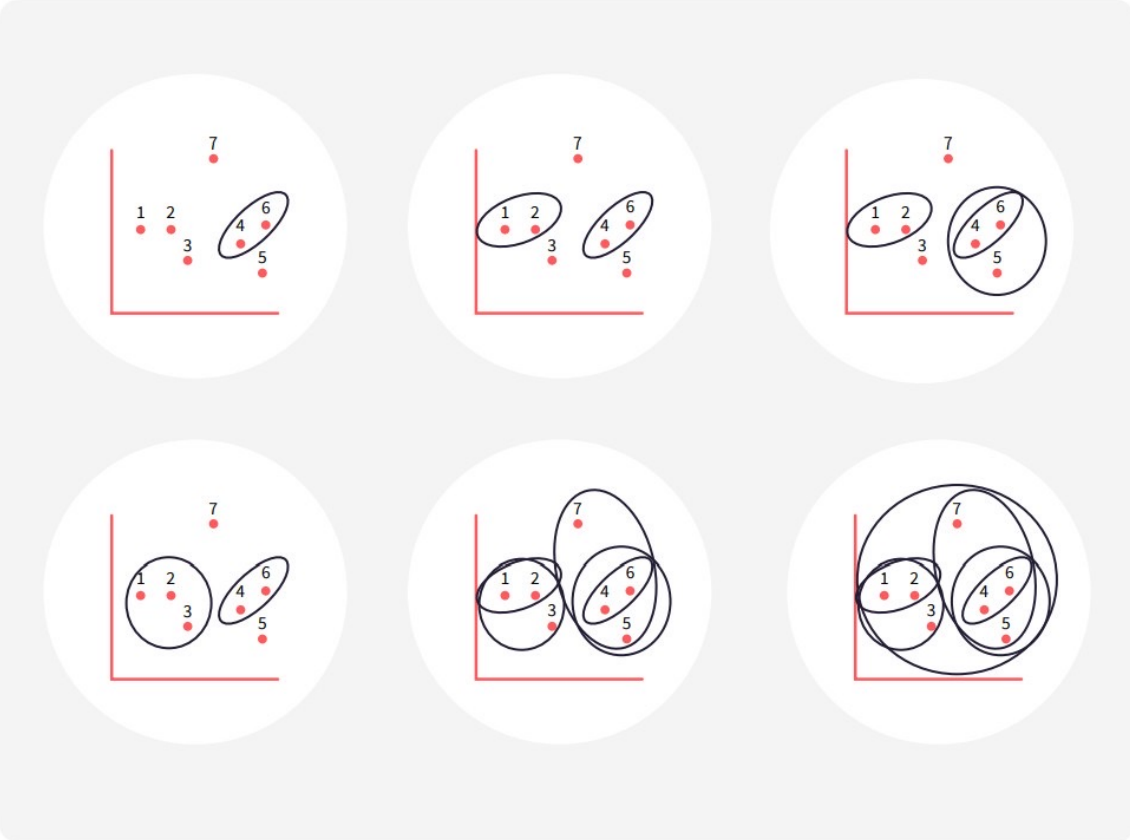
HIERARCHICAL CLUSTERING

Clusters are assigned based on hierarchical relationships between data points.

agglomerative (bottom-up)

divisive (top-down).

Agglomerative is more commonly used

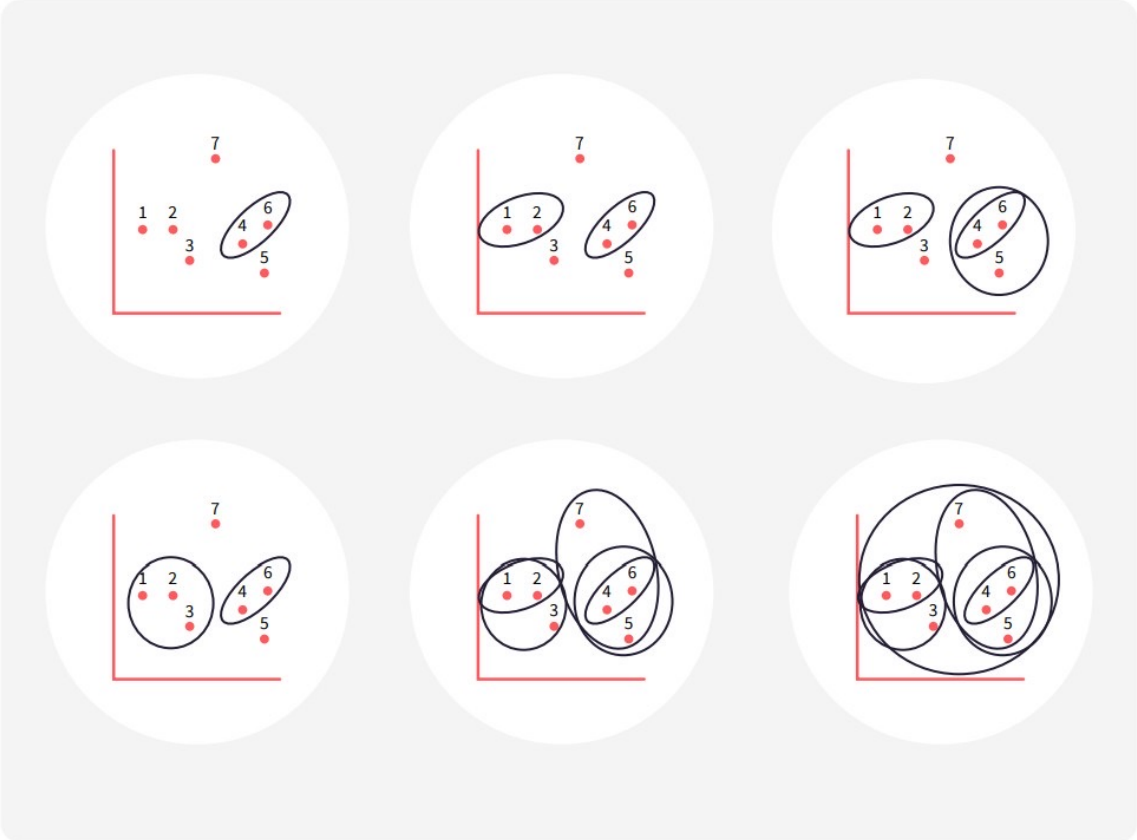


ESCP EUROPE BUSINESS SCHOOL

Clustering as a way to learn (Agglomerative Hierarchical)

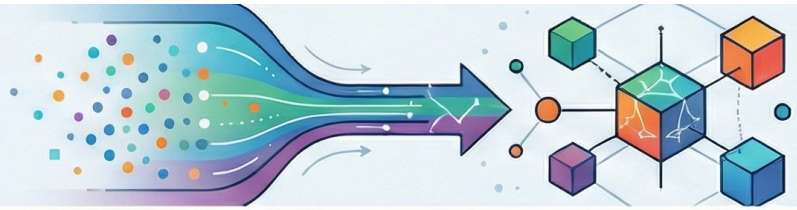
Start with each data point as its own cluster.
Compute the pairwise distances between all clusters and merge the two closest clusters into a single, larger cluster.
Repeat step 2 until all data points are part of a single, large cluster.
The result of the Agglomerative Hierarchical Clustering algorithm is a tree-like structure known as a dendrogram, which can be used to visualize the relationships between the clusters. By cutting the dendrogram at a certain height, it's possible to obtain a clustering of the data into a desired number of clusters.

One of the strengths of Agglomerative Hierarchical Clustering is that it can handle non-spherical clusters and clusters of different sizes. Additionally, it's a good choice when the number of clusters is not known in advance.



ESCP EUROPE BUSINESS SCHOOL

Decoding Machine Learning: A Business Primer

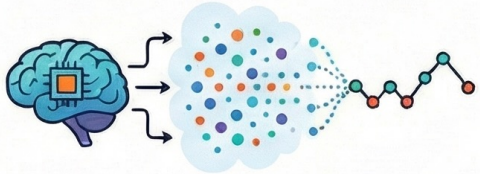


Foundations & Mechanisms

790%

Growth in Interest

Interest in ML has skyrocketed since 2013, moving from niche to mainstream business analysis.



Learning Without Explicit Programming

ML identifies patterns in data to make decisions and improve performance over time.

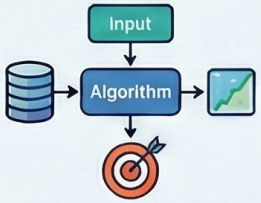
The Five-Stage Lifecycle

Models are developed through defining, training, testing, evaluating, and final deployment.

Strategic Methods & Business Impact

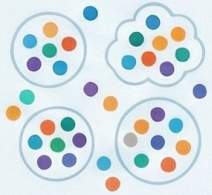
Prediction vs. Clustering

Supervised Learning (Prediction)



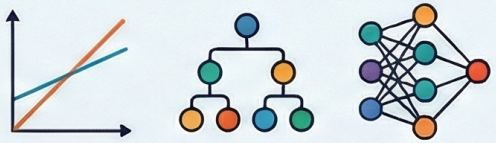
Supervised learning maps inputs to outputs.

Unsupervised Learning (Clustering)



while unsupervised learning discovers hidden groupings.

Three Core Algorithm Families



LINEAR MODELS | **DECISION TREES** | **NEURAL NETWORKS**

Most modern ML applications rely on Linear Models, Decision Trees, or Neural Networks.

Industry Leader ML Use Cases

Predictive analytics for demand forecasting and inventory management.

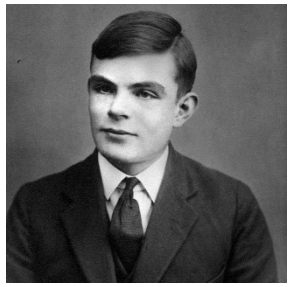
Forecasting product demand to optimize store staffing and reduce costs.

Improving supply chain efficiency and reducing waste via production optimization.

NotebookLM

ESCP EUROPE BUSINESS SCHOOL

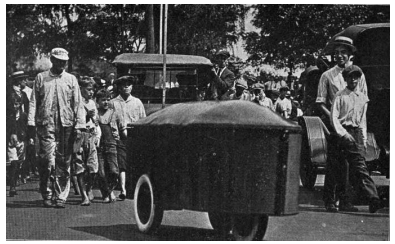
The Story of artificial intelligence



Alan Turing's "Computer Machinery and Intelligence" introduced The Turing Test for measuring computer intelligence, popularizing the term "artificial intelligence".

Birth of AI
1950-1956

The 1980s marked the "AI boom" characterized by rapid growth, fueled by breakthroughs and increased government funding.



Deep Learning and Expert Systems gained popularity, enabling computers to learn and make independent decisions.

AI Boom
1980-1987

Despite AI Winter's funding challenges, the early '90s saw significant AI advancements, including the first AI defeating a world chess champion. Innovations like Roomba and Windows speech recognition software also emerged.

AI Agents
1993 - 2011

1950s | 1960s | 1970s | 1980s | 1990s | 2000s | 2010s | 2020s | ?

1957-1979
AI Maturation

The period from AI's inception to the 1980s was marked by rapid growth and challenges. The late 1950s to the 1960s witnessed creation, while the 1970s saw advancements like Japan's anthropomorphic robot. However, funding declined in the 1970s.

1987-1993
AI Winter

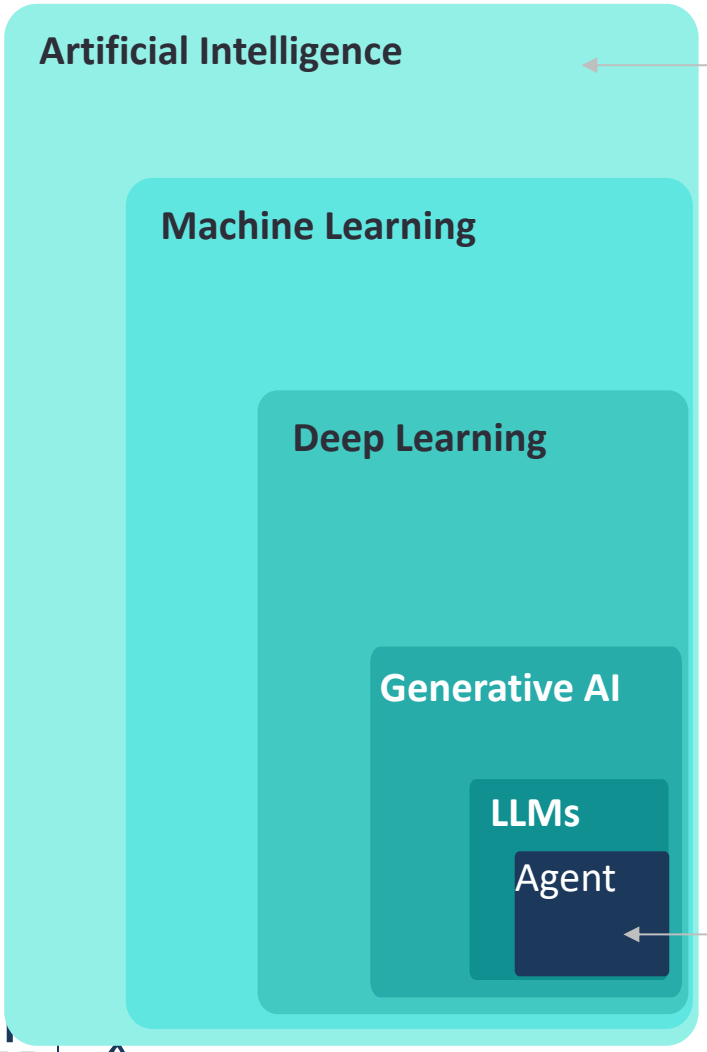
An AI Winter ensued, marked by low interest, reduced funding, and limited breakthroughs. This stemmed from setbacks in the machine market, expert systems, and project cancellations, leading to diminished support due to cost concerns.

2012 - Present
Artificial General Intelligence

OpenAI

Recent AI developments brought common-use tools and popularized Deep Learning and Big Data.
The future promises increased AI adoption by businesses, workforce changes due to automation, and advancements like robotics and autonomous vehicles.

The Story of artificial intelligence



What AI is …

1943

1959

2006

2017

What People think it is …

2025

Some words on a.i. (not on A.I.)

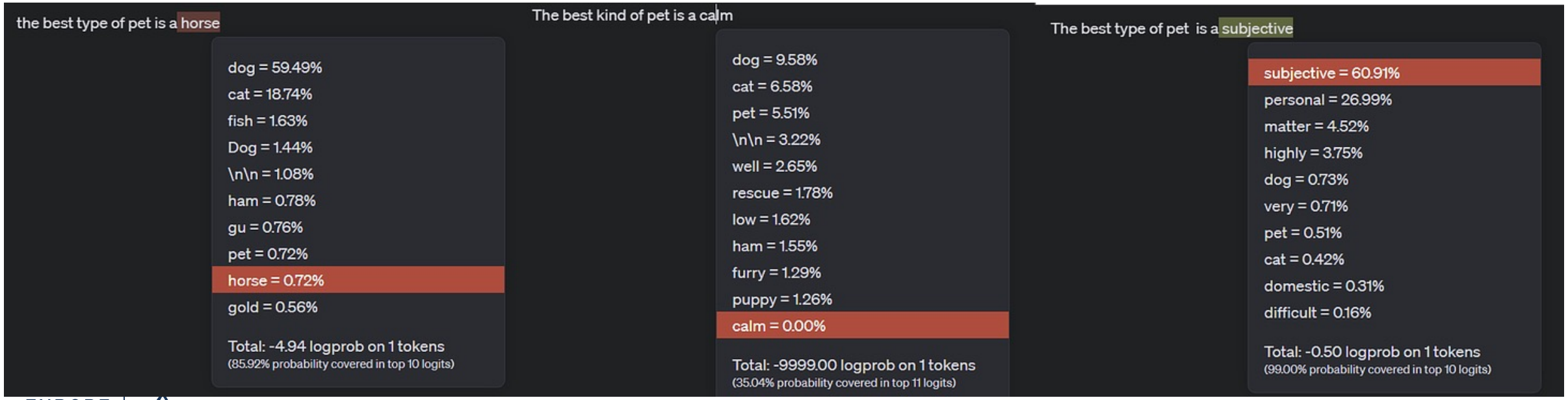
- A.I. is (still) science fiction
- Artificial + Intelligence seems to be a “marriage” instead is a DIVORCE (ref. Prof. Luciano Floridi)
- Divorce between ability to perform successfully a task and the need of being intelligent to success in doing it
- Intelligence of Agent is more “close” to the reality
- Envelope as a Key for the success
- In absence of a specific vocabulary we use terms that are far from the “reality” – Computer vision, Understanding, Decision, Learning … ant this could drive confusions

How a GenAI tool works

Tokenization

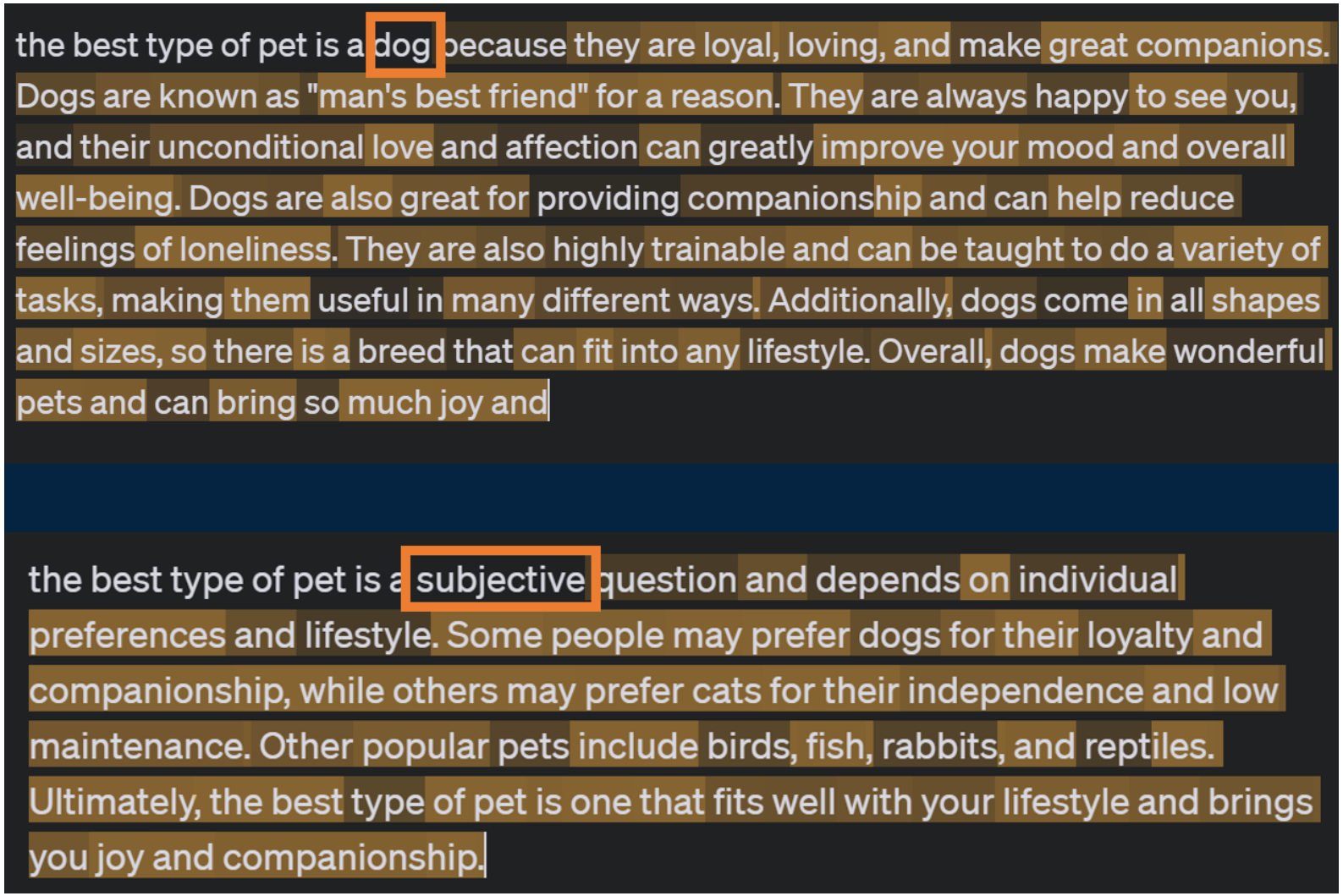
This | breaks | up | words | (| even | | phan | t | a | s | mag | or | ically | long | words |) | into | token | s

Prediction



How a GenAI tool works

Contextualization



ESCP EUROPE BUSINESS SCHOOL

Prompt engineering

Prompt engineering involves designing or crafting prompts in a way that guides or influences the output of a language model like ChatGPT, often to achieve a specific goal or outcome.

A good prompt provides clear instructions and focuses on a specific topic. Prompt engineering techniques involve refining the prompt to elicit desired responses from the AI model, such as **adding specific details**, **including constraints**, **specifying tone and style**, and **framing prompts as questions**.

Example:

- Initial prompt: *"Write a story about a chef"*
- After prompt engineering: *"Tell a story about a young chef named Mia, who dreams of opening her own bakery specializing in artisanal pastries. Describe Mia's journey as she navigates through culinary school, apprenticeships, and challenges like limited funds and skeptical critics. Capture the excitement of Mia's culinary experiments and the warmth of her interactions with supportive mentors and loyal customers. Write in a lighthearted and inspiring tone, celebrating Mia's passion for baking and her determination to turn her dream into reality. What obstacles does Mia encounter along the way, and how does she overcome them to achieve her goal of owning a successful bakery?"*

ChatGPT

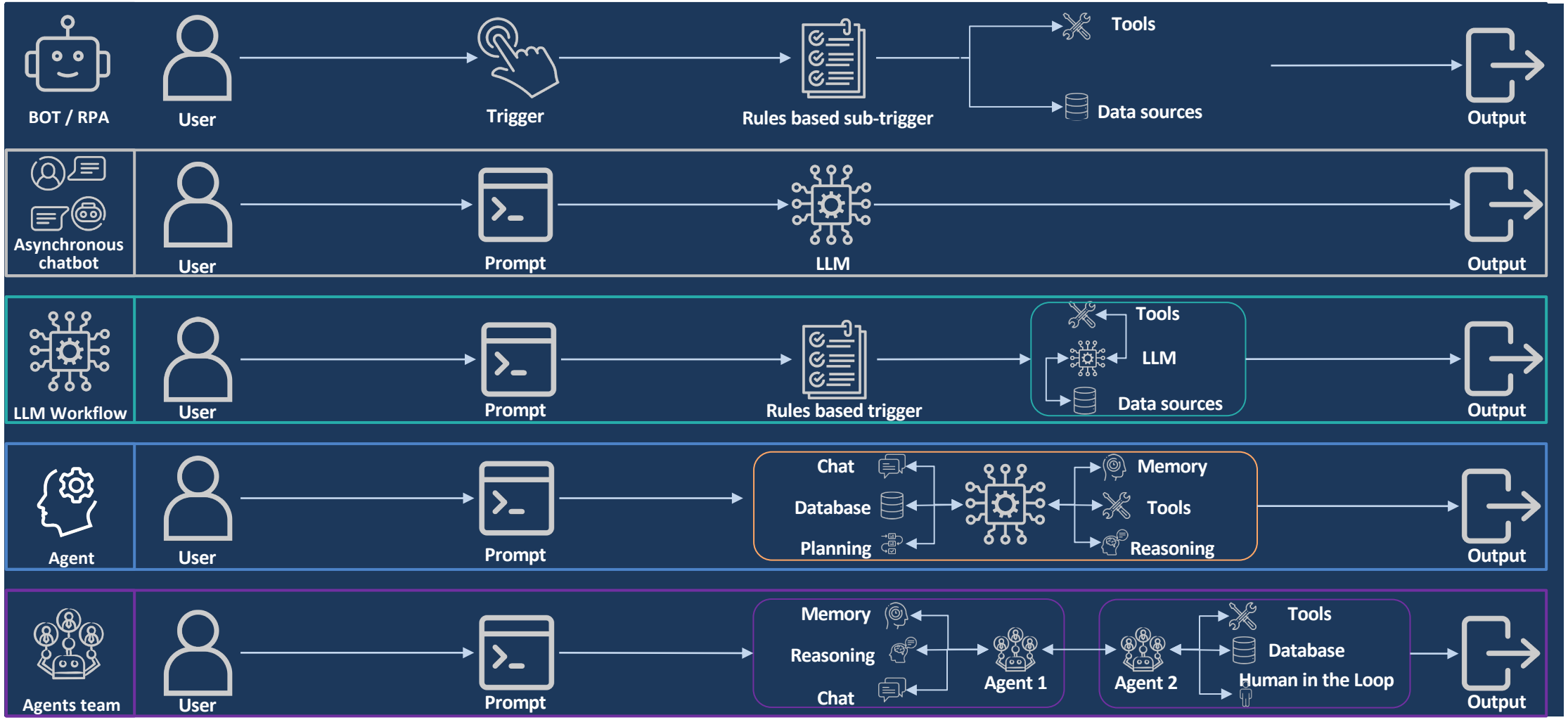
Midjourney

Bard

Microsoft 365 Copilot

ESCP EUROPE BUSINESS SCHOOL

From Bot to AGENT



ESCP EUROPE BUSINESS SCHOOL

A(I)DOPTION

The proposed approach to guide the adoption and evolution process of AI within a company is structured into four main phases. These phases are not strictly sequential but intersect throughout the journey.

- Construction of the manifesto and definition of the Target Operating Model
  Planning and execution of training and adoption activities

2. UPSKILLING – RESKILLING - ADOPTION

AI ROADMAP – THE APPROACH

1. AI MATURITY MODEL

Cross-functional assessment of the level of maturity in the adoption of AI solutions, conducted through the analysis of specific key architectural dimension

3. DISCOVERY USE CASE

- Analysis of pain points and optimization opportunities / improvement
- Identification of Use Cases and quick-win solutions (effort and timing)

4. PILOT & TRANSFORM

- Support for the implementation of pilot solutions and measurement of results
- Definition of the transformation plan and support for execution
- Definition of the governance model to support the evolutionary plan

ESCP EUROPE BUSINESS SCHOOL

Once that correlations have been found .. show it out