Machine Learning

Salvador Ruiz Correa

September 30, 2025

MACHINE LEARNING is a data-driven approach to building predictive models and intelligent systems, enabling computers to learn patterns from experience rather than relying on explicit programming. This lecture introduces the foundational concepts of machine learning and its four primary paradigms: supervised learning, unsupervised learning, self-supervised learning, and reinforcement learning. Each paradigm addresses different types of learning tasks, from classification and regression to pattern discovery and decision-making under uncertainty. The lecture also highlights the role of deep learning as a subfield of machine learning, which is particularly effective for handling high-dimensional and unstructured data. By exploring these frameworks, participants gain a conceptual understanding of how machine learning algorithms operate, adapt, and support real-world applications across various domains, including public health, robotics, and natural language processing.

AGENDA:

Machine Learning

Machine learning is a powerful and rapidly evolving field that enables computers to *learn from data and make predictions or decisions without being explicitly programmed for every task*. Instead of writing rules by hand, we allow algorithms to discover patterns and relationships by analyzing examples. This data-driven approach has revolutionized industries ranging from healthcare and finance to transportation and public health, where predictive models can anticipate disease outbreaks or optimize resource allocation.

There are several *distinct paradigms* within machine learning, each suited to different types of problems and data structures. The most commonly used is supervised learning. In this approach, the model is trained on labeled data, where each input is paired with a known output. For example, a model might learn to predict dengue incidence based on past case counts and weather conditions. Supervised learning is ideal for tasks such as classification (e.g., diagnosing diseases) and regression (e.g., forecasting case numbers), where the goal is to learn a mapping from inputs to outputs.

Unsupervised learning, by contrast, deals with unlabeled data. Here, the algorithm seeks to uncover hidden patterns or groupings within the dataset. It's often used for clustering, anomaly detection, or dimensionality reduction. For instance, unsupervised learning may reveal clusters of municipalities with similar transmission dynamics, enabling epidemiologists to target interventions more effectively.

This paradigm is exploratory and helpful in discovering structure in complex datasets. Recently, Ilya Sutskever (cofounder of OpenAI) proposed some ideas concerning generalization in the context of unsupervised learning. These ideas rely on concepts from complexity theory (https://simons.berkeley.edu/talks/ilya-sutskever-openai-2023-08-14).

Self-supervised learning is a relatively new and increasingly influential paradigm, particularly in the field of deep learning. It involves generating labels from the data itself, allowing models to learn useful representations without external supervision. A typical example is predicting the next word in a sentence or reconstructing missing values in a time series. Self-supervised learning is often employed to pretrain models on large, unlabeled datasets before fine-tuning them on specific tasks, thereby improving performance and data efficiency.

Finally, *reinforcement learning* is about learning through interaction. The model operates in an environment, makes decisions, receives feedback in the form of rewards or penalties, and adjusts its strategy to maximize long-term reward. This approach is used in robotics, game playing, and adaptive control systems. In public health, reinforcement learning could be applied to optimize mosquito control strategies or allocate resources dynamically based on changing risk levels.

At its core, machine learning encompasses a family of techniques designed to extract insights from data. One of its most prominent subfields is *deep learning, which focuses on neural networks with many layers and is particularly effective for tasks involving images, speech, and complex time series*. Deep learning has driven significant breakthroughs in natural language processing, computer vision, and biomedical diagnostics, yet it remains an integral part of the broader machine learning landscape.

Together, these paradigms—and the deep learning techniques that often enhance them—form the foundation of modern artificial intelligence. Understanding their differences and strengths is essential for designing practical solutions to real-world problems, whether you're building a disease forecasting platform or training autonomous systems.

Supervised Learning

Supervised learning is a foundational paradigm in machine learning where models are trained on labeled datasets—that is, data where each input is paired with a known output. The goal is to learn a mapping function that can accurately predict outputs for new, unseen inputs. This approach is efficient for tasks such as classification (e.g., diagnosing whether a patient has dengue based on symptoms) and regression (e.g., forecasting the number of cases in a municipality). During training, the model iteratively adjusts its internal parameters to minimize the difference between its predictions and the actual labels, typically using metrics such as mean squared error or cross-entropy loss. One of the strengths of supervised learning is its ability to generalize from historical patterns, provided the training data is representative and sufficiently diverse. However, its performance heavily depends on the quality and quantity of labeled data, making data curation and preprocessing critical steps. In public health applications, supervised learning enables the development of early warning systems, risk stratification, and resource allocation by leveraging historical surveillance data to anticipate future trends with precision and actionable insights.

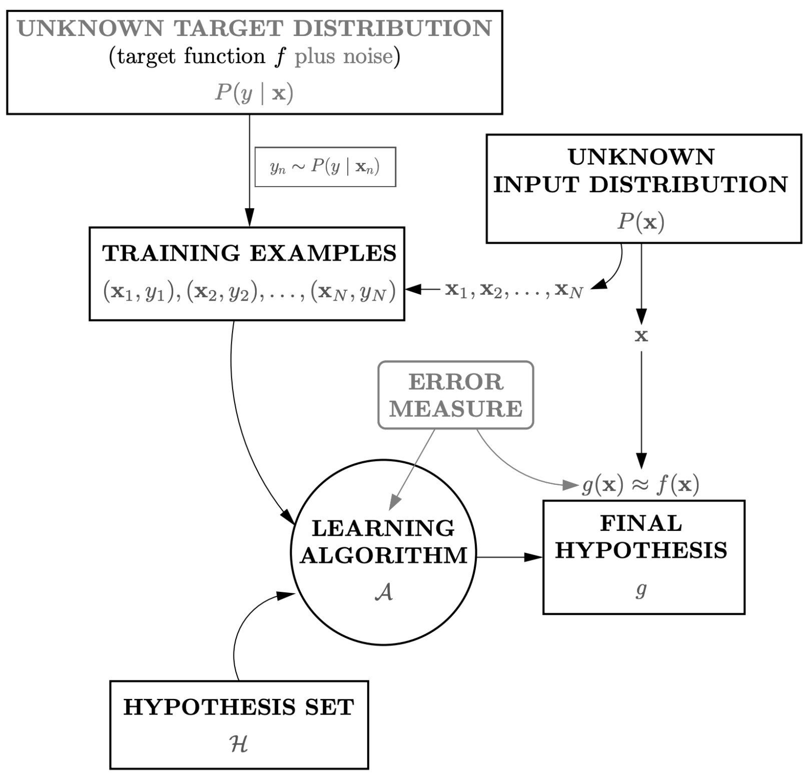


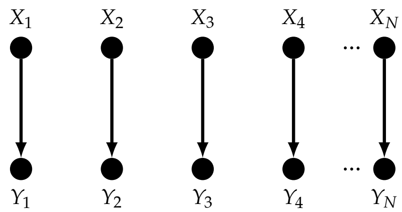
Figure 1: Statistical learning framework. From: https://www.cs.rpi.edu/ magdon/courses/learn/slides.html

Supervised learning is based on the framework summarized in Figure 1. A key element of the framework is the joint probability distribution function

$$P(\mathbf{x}_n, y) = P(y \mid \mathbf{x}_n) p(\mathbf{x}_n),$$

$n = 1, \ldots, N$. This function is called the *generating probability distri-*

bution function. A key assumption in statistical learning from data is that the *training examples are independent and identically distributed according to the joint probability generating function*. The corresponding graphical model is the following.



Box 1: Supervised Learning Techniques

- *Linear Regression*: Predicts continuous outcomes using a linear combination of input features.
- *k-Nearest Neighbors (k-NN)*: Instance-based method that classifies based on the majority label among the k closest training examples.
- *Naive Bayes*: Probabilistic classifier based on Bayes' theorem, assuming feature independence. Often used in text classification and spam detection.
- *Logistic Regression*: Used for binary classification problems, such as disease presence vs. absence.
- *Decision Trees*: Simple, interpretable models that split data based on feature thresholds.
- *Support Vector Machines (SVM)*: Effective for classification tasks, especially with high-dimensional data. Can use linear or non-linear kernels such as radial basis function (RBF) or polynomial kernels.
- *Gaussian Process:* A Gaussian Process (GP) in machine learning is a powerful, non-parametric Bayesian approach used primarily for regression and probabilistic modeling. Instead of assuming a fixed functional form (like a linear or polynomial model), a GP defines a distribution over possible functions that fit the data, allowing for flexible, uncertainty-aware predictions.
- *Feedforward Neural Networks (FNN)*: Basic architecture for regression and classification tasks.
- *Convolutional Neural Networks (CNN)*: Specialized for image and spatial data.
- *Recurrent Neural Networks (RNN)*: Designed for sequential data such as time series or text.
- *Bagging*: Combines multiple models trained on different subsets of the data (e.g., Random Forest).
- *Boosting*: Builds models sequentially to reduce bias (e.g., AdaBoost, Gradient Boosting).
- *Stacking*: Combines predictions from multiple models using a meta-model.
- *Transformers*: Another powerful supervised learning architecture is the Transformer, originally introduced for natural language processing but now widely applied across domains including time series forecasting, image classification, and genomics. Transformers use self-attention mechanisms to model relationships between elements in a sequence, allowing them to capture long-range dependencies more effectively than traditional recurrent networks.

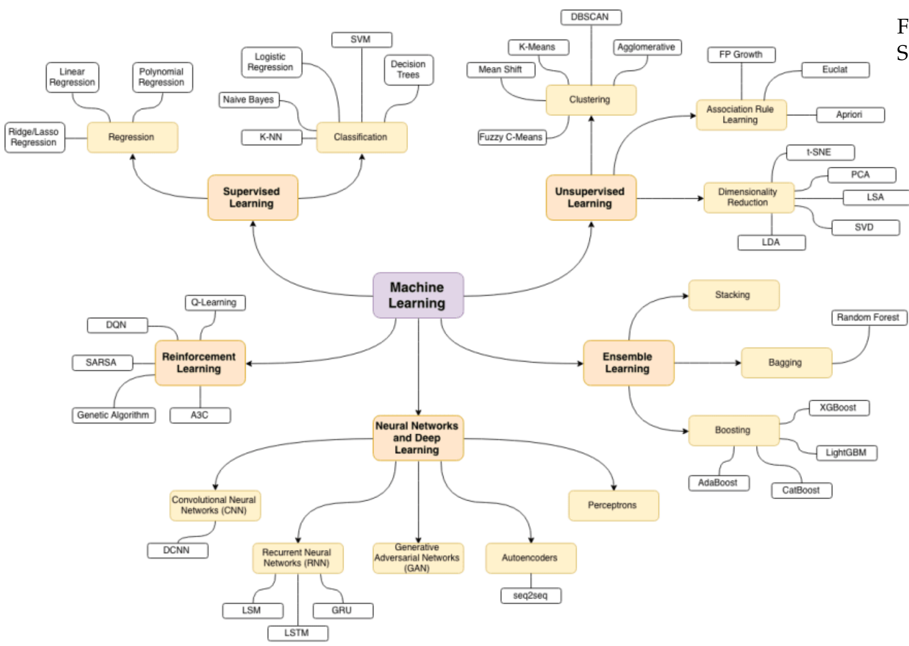


Figure 2: Machine learning techniques. Source: https://freezealar.com

Artificial Intelligence

The Birth of Artificial Intelligence (1956)

The field of Artificial Intelligence was born during a two-month workshop that took place in *Dartmouth College* in the summer of 1956. The workshop was organized by Jhon McCarthy (*Dartmouth College*) who convinced Marvin Minsky (*Harvard University*), Claude Shannon (*Bell Telephone Laboratories*), and Nathaniel Rochester (*I.B.M. Corporation*) to help him bring together U.S. researchers interested in automata theory, neural networks, and the study of intelligence. The original workshop proposal, written by McCarthy and colleagues, states:

Figure 3: Dartmouth College workshop (1956) From left to right: Oliver G. Selfridge, Nataniel Rochester, Ray Solmonoff, Marvin Minsky, Trenchard Moore, John McCarthy, Claude E. Shanon.

> We propose that a 2-month, 10-man study of artificial intelligence be carried out during the summer of 1956 at *Dartmouth College* in Hanover, New Hampshire. The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.

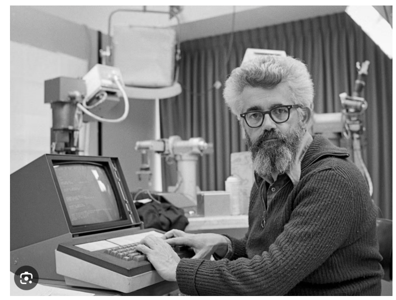


Figure 4: John McCarthy (1927-2011) was an American computer scientist and cognitive scientist. He was one of the founders of the discipline of artificial intelligence. He co-authored the document that coined the term "artificial intelligence" (AI), developed the programming language family Lisp, significantly influenced the design of the language ALGOL, and popularized computer time-sharing.

° McCarthy, J., Minsky, M. L., Rochester, N., and Shannon, C. E. (1955). Proposal for the Dartmouth summer research project on artificial intelligence. Tech. rep., Dartmouth College.

There were 10 attendees in all, including Trenchard More from *Princeton*, Arthur Samuel from *IBM*, Ray Solomonoff and Oliver Selfridge from the *MIT*, and Allen Newell and Herbert Simon from *CMU*. Arthur Samuel coined the term *machine learning* (ML) or *self-teaching computers*) in 1959.

Artificial Intelligence as a separate field

Looking at the proposal for the Dartmouth workshop (McCarthy et al., 1955), we can see why it was necessary for AI to become a separate field.

1. *Artificial Intelligence* from the start embraced the idea of duplicating human faculties such as creativity, self-improvement, and language use. None of the other fields such as mathematics, control theory, operations research or decision address these issues.
2. *Artificial Intelligence* is the only one of these fields that is clearly a branch of computer science (although operations research does share an emphasis on computer simulations).
3. *Artificial Intelligence* is the only field to attempt to build machines that will function autonomously in complex, changing environments.

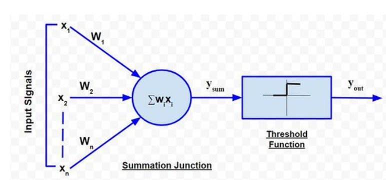


Figure 5: Warren McCulloch and Walter Pitts model (1943). https://www.geeksforgeeks.org/implementing-models-of-artificial-neural-network/

Artifficial Intelligence Definition

Informally speaking, Artificial Intelligence refers to the simulation of human intelligence in machines or computer systems. It involves the development of algorithms, software, and hardware that enable machines to perform tasks that typically require human intelligence. These tasks include reasoning, problem-solving, learning from experience, understanding natural language, and perceiving and interacting with the environment.

Artificial Intelligence historically evolved from disciplines that contributed ideas, viewpoints, and techniques. These include philosophy, mathematics, economics, neuroscience, psychology, computer engineering, control theory, cybernetics, and linguistics.

The first work that is now generally recognized as *Artificial Intelligence* was done by Warren McCulloch and Walter Pitts (1943). They drew on three sources: knowledge of the basic physiology and function of neurons in the brain; formal analysis of propositional logic due to Russell and Whitehead; and Turing's theory of computation. They proposed a model of artificial neurons in which each neuron is characterized as being "on" or "off," with a switch to "on" occurring in response to stimulation by a sufficient number of neighboring neurons. The state of a neuron was conceived of as "factually equivalent to a proposition which proposed its adequate stimulus." They showed, for example, that any computable function could be computed by some network of connected neurons, and that all the logical connectives could be implemented by simple net structures. McCulloch and Pitts also suggested that suitably defined networks could learn. Donald Hebb (1949) demonstrated a simple updating rule for modifying the connection strengths between neurons. His rule, now called Hebbian learning, remains an influential model to this day.

John McCarthy coined the term *Artificial Intelligence* as: "the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to biologically observable methods.

Artificial Intelligence can also be defined according to several criteria, namely thought processes, reasoning, and behavior: thinking humanly, acting humanly, thinking rationally and acting rationally.

1. **Acting humanly, The Turing approach.** The Turing Test, proposed by Alan Turing (1950), was designed to provide a satisfactory operational definition of intelligence. A computer passes the test if a human interrogator, after posing some written questions, cannot tell whether the written responses come from a person or from

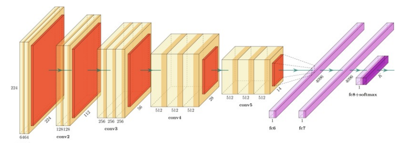


Figure 6: *Artificial Intelligence* refers to the simulation of human intelligence in machines or computer systems. It involves the development of algorithms, software, and hardware that enable machines to perform tasks that typically require human intelligence used in *AI* computer vision applications. The figure shows a convolutional neural network architecture. https://doi.org/10.1016/j.ejmp.2017.05.071

Figure 7: "Alan Mathison Turing (1912-1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist. Turing was highly influential in the development of theoretical computer science, providing a formalization of the concepts of algorithm and computation with the Turing machine, which can be considered a model of a general-purpose computer. He is widely considered to be the father of theoretical computer science and artificial intelligence.

a computer. Turing's test deliberately avoided direct physical interaction between the interrogator and the computer, because physical simulation of a person is unnecessary for intelligence. However, the so-called total Turing Test includes a video signal so that the interrogator can test the subject's perceptual abilities, as well as the opportunity for the interrogator to pass physical objects "through the hatch." To pass the total Turing Test, the computer will need to perceive objects, manipulate objects, and move about (Russel and Norvig, 2010). The following capabilities are required:

- Natural language processing (NLP) to enable it to communicate successfully in English.
- Knowledge representation to store what it knows or hears.
- Automated reasoning to use the stored information to answer questions and to draw new conclusions.
- Machine learning to adapt to new circumstances and to detect and extrapolate patterns.
- Computer vision, to perceive objects.
- Robotics, to manipulate objects and move about.

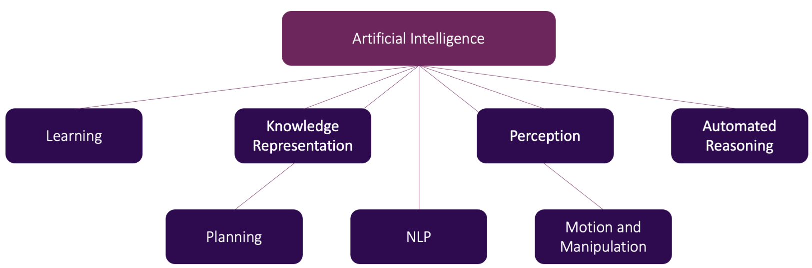


Figure 8: Acting humanly. Several computer capabilities are required to pass the Turing Test: natural language processing, knowledge representation, automated reasoning, and machine learning to adapt to new circumstances and to detect and extrapolate patterns.

2 **Thinking humanly.** If we are going to say that a given program thinks like a human, we must have some way of determining how humans think. We need to get inside the actual workings of human minds. There are three ways to do this: through introspection—trying to catch our own thoughts as they go by; through psychological experiments—observing a person in action; and through brain imaging—observing the brain in action. Once we have a sufficiently precise theory of the mind, it becomes possible to express the theory as a computer program. If the program's input–output behavior matches corresponding human behavior, that is evidence that some of the program's mechanisms could also be operating in humans (Russel and Norvig, 2010).
The interdisciplinary field of cognitive science brings together computer models from AI and experimental techniques from

Figure 9: Avram Noam Chomsky (1928) is an American professor and public intellectual known for his work in linguistics, political activism, and social criticism. Sometimes called "the father of modern linguistics", Chomsky is also a major figure in analytic philosophy and one of the founders of the field of cognitive science.

psychology to construct precise and testable theories of the human mind. The field can be said to have started at a workshop in September 1956 at MIT. At the workshop, George Miller presented The Magic Number Seven, Noam Chomsky presented Three Models of Language, and Allen Newell and Herbert Simon presented The Logic Theory Machine. These three influential papers showed how computer models could be used to address the psychology of memory, language, and logical thinking, respectively.

3 **Thinking rationally.** Logicians in the 19th century developed a precise notation for statements about all kinds of objects in the world and the relations among them. By 1965, programs existed that could, in principle, solve any solvable problem described in logical notation. (Although if no solution exists, the program might loop forever.) The so-called logicist tradition within artificial intelligence hopes to build on such programs to create intelligent systems. There are two main obstacles to this approach. First, it is not easy to take informal knowledge and state it in the formal terms required by logical notation, particularly when the knowledge is less than 100% certain. Second, there is a big difference between solving a problem "in principle" and solving it in practice.

4 **Acting rationally: The rational agent approach.** An agent is just something that acts (agent comes from the Latin agere, to do). Of course, all computer programs do something, but computer agents are expected to do more: operate autonomously, perceive their environment, persist over a prolonged time period, adapt to change, and create and pursue goals. A rational agent is one that acts so as to achieve the best outcome or when there is uncertainty, the best expected outcome.

The rational-agent approach has two advantages over the other approaches. First, it is more general than the "laws of thought" approach because correct inference is just one of several possible mechanisms for achieving rationality. Second, it is more amenable to scientific development than approaches based on human behavior or human thought. The standard of rationality is mathematically well defined and completely general and can be "unpacked" to generate agent designs that provably achieve it. Human behavior, on the other hand, is well adapted to one specific environment and is defined by, well, the sum total of all the things that humans do.

An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators. This simple idea is illustrated in Figure 2.1. A human agent has eyes, ears, and other organs for sensors and hands, legs,

Figure 11: In artificial intelligence, an expert system is a computer system emulating the decision-making ability of a human expert. Expert systems are designed to solve complex problems by reasoning through bodies of knowledge, represented mainly as if-then rules rather than through conventional procedural code. The first expert systems were created in the 1970s and then proliferated in the 1980s.

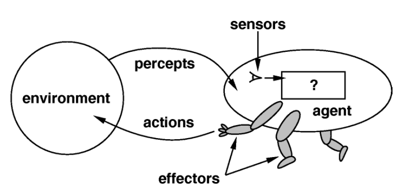


Figure 12: An agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators.

vocal tract, and so on for actuators. A robotic agent might have cameras and infrared range finders for sensors and various motors for actuators. A software agent receives keystrokes, file contents, and network packets as sensory inputs and acts on the environment by displaying on the screen, writing files, and sending network packets.

Figure 13: Tesla cars come standard with advanced hardware capable of providing Autopilot features, and full self-driving capability through software updates designed to improve functionality over time. https://www.tesla.com/es_PR/autopilot.

Causal Theory (Perl, 2006)

Judea Pearl's causal theory, often referred to as the "Pearlian causal framework," is a significant contribution to the field of causal inference and artificial intelligence. It seeks to understand and model cause-and-effect relationships in data and complex systems. Here's a summarized overview of Judea Pearl's causal theory:

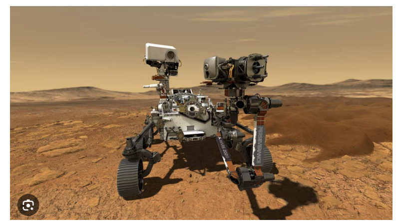


Figure 14: NASA's Perseverance Mars rover used an artificial intelligence software called Autonomous Exploration for Gathering Increased Science (AEGIS) to select and target the rock seen in close-up here. It's one of two rocks that the AI for the first time helped Perseverance study without direction from the mission's team back on Earth.

The Ladder of Causality

Pearl's ladder of causality is a graphical representation introduced by Judea Pearl, a prominent computer scientist and statistician, in his work on causal inference. This ladder is a conceptual framework that helps illustrate the different levels of causation and the types of questions we can ask about causal relationships. It provides a systematic approach to thinking about causality and inferring causation from data.

Pearl's ladder of causality consists of four rungs, representing progressively more complex and sophisticated levels of causal understanding:

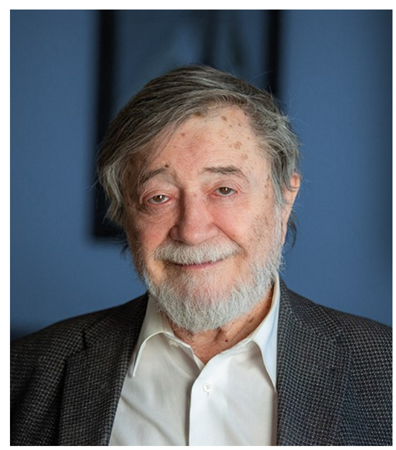


Figure 15: Judea Pearl (born September 4, 1936) is an Israeli-American computer scientist and philosopher, best known for championing the probabilistic approach to artificial intelligence and the development of Bayesian networks. He is also credited with developing a theory of causal and counterfactual inference based on structural models. In 2011, the Association for Computing Machinery (ACM) awarded Pearl with the Turing Award, the highest distinction in computer science, "for fundamental contributions to artificial intelligence through the development of a calculus for probabilistic and causal reasoning".

1. Association: At the lowest rung of the ladder, we have associations or correlations between variables. This level merely identifies statistical relationships between variables but doesn't imply causation. For example, if we observe that as ice cream sales increase, the number of drowning incidents also increases, this is an association but not necessarily a causal relationship.
2. Intervention: The second rung involves interventions or controlled experiments. Here, we can establish causation by manipulating one variable (the independent variable) and observing the effect on another variable (the dependent variable). Randomized controlled trials (RCTs) are a common example of this level, where researchers actively manipulate variables to assess causation.
3. Counterfactuals: Moving up the ladder, we reach the counterfactual level. This level involves comparing what actually happened to what would have happened under different conditions. It addresses questions like "What would have happened if we hadn't intervened?" Counterfactuals help us understand causation by

considering both the observed outcome and the unobserved alternative outcomes.

4 Structural Models: At the highest rung of the ladder, we have structural causal models. These models use mathematical and graphical representations to capture the complex causal relationships between variables in a system. They incorporate causal diagrams, which illustrate how variables influence each other. Structural models allow for the estimation of causal effects even when experiments or counterfactuals are not feasible or ethical.

5 The ladder of causality highlights that causation is a complex and multi-faceted concept. Depending on the level of understanding and the available data, we may be limited to inferring association, conducting experiments, considering counterfactuals, or employing sophisticated causal modeling techniques to understand and quantify causal relationships in different contexts.

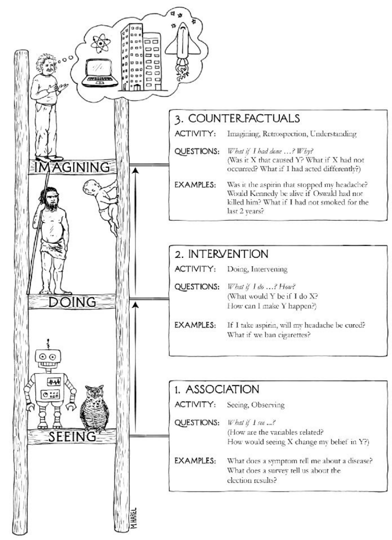


Figure 16: The ladder of causality is a conceptual framework that helps illustrate the different levels of causation and the types of questions we can ask about causal relationships. It provides a systematic way to think about causality and infer causation from data.

Causality as a Fundamental Concept

Pearl argues that causality is a fundamental concept in understanding the world, going beyond correlations and statistical associations.

1 Causal Graphs: Pearl introduced the use of causal graphs or Bayesian networks to represent causal relationships visually. In these graphs, nodes represent variables, and directed edges represent causal relationships between variables. The graph structure helps elucidate the direction and strength of causation. Three Types of Variables:

2 Pearl distinguishes between three types of variables:
   - Endogenous Variables: Affected by other variables within the system.
   - Exogenous Variables: External factors that influence the system but are not influenced by it.
   - Latent Variables: Unobservable factors that can affect the system. Judea Pearl's causal theory, often referred to as the "Pearlian causal framework," is a significant contribution to the field of causal inference and artificial intelligence. It seeks to understand and model cause-and-effect relationships in data and complex systems. Here's a summarized overview of Judea Pearl's causal theory:

3 Structural Causal Models (SCMs): SCMs are mathematical models used to formalize causal relationships in a system. They consist of three components: a causal graph, a set of equations representing the functional relationships between variables, and a set of exogenous variables.

4 Do-Calculus: Pearl developed the Do-Calculus, a set of rules and

operations for inferring causal effects from observational and interventional data. It enables researchers to answer counterfactual questions, such as "What would happen if we intervened to change a specific variable?"

5 Counterfactuals: Counterfactuals refer to statements about what would have happened if the world had been different. Pearl's framework enables the formal representation and reasoning about counterfactuals, which are crucial for understanding causation. Potential Outcomes:

6 Pearl's theory incorporates potential outcomes, where each unit or entity can have multiple possible outcomes depending on different interventions or treatments.

7 Causal Inference: Judea Pearl's causal theory has had a profound impact on the field of causal inference, particularly in applications like epidemiology, economics, and machine learning. It provides a rigorous framework for drawing causal conclusions from observational and experimental data.

In summary, Judea Pearl's causal theory provides a systematic and mathematical foundation for understanding causality, incorporating causal graphs, structural causal models, Do-Calculus, counterfactuals, and potential outcomes. It has wide-ranging applications in various fields and has advanced our ability to make causal inferences from data. Structural Causal Models (SCMs):

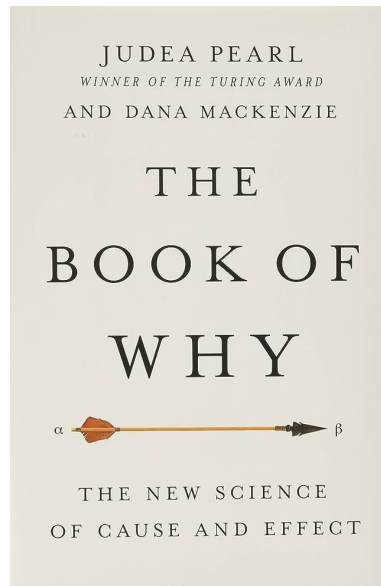


Figure 17: "The Book of Why: The New Science of Cause and Effect" is a book written by Judea Pearl and Dana Mackenzie. In this book, Pearl, a computer scientist and pioneer in the field of causal inference, explores the fundamental concept of causality and its role in understanding the world around us.