Understanding and Changing the World – From Information to Knowledge and Intelligence

SUSTECH – RITAS
October 26, 2023

Joseph Sifakis

Introduction – Motivation for Writing the Book

- In the early 1980s, I came to realize that my research into applied logic and mathematical language theories was directly related to philosophical problems, particularly the problem of consciousness and language.

- I thus started reading and became interested in philosophy. I must say that, despite my best efforts, I could not find a path through the bewildering maze of philosophical theories.

- I believe that any attempt to formulate our knowledge must respect the laws of an economy of thought, must be based on a set of well-founded concepts that enjoy clearly defined semantic relationships with each other.

- I wanted to gradually build an outlook of the world where I always distinguish what I can talk about with some degree of certainty from what I do not know.

- Our aim must be to push the boundaries of knowledge as far as possible, while remaining aware of our limitations.

- I quickly identified the issues of knowledge and, therefore, consciousness and language as the key themes.

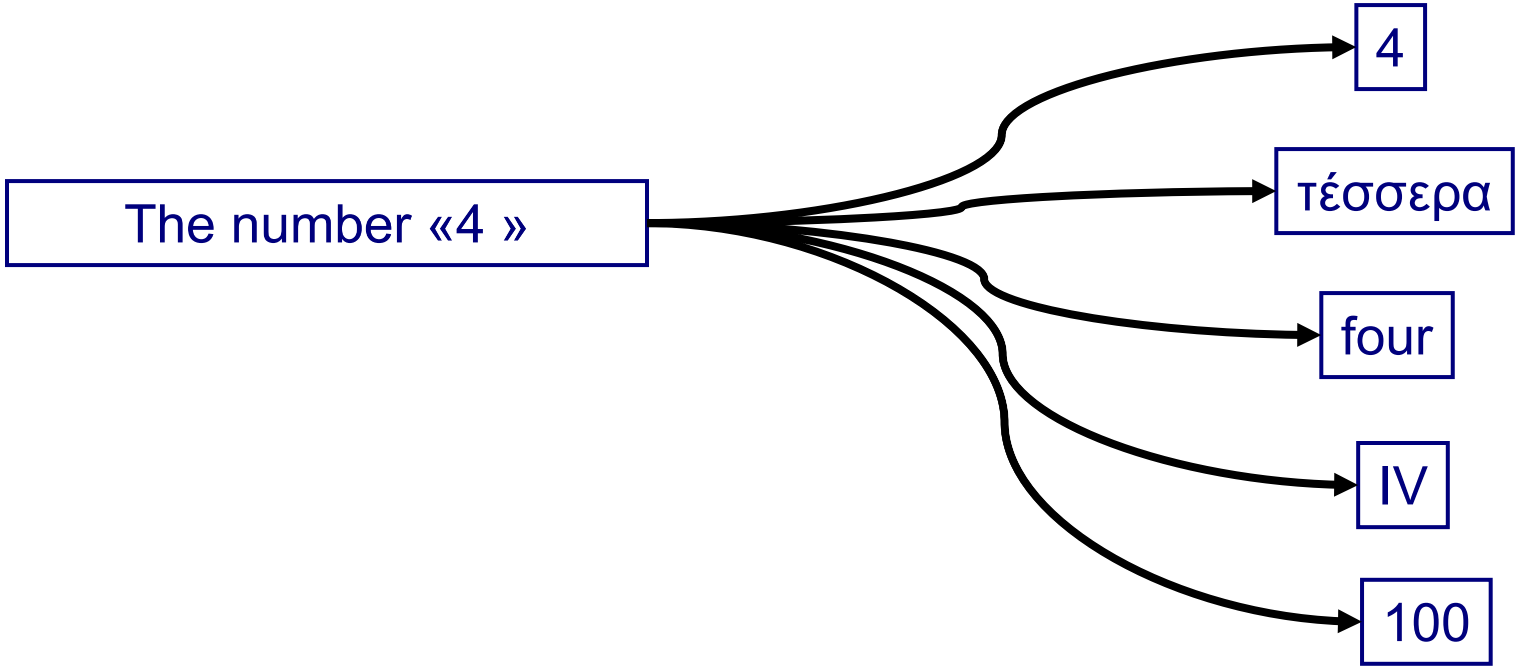
    - What are the mechanisms for producing and applying knowledge, especially scientific and technical knowledge?
    - How can human and machine intelligence be compared?
    - Can human intelligence, and consciousness in particular, be studied through the prism of Computing?

Overview

- ❑ Information and Computation
- ❑ About Knowledge
- ❑ About Intelligence
- ❑ The Impact of intelligent Systems
- ❑ Discussion

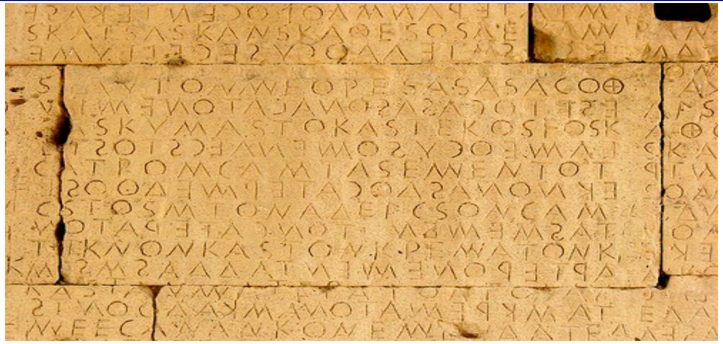
Information and Computation – What is Information?

Information is a relationship between the syntax and the semantic domain of a given language



Information and Computation – Information is in the Mind of the Beholder

No information



Information for an archeologist

$$\oint \mathbf{E} \cdot d\mathbf{A} = \frac{q_{enc}}{\varepsilon_0}$$
$$\oint \mathbf{B} \cdot d\mathbf{A} = 0$$
$$\oint \mathbf{E} \cdot d\mathbf{s} = -\frac{d\Phi_B}{dt}$$
$$\oint \mathbf{B} \cdot d\mathbf{s} = \mu_0 \varepsilon_0 \frac{d\Phi_E}{dt} + \mu_0 i_{enc}$$

Information for a Physicist

Information

Information

- is an <u>entity</u> different from matter/energy
- is non-physical although it needs media for its representation.
- is not subject to physical space-time constraints - the Theory of Computation is time-ignorant

Information and Computation – Syntactic Information

Syntactic information is measured as the quantity of symbols, pixels, bits needed for a representation of information.
According to Shannon's Theory, it

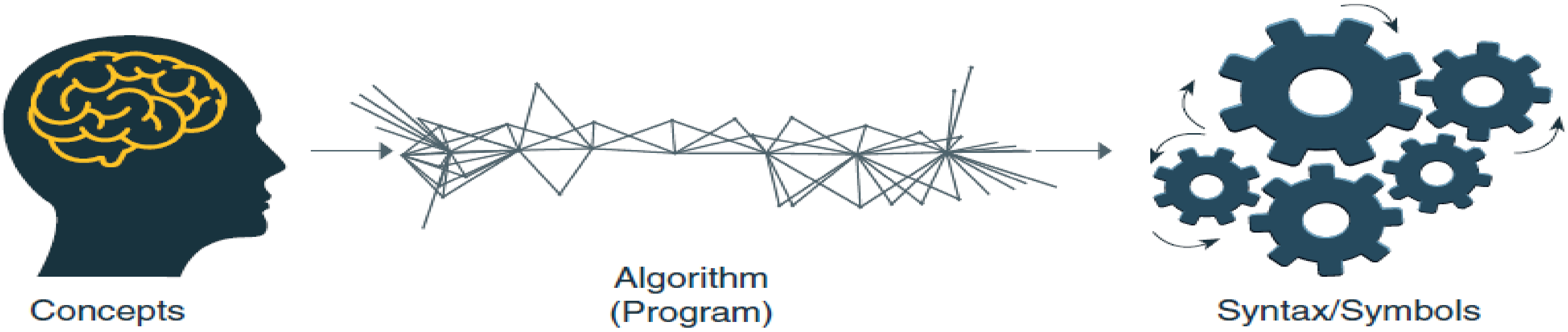
- characterizes the content of a message, not its meaning
- is *nlog(b)*, the number of yes/no questions one would have asked to completely resolve ambiguity for a word of length *n* on an alphabet of *b* symbols

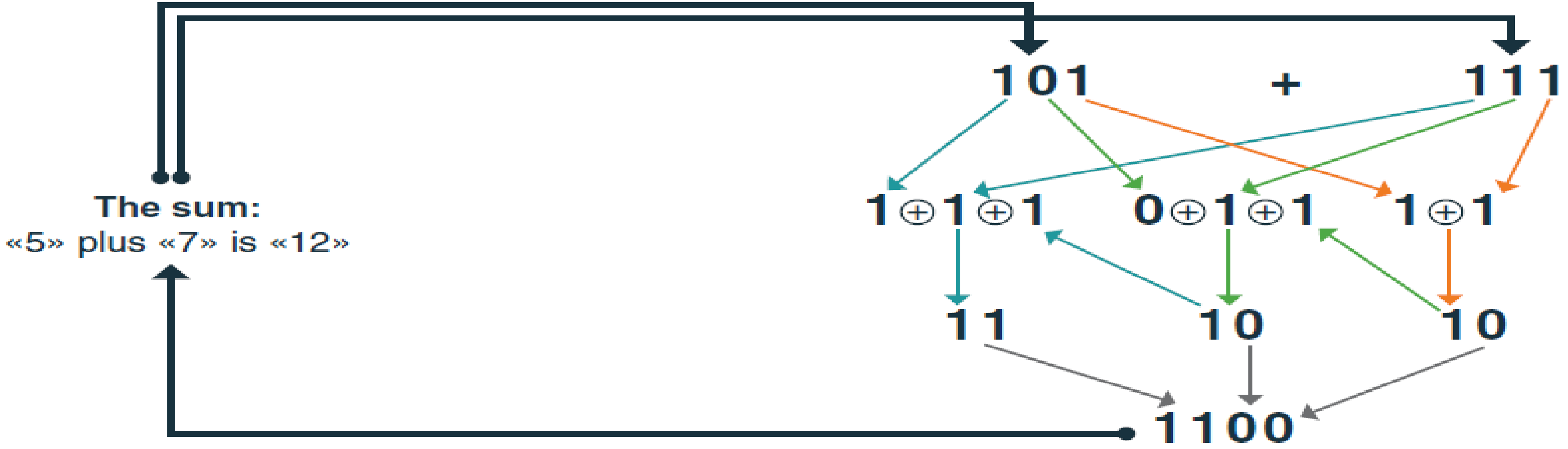
g	o	o	d	m	o	r	n	i	n	g
g	d	r	d	o	m	o	n	o	g	i

Syntactic information theory e.g. Shannon, Kolmogorov

- finds application in data compression, channel coding, information representation techniques
- ignores meaning - It is like saying that one kilo of gold and one kilo of lead are equivalent!

Information and Computation – Algorithms





- Gödel's incompleteness theorem

For any formal system based on arithmetic, there will always be statements about the natural numbers that are true, but that are unprovable within the system

Examples of propositions not provable mechanically: *"a program terminates"*, *"the variable x of a program is bounded by some value"*

This theorem sets limits to mathematical knowledge – a kind of "uncertainty principle" for Mathematics.

- Complexity:

  Each problem admitting an algorithmic solution is characterized by its complexity, that is the amount of resources in time and memory needed for its solution.

  - The same problem can be solved by algorithms of different complexity
  - The complexity of a problem cannot be lower than a certain bound

Overview

- Information and Computation
- About Knowledge
- About Intelligence
- The Impact of intelligent Systems
- Discussion

Knowledge – Understanding and Changing the World

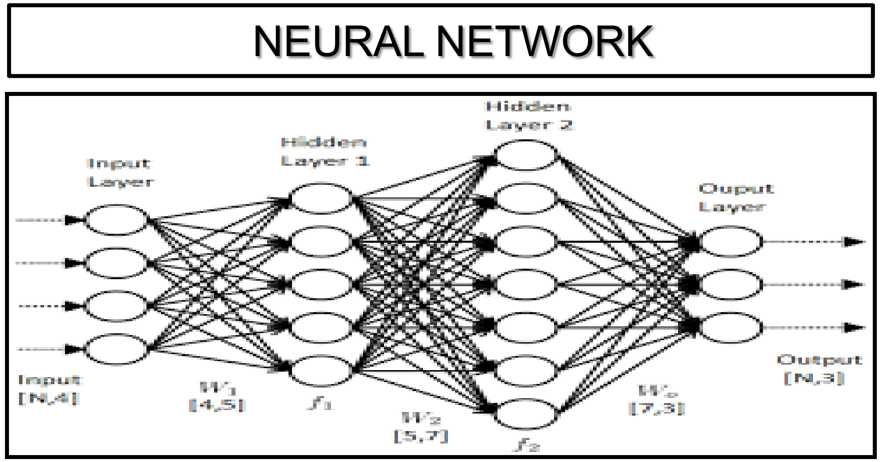
Knowledge – An Interesting Analogy

Fast thinking vs. Slow thinking (D. Kahneman's "Thinking Fast and Slow")

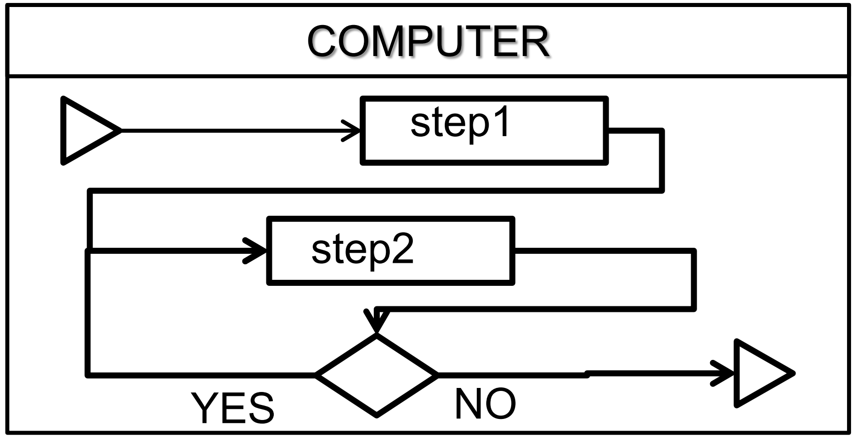
System 1: "Fast" Thinking
- Non-conscious – automatic – effortless;
- Without self-awareness or control;
- Handles all kind of empirical implicit knowledge e.g. walking, speaking or playing the piano.

System 2: "Slow" Thinking
- Conscious – controlled– effortful;
- With self-awareness and control;
- Is the source of any reasoned knowledge e.g. mathematical, scientific, technical.

Neural Networks vs. Conventional Computers

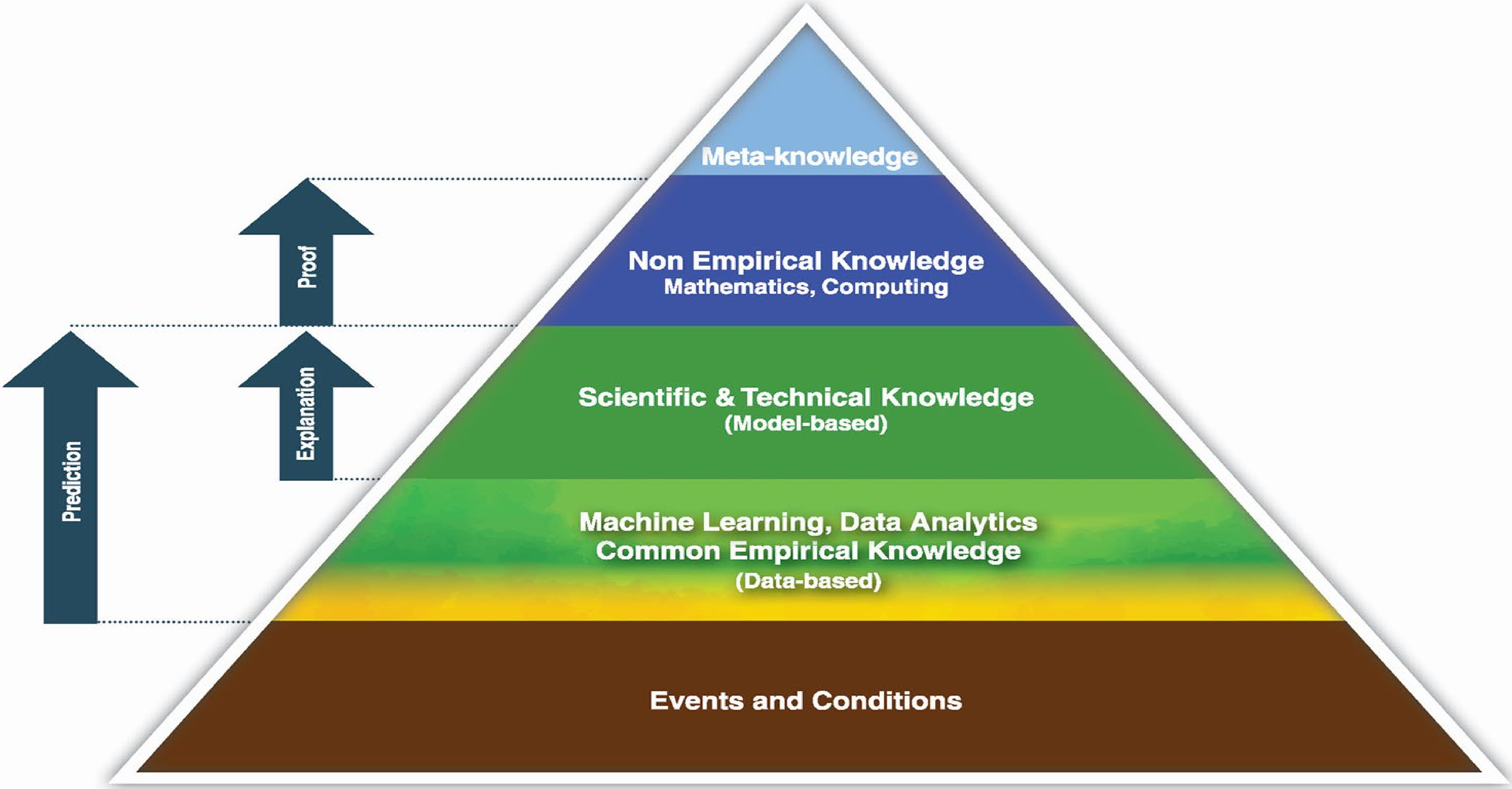


- Generate empirical knowledge after training (<u>Data-based</u> knowledge) – do not execute algorithms
- Distinguish "cats from dogs" exactly as kids do – Cannot be verified!

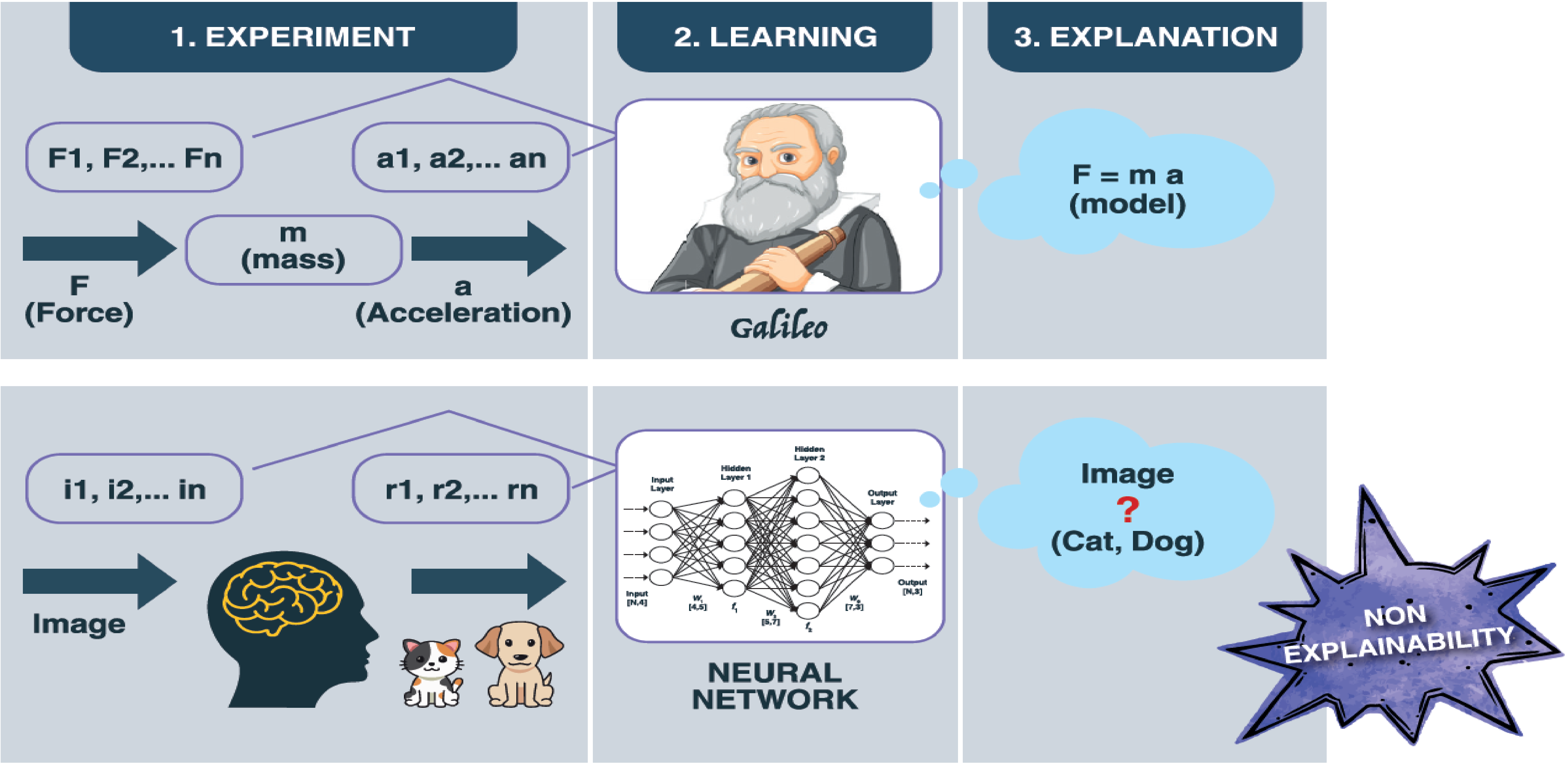


- Execute algorithms (<u>Model-based</u> knowledge).
- Deal with explicitly formalized knowledge
- Can be understood and verified!

Knowledge – The Knowledge Hierarchy



Knowledge – Scientific vs. ML-generated Knowledge



Knowledge – Abstraction and Modularity

- Reality has depth and breadth. To cope with complexity, we study the physical world at different levels of abstraction, in scales
  - From 10^{-35}m, the Planck length
  - To 10^{25} m, size of the observed universe

- Abstraction is a holistic way to break complexity by revealing relevant features of the observed reality
  "Being abstract is something profoundly different from being vague … The purpose of abstraction is not to be vague, but to create a new semantic level in which one can be absolutely precise" E.W. Dijsktra

- Modularity: Complex systems can be built from a relatively small number of types of components (bricks, atomic elements) and glue (mortar) that can be considered as a composition operator.
  Basic assumptions (compositionality principle):
  1. Any system can be built as the composition of a finite set of predefined types of components.
  2. The behavior of each component can be studied separately.
  3. The behavior of a composite component can be inferred by composing the behavior of its constituents
     This assumption is valid in classical Physics but fails for bio-systems, programs, linguistic systems, etc.

Knowledge – Abstraction Hierarchies

PHYSICS	INFORMATICS	BIOLOGY
The Universe	Cyber-world	Ecosystem
Galaxy	Networked Systems	Organism
Solar System	Embedded Systems	Organs
Electro-mechanical Systems	Virtual Machine	Tissues
Solids – Liquids – Gases	Hardware Architecture	Cells
Molecules	Integrated Circuit	RNA networks
Atoms	Logic Gate	Protein and RNA
Particles	Transistor	Genes

Knowledge – Abstraction Hierarchies

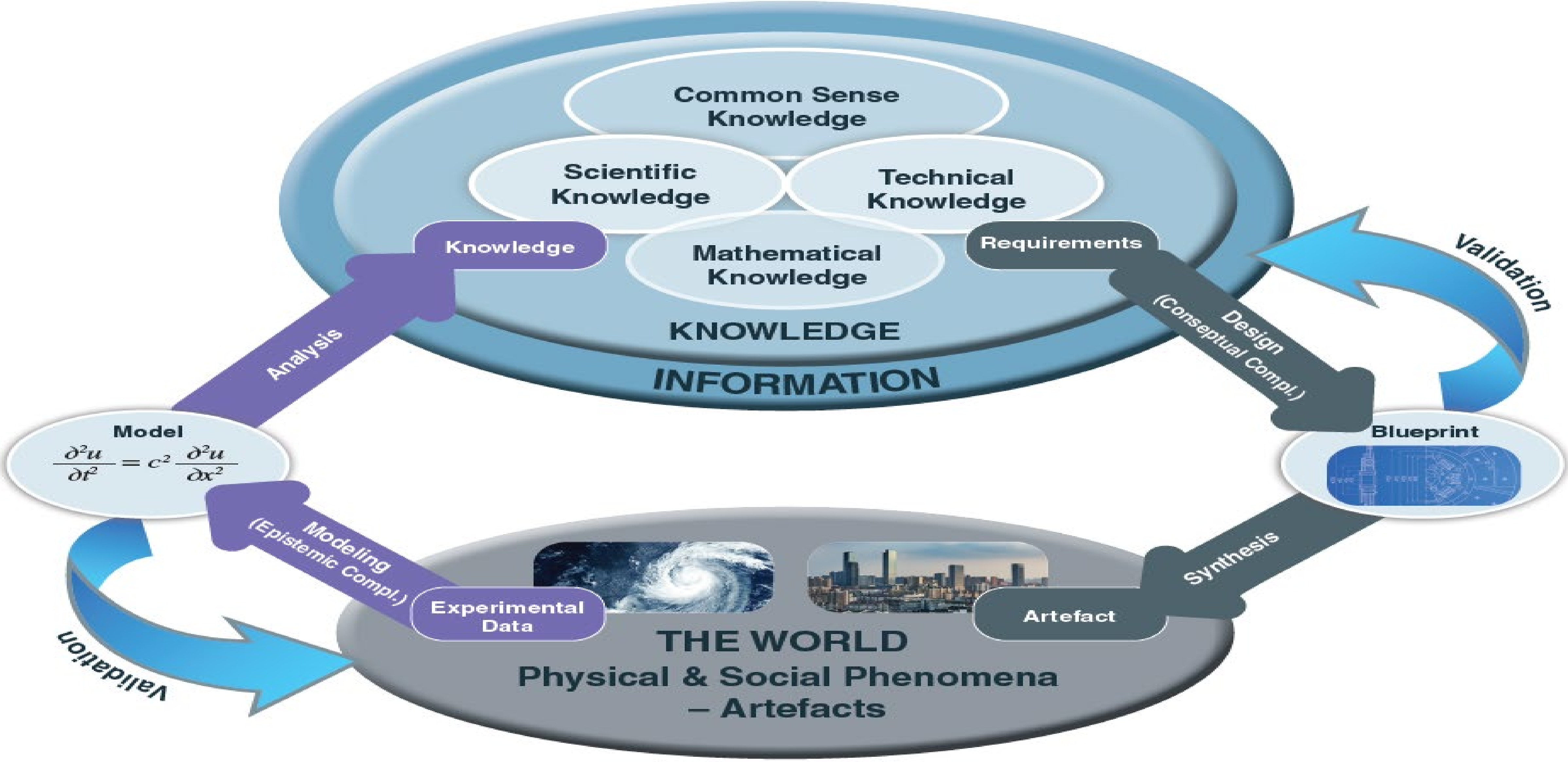
- Abstraction hierarchies are a methodological simplification used to determine successive levels of granularity of observation at which relevant system properties can be studied.
  - The models of the hierarchy should be related through some adequate abstraction relation.
  - The abstraction relation should link the laws and properties at one layer to laws and properties of the upper layers.

- Is it possible to unify knowledge in a domain using the compositionality principle: knowing the properties of components at one layer, is it possible to infer global properties of composite components at a higher level?
  - *properties of water from properties of the atoms of hydrogen and oxygen and rules for their composition?*
  - *properties of neural network from behavioral properties of its nodes?*
  - *properties of mental processes from behavioral properties of components (neurons) of the brain?*

These questions are of the same nature, and will probably find no answers!

- A specific problem for computing systems is *component heterogeneity* - This is a key limitation to mastering component-based construction of software

Knowledge – Science and Engineering



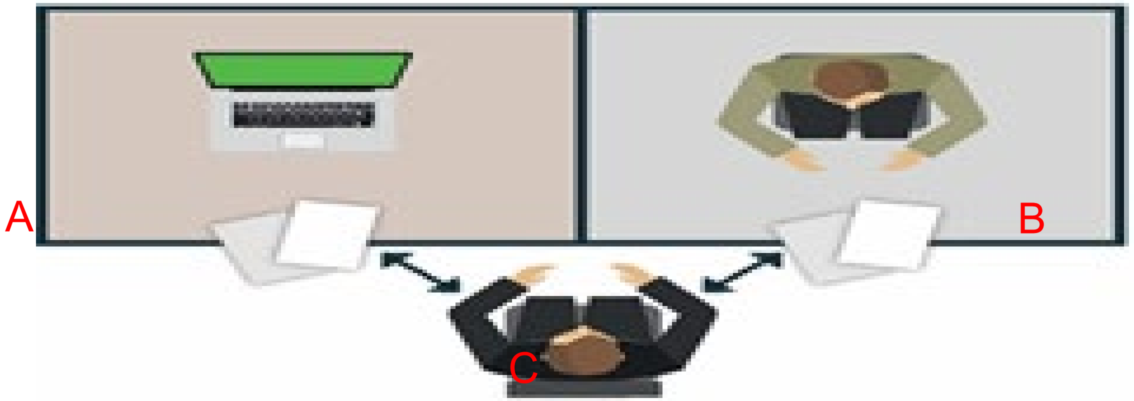
Overview

- Information and Computation
- About Knowledge
- About Intelligence
- The Impact of intelligent Systems
- Discussion

Intelligence – The Turing Test

"Can we experimentally distinguish a computer from a human by analysing their answers to a series of questions?"

- Turing Test (Imitation Game):
  1. C sends questions to A and B who, in turn, provide a corresponding answer to each question.
  2. If C cannot tell which is the computer and which the person, then A and B are equally intelligent.



- Criticism:
  - Success depends on human judgement (subjective) and the choice of the test cases (questions).
  - The test cannot be a question/answer game - much of human intelligence is expressed by interaction with the environment (speech, movement, social behavior, etc.)

- Replacement test: *An agent A (indifferently machine or human) is as intelligent as an agent B performing a given task characterized by given well-founded success criteria, if A can successfully replace B. e.g.*
  - a machine is as intelligent as a human driver is if it can successfully replace the driver.
  - a human is as intelligent as a janitor robot if it can successfully replace the robot according to given cleaning criteria.

Note that the Turing test is a special case, provided of course that its success can be rigorously specified.

Intelligence – AI Systems

- Game playing systems mark AI's first big success:
  Deep Blue (IBM 1997), WATSON (IBM 2011), AlphaGo (Google DeepMind 2016)
  beat human players by calculating much faster and more accurately in complex data spaces

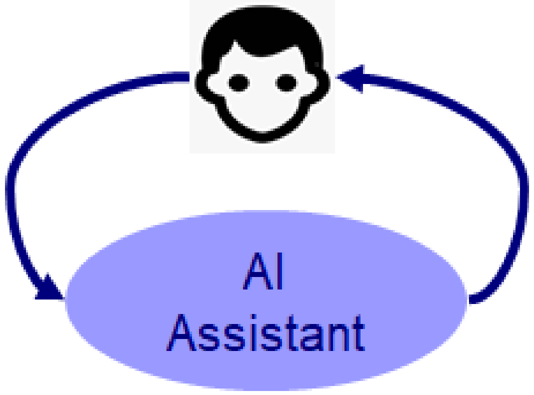
- The advent of LLMs (Large Language Models) such as ChatGPT marks a paradigm shift, as they provide a pragmatic solution to the long-standing problem of natural language processing
  - LLMs consider that the meaning of a word is defined by the set of its contexts of use.
    - Machine learning estimates a probability distribution over words to predict the next most likely word in a sentence.
  - LLMs are able to answer, often in a relevant way, almost all natural language queries.
    - But if they do not have the information in their learning corpus, they make up the most plausible and for precise questions they may produce fakes that look like the real thing.

Intelligence – Possible Uses for AI systems

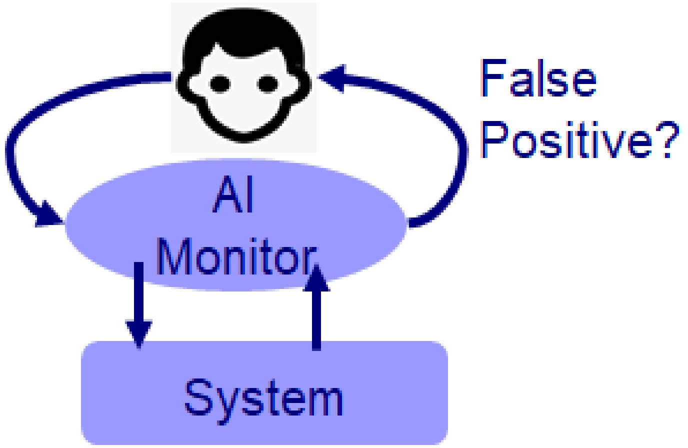
- Big Tech companies are focusing on building Intelligent Assistants that interact with a user to provide a service, for example in question-and-answer mode.
  We see three different ways of using AI systems:

  1. Oracles (assistants) that tight interaction with a user provide a given service
  2. Monitors that analyze the behavior of a system and make predictions about relevant properties such as safety, security or performance
  3. Controllers that continuously interact with an environment to perform a given task characterized by a set of goals e.g. driving a car – specific recovery mechanisms are needed in case of failure!

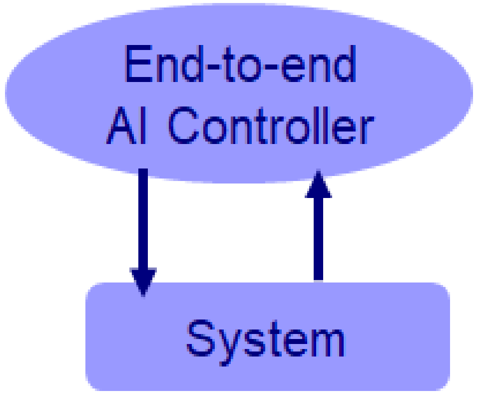
Note that for 2 and 3 we need to adapt existing technology and exploit its potential to solve the specific problems.



Conversational Interaction



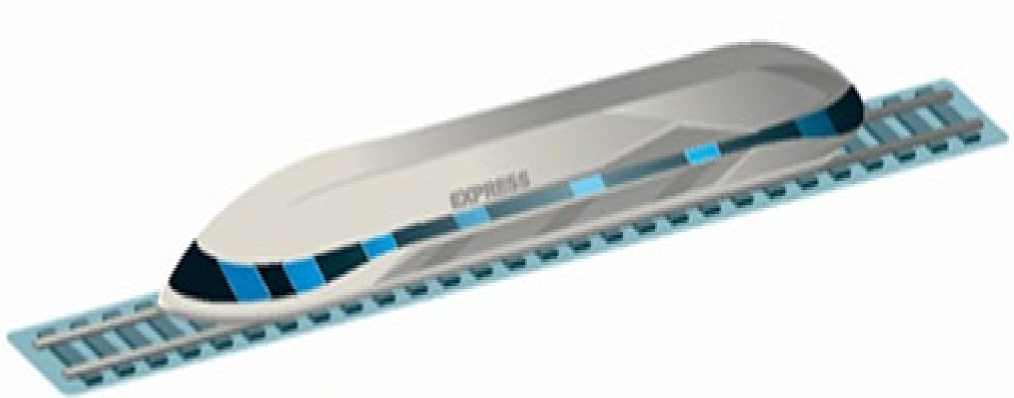
Prediction/detection of events



Autonomous System

Intelligence – Autonomous Systems

Thermostat


Automated Shuttle

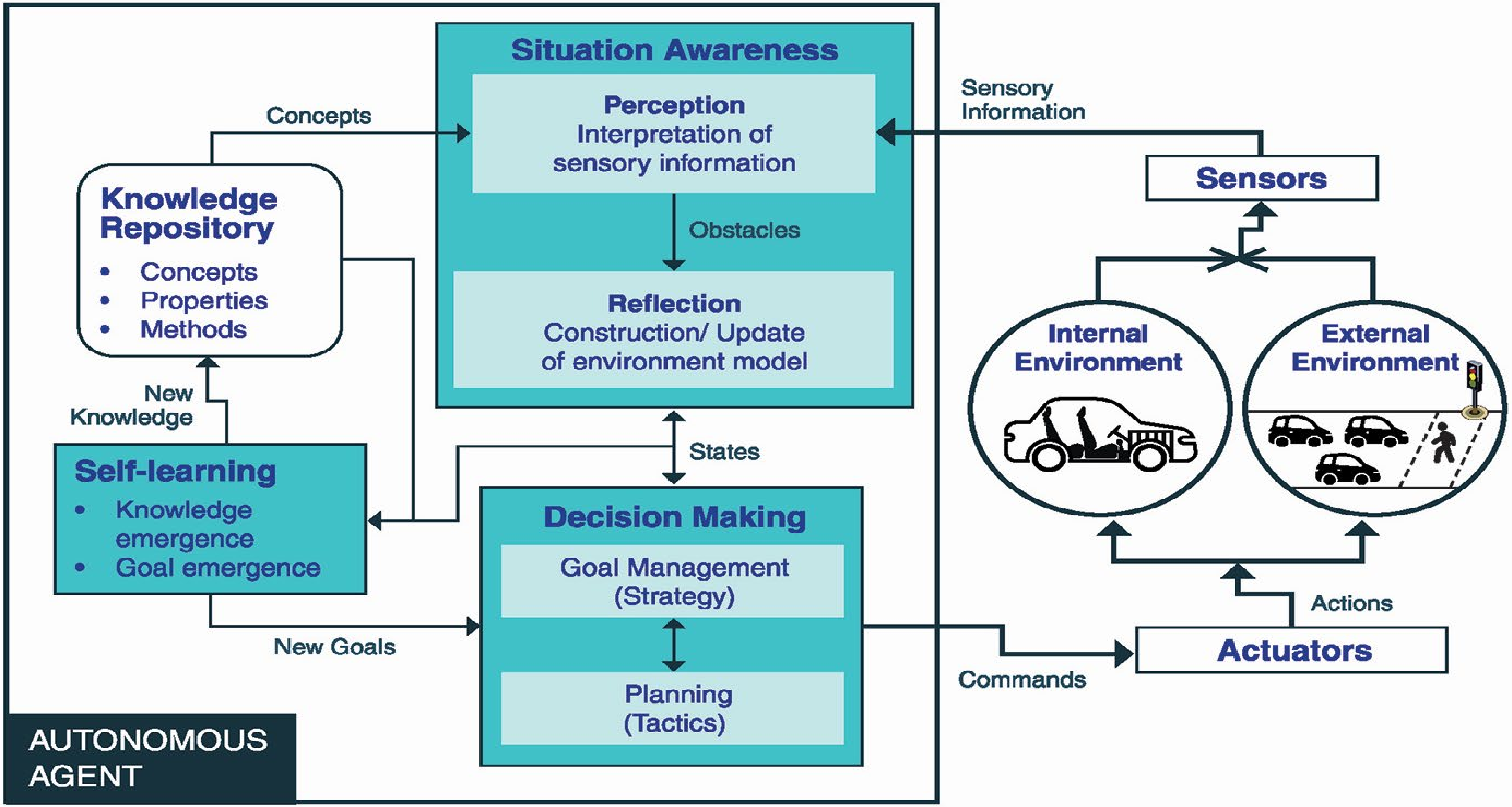

Chess Robots


Football Robots


Autonomous Car

	Thermostat	Shuttle	Chess Robot	Football Robot	Autonomous Car
SITUATION AWARENESS	Temperature (number)	Distance from next stop (number)	Pawns on the board (static image)	Players on the pitch (dynamic image)	Obstacles on the road (dynamic image)
DECISION MAKING	Static goals <18 → ON >22 → OFF	Static goals ▪ Stop ▪ Accelerate ▪ Decelerate	Static well-defined goals Dynamic planning of goals	Dynamic multiple goals Dynamic planning of goals	Dynamic multiple conflicting goals Dynamic planning of goals

Intelligence – Basic Autonomic Functions



Intelligence – The Autonomy Challenge

- Autonomous agents rely on computational intelligence to overcome complexity limitations

  - Complexity of perception due to the difficulty to interpret stimuli (cope with ambiguity, vagueness) and to timely generate corresponding inputs for the agent environment model.

  - Complexity of uncertainty due to situations involving imperfect or unknown information implying lack of predictability about the environment such as dynamic change caused by physical or human processes, rare events, critical events such as failures and attacks.

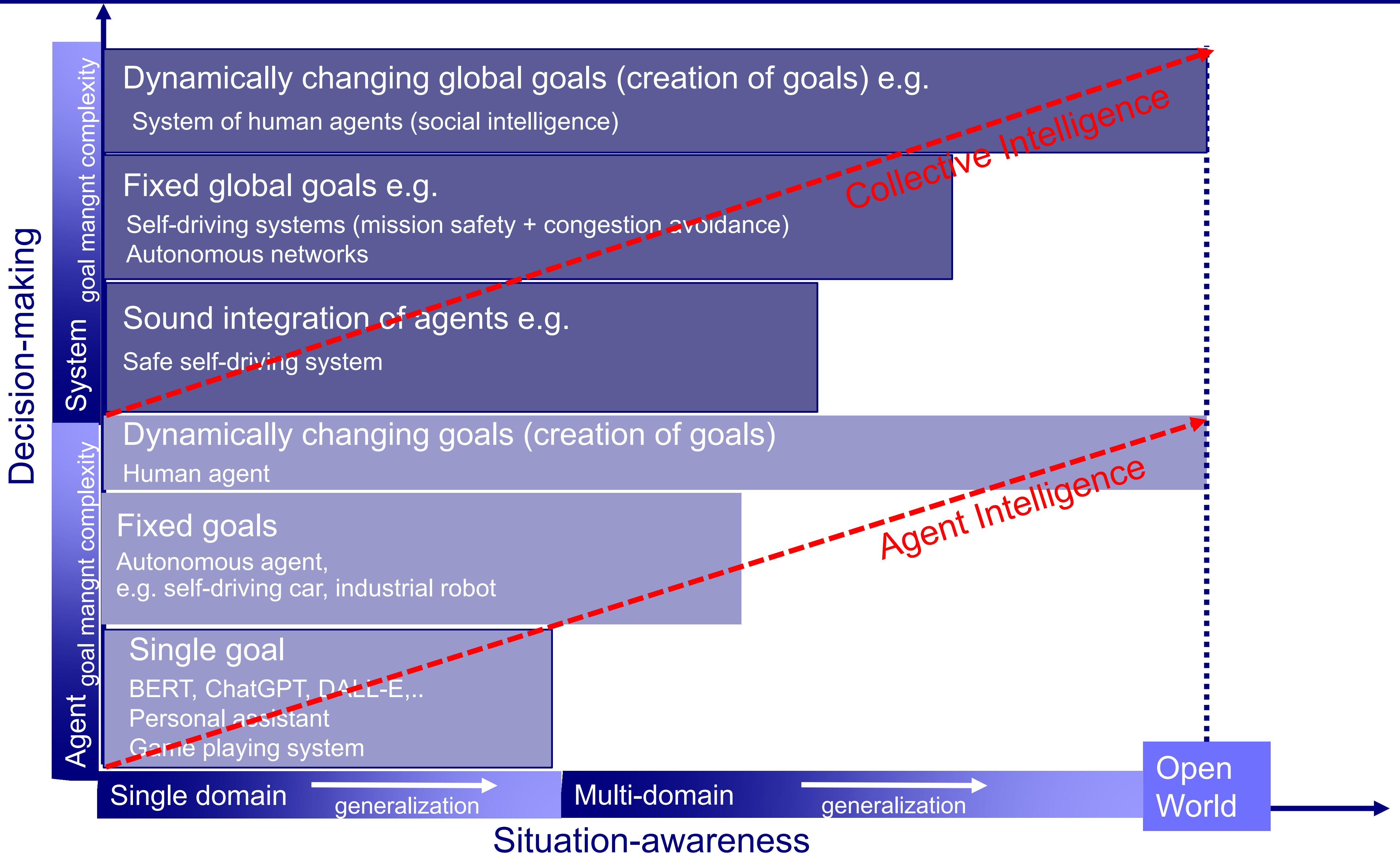
  - Complexity of decision reflected in the complexity of the agent's decision process (goal management and planning) and impacted by factors such as diversity of goals and size of the space of solutions for planning.

- However, building autonomous systems involves difficult systems engineering problems that are not related to the fact that agents are intelligent -- problems that could explain the setbacks of autonomous car industry,

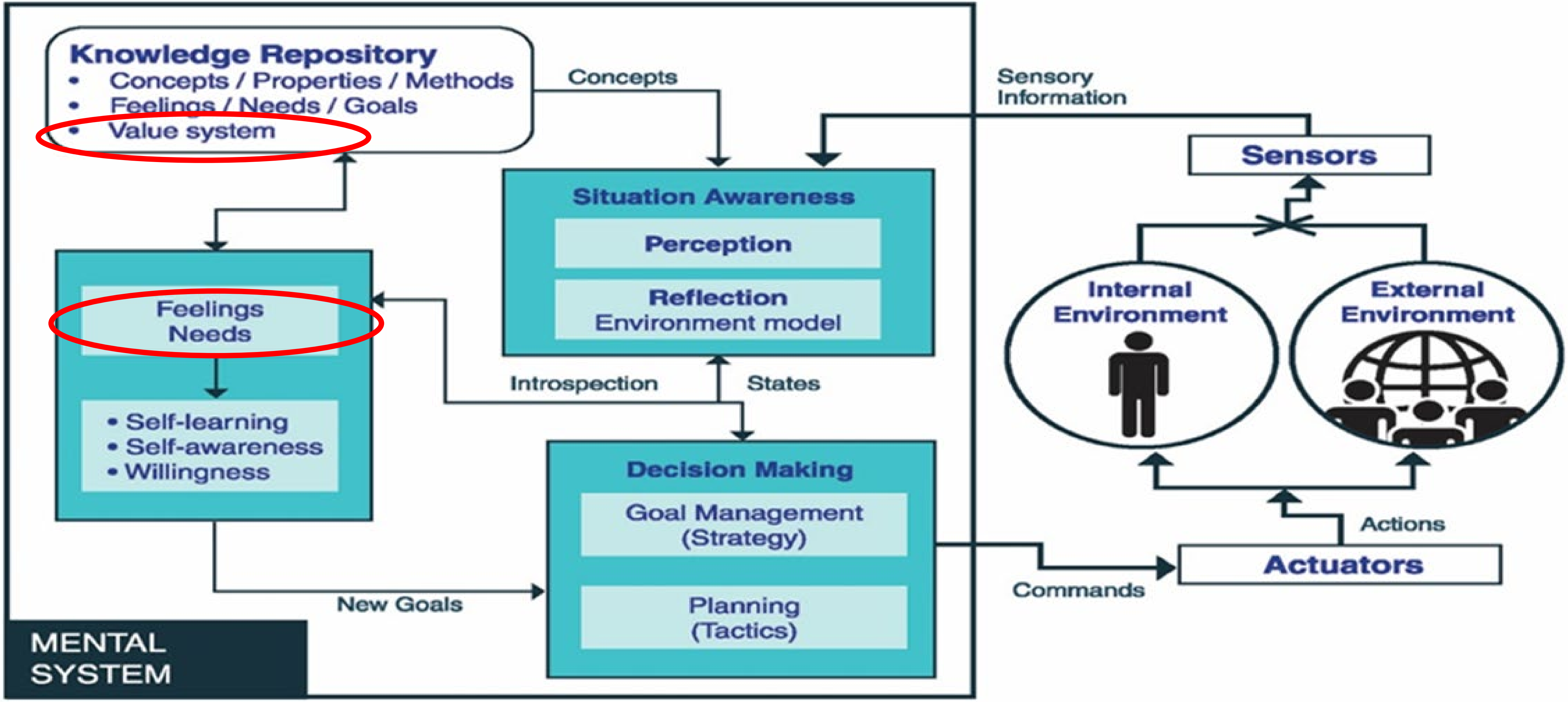
  - System agents should be
    - integrated in complex cyber physical environments systems e.g. electromechanical systems
    - be able to harmoniously collaborate with human operators – It's not just an HMI problem!

  - System agents should be adequately coordinated to achieve
    - Symbiosis: the coordination of agents does not impede the achievement of their individual goals
    - Synergy: agents collaborate to achieve global system goals by demonstrating collective intelligence.

Intelligence – From Agent Intelligence to Collective Intelligence



Intelligence – Mental Systems



- A mental system is a behavioral model (a conceptual construction),
  - whose behavior can characterize aspects of human intelligence as the manifestation of mental phenomena.
  - showing how the functionality of an autonomous system can be extended to approach human intelligence.

Intelligence – Mental Value Systems and Feelings

- Organisms make choices among several alternatives based on the subjective value they place on them – A mental system, decides on the basis of:
  1. An *individual value system* used to associate a value balance to actions - positive value means benefit, recognition, satisfaction while negative value means loss of money, punishment, shame, lack of esteem, etc.
  2. The intensity of *feelings* emerging as the manifestation of specific needs.

1. An individual value system determines the way the mental system interacts with its environment, i.e. its "social behavior"
   - reflects a common value system fostered by organizations to achieve global goals by encouraging positively valued actions.
   - is a collection of value scales for the different fields of action e.g. economic (monetary values), legal (penal values), educational (excellence), military (readiness, self-sacrifice,,), esthetic (beautiful, pleasant).

2. Feelings are abstractions of mental states
   - Reflect a specific need and emerge either automatically or as the result of an introspection process e.g. biological (thirst, hunger, sleep, fatigue..), safety (fear, anxiety), intellectual (compassion, solidarity, temperance, honesty), etc.
   - Have intensity represented by a state variable of the mental system taking values in some (subjective) scale of values e.g. hunger varies in [satiety, starvation].

Intelligence – Value-based Decision Making

- A mental system can be conceived as a machine including the semantic model, with a very large set of states defined by the values of three types of variables:
  1) External environment variables representing states of the physical, economic and social contexts;
  2) Internal environment variables characterizing the body's state according to physiological parameters such as temperature, heart rate, and blood pressure;
  3) Variables measuring the intensity of feelings by taking the values of the respective scale.

- Actions are state changes performed either by the system (controllable) or by its environment (non-controllable)
  - An action can be executed if its pre-conditions are met e.g. I have money to buy food, I feel hungry.
  - Actions have value balance e.g. "Buy/eat a hamburger", decreases the money available and gives satiety. "change of room temperature" can affect our feeling of comfort.

- *Conflicts* between actions arise when two actions have common preconditions and choosing one may cancel the preconditions of the other
  - "Free will" is exercised when the system chooses between actions based on the respective value balance.
  - Choosing between enabled actions boils down to solving an optimization problem in a two-party game.

Intelligence – Common Sense Knowledge

Tesla's Autopilot Feature Mistakes Moon for Yellow Traffic Light, Watch Video

he autopilot system of a Tesla car can be seen confusing the yellow moon in the sky. (Credits: Twitter/ Jordan Nelson)

In a viral video shared by Twitter user Jordon Nelson, the autopilot system of a Tesla car can be seen confusing the yellow moon in the sky with a yellow traffic light.

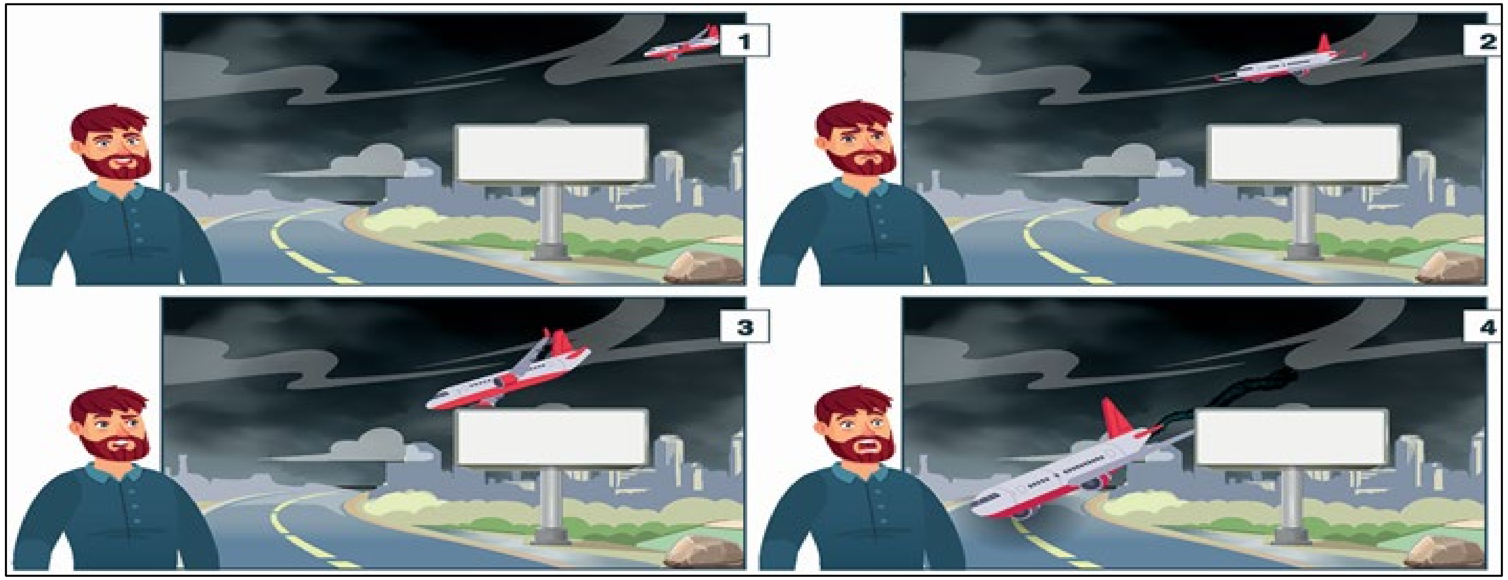
- TRENDING DESK
- LAST UPDATED: JULY 26, 2021, 17:02 IST
- FOLLOW US ON: Facebook Twitter Instagram Telegram Google News

BUZZ STAFF

- ❑ To match human-level performance, systems should be able to deal with common sense knowledge.
- ❑ Human mind is equipped with a semantic model of the world:
  - a vast network of knowledge progressively built and automatically updated throughout life by learning and reasoning, and involving concepts, cognition rules and patterns;
  - used to interpret sensory information and natural language.
- ❑ Human understanding combines:
  - bottom-up reasoning from sensor level to the semantic model of the mind;
  - and top-down reasoning from the semantic model to perception.

Intelligence – Human Situation Awareness

- We recognize a stop sign partially covered with snow because the image matches a conceptual model of a stop sign with its properties (size, color and vertical position).
- In contrast, neural networks must be trained to recognize stop signs in all possible weather conditions.



- This sequence of images is almost instantaneously interpreted as an aircraft accident using common sense knowledge.

- In contrast, a machine although it may be able to recognize the objects in each frame lacks knowledge allowing to infer the same conclusion.

❑ For machines to match human situational awareness, they must be able

- to progressively develop knowledge about their environment, in particular to understand entirely new situations;
- to combine learning and reasoning as well as concrete sensory information with symbolic knowledge.

This is probably the most difficult problem to solve, as shown by the poor progress made so far in the semantic analysis of natural languages.

Overview

- Information and Computation
- About Knowledge
- About Intelligence
- The Impact of intelligent Systems
- Discussion

The Impact of Intelligent Systems – Social Impact

Social acceptance depends on the role of institutions that largely contribute to shaping the public opinion, e.g. about what is true, right, safe or secure.

- The trustworthiness of infrastructure and of all kinds of artefacts, from toasters, to toys, buildings, planes and cars.
  - is determined by standards relying on scientific and technical knowledge,
  - is controlled by independent bodies overseen by government agencies, e.g., in the US, FDA, FAA, NHTSA.

- Unfortunately, ICT systems and applications are not subject to this general rule requiring security and safety guarantees.
  - Exceptions are some critical applications (transport, nuclear power plants...).
  - Today, for AI applications, the lack of standards is compounded by permissive policies e.g. competent US authorities accept, for autonomous cars and medical devices, "self-certification" by manufacturers!

- Freedom of choice vs. performance: not to give decision-making power to systems if we are not sure that
  - they use reliable information in an unbiased and neutral way;
  - the gain in performance is commensurable with the lack of human control.

- Division of work between human and machine: technological progress and innovation imply a loss of skills
  - The use of levers has made muscle power less necessary for our survival;
  - The lack of muscle power is not as dramatic as the loss of the skill to produce knowledge and act responsibly - which is the essence of human nature.

  - I do not believe that computers can surpass the intelligence of their creators.
    But it would be possible for creators to be enslaved to computers because they would be overtaken by the complexity of their management or because of laziness and "for convenience" .
    And that would be a catastrophic scenario for humanity!

The Impact of Intelligent Systems – A New Kind of Science

- We can leverage the complementarity between humans and machines to accelerate the development of knowledge.

  - Humans are limited by *cognitive complexity* in extracting knowledge from high-dimensional data, whereas they can develop symbolic knowledge using powerful abstraction mechanisms, e.g. induction, metaphors, analogies, creative thinking.

    $\Longrightarrow$ Useful scientific theories involve a small number of independent concepts and parameters. Often to study complex phenomena e.g. economic, we do simplifications that may prove to be unrealistic.

  - Machines are capable of computational intelligence by creating knowledge from high-dimensional data, but their current ability to create and apply symbolic knowledge is limited.

- We can generate a new type of knowledge, between scientific knowledge and implicit empirical knowledge, allowing predictability and only limited understanding.
  - Thanks to supercomputers and AI, we can build *neural oracles* allowing predictability of complex phenomena e.g. geophysical, economic, social, etc.
  - If a neural oracle is explainable, then its models can serve as a basis for the construction of a "theory" explicating the relations between the observables.
  - The knowledge so produced is entirely discovered by machines and impossible to apprehend conceptually.
  - Using this kind of knowledge, especially to make critical decisions, should give us pause.

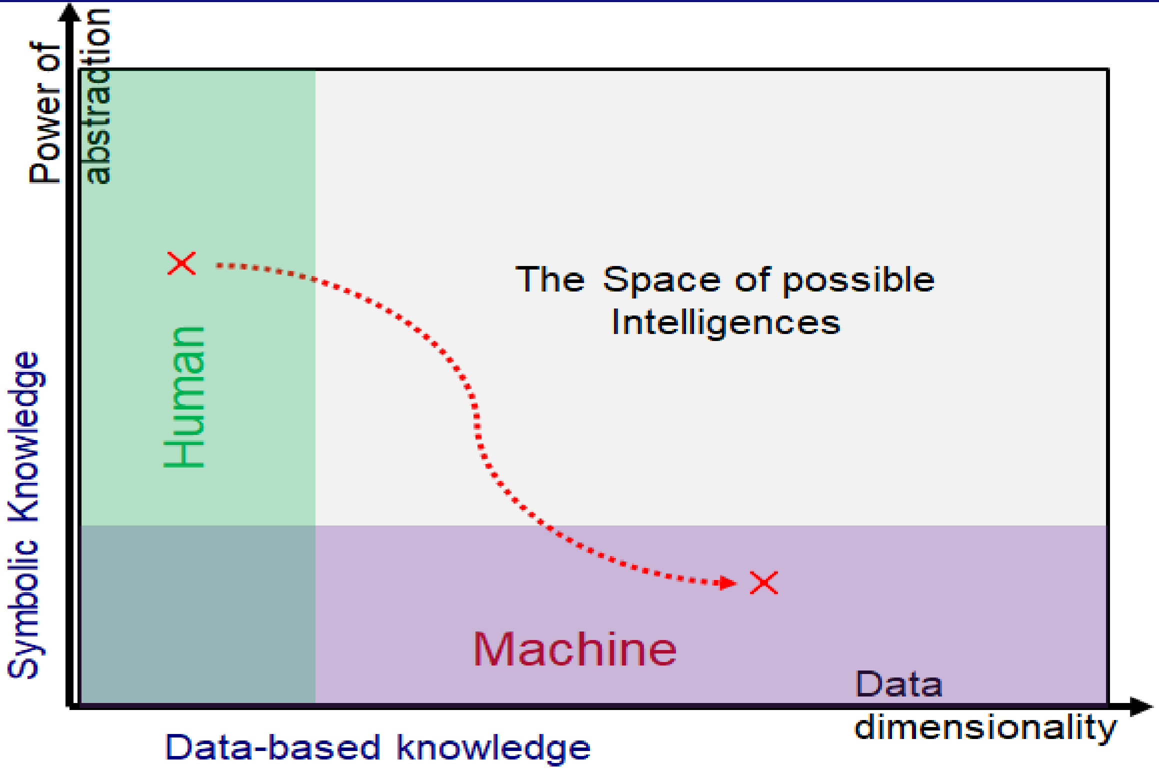
- Knowledge production/application is no longer the preserve of humans – two possible scenarios:
  - **Symbiosis**: human-centred machine-assisted development of science and technology
  - **Divergence**: parallel development of a "para-science “

The Impact of Intelligent Systems – The Space of Possible Intelligences

- The replacement test relativizes and generalizes the concept of intelligence suggesting that intelligence is not one-dimensional
  - To say that "S1 is smarter than S2" is meaningless without specifying the task(s) and the criteria for success. There are <u>multiple intelligences</u>, each characterizing the ability to perform tasks in different contexts;
  - Human intelligence is not "general purpose"; it is the result of historical evolution in a given physical environment. If <u>human intelligence is the benchmark</u>, AGI should be able to perform/coordinate a set of tasks characterizing human skills.

- The <u>space of possible intelligences</u>: equivalent systems may use very different creative processes.
  - Humans are limited in analysis of multidimensional data, but are capable of common sense, abstraction and creativity.
  - AI systems outperform humans in learning multidimensional data, but fail to link symbolic to data-based knowledge.

- We need to explore the vast space of intelligences, particularly by delving into the various aspects of human symbolic intelligence and their relationship to data-driven intelligence.
  - Can we bridge the gap between symbolic and concrete knowledge exclusively by using neural networks?
  - Is it possible to trade symbolic reasoning capability for data-based learning as shown by LLM's opening the way to efficient solutions to symbolic reasoning problems e.g. MathPrompter



The Impact of Intelligent Systems – Exploring Human Consciousness

- Our progress in building intelligent systems will ultimately depend on our ability to elicit the mind-brain relationship.
  - Mental phenomena are essential for understanding human behavior, just as software is very important for understanding what a computer does!
  - Many will claim that the mind-brain problem can only be the subject of philosophical inquiry, and that a strict “scientific” approach is impossible or even misplaced.
  - Human intelligence cannot be unravelled if we focus exclusively on the study and simulation of brain processes and ignore mental phenomena, see the Human Brain Project.

- Questions about cosmogony or the evolution of species seem easier in the face of the "big bang of consciousness”:
  - *How did cognition and language appear through evolution? How do we comprehend by linking phenomena to analogies and metaphors? How do we ascribe meaning to symbols? How do we create? How does awareness function as a system that combines wishes, motives and will? How do we choose goals and act on them? …*

- Linking mental functions and brain networks is a worthwhile, imperative endeavor where interdisciplinary cooperation between informatics, biology and medicine, could be decisive.

- Let's hope that one day we will be able to unravel the mystery of the “big bang of consciousness", and that physical cosmology will be complemented by a parallel one, which will shed light on our unexplored self and make us climb a little higher on the marvellous scale of self-knowledge.

Overview

- Information and Computation
- About Knowledge
- About Intelligence
- The Impact of intelligent Systems
- Discussion

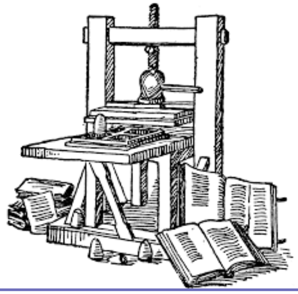
Conclusion

The information and knowledge revolution initiated by ICTs and propelled by their marriage with AI

- Puts computers at the heart of our efforts to understand and change the world.
- Reverses the positivist idea that the only established knowledge is "scientific", and places the emphasis on
  - the combination of different types of implicit and explicit knowledge depending on its degree of validity and generality
  - the unification of languages from formal to natural languages
- Paves the way for the unveiling of the "big bang" of consciousness, through the lens of computational intelligence
- Raises deeper philosophical questions on the role of knowledge
  - the way it is developed and applied – how and by whom is its validity controlled and certified?
  - Its social dimension as a common good and the challenges of using it for our well-being

Conclusion – For a Harmonious Symbiosis between Humans and Machines

- The advent of computers is a major step forward in the conquest of knowledge.


- Just as machines multiply man's muscle power, computers multiply our mental abilities.


- But be careful, just as a bird is not an aeroplane, so the human mind is not a computer.

“Intelligence is what you use when you don't know what to do.“

Jean Piaget, Swiss Psychologist (1896-1980)

Thank you