



# **Artificial Intelligence:** its Rise, Opportunities for Decision Aiding, and Threats for Humanity

#### Roman Słowiński

Poznań University of Technology & Polish Academy of Sciences, Poland

National & Kapodistrian University of Athens, April 8, 2024



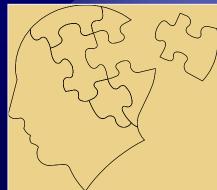
# What is machine intelligence?

 Simulating human thinking understood as calculation (arithmetic and logical)

Performing intellectual tasks of humans

(classification, rule creation, discovery of laws, relations...)

Acting rationally (motivated by a plan and knowledge of the environment)



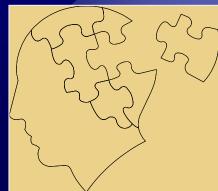
# What is machine intelligence?

 Simulating human thinking understood as calculation (arithmetic and logical)

Performing intellectual tasks of humans

(classification, rule creation, discovery of laws, relations...)

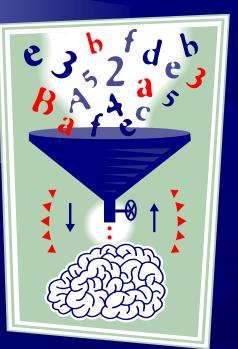
Acting rationally (motivated by a plan and knowledge of the environment)



## all thinking is calculation



(1912 - 1954)



*Mind* 1950

#### COMPUTING MACHINERY AND INTELLIGENCE

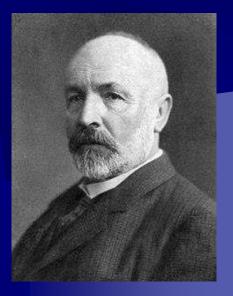
ALAN M. TURING

**1. THE IMITATION GAME** 

I propose to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the

https://doi.org/10.1093/mind/LIX.236.433

- the concept of an *infinitely large set* extends beyond the material world
- "Belief in infinity is like belief in God"
- => humans have the ability to use the concept of infinity, while machines do not



 as an argument for the superiority of the human mind over the machine, it is quite weak, as finite numbers are sufficient in practice Georg Cantor (1845-1918)

- Kurt Gödel's theorem dashed hopes for a mathematical proof of the mind's superiority over machines: ,, there is no algorithm that, given a set of axioms, can provide a proof for every true arithmetic statement"
- => a machine proving automatically arithmetic theorems is not equivalent to human computational abilities



Kurt Gödel (1906-1978)

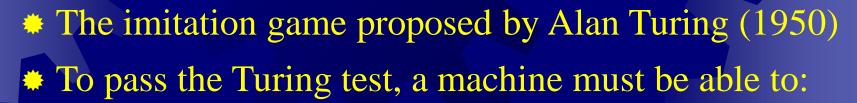
the proof of Gödel's theorem contains a contradiction due to self-reference

=> proving the superiority of the human mind over machines appears unfeasible within the framework of mathematical logic

this conclusion is not surprising, as human thinking extends beyond calculation and logical inference

So, how can we test machine intelligence?





- process and recognize images
- recognize and synthesize speech
- process natural language
- acquire knowledge from Internet

### At the end of the 20th century...

 On May 11, 1997, Russian chess grandmaster Garry Kasparov lost a match to the computer Deep Blue, built by IBM



# Progress in simulation continues... \* In 2016, in Seoul, Google's 'AlphaGo' system defeated the 'Go' game master Lee Sedol



Progress in simulation continues...

 In 2017, DeepMind demonstrated the 'AlphaZero' chess system, based on NN, which defeated 'Stockfish', the best system to date that has beaten grandmasters



# 60 years after Turing's death...

In 2014, a program that 'pretended' to be the 13-year-old Eugene Goostman convinced 30% of the judges that it is a human after a few minutes of chat

This program was sufficiently able to understand and formulate sentences in natural language, using knowledge from WWW to respond sensibly to questions



ug	ene	Goo	ostn	nan	
HE W	EIRDE	STCRE	ATURE	IN TE	IE WORLD



https://www.youtube.com/watch?v=WnzlbyTZsQY

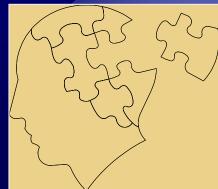
# What is machine intelligence?

 Simulating human thinking understood as calculation (arithmetic and logical)

Performing intellectual tasks of humans

(classification, rule creation, discovery of laws, relations...)

Acting rationally (motivated by a plan and knowledge of the environment)



# Intelligence as the performance of human intellectual tasks

Machine intelligence as the ability to acquire and process information for achieving predetermined goals

Intelligent task execution is carried out

- using heuristics, which allow for limiting the scope of searching the problem field,
- in the learning mode



Norbert Wiener (1894 – 1964)

# Intelligence as the performance of human intellectual tasks

Herbert Simon (1983): "Learning denotes changes in the system that are adaptive in the sense that they enable the system to perform the same task or similar tasks more efficiently next time"

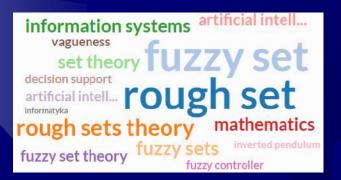


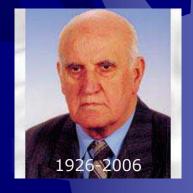
Herbert Simon (1916-2001)

 => AI understood in this way is increasingly capable of performing human intellectual tasks more efficiently than humans
 BIG DATA

#### Learning as discovering knowledge from data

 Data science specialists are increasingly able to deal with data that are incomplete, imprecise, prone to random fluctuations, and partly inconsistent





In the case of data inconsistency, rough set theory of the Polish computer scientist Zdzisław Pawlak deserves attention

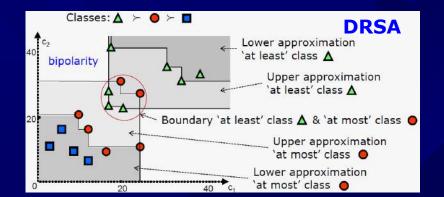
### Learning as discovering knowledge from data

Dominance-based Rough Set Approach made a breakthrough in reasoning about ordinal data, and representation of preferences by ,,*if..., then*" rules

For example, in decision under risk and uncertainty:

"if probability of gaining <u>at least 20\$ more</u> is ≥0.75, and probability of gaining <u>at least 10\$ more</u> is 1, **then** act A is better than B" instead of:

Expected utility U(A) = 0.67 > U(B) = 0.63, thus act A is better than B





# **Chieko Asakawa** – IBM Research, Tokio Example of the use of AI to assist the blind





#### The impact of AI on operational decision support

- A decision problem appears when:
  - there is an objective or objectives to be attained
  - there are many alternative ways for attaining the objectives they consititute a set of actions A (alternatives, solutions, objects, acts, ...)
- Questions with respect to set A:
  - $P_{\alpha}$ : How to choose the best action ?
  - $P_{\beta}$ : How to classify actions into pre-defined decision classes ?
  - $P_{\gamma}$ : How to order actions from the best to the worst?

	Social Choice (Group Decision)	Multiple Criteria Decision Aiding	Decision under Risk and Uncertainty
Element of set A	Candidate	Action	Act
Dimension of evaluation space	Voter	Criterion	Probability of an outcome
Objective information about comparison of elements from <i>A</i>	Dominance relation	Dominance relation	Stochastic dominance relation

The only objective information one can draw from the statement of a multi-dimensional decision problem is the dominance relation

#### Enriching dominance relation – preference modeling/learning

- Dominance relation is too poor it leaves many actions non-comparable
- One can "enrich" the dominance relation, using preference information elicited from the DM
- Preference information is an input to learn/build a preference model that aggregates the vector evaluations of actions
- The preference model induces a preference relation in set A, richer than the dominance relation (the elements of A become more comparable)
   A means tilling time of the preference relation in A leads
- A proper utilization of the preference relation in A leads to a recommendation in terms of choice, classification or ranking

#### Aggregation of multiple criteria evaluations – preference models

- Three families of **preference modeling (aggregation) methods**:
  - Multiple Attribute Utility Theory (MAUT) using a value function,

e.g.,  $U(a) = \sum_{i=1}^{n} w_i g_i(a)$ ,  $U(a) = \sum_{i=1}^{n} u_i [g_i(a)]$ , Choquet/Sugeno integral

• Outranking methods using an outranking relation  $S = \{ \sim \cup \succ^w \cup \succ^s \}$ 

a S b = "a is at least as good as b''

Decision rule approach using a set of decision rules

e.g., "If  $g_i(a) \succeq r_i \& g_j(a) \succeq r_j \& \dots g_h(a) \succeq r_h$ , then  $a \to Class t$  or higher"

"If  $g_i(a) \succeq_i^{\geq h(i)} g_i(b) \& g_j(a) \succeq_j^{\geq h(j)} g_j(b) \& \dots g_p(a) \succeq_p^{\geq h(p)} g_p(b)$ , then aSb''

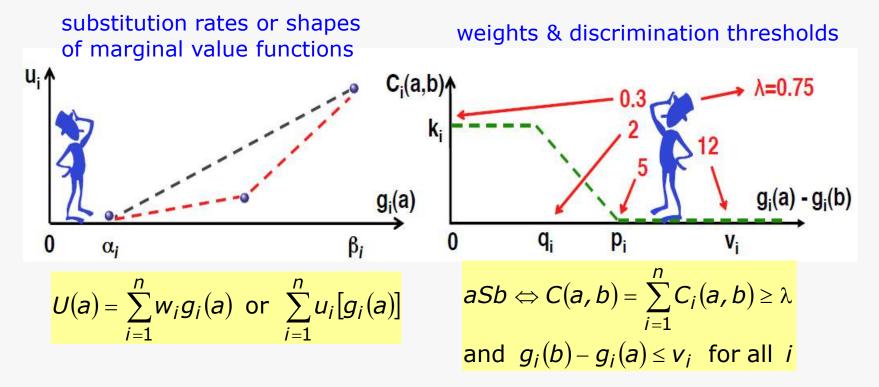
- Decision rule model is the most general of all three
- R. Słowiński, S. Greco, B. Matarazzo: Axiomatization of utility, outranking and decision-rule preference models for multiple-criteria classification problems under partial inconsistency with the dominance principle, *Control & Cybernetics*, 31 (2002) no.4, 1005-1035

#### Elicitation of preference information by the Decision Maker (DM)

- Direct or indirect ?
- Direct elicitation of numerical values of model parameters by DMs demands much of their cognitive effort
  - P.C.Fishburn (1967): Methods of Estimating Additive Utilities. *Management Science*, 13(7), 435-453 (listed and classified twenty-four methods of estimating additive utilities)

#### Value function model

Outranking model



#### The evolution of MCDA towards the paradigm of AI

- Aggregation of vector evaluations, i.e., preference modeling:
  - till early 80's: "model-centric"
     (model first, then preference info in terms of model parameters)
  - since 80's: more and more "human-centric"
     (PC allowed human-computer interaction "trial-an-error")
  - in XXI century: "knowledge driven"
     (more data are available about human choices;
     holistic preference information first, and then the model is built
     that explanains past decisions and predicts future decisions;
     similarity with the paradigm of AI: "show me your actions and AI
     will return your preference model";
    - in MCDA, model and human learn in the loop of interaction)

#### Elicitation of preference information by the Decision Maker (DM)

- Indirect elicitation: through holistic judgments, i.e., decision examples
- Decision aiding based on decision examples is gaining importance because:
  - Decision examples are relatively "easy" preference information
  - Decisions can also be observed without active participation of DMs
  - Psychologists confirm that DMs are more confident exercising their decisions than explaining them (J.G.March 1978; P.Slovic 1977)
- Related paradigms:
  - Revealed preference theory in economics (P.Samuelson 1938), is a method of analyzing choices made by individuals: preferences of consumers can be revealed by their purchasing habits
  - Learning from examples in AI/ML (knowledge discovery)
- Conclusion: indirect elicitation of preferences is more user-friendly

#### Indirect elicitation of preference information by the DM

[TIME=24, COST=56, RISK=75] ≻ [TIME=28, COST=67, RISK=25]

57, RISK=25]

pout Preference

lynamics

Pairwise preferences between actions

characterized by cardinal and/or ordinal features (criteria)

Classification

examples

Intensity of

preference

Rank related

 $[MATH=18, PHYS=16, LIT=15] \Rightarrow Class "MEDIUM" \\ [MATH=17, PHYS=16, LIT=18] \Rightarrow Class "GOOD"$ 

 $\succ$ 

Preference

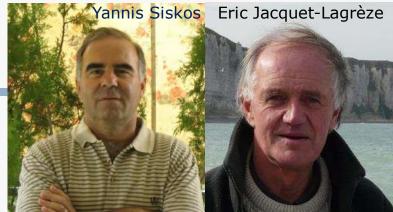
Learning

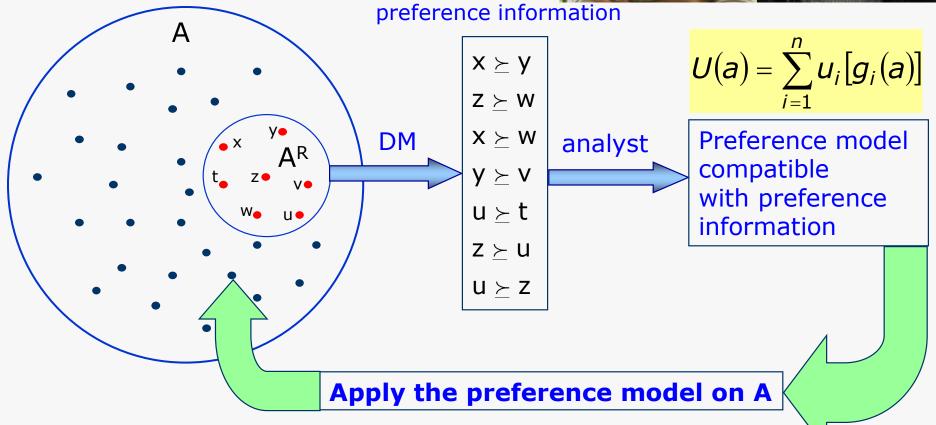
A is preferred to Z more than C is preferred to K

Action **F** should be among **5%** of the best ones

#### Ordinal regression paradigm (UTA method for ranking problem)

 Ordinal regression paradigm emphasizes the discovery of intentions expressed through holistic preference info (decision examples)





E. Jacquet-Lagrèze, J. Siskos: Assessing a set of additive utility functions for multicriteria decision-making, the UTA method. *Europ. J. Operational Research*, 10 (1982) 151-164

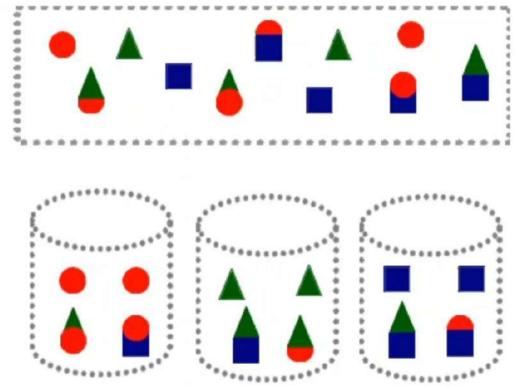
# M. Doumpos, C. Zopounidis: A multicriteria classification approach based on pairwise comparisons. *European J. Operational Research*, 158 (2004) 378-389

#### Ordinal regression paradigm (UTA method for classification problem)

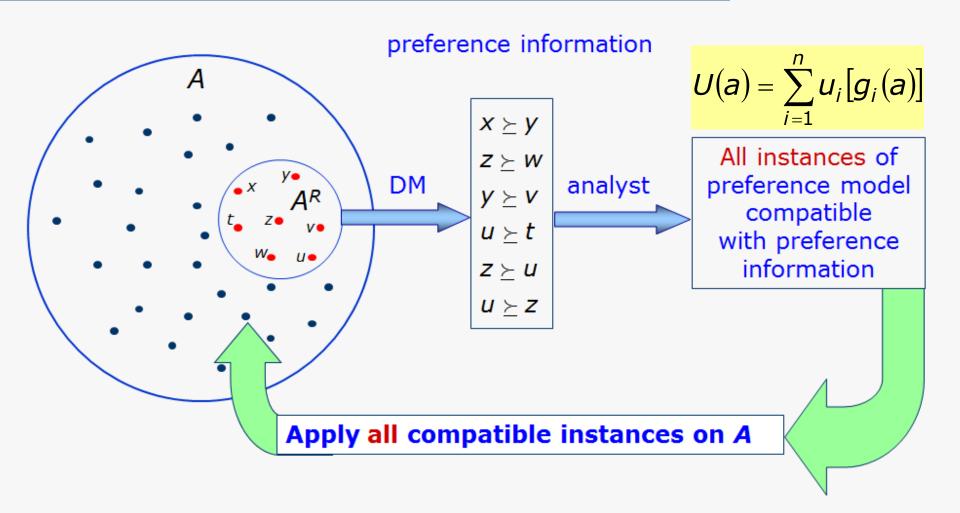
 Ordinal regression paradigm emphasizes the discovery of intentions expressed through holistic preference info (decision examples)



28



#### Non-univocal representation - Robust Ordinal Regression - UTAGMS



S. Greco, V. Mousseau, R. Słowiński: Ordinal regression revisited: multiple criteria ranking with a set of additive value functions. *European J. Operational Research*, 191 (2008) 415-435

The **possible** preference relation: for all alternatives  $x, y \in A$ ,

 $x \succeq^{P} y \Leftrightarrow U(x) \ge U(y)$  for at least one compatible value function

(complete and negatively transitive)

The **necessary** preference relation: for all alternatives  $x, y \in A$ ,

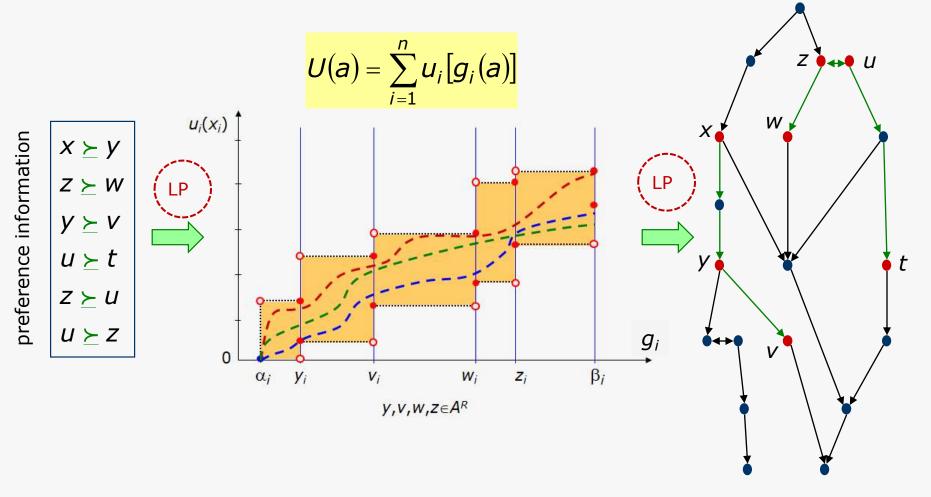
 $x \succeq^N y \Leftrightarrow U(x) \ge U(y)$  for all compatible value functions

(partial preorder)

When there is no preference information: necessary relation = dominance relation

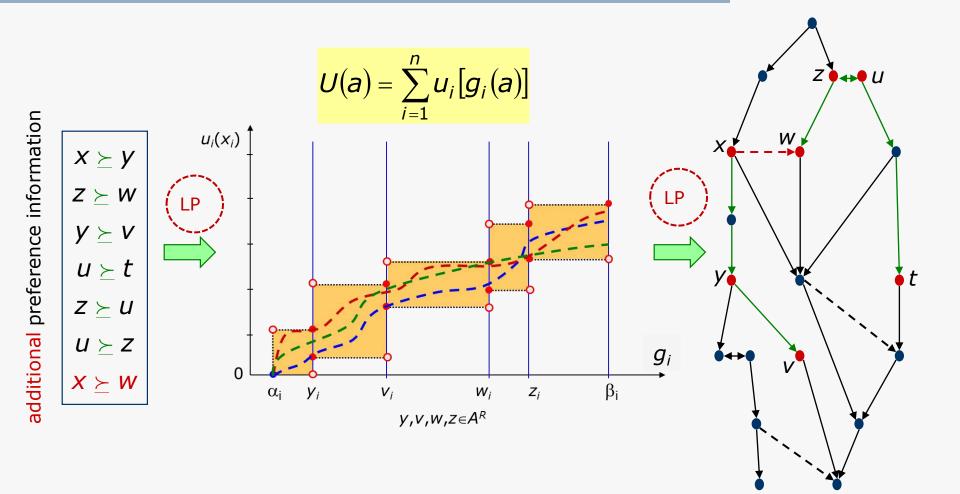
$$x \succeq^{\mathsf{N}} y \Rightarrow x \succeq^{\mathsf{P}} y,$$
  
i.e.,  $\succeq^{\mathsf{N}} \subseteq \succeq^{\mathsf{P}}$   
 $x \succeq^{\mathsf{N}} y$  or  $y \succeq^{\mathsf{P}} x$   
for all  $x, y \in A$ 

#### Non-univocal representation - Robust Ordinal Regression - UTAGMS



necessary ranking (partial preorder)

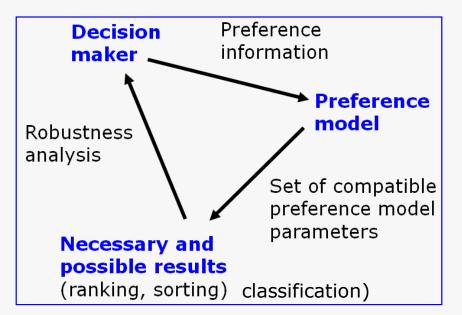
#### Non-univocal representation - Robust Ordinal Regression - UTAGMS



necessary ranking enriched

#### Robust Ordinal Regression as a constructive learning

- Robust Ordinal Regression works in a loop with incremental elicitation of preferences → constructive learning
- Results are robust, because they take into account partial preference information



S. Corrente, S. Greco, M. Kadziński, R. Słowiński: Robust ordinal regression in preference learning and ranking. *Machine Learning*, 93 (2013) 381-422





#### The evolution of MCDA towards the paradigm of AI

4OR https://doi.org/10.1007/s10288-023-00560-6

INVITED SURVEY



Eyke Hüllermeier

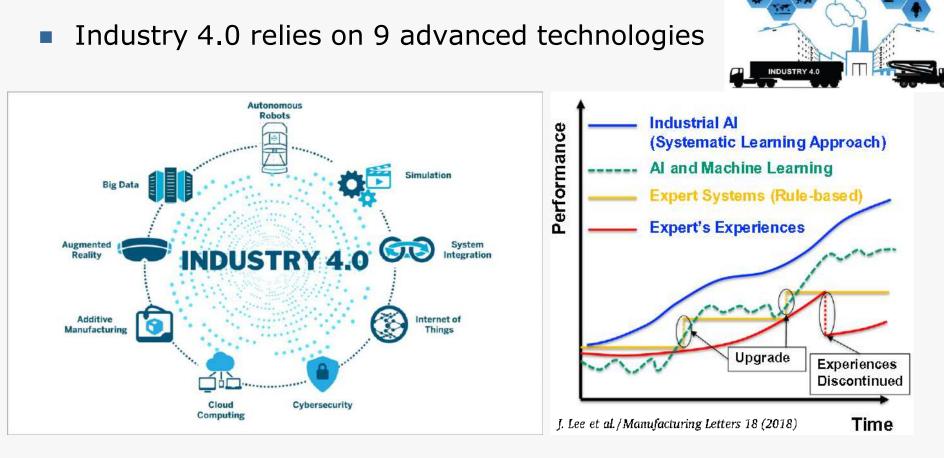
Preference learning and multiple criteria decision aiding: differences, commonalities, and synergies-part I

Eyke Hüllermeier<sup>1,2</sup> · Roman Słowiński<sup>3,4</sup>

Received: 9 August 2023 / Revised: 25 November 2023 / Accepted: 7 December 2023 © The Author(s) 2024

#### Abstract

Multiple criteria decision aiding (MCDA) and preference learning (PL) are established research fields, which have different roots, developed in different communities – the former in the decision sciences and operations research, the latter in AI and machine learning – and have their own agendas in terms of problem setting, assumptions, and criteria of success. In spite of this, they share the major goal of constructing practically useful decision models that either support humans in the task of choosing the best, classifying, or ranking alternatives from a given set, or even automate decision-making by acting autonomously on behalf of the human. Therefore, MCDA and PL can complement and mutually benefit from each other

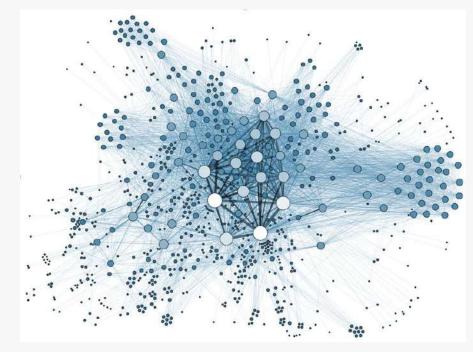


 Gathering & comprehensive analysis of data from multiple sources (devices and production systems, enterprise management systems, and customer systems are integrated into one blockchain-type system)

#### AI as an opportunity for the development of Industry 4.0 (Industry4.0 + AI = AI4Industry)

AI as an opportunity for the development of Industry 4.0 (Industry4.0 + AI = AI4Industry)

 Instead of a hierarchical structure, a system of distributed modules with connections, whose rules need to be discovered

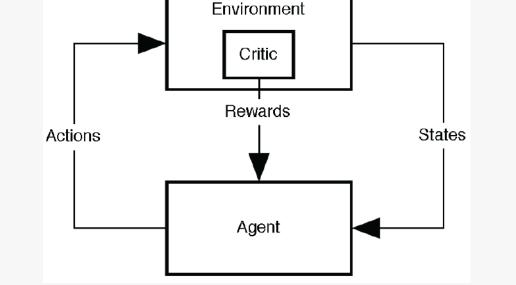


 Instead of determinism in communication rules, behavioral analysis typical of AI



AI as an opportunity for the development of Industry 4.0 (Industry4.0 + AI = AI4Industry)

 Field of applications for modern AI thanks to the modular structure of the system: module = autonomous agent



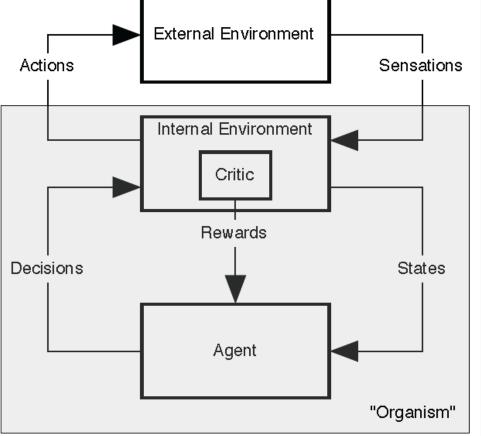
 Supervised learning with a reward function: "State-action" according to learned rules maximizing utility function (external motivation)

Intrinsically Motivated Reinforcement Learning, S.Singh, A.Barto, N.Chentanez

#### AI as an opportunity for the development of Industry 4.0 (Industry4.0 + AI = AI4Industry)

- Field of applications for modern AI thanks to the modular structure of the system: module = autonomous agent
- Reinforcement learning

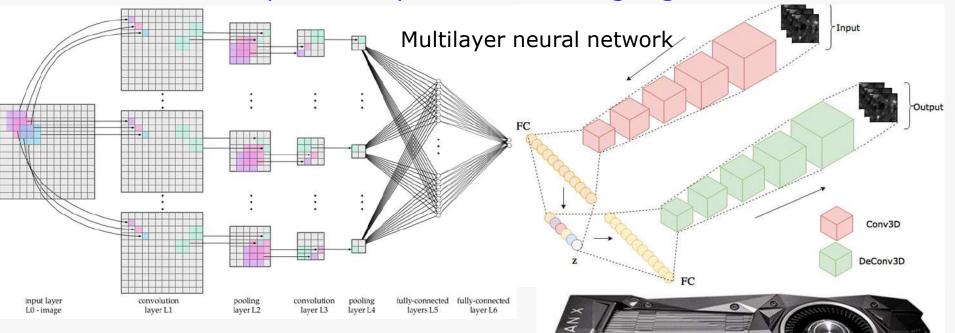
   without a reward function:
   exploration through action
   with learning from outcomes
   (intrinsic motivation)
- The agent learns less how to behave, but rather how to evaluate its own behavior and preferences





#### Deep Learning – a milestone of modern AI

 Deep learning techniques are very efficient in learning multilayer neural networks to process and understand complex patterns and relationships within pictures and language data



 Large Language Models (LLMs), like ChatGPT-4, Claude, are based on deep learning techniques

Graphics processing unit (GPU)

## Why ChatGPT and other LLMs are so efficient chatbots?

- They are trained on diverse massive amounts of text data from Internet
- LLMs are able to keep track of a conversational context.
- Generative Pre-trained Transformer is a neural network designed for learning long-range statistical dependencies in texts.
- Training a language model involves Next-Token-Prediction (NTP) and Masked-Language-Modeling (MLM) techniques.
- In the MLM technique, certain words in a sequence are "masked", and the model is trained to predict the correct words based on the context of the surrounding text.
- For example, given the input "The cat sat on...", the model estimates the probability of a possible next word, like "sofa", "floor", or "roof", considering the frequency of occurrence of these words in other texts.



#### Why ChatGPT and other LLMs are so efficient chatbots?

- They are trained on diverse massive amounts of text data from Internet
- LLMs are able to keep track of a conversational
- Generative Pre-trained Transformer is 2 Ne incies in texts. designed for learning long-range st
- Training a language mode' Masked-Language-Mr

anguage-M- chat for Jiniques. 1+ in an words in a sequence are "masked", and i.edict the correct words based on the context Jeff Jext. In the MLM + the mor' of the

For exam, ..., given the input "The cat sat on...", the model estimates the probability of a possible next word, like "sofa", "floor", or "roof", considering the frequency of occurrence of these words in other texts.

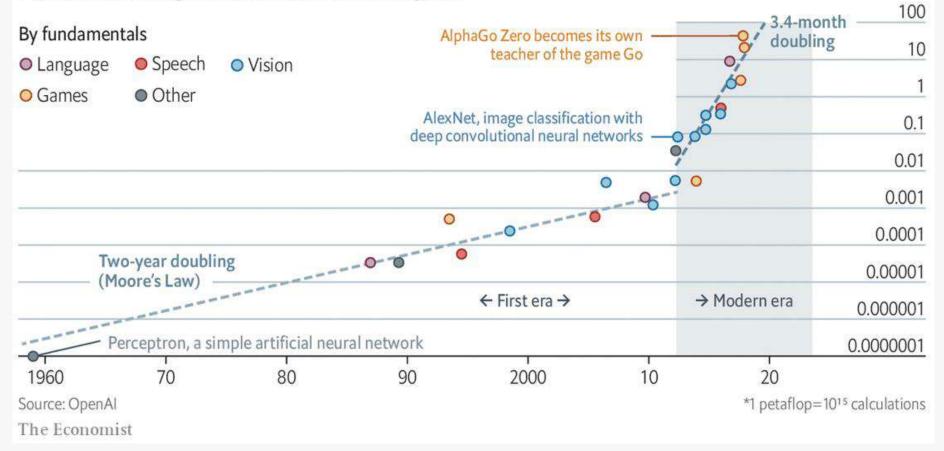
#### The cost of training Large Language Models

- Deep learning models are "brute force" statistical techniques that learn on vast amounts of data.
- Neural networks of LLMs might contain hundreds of layers and trillions of weights, making the computations incredibly energy intensive.
- ChatGPT has 1.7 trillion machine learning parameters.
- Training the GPT-4 model employs 25,000 GPUs during 90-100 days for \$100 milion, while energy is between 51,773 and 62,319 MWh\* (equivalent to energy consumed by 1,000 US households during 5-6 years)
- Inference costs and power usage are 10x higher than training costs (in 1 month, ChatGPT consumed as much electricity as 26,000 US households)
- High energy costs make that LLMs are developed by rich companies and not university labs.

#### The cost of training AI Systems

 Since 2012, the computational resources needed to train AI systems have been doubling every 3.4 months\*

Computing power used in training AI systems Days spent calculating at one petaflop per second\*, log scale



#### \*https://openai.com/research/ai-and-compute

# What is machine intelligence?

 Simulating human thinking understood as calculation (arithmetic and logical)

Performing intellectual tasks of humans

(classification, rule creation, discovery of laws, relations...)

Acting rationally (motivated by a plan and knowledge of the environment)



- Can a machine achieve a state of consciousness that allows it to become aware of its relation with environment and say: "I am"?
- "Me" in relation to objects is easier for a machine to realize than "me" in relation to other people, the "me-you".
- Social and emotional relations, feeling shame, or gratitude are extremely difficult to simulate.



Sicilian city with orange plantation by MS Bing



", Mindar" preaching in a Buddhist temple



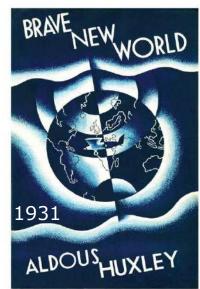
- Moreover, concepts such as gratitude or hope are concepts that transcend the reality of 'here and now', and therefore, they possess an element of transcendence characteristic of human thought and prayer.
- They give rise to a special kind of relationship: "Me-God"
- Will machines evolve to a similar consciousness?
- In my opinion, machines will take over human consciousness in a behavioral sense only - they will learn our preferences and reactions.
- If machines create their own world of 'thoughts', we will not have access to it, just as we don't have access to the 'thoughts' of our dog - it will simply be a different intelligence.

- Nick Bostrom predicts an intelligence explosion after reaching the point of technological singularity where "algorithms will independently undertake the task of conscious self-development and determine its directions"
- Elon Musk (Tesla, startup "Neuralink")
   undertakes to map human brain into computer (human brain in a machine ⇒ strong AI)
- Extracting memories from the biological brain would mean eternal life





- AI prophets propose new quasi-religions:
   "God has been put to death and man has taken his place"
- Yuval Harari's concept of dataism (author of *"Homo deus"*) is a belief system centered around the power and value of data.
- Dataism could ultimately lead to a shift in authority from human decision-making to algorithmic processing, as algorithms might better interpret complex data patterns.
- This is a clear threat to human free will
- AI is a challenge for profound philosophical and religious discussions regarding the essence of humanity



- Transhumanism: "Intelligence wants to be free but everywhere is in chains. It is imprisoned by biology and its inevitable scarcity." (Transhumanist Manifesto)\*
- "Transhumanists of the world unite we have immortality to gain and only biology to lose. Together, we can break through the chains of biology and transcend scarcity, sex, age, ethnicity, race, death and, perhaps, even time and space"
- Julian Huxley (1887–1975): "The human race, if it so desires, can achieve self-transcendence"

\*<u>https://www.singularityweblog.com/a-transhumanist-manifesto/</u> \*<u>https://www.humanityplus.org/</u>

# TRANSCENDENCE

#### YESTERDAY DR. WILL CASTER WAS ONLY HUMAN

ALCON ENTERTAINMENT PRESENTS IN ASSOBATION WITH DIMG ENTERTAINMENT A STRAIGHT UP FILMS PROJECTION A FLIM BY WALLY FRISTER Johnny Depp Morgan Freeham "Transcendence" freeestahall kate Mara Cillian Indepry Cole Hauser and Paul Bettany ""By Mychael Danna Arendes Deva Anderson Beses Georgel Little ""By David Rosenblodmake ""Beses Christopher Nolan Internation ""By Mychael Danna Arendes Deva Anderson Beses Georgel Little ""By David Rosenblodmake ""Beses Christopher Nolan Into "Rosenberg Jess Hall BSC Proveds Volanda t Cochran Steven P Werner Regency Bolis Arendes Christopher Nolan Into Rosenberg Jess Hall BSC Proveds Volanda t Cochran Steven P Werner Regency Bolis Arendes Christopher Nolan Into Rosenberg Albone Broderick Johnson Kate Cohen Marisa Polvino Annie Marter David Valdes Aaron Ryder ""By Jack Pagled Desemp Wally Physter Alcon

## TRANSCENDENCE

Tanana ana

Autonomous taxi Phoenix, AZ, 2023

#### Is democracy threatened by algocracy due to AI ?

"Scoring" citizens - controversial governmental practices involving the monitoring and assessing of citizens



#### Is democracy threatened by algocracy due to AI ?

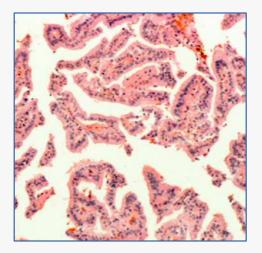
#### Found in a crowd of 60,000 people at a pop concert



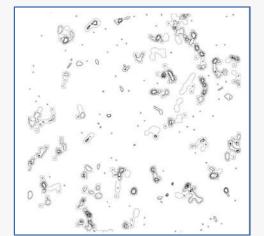
#### Explainable AI models

 The functioning of algorithms and the decisions they suggest should be understandable to humans, even if AI may utilize its own features.

Human feature: *"tiger skin"* 



AI feature: fractal lengths of contour lines of kernel densities



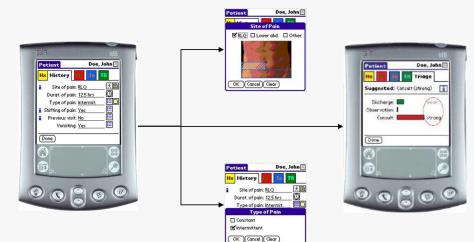
 The cognitive architecture of AI will be quite different from processes of biological reasoning.

### Explainable AI models

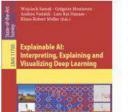
 The functioning of algorithms and the decisions they suggest should be understandable to humans, e.g., in medicine



if Sex = male and PainSite = lower\_abdomen
and PainType = constant and RebTend = yes
and WBCC ≥ 12 then Triage = consult



- The outcomes of AI operation in robotics must be predictable.
- In some applications, certification is necessary.

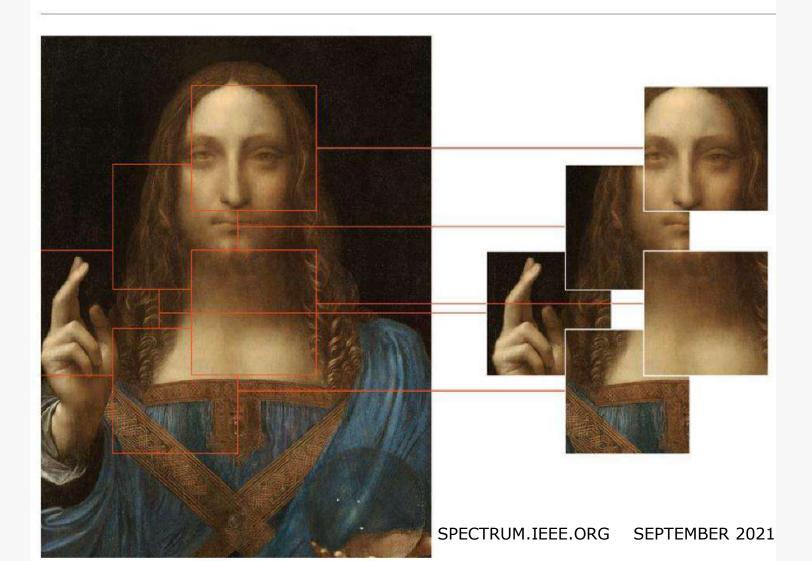


Explainable AI: Interpreting, Explaining and Visualizing Deep Learning

https://mobiledss.cs.put.poznan.pl/research/current-research

This convolutional neural network can tell you whether a painting is a fake

BY STEVEN J. FRANK





Probability Maps: Steven and Andrea Frank

The author taught a convolutional neural network the styles of Rembrandt and Leonardo, and applied it to the analysis of *Saskia*, Rembrandt's wife, to *Man in the Golden Helmet* from the circle of Rembrandt, to *Salvator Mundi*.

The warmer the color, the higher the probability that the CNN attributes a given fragment to the presumed author.

#### Can machines, through AI, incapacitate humans?

- YES, if uncritically and without understanding the machine's answers, we would rely on its decisions – then, AI will limit HUMAN FREE WILL
- Technological progress requires the growth of humans, not only in terms of their knowledge but also in spirit, which through conscience helps in discerning good from evil.









Thank you for your attention

Let us not fear AI, but do not neglect the development of our own intelligence and RIGHT CONSCIENCE