TRIPLEHELIX CONFERENCE MONTERREY2025

IA 4 AI

Innovation Assessment 4 Artificial Intelligence

Monterrey 28-31 October, 2025

Riccardo Viale

President **Behavioral Insights Bicocca-Ciseps, Università di Milano Bicocca**

Scientific Head **Italy Behavioral Insights Team, Prime Minister Department of Civil Services**

President Scientific Committe **COTEC-Fondazione per l'Innovazione**

with DOMINIQUE FORAY (Lousanne Polytechnique), LUC SOETE (Maastricht University) and MERLE JACOB (Lund University)

SNAPSHOT

- A) WHAT IS INNOVATION ASSESSMENT AND WHY IS IT DIFFERENT FROM TECHNOLOGY ASSESSMENT?
- B) WHY DO WE NEED INNOVATION ASSESSMENT OF CURRENT AI?
- C) THE CHANGE FROM PSYCHOLOGY BASED SIMULATIVE AI TO CONNECTIONIST EMULATIVE AI HAS GENERATED UNCERTAINTY ABOUT THE DEVELOPMENT OF AI.
- D) HOW TO DO INNOVATION ASSESSMENT OF ARTIFICIAL INTELLIGENCE

REFERENCES

1) Dominique Foray, Merle Jacob, Luc Soete and Riccardo Viale (2025)

"**On the need for a new innovation assessment discipline: Artificial Intelligence as unbounded innovation"**Paper prepared for the COTEC Workshop, Rome, 27th March 2025.

2) Valerio Capraro, Austin Lentsch, Daron Acemoglu, Selin Akgun, Aisel Akhmedova,Ennio Bilancini, Jean-François Bonnefon, Pablo Brañas-Garza, Luigi Butera, Karen M.Douglas, Jim A.C. Everett1, Gerd Gigerenzer, Christine Greenhow, Daniel A.Hashimoto, Julianne Holt-Lunstad1, Jolanda Jetten, Simon Johnson, ChiaraLongoni, Pete Lunn, Simone Natale, Iyad Rahwan, Neil Selwyn, Vivek Singh,Siddharth Suri, Jennifer Sutcliffe, Joe Tomlinson, Sander van der Linden, Paul A. M.Van Lange, Friederike Wall, Jay J. Van Bavel and Riccardo Viale (2024).

«**The impact of generative artificial intelligence on socioeconomic inequalities and policy making**», *PNAS Plus* Volume 3, Issue 6, June 2024, pgae191

A) INNOVATION ASSESSMENT

Understanding innovation

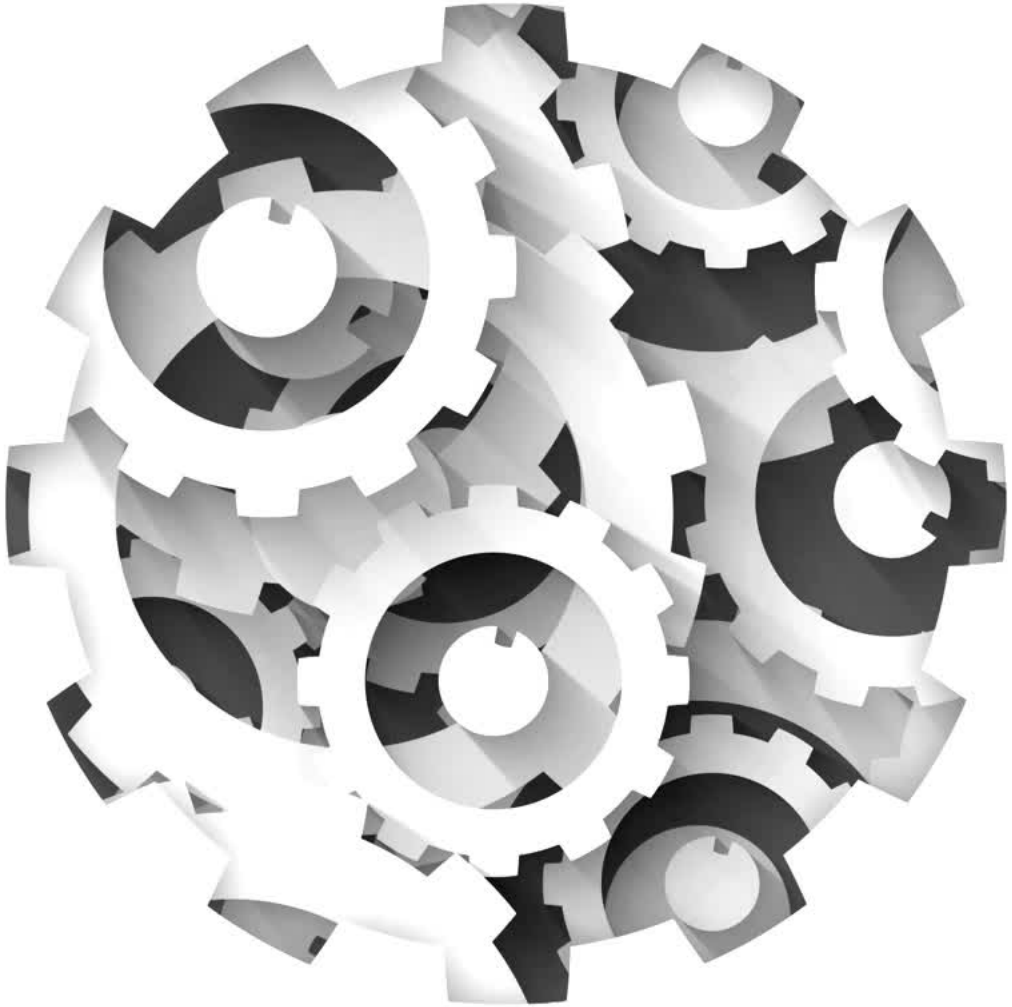
Proposition 1

- Understanding an innovation requires its full deployment and full deployment allows one to discover:
  - The ‘’economic properties’’ of the innovation such as new and breakthrough economic knowledge
    - For example when innovation emerges at the intersection between the economics of privately owned mid-range goods (such as car), multisided platforms and digital technologies, significant economic knowledge is generated
  - The ‘’social properties’’
    - Only real experiments in society will help to understand the magnitude of negative externalities, creative destruction and destructive creation as well as the positive effects

Innovation assessment: Proposition 2

- Assessing innovation should in most cases be done *ex post* – e.g. to let the considered innovation to be fully deployed and adopted within a significant time frame

  - *IA is a measure by which policy abstains from restricting the evolution of a particular intervention until potential differential impacts are discernible*



IA emphasizes the **social properties of innovation**: how it affects identity, trust, inclusion, equality, fairness, and power dynamics.

Some innovations generate **creative destruction**, improving efficiency while displacing older systems (from Marx and Schumpeter to Aghion and Howitt).

Others create **destructive creation**, where ultimately social goods will be lost (e.g., affordable housing, cultural norms, civic trust) without clear overall societal benefits.

- IA helps distinguish between these two trajectories and offers a basis for **normative judgment** in regulation.

IA *versus* TA: Proposition 3

IA is a new, complementary intervention that is different from TA

TA is a policy project which is dedicated to solve ‘’objective problems’’ of security, safety, environmental and public health issues.

- It carries out ‘’services’’ of certification through laboratory tests and experiments
- TA is undertaken *ex ante* before the technology enters the economy

IA is intended to function as a complementary policy measure that would allow policy makers to observe innovation deployment, assess effects and impacts and eventually regulate ex-post

IA dilemmas: Proposition 4

IA as an *ex post* policy process involves a series of dilemmas

- These dilemmas do not all apply to TA

Speed and time

- when IA qua ex post policy is acceptable and recommendable (what is truly irreversible in terms of effect, what is not)
- to what extent the speed of IA varies across cases - depending upon of the severity of the social problems that are created during deployment

The blind giant's quandary (Paul David)

- *Ex ante* the giant has a lot of power to regulate but s/he is blind, *ex post s/*he understands the problem well but has much less power

Proposition 4 – cont'd

What is good, what is bad?

- The nature of IA makes the exercise highly ambiguous (the opposite of TA) and it is often difficult to identify between the bad and the good innovations

IA external validity is poor

- Contrary to TA, the IA knowledge and findings are local and context-dependent: some economic/social properties which are desirable somewhere are undesirable in another place or at another time
- IA findings are subject to temporal limitations i.e what is produced at a certain time and in certain circumstances may not be replicated at another time or can lead to poor results
- The implication is while TA knowledge does not need to be produced more than once, this is not the case for IA knowledge – thus the production of IA information is not like a fixed cost in production (and so is likely to be a weak source of scale economies in IA operations)

Institutions and Key Domains

Proposition 5

IA institutional framework

Agencies which observe, support ex post assessment, decide about regulatory principles – e.g EU HTA (health technology assessment),Office of technology evaluation (OTE) testbeds, demonstration projects

Sandbox – enabling to test innovation in a controlled ‘’real-world’’ environment (cases of innovation for which potentially negative social effects can be predicted – in finance)

Institutional Tools

- To operationalize IA, public authorities need new tools and institutional capacities:
- **1) REGULATORY SANDBOXES**
  These are **controlled real-world environments** where innovators can test products or services **under temporary regulatory relaxations**.
  Key characteristics include:
- Predefined timeframes and scope
- Real-time data sharing with regulators
- Clear exit strategies and evaluation criteria
- Sandboxes are particularly useful in sectors like **fintech, mobility, and public services**, where innovation interacts heavily with regulation. However, for **rapidly evolving and general-purpose technologies** like Generative AI, traditional sandboxes may lack the agility required.

- **2) OBSERVATION AND MONITORING UNITS**
  A more sustainable approach is to create **dedicated IA observatories or units** within ministries, agencies, or statistical offices. These can:
- Track innovation diffusion and usage metrics
- Gather citizen and stakeholder feedback
- Assess long-term impacts on employment, equity, and trust
- Such units could be modelled after existing foresight or policy evaluation bodies, but with a focus on **post-deployment innovation governance**.

Key Assessment Domains

A robust IA framework should cover the following domains:

- **Social Inclusion and economic equality**: Access disparities, bias amplification, cultural impact, increased inequality and digital divide.
- **Legal and Institutional Fit**: Compatibility with existing laws, regulatory gaps, and needs for reform.
- **Environmental Impact**: Resource intensity, emissions, lifecycle effects (especially for digital and AI tools).
- **Cultural impact**: Behavioral changes, risk propensity, time discounting, social norms, overconfidence.
- **Governance and Ethics**: Accountability, transparency, user autonomy, and fairness.

These domains can be weighted and applied differently depending on the innovation type—e.g., fintech vs. public health AI.

TO SUMMARIZE

IA a new discipline (different from TA)

- - IA enables policy learning after full-scale deployment because it allows insights on the materialization of real economic and social impacts before introducing regulation.
- - It allows for entrepreneurial innovative dynamism while preserving the possibility of post-deployment regulation.
- - It is essential for responsible Innovation governance and for guiding public and private innovation in democratic societies.

WHY Application to AI?

B) WHY DOES AI NEED IA?

- Because current AI technology is too fast to allow Technology Assessment (the evolution of deep learning towards always more powerful LLM)
- Because current AI technology is too uncertain to do Technology Foresight (previous examples: the failure of the foresight of Fifth Generation Computer Systems in Japan in 1982)
- Because current AI technology has too complex systemic impact to rely only on sandbox live tests (the deep learning impact on cyber security system or the LLM effect on education)

The role of INNOVATION ASSESSMENT of current AI is particularly justified:

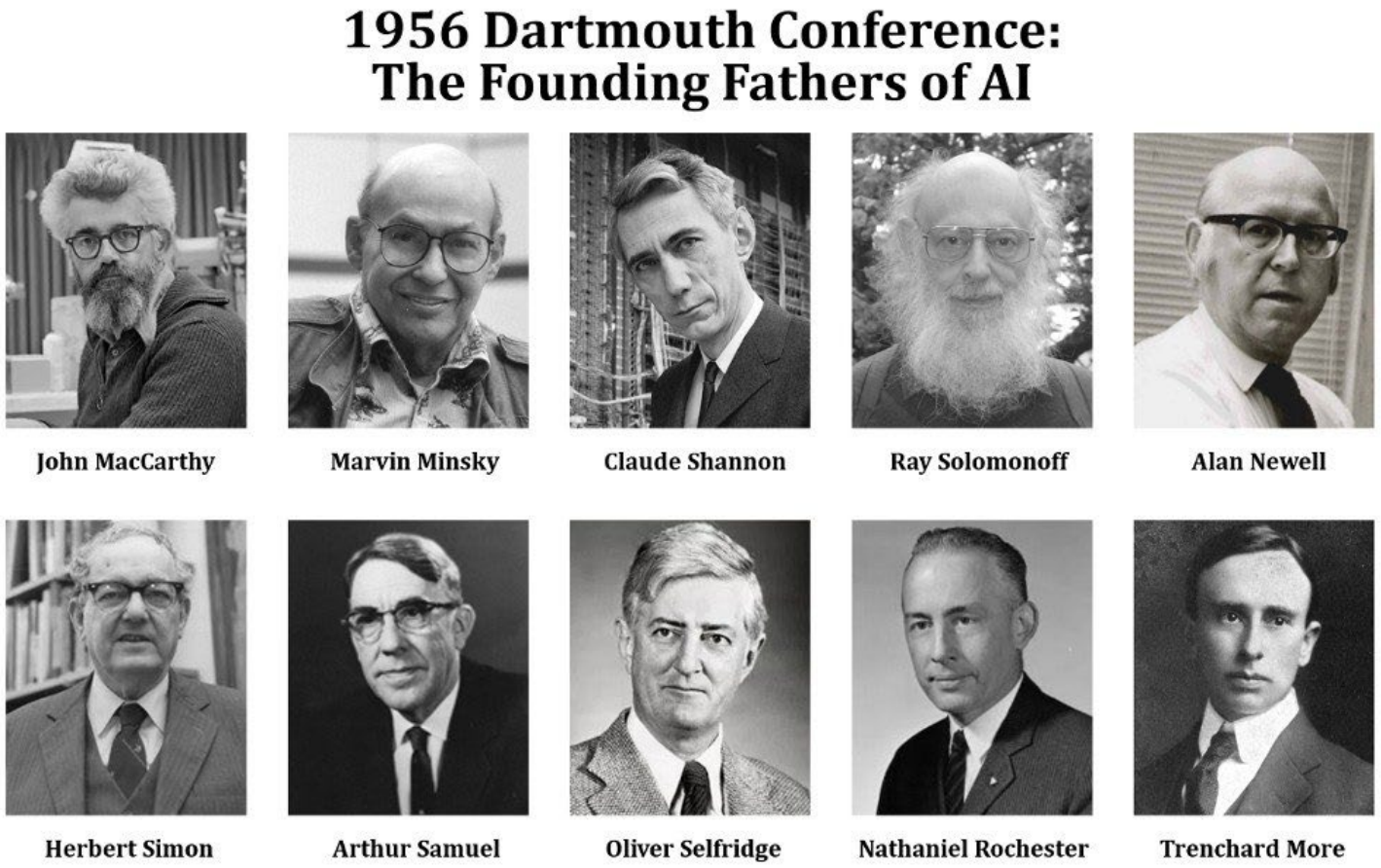
for the umpredicatable and opaque features of connectionist emulative AI

that is very different from the predicatable and transparent features of psychology based simulationist AI

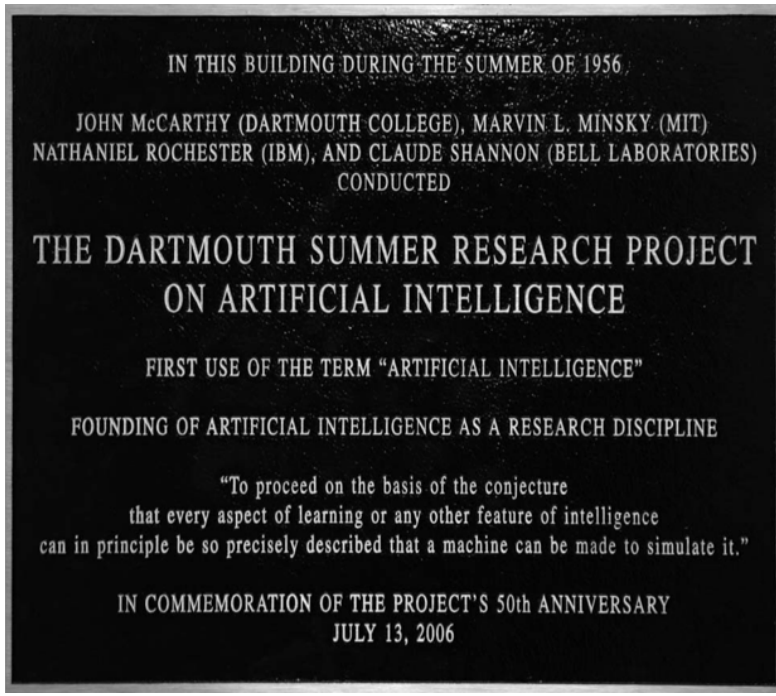
WHY? BECAUSE OF A RADICAL CHANGE OF AI

2 MAIN PROGRAMMES OF AI

1) SIMULATIVE AI



THE DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE


IN THIS BUILDING DURING THE SUMMER OF 1956
JOHN McCARTHY (DARTMOUTH COLLEGE), MARVIN L. MINSKY (MIT)
NATHANIEL ROCHESTER (IBM), AND CLAUDE SHANNON (BELL LABORATORIES)
CONDUCTED
THE DARTMOUTH SUMMER RESEARCH PROJECT
ON ARTIFICIAL INTELLIGENCE
FIRST USE OF THE TERM "ARTIFICIAL INTELLIGENCE"
FOUNDING OF ARTIFICIAL INTELLIGENCE AS A RESEARCH DISCIPLINE
"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."
IN COMMEMORATION OF THE PROJECT'S 50th ANNIVERSARY
JULY 13, 2006

THE BIRTH OF ARTIFICIAL INTELLIGENCE

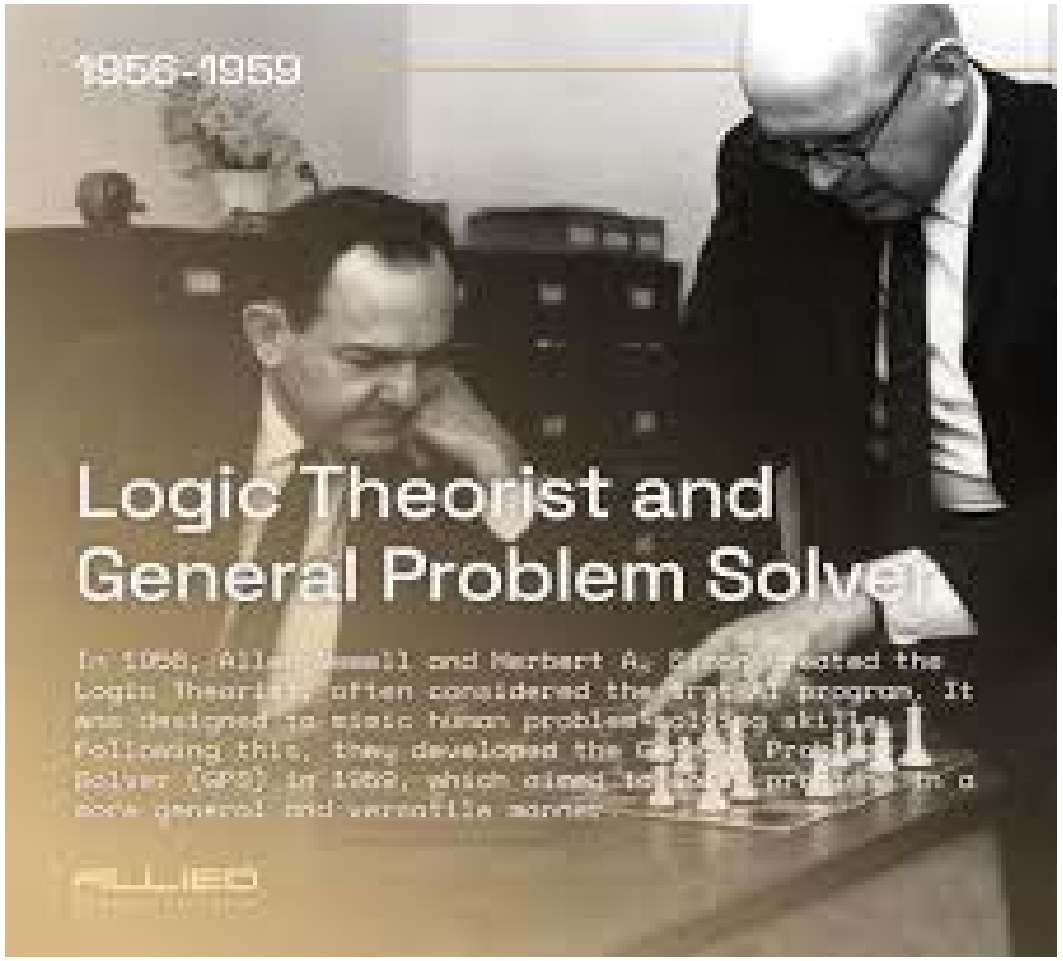
(SIMON, MINSKY, NEWELL, MC CARTHY,… Dartmouth Conference, 1956)

- AI tries to simulate and implement in the computer by advanced software the most important cognitive and sensory human functions: planning, language understanding, sensory recognition, learning, problem solving, etc..

SIMULATIVE ARTIFICIAL INTELLIGENCE:

- DESCRIPTIVE PROTOCOLS FOR COMPUTER SIMULATION OF HUMAN PROBLEM SOLVING AND THINKING
- INFORMATION PROCESSING PSYCHOLOGY
- SUCCESSES: Expert Systems, Problem Solving, Ex Post Simulation of Scientific Discoveries.
- FAILURES: machine translation, image and speech processing, artificial movement….

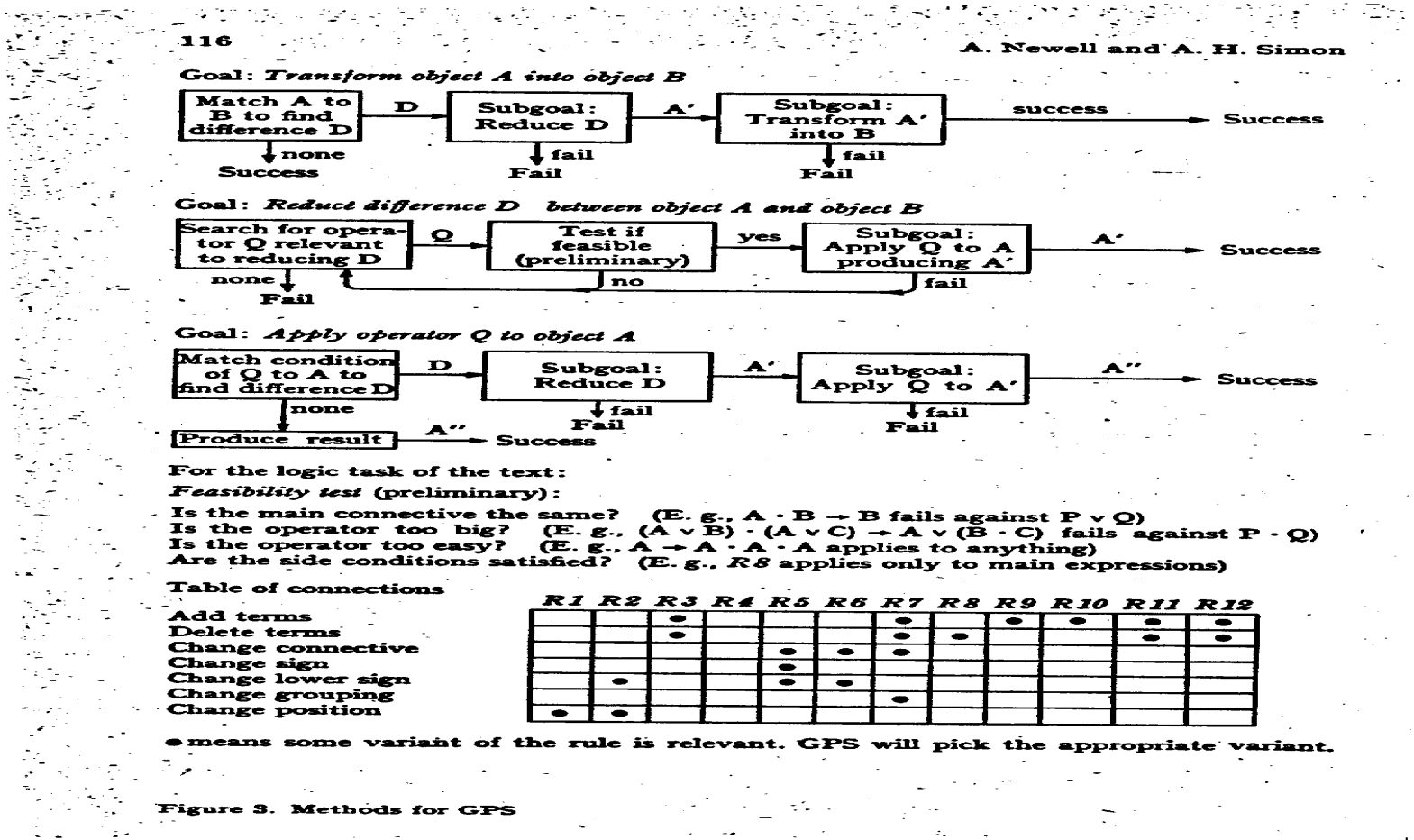
GREAT SUCCESSES: LOGIC THEORIST AND GPS OF HERBERT SIMON and ALAN NEWELL



- **Logic Theorist** is a computer program written in 1956 by Allen Newell, Herbert . Simon, and Cliff Shaw. It was the first program deliberately engineered to perform automated reasoning, and has been described as "the first artificial intelligence program".
- Logic Theorist proved 38 of the first 52 theorems in chapter two of Whitehead and Bertrand Russell's *Principia Mathematica*, and found new and shorter proofs for some of them.
- Logic Theorist explored a search tree: the root was the initial hypothesis, each branch was a deduction based on the rules of logic.

- Newell and Simon realized that the search tree would grow exponentially and that they needed to "trim" some branches, using "rules of thumb" to determine which pathways were unlikely to lead to a solution. They called these *ad hoc* rules "heuristics", using a term introduced by George Pólya in his classic book on mathematical proof, *How to Solve It.*
- Heuristics would become an important area of research in artificial intelligence and remains an important method to overcome the intractable combinatorial explosion of exponentially growing searches.

General Problem Solver (GPS) is a computer program created in 1957 by Herbert A. Simon, J. C. Shaw, and Allen Newell (RAND Corporation) intended to work as a universal problem solver machine. In contrast to the former Logic Theorist project, the **GPS** works with means–ends analysis heuristic.



Goal: *Transform object A into object B*

Goal: *Reduct difference D between object A and object B*

Goal: *Apply operator Q to object A*

For the logic task of the text:

Feasibility test (preliminary):

Is the main connective the same? (E. g., A · B → B fails against P ∨ Q)
Is the operator too big? (E. g., (A ∨ B) · (A ∨ C) → A ∨ (B · C) fails against P · Q)
Is the operator too easy? (E. g., A → A · A · A applies to anything)
Are the side conditions satisfied? (E. g., *R8* applies only to main expressions)

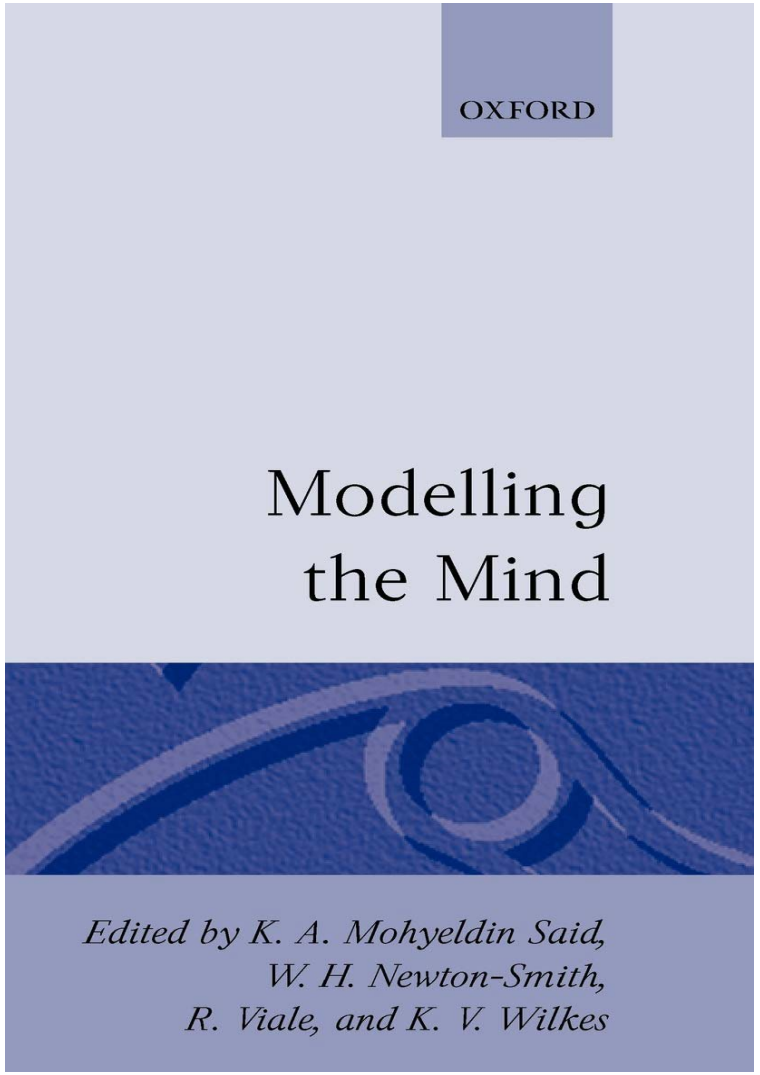
Table of connections

	R1	*R2*	*R3*	*R4*	*R5*	*R6*	*R7*	*R8*	*R9*	*R10*	*R11*	*R12*
Add terms			•				•		•	•	•	•
Delete terms			•				•	•			•	•
Change connective					•	•	•					
Change sign					•							
Change lower sign		•			•	•						
Change grouping							•					
Change position	•	•										

• means some variant of the rule is relevant. GPS will pick the appropriate variant.

Figure 3. Methods for GPS

Riccardo Viale (ed.) Feltrinelli (1987) and Oxford Uuniversity Press (1989): Contributions of Minsky, Dennett, Pribram, Poggio, Davidson, Braitenberg, Johnson Laird, Boden, Blakemore, Wilkes, etc..

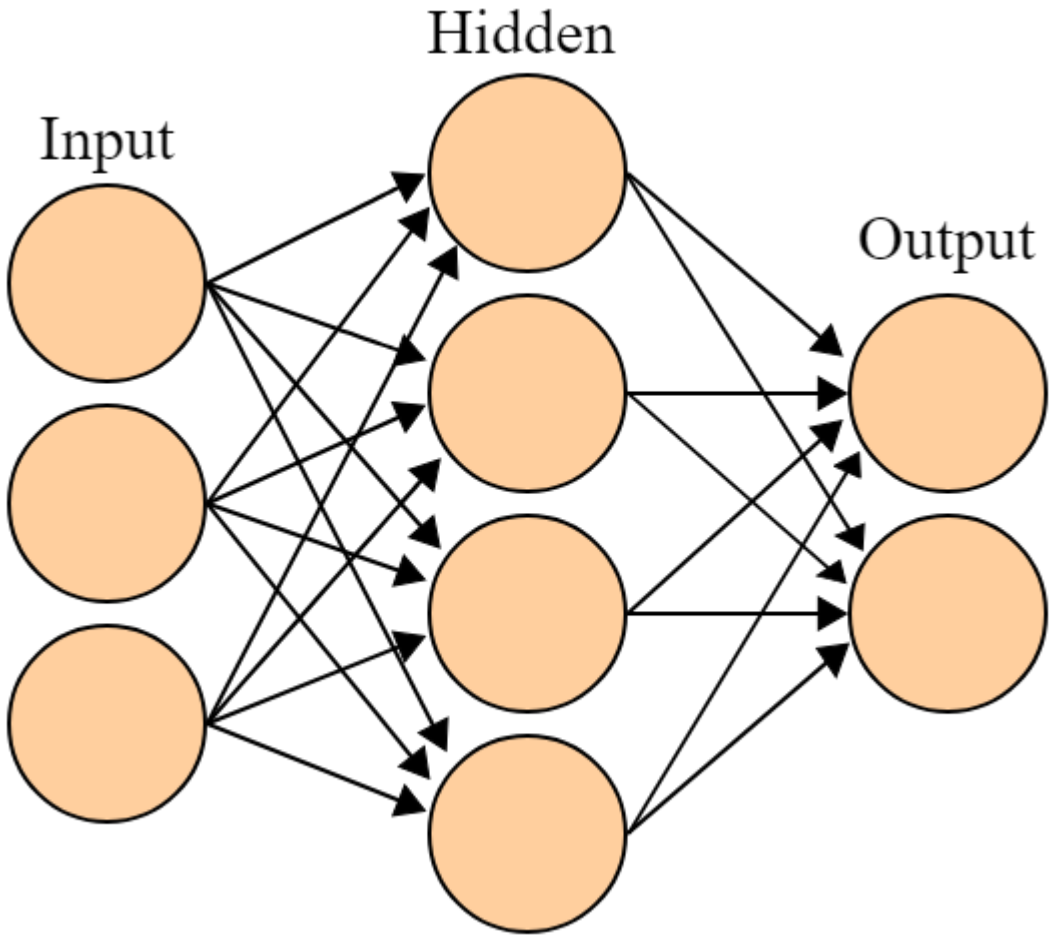


MAIN FEATURES of SIMULATIVE AI

- AI is psychology based: it simulates how the humans reasons and solve problems.
- It relies on a Cartesian dualism between mind and body (software and hardware): Detachment from the embodied dimension
- Transparent
- Predictable

2) EMULATIVE AI: CONNECTIONISM (3WAVES)

CONNECTIONISM is an approach to the study of human mental processes and cognition that utilizes mathematical models known as connectionist networks or artificial neural networks (ANN)



- **THE FIRST WAVE** appeared 1943 with Warren Sturgis McCulloch and Walter Pitts both focusing on comprehending neural circuitry through a formal and mathematical approach, and Frank Rosenblatt who published the 1958 paper "The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain" in *Psychological Review*, while working at the Cornell Aeronautical Laboratory.
- The first wave ended with the 1969 book about the limitations of the original perceptron idea, written by Marvin Minsky and Seymour Papert, which contributed to discouraging major funding agencies in the US from investing in connectionist research.
- With a few noteworthy deviations, most connectionist research entered a period of inactivity until the mid-1980s.
- The term *connectionist model* was reintroduced in a 1982 paper in the journal *Cognitive Science* by Jerome Feldman and Dana Ballard.

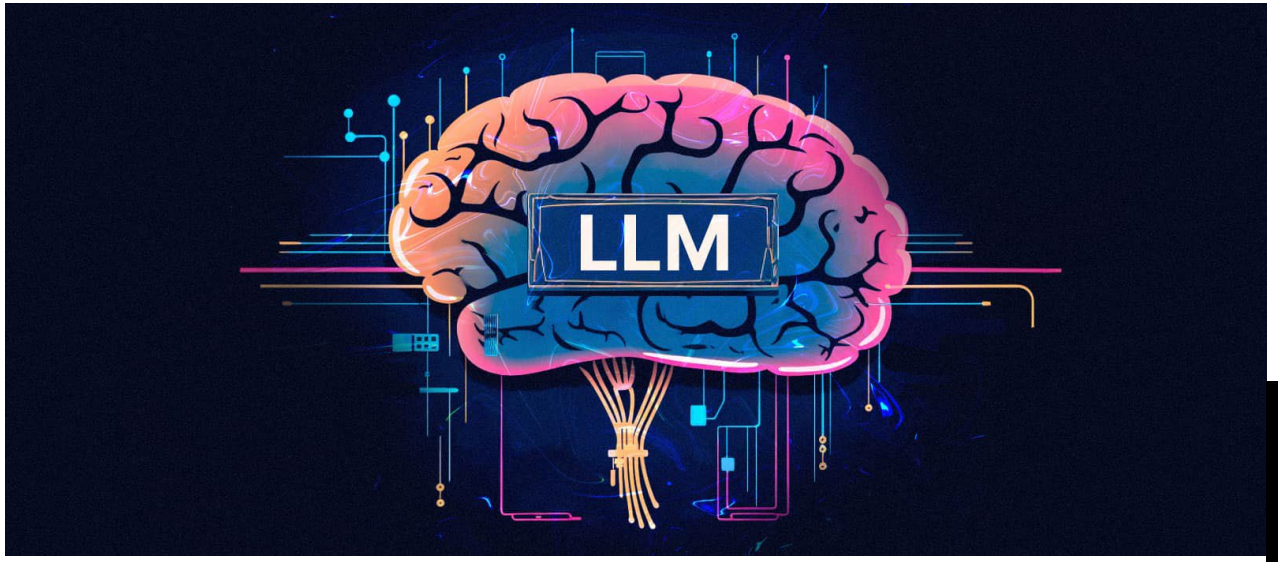
- The **SECOND WAVE** blossomed in the late 1980s, following a 1987 book *Parallel Distributed Processing* by James L. McClelland, David E. Rumelhart et al., which introduced a couple of improvements to the simple perceptron idea, such as intermediate processors (now known as "hidden layers") alongside input and output units, and used a sigmoid non linear activation function (between 0 and 1) instead of the old "all-or-nothing" linear function (0 V 1).
- Some advantages of the second wave connectionist approach included its applicability to a broad array of functions, structural approximation to biological neurons, low requirements for innate structure, and capacity for graceful degradation.
- Its disadvantages included the difficulty in deciphering how ANNs process information or account for the compositionality of mental representations, and a resultant difficulty explaining phenomena at a higher level.

- The **CURRENT (THIRD) WAVE** has been marked by advances in deep learning, with the introduction of ReLU (Rectified Linear Unit) rectified activation function that is a non-linear activation function that outputs the input if it's positive, and zero otherwise.
- ReLu allows to learn complex relationships in data that a simple linear model cannot. It has made possible the creation of LARGE LANGUAGE MODELS that learn through statistical inference on Big Data

- The success of deep-learning networks in the past decade has greatly increased the popularity of this approach, but the complexity, opacity, unpredictability and scale of such networks has brought with them increased interpretability, and managing problems.

From LLM to SPIKINGBRAIN: A FOURTH WAVE?

It mimics the human brain by using "spikes"—discrete, event-driven signals—instead of continuous data processing. This approach enables SpikingBrain models to be more energy-efficient and faster, particularly with long sequences of text.



MACHINE LEARNING

INPUT

FEATURE EXTRACTION

CLASSIFICATION

CAR
NOT CAR

OUTPUT

DEEP LEARNING

INPUT

HIDDEN
NEURONS

FEATURE EXTRACTION + CLASSIFICATION

CAR
Color: Red
Make: Ford
Model: Mustang
NOT CAR

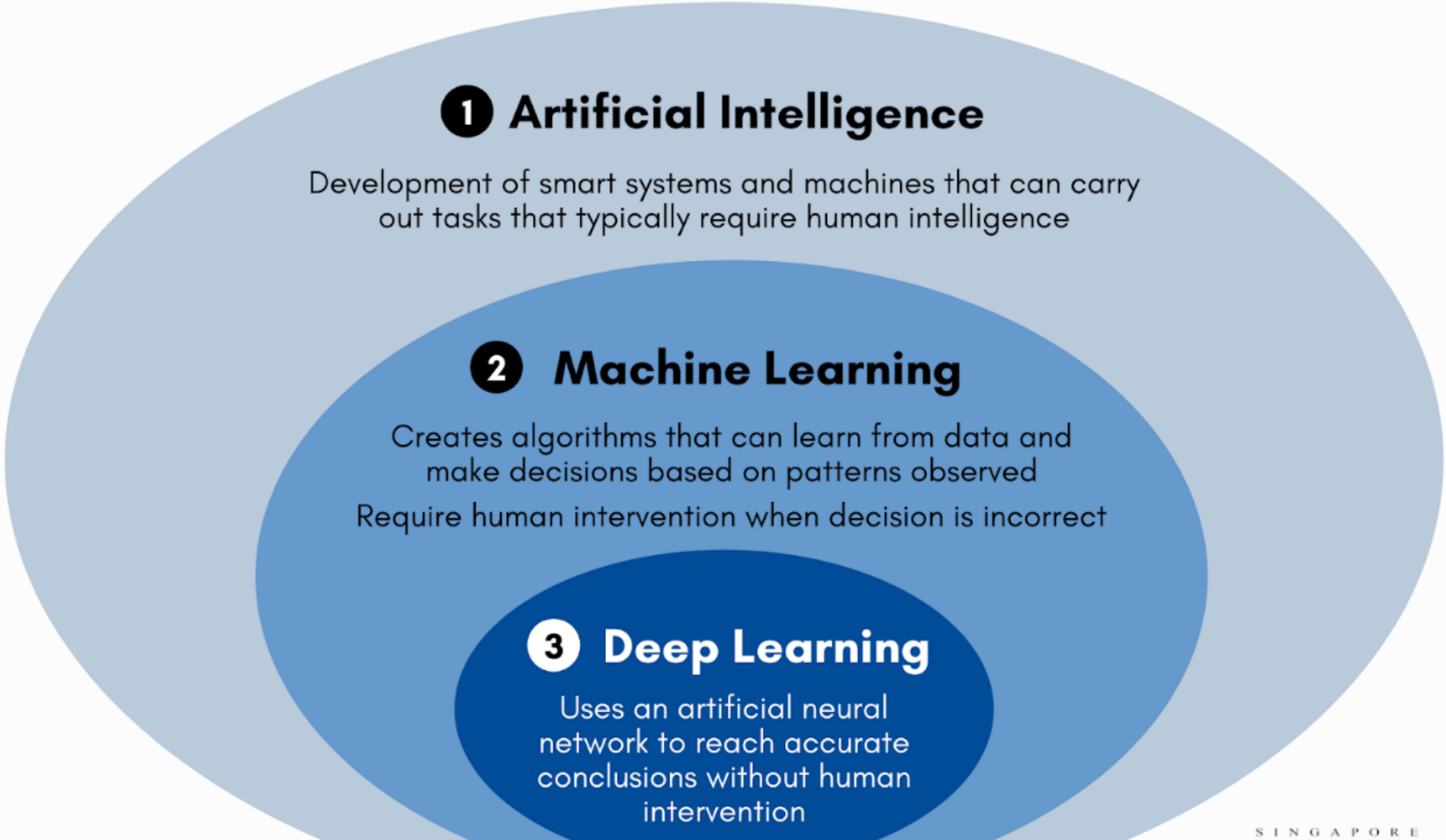
OUTPUT

DLT LABS

MAIN FEATURES of EMULATIVE AI

- Imitation of bottom up development of neural networks in humans: the feedbacks from the environment teach how the neural network should develop
- (But newborns are born with innate principles of thinking: They are not Tabula Rasa.)
- Opacity of the process.
- Umpredicatability of the output

ARTIFICIAL INTELLIGENCE VS MACHINE LEARNING VS DEEP LEARNING



SINGAPORE

TO SUMMARIZE

- SYMBOLIC AI (Simulative): Symbolic Systems where a lot of knowledge is inserted through sequential instructions by a human and that are able to plan sequences of actions to achieve a goal.
- SUBSYMBOLIC AI (Emulative): artificial neural networks in which the networks work in parallel on multiple layers of artificial neurons interconnected with each other (PDP-Parallel Distributed Processing)
- SYMBOLIC AI (Simulative) based on sequential instructions, logic and top-down processing: EXPERT SYSTEMS
- SUBSYMBOLIC AI (Emulative) based on parallelism, probability, and bottom-up processing: MACHINE LEARNING

C) NEED FOR EX-ANTE REGULATION or FOR IA 4 AI?

- Today Generative Artificial Intelligence (GenAI) has expanded the boundaries of innovation beyond traditional industrial applications into nearly all spheres of human activity, from healthcare and education to governance and social interactions.
- This transformation raises questions about the adequacy of traditional policy tools, particularly Technology Assessment (TA), which has historically focused on safety, environmental impacts, and technical performance before technological innovation is applied or as it is being applied.
- TA, it may be argued, is intended to be precautionary and takes its point of departure in the principle of reducing risk levels to as low as readily achievable. But what about radical uncertainty of a fast changing technology as GenAI? TA can not work!

1) Need for AI regulation: polarization of views

- In April 2021, the European Commission proposed the first EU artificial intelligence law, establishing a risk-based AI classification system. The EU AI Act (introduced on 1-08-2024), regulating AI applications *ex ante* was designed before the advent of ChatGPT. It did not provide for any regulations in new application areas such as "General Purpose" AI (GPAI) models.
- California AI bill (vetoed in September 2024) enhanced the EU's AI regulation for GPAI models but was withdrawn under public and industry pressure.
- Most recent EU "Code of Practice" on GPAI will provide the most detailed legally-backed guidelines for GPAI providers (entering its final phase).
- Yoshua Bengio chair of the group for technical risk mitigation on the EU Code of Practice: a CoP that will stand the test of time and **enable AI-driven innovation, while also stimulating innovation in AI safety practices**.

2. Regulatory quicksand boxes in AI?

- Three major battle lines:
  - "AI Safety"-focused concerns about future-scale AI risks. Similar to TA but “Deep tech” AI technologies difficult to assess in a scientific manner, often involve science “fiction”. Also mental health implications unknown.
  - “Human rights”-focused concerns about risks to fundamental rights. E.g. AI innovation in areas such as education, security, health, but also public administration. Gains (profits) are not private but societal, sometimes “political” with potential for increased control, ethnic profiling, opinion manipulation, etc.…
  - Intellectual rightsholders fighting for strict application of copyright rules. Debates focuses on need to publish detailed summaries of algorithm’s training data.

3. Opening the AI "black box" of data used

- "Good practices in data stewardship and analysis could be more effective than regulation": role of regulators should be on opening "black box" of data used for training the underlying algorithms
- Data used in generative AI models mixes:
  - raw data processed within context of a particular experiment "*experimental data*";
  - data based on statistically agreed regular observations "*observational data*";
  - data analyzed and presented in the form of (peer reviewed) '*established knowledge* (EK)'.
- "Need for FAIR (Fully AI Ready) data" Barend Mons (Leiden Un.)

4. GovTech as European AI specialization

- Develop local areas of technological software expertise fed by the huge demand for GovTech in Europe's multi-governance world.
- AI offers EU possibility to overcome public so to say "natural" inefficiencies to achieve productivity gains in non-tradable sectors: a "competitiveness transformation".
- At same time access to FAIR data is now essential in creating trust and acceptance of AI applications in government services.
- Possible new role of statistical offices in line with legislation forcing companies to provide information often involving detailed, confidential information about individual firms' activities.

5. The Role of the Public Sector (TripleHelix)

- Finally, IA must also consider the **state as an innovator and regulator**. In Europe, where public services comprise a significant share of GDP and social infrastructure, innovations like AI in health, education, and justice demand **proactive IA frameworks** that reflect public values and democratic oversight.
- The rollout of AI diagnostic tools in hospitals should be assessed not only for clinical accuracy but also for their **impact on equity**, **medical trust**, and **professional accountability**.
- Smart city applications must be evaluated in terms of **privacy**, **citizen agency**, and **accessibility**—not only technical efficiency
- Particularly in education, proactive AI frameworks will be needed sometimes drastically limiting the use of AI, sometimes enabling it to **reduce cognitive barriers** and improve a **more "sartorial"** approach to teaching and learning

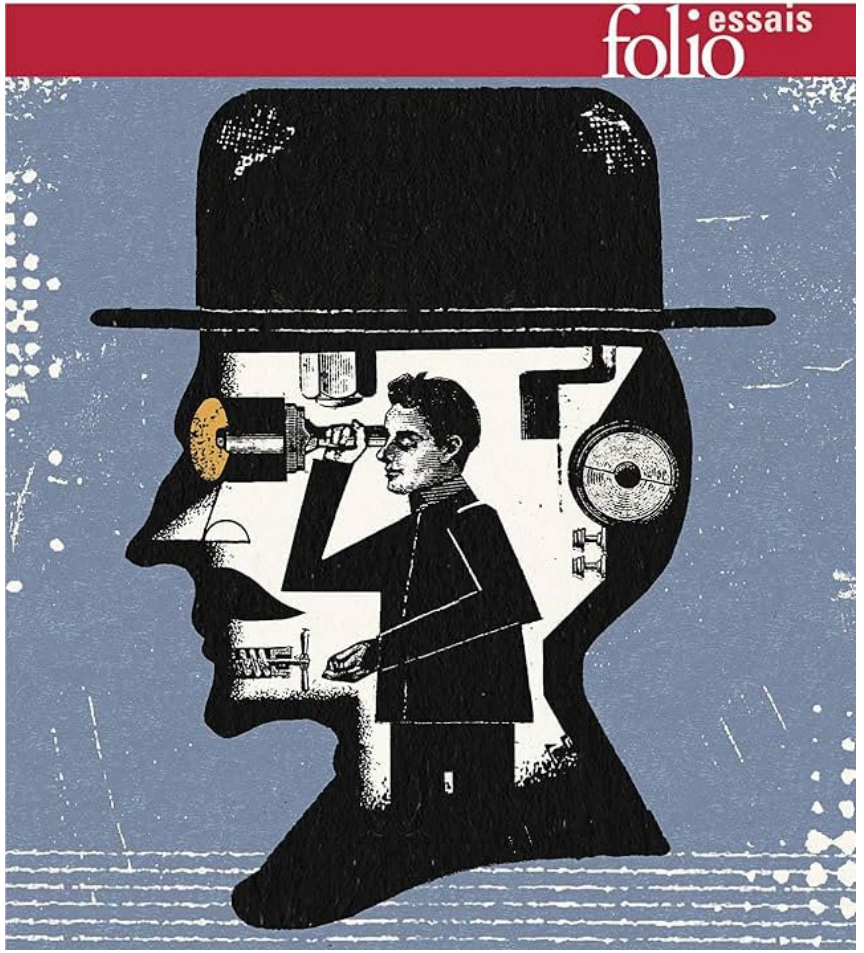
D) POLICY RECOMMENDATIONS AND CONCLUSIONS

- Draghi proposal for a "AI Vertical Priorities Plan" on the acceleration of the introduction of AI development in the Euin 10 industrial sectors with "the opening up, EU-wide coordination and harmonisation of national "AI Sandbox regimes" to companies participating in the plan".
- Focus here on public sectors which have not been integrated/harmonized and represent natural limit to achievement of European productivity growth.
- Applications of generative AI concern personal/private data including data tracking individual online behaviour.
- Scope for specific AI applications dealing less with the creation of patterns of user behaviour, but with government "Government as a Platform" approaches.

- The EU should lead globally in developing and institutionalizing Innovation Assessment as a **complementary policy discipline** to Technology Assessment. This requires building flexible, reflexive, and democratic innovation governance systems.
- **Recommendations:**
- **Institutionalize Innovation Assessment across EU and Member States**
  Establish IA units tasked with monitoring post-deployment impacts, conducting social and economic evaluations, and feeding findings into adaptive regulatory processes.
- **Differentiate TA and IA in Policy Instruments**
  Clarify in EU legislation when and how TA vs. IA applies—particularly for digital platforms, AI, and business model innovations.

- **Promote Data Transparency for AI Assessment**
  Apply FAIR (Findable, Accessible, Interoperable, Reusable) principles to AI training data. Use national statistical offices to steward non-personal observational data under clear legal protections.
- **Support Real-World Experimentation through Regulated Sandboxes**
  Where appropriate, use innovation sandboxes to allow temporary, conditional experimentation—especially in sectors like mobility, fintech, or GovTech.
- **Prioritize Public Sector AI (GovTech) Innovation**
  Encourage AI use in healthcare, education, and public administration to address Europe's productivity gap while upholding public trust through transparent IA frameworks, and support private entrepreneurship in AI.

Julien Offroy
de La Mettrie

L'Homme-Machine

Thanks for your attention