UNI

UN MONDO FATTO BENE

Il ruolo della Infrastruttura per la Qualità nell’Intelligenza Artificiale
Centro Studi sulla Normazione

Daniele Gerundino

Intelligenza Artificiale tra valutazione del rischio e certificazione accreditata – Roma, 14 ottobre 2024

Origins of AI – summer 1956

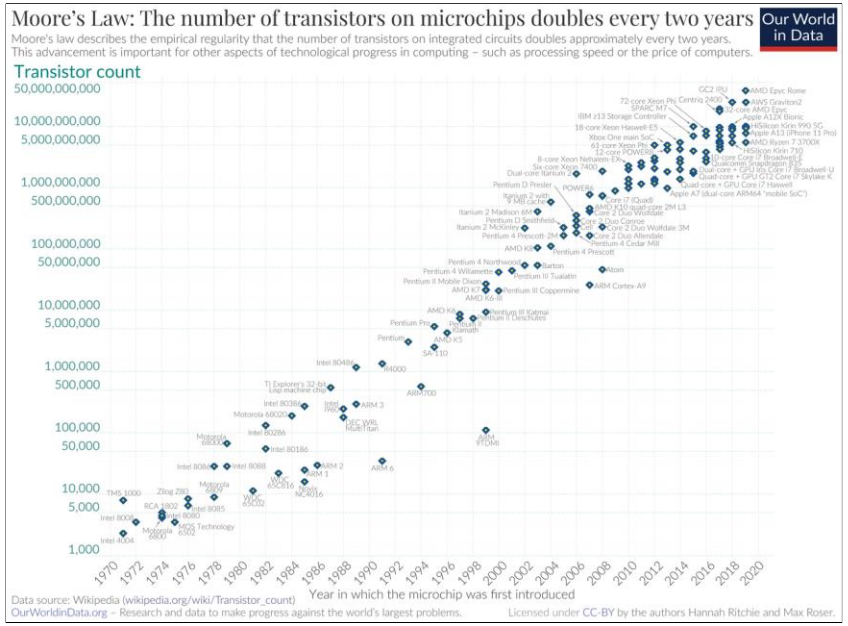
The Dartmouth Summer Research Project on Artificial Intelligence, held from **18 June through 17 August of 1956**.

Organized by **John McCarthy, Marvin Minsky, Claude Shannon, and Nathaniel Rochester**, it brought together many of the leading thinkers in the emerging fields of computer science and information theory.

They proposed a study 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***."

Fast forward to the 2000s

Moore’s Law

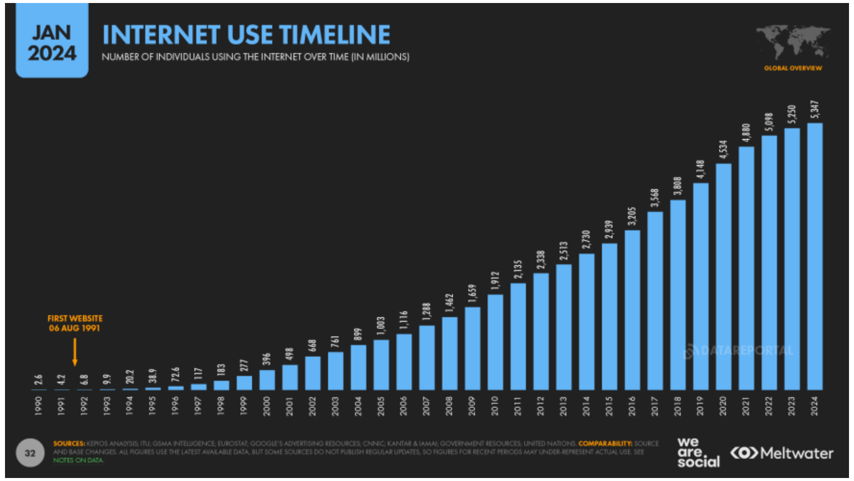


Moore’s Law: The number of transistors on microchips doubles every two years

Moore’s law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years. This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.

Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)
OurWorldinData.org – Research and data to make progress against the world’s largest problems. Licensed under CC-BY by the authors Hannah Ritchie and Max Roser.

Internet Growth worldwide

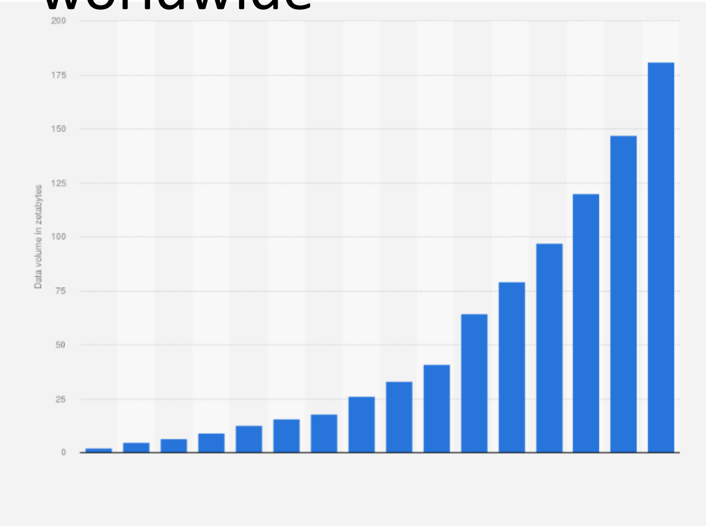


JAN 2024 INTERNET USE TIMELINE

NUMBER OF INDIVIDUALS USING THE INTERNET OVER TIME (IN MILLIONS)

SOURCES: KEPIOS ANALYSIS; ITU; GSMA INTELLIGENCE; EUROSTAT; GOOGLE’S ADVERTISING RESOURCES; CNNIC; KANTAR & IAMAI; GOVERNMENT RESOURCES; UNITED NATIONS. COMPARABILITY: SOURCE AND BASE CHANGES. ALL FIGURES USE THE LATEST AVAILABLE DATA, BUT SOME SOURCES DO NOT PUBLISH REGULAR UPDATES, SO FIGURES FOR RECENT PERIODS MAY UNDER-REPRESENT ACTUAL USE. SEE NOTES ON DATA.

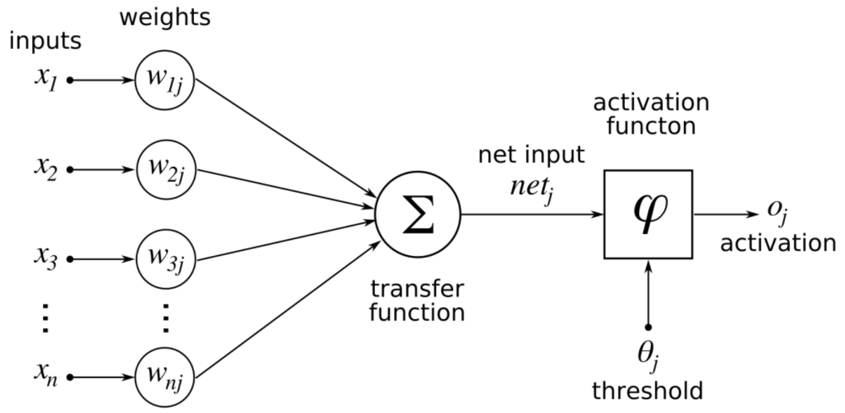
Digital data Growth worldwide



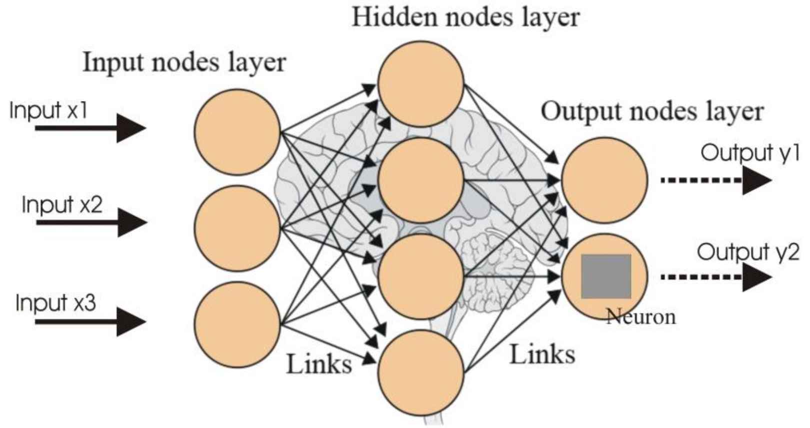
The exponential growth of digitalization since the end of 1960s has created the basis for today’s Artificial intelligence

Today’s dominant forms of AI

Artificial Neural Networks (ANNs)

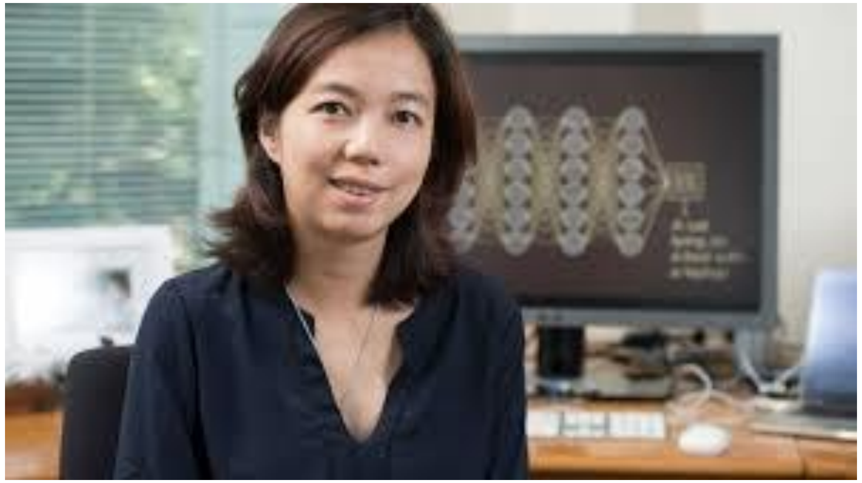


Machine Learning and Deep Learning



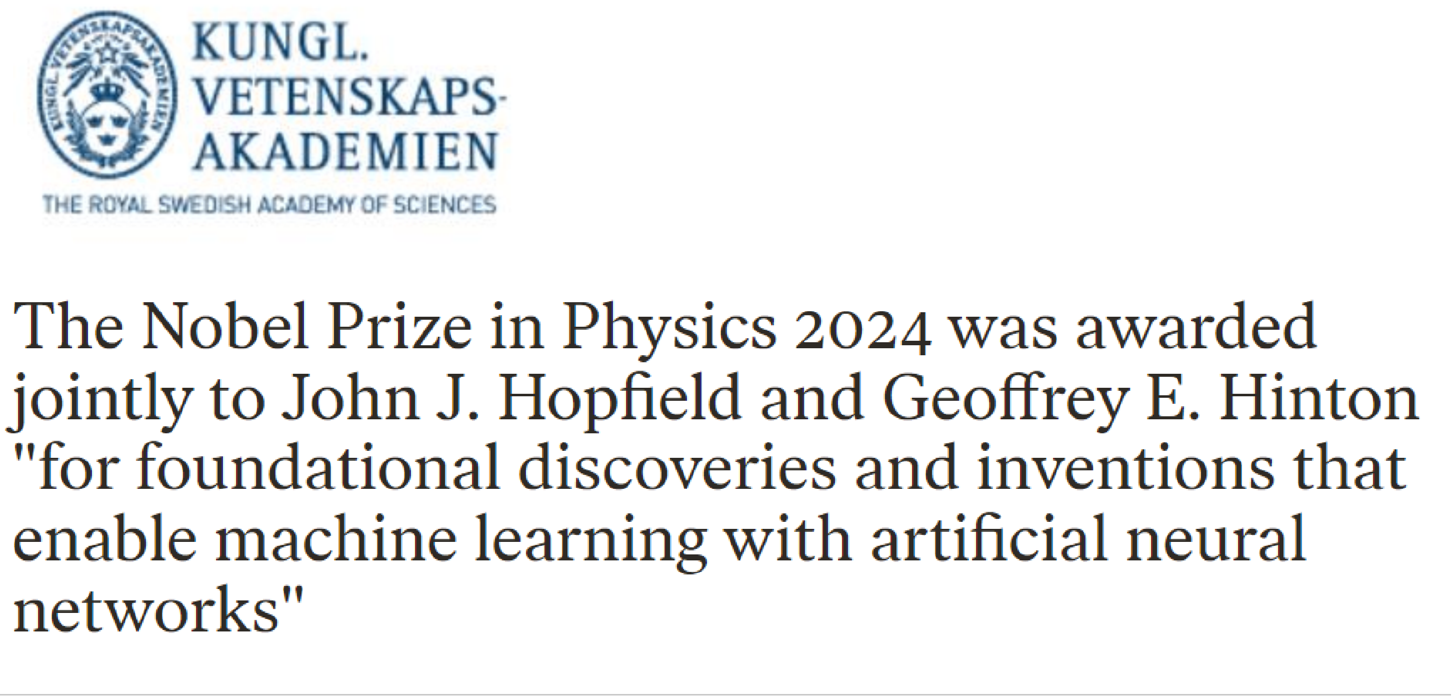
Deep Learning breakthrough – 2012

“*The* ***deep-learning breakthrough happened in 2012****, when two parallel ideas merged during an AI competition called the* ***ImageNet challenge****. Professor* ***Fei-Fei Li*** *had spent years collecting and organizing images with the idea that showing algorithms more data was more important than crafting the perfect learning algorithm. At the University of Toronto, professor* ***Geoffrey Hinton*** *and Ph.D students* ***Alex Krizhevsky*** *and* ***Ilya Sutskever*** *used Li’s data to supercharge their neural networks.*”
Quartz, October 2018



Geoff’s Hinton ***Convolutional Neural Network*** *and* ***Backpropagation*** *algorithm have proven to be extraordinary successful*
*(*However, these neural networks were still trained as *discriminative* models performing primarily classification tasks)

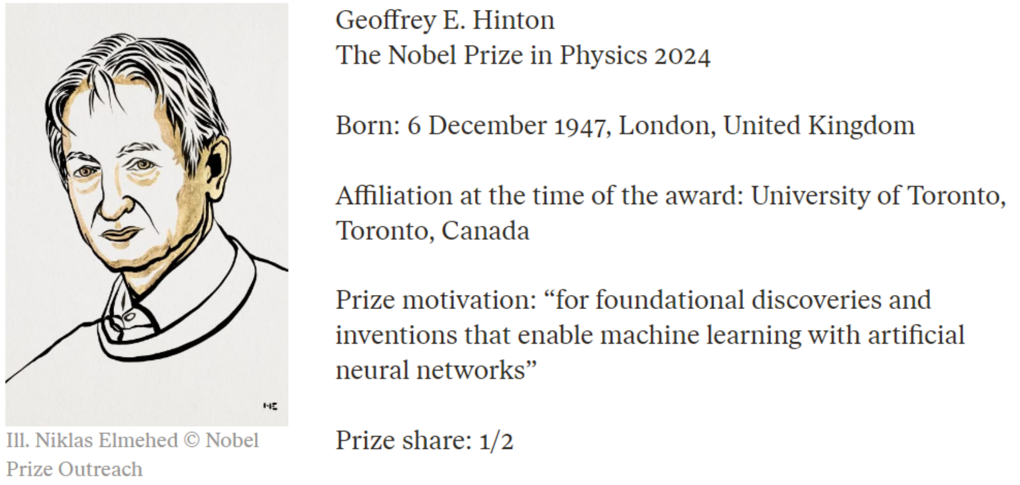
By the way…



KUNGL. VETENSKAPS-AKADEMIEN

THE ROYAL SWEDISH ACADEMY OF SCIENCES

The Nobel Prize in Physics 2024 was awarded jointly to John J. Hopfield and Geoffrey E. Hinton "for foundational discoveries and inventions that enable machine learning with artificial neural networks"



Ill. Niklas Elmehed © Nobel Prize Outreach

Geoffrey E. Hinton
The Nobel Prize in Physics 2024

Born: 6 December 1947, London, United Kingdom

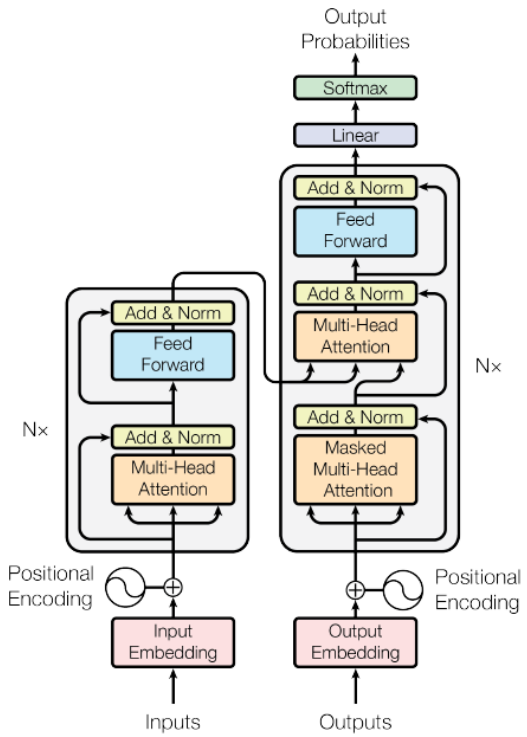
Affiliation at the time of the award: University of Toronto, Toronto, Canada

Prize motivation: “for foundational discoveries and inventions that enable machine learning with artificial neural networks”

Prize share: 1/2

Generative AI breakthroughs – 2018

The Transformer - model architecture – 2017



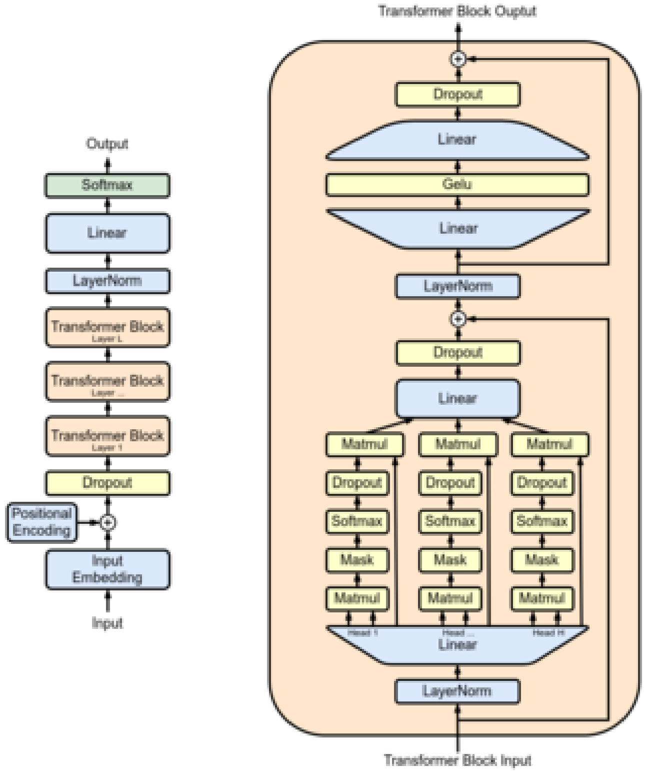
A transformer model is a neural network that learns context and thus meaning by tracking relationships in sequential data (like the words in this sentence). Transformer models apply an evolving set of mathematical techniques, called **attention or self-attention**, to detect subtle ways even distant data elements in a series influence and depend on each other.

It is especially suited to deal with Natural Language Processing

Attention Is All You Need

Paper published in 2017 by a team of researchers at ***Google Brain****. Transformers are among the newest and most powerful classes of models invented to date.*

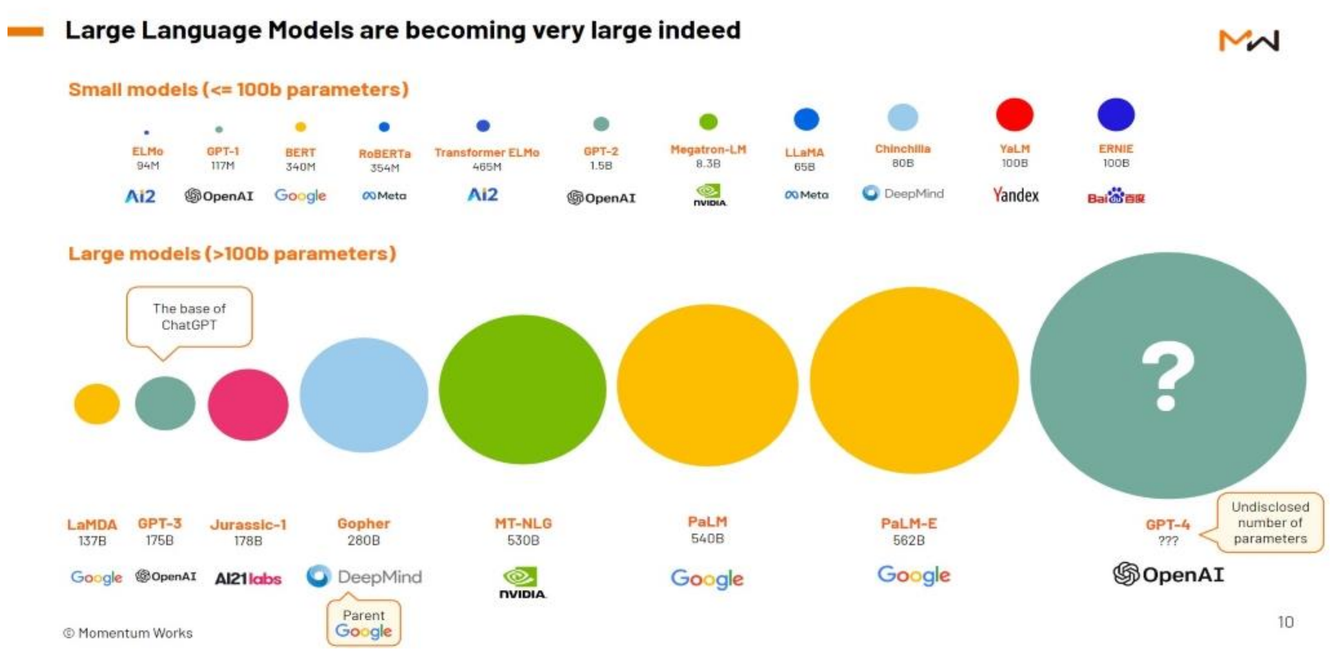
Generative pre-trained transformers (GPT) – 2018

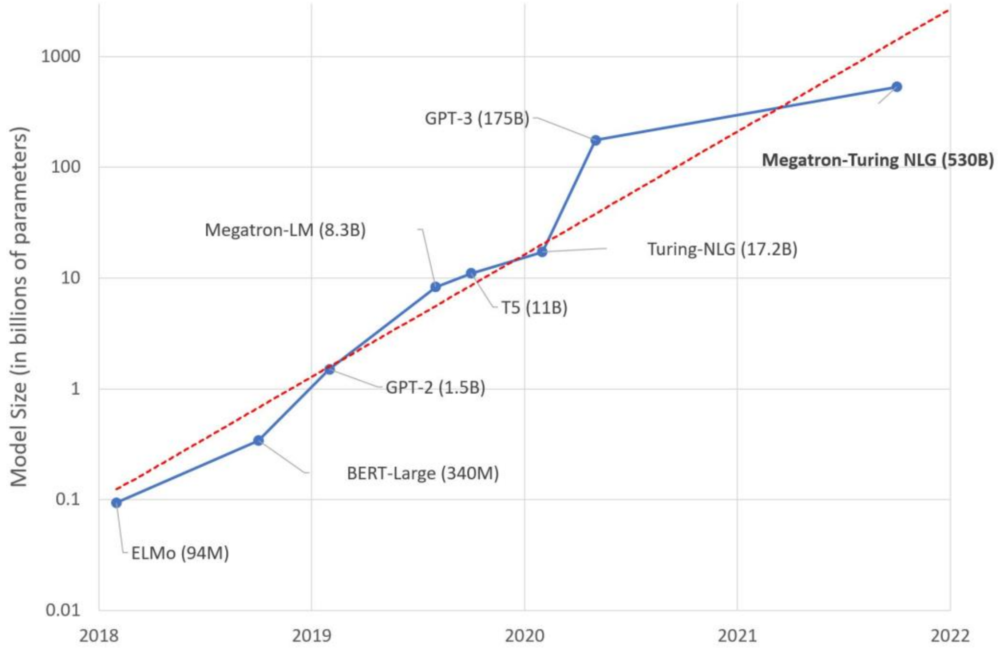


GPT models are artificial neural networks that are based on the transformer architecture, pre-trained on large data sets of unlabelled text, and able to generate novel human-like content.

The first ***GPT*** *was introduced in 2018 by the American artificial intelligence (AI) company* ***OpenAI****. The models released by Open AI since 1918 have been incredibly influential*

Large Language Models (LLMs)





LLMs – a sort of new Moore’s Law?

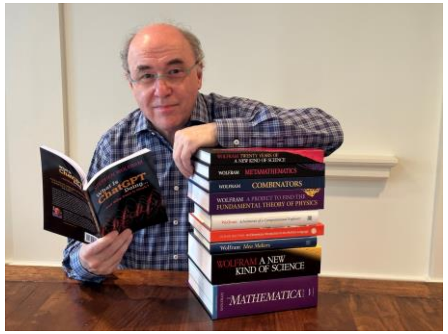
- A large language model (LLM) consists of a neural network with many parameters (typically billions of weights or more), trained on large quantities of unlabeled text using self-supervised learning or semi-supervised learning. LLMs are general purpose models which excel at a wide range of tasks and a form of Generative AI.
- LLMs are typically based on Transformer architectures

These technologies are demonstrating unprecedented capabilities in a broad variety of areas and are re-shaping research and development activities, business processes and professions.

Their transformational potential and substantial benefits to society are only starting to become clear.

By the way: there might be (much!) more…

Stephen Wolfram



"…the remarkable thing is that all these operations – individually as simple as they are – can somehow together manage to do such a good "human like" job of generating text. […] There's no "ultimate theoretical reason" why anything like this should work. I think we have to view this as a – potentially surprising – scientific discovery: that somehow in a neural net like ChatGPT it's possible to capture the essence of what human brains manage to do in generating language…"

…in its training ChatGPT has somehow "implicitly discovered" whatever regularities in language (and thinking) make this possible. The success of ChatGPT is, I think, giving us evidence of a fundamental piece of science: it's suggesting that we can expect to be major new "laws of language" – and effectively "laws of thought" – out there to discover.

If we could somehow make these laws explicit, there is the potential to do the kinds of things ChatGPT does in vastly more direct, efficient – and transparent – way."

AI is a key enabler of Industry 4.0

The idea is to establish **flexible manufacturing operations** (that can easily be re-tooled to produce variations of existing products or entirely new products), where:

cyber-physical systems (robots, actuators and sensors) are able to **manage autonomously** factory shop-floor activities (including handling of materials and parts, the execution of tasks and monitoring and control activities)

capturing and exchanging data efficiently and securely among themselves and with local or remote design and control centers

producing digital copies of product items at any relevant stage, copies that can be validated or modified by designers whenever needed.

…and a key driver of a "Smart QI"

UNIDO UNITED NATIONS INDUSTRIAL DEVELOPMENT ORGANIZATION

SUSTAINABLE DEVELOPMENT GOALS

SMART QUALITY INFRASTRUCTURE

Shaping a sustainable future

There is a symbiotic relationship between the technologies for Industry 4.0 (including AI) and the Quality Infrastructure (QI):

- QI is an important enabler of their development
- The technologies shaping Industry 4.0 are deeply transforming the QI itself

Areas where AI can bring substantial benefits:

- Medical research and health care
- Technology and product design
- Process re-engineering
- Transportation
- Use of resources
- Efficiency of business processes
- …and much more

At the same time, AI poses substantial threats to the fabric of society by contributing to:

- violating fundamental human rights
- increasing inequalities
- spreading disinformation
- altering democratic processes
- damaging minorities and disadvantaged people
- violating intellectual property rights and
- bringing vast and perhaps irreversible harm (e.g. autonomous weapons).

In the live presentation were highlighted a few examples concerning:

- Impact of AI on jobs and business functions
- AI and intellectual property rights
- Autonomous weapons: probably the most severe and urgent threat from AI facing humanity

The UN and a significant number of countries are advocating action to control the AWS race

Humanity at the Crossroads:
Autonomous Weapons Systems
and the Challenge of Regulation
Vienna 2024

Autonomous Weapons Systems (AWS), which - once activated - select targets and apply force without further human intervention, raise concerns from legal, ethical and security perspectives. Fundamental challenges relate to the nature of human control, accountability and the overall compatibility of such systems with international law, including international humanitarian law (IHL) and international human rights law (IHRL).

It is also important to bear in mind that:

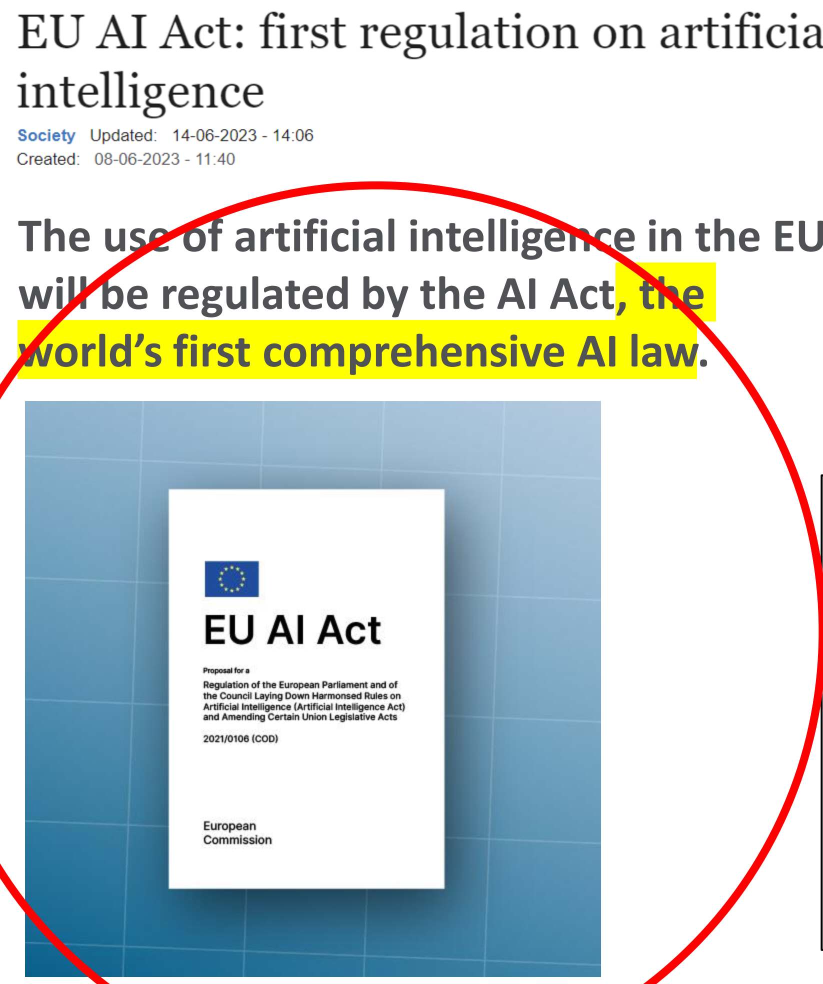
...the world leading AI systems are in the hands of an oligopolistic framework driven by the quest for power and profit…

What can be done to manage the AI threats and promote its benefits for all

- Urgent actions by governments – by means of comprehensive policy frameworks that can be supported/complemented by :

  - **Codes of conduct and voluntary standardization initiatives**
  - Engagement of expert groups with important consultative roles
  - Focused investments (Public and/or o PPP)
  - Scientific collaboration dealing with critical issues, such as technologies and standards for AI trustworthiness and security, models and approaches to embed human values, and to ensure understandability, control and transparency of AI systems

Legislative initiatives in key jurisdictions



EU AI Act: first regulation on artificial intelligence

Society Updated: 14-06-2023 - 14:06
Created: 08-06-2023 - 11:40

The use of artificial intelligence in the EU will be regulated by the AI Act, the world's first comprehensive AI law.

EU AI Act

Proposal for a
Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts

2021/0106 (COD)

European
Commission

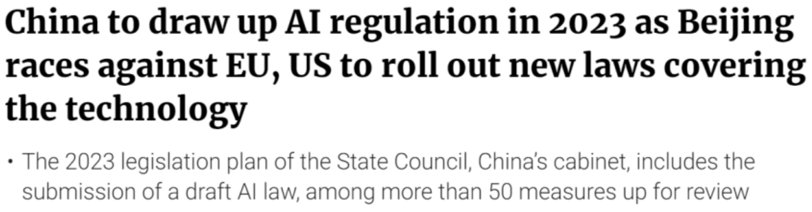
The White House Executive Order (E.O.) 14110 on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

The E.O. directs over 50 federal entities to engage in more than 100 specific actions to implement the guidance set forth across eight overarching policy areas

OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

BRIEFING ROOM › PRESIDENTIAL ACTIONS



China to draw up AI regulation in 2023 as Beijing races against EU, US to roll out new laws covering the technology

- The 2023 legislation plan of the State Council, China's cabinet, includes the submission of a draft AI law, among more than 50 measures up for review

The G7 AI Hiroshima Process

G7 2023 HIROSHIMA SUMMIT
Digital and Tech Ministers' Meeting
in Takasaki, Gunma April 29-30, 2023
JAPAN
JAPAN
JAPAN

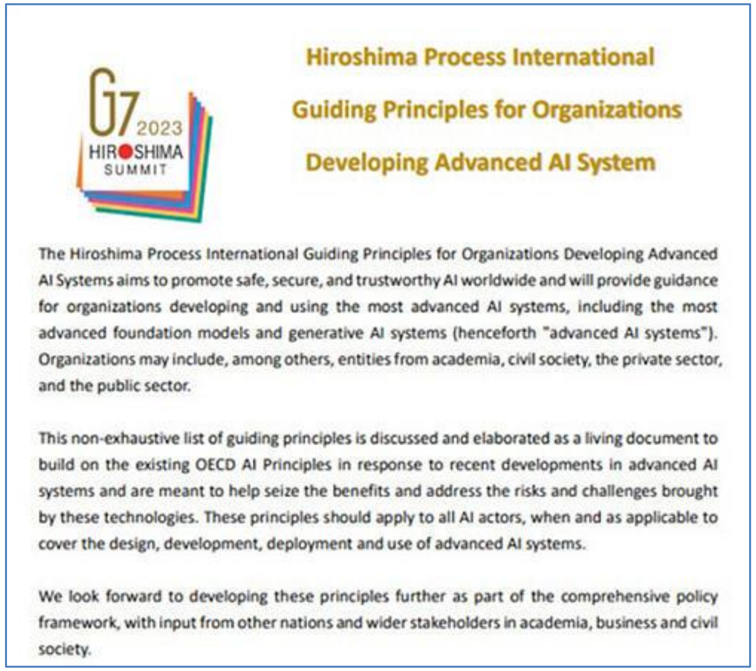
The Hiroshima AI Process was launched in May 2023, following the Leaders’ direction at the G7 Hiroshima Summit, with the objective of discussing the opportunities and risks of AI.

After continuous discussions including at an interim minister-level meeting in September and a multi-stakeholder high-level meeting at IGF Kyoto 2023 in October, "the Hiroshima AI Process Comprehensive Policy Framework“ was successfully agreed upon at the G7 Digital & Tech Ministers’ Meeting in December 2023 and was endorsed by the G7 Leaders in the same month.

UNI
UN MONDO FATTO BENE

The Hiroshima AI Process Comprehensive Policy Framework

G7 Hiroshima Process International Guiding Principles



G7 2023
HIROSHIMA
SUMMIT

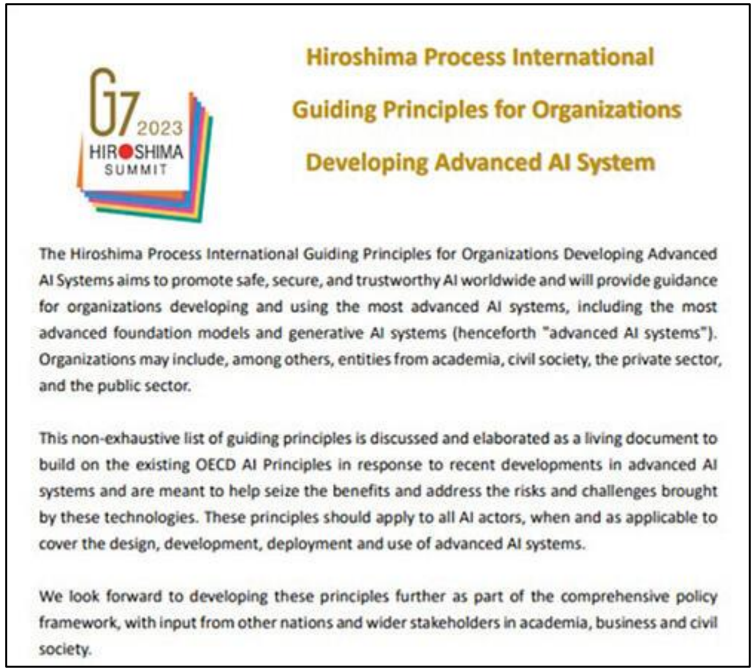
Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System

The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide guidance for organizations developing and using the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems"). Organizations may include, among others, entities from academia, civil society, the private sector, and the public sector.

This non-exhaustive list of guiding principles is discussed and elaborated as a living document to build on the existing OECD AI Principles in response to recent developments in advanced AI systems and are meant to help seize the benefits and address the risks and challenges brought by these technologies. These principles should apply to all AI actors, when and as applicable to cover the design, development, deployment and use of advanced AI systems.

We look forward to developing these principles further as part of the comprehensive policy framework, with input from other nations and wider stakeholders in academia, business and civil society.

G7 Hiroshima Process International Code of Conduct



G7 2023
HIROSHIMA
SUMMIT

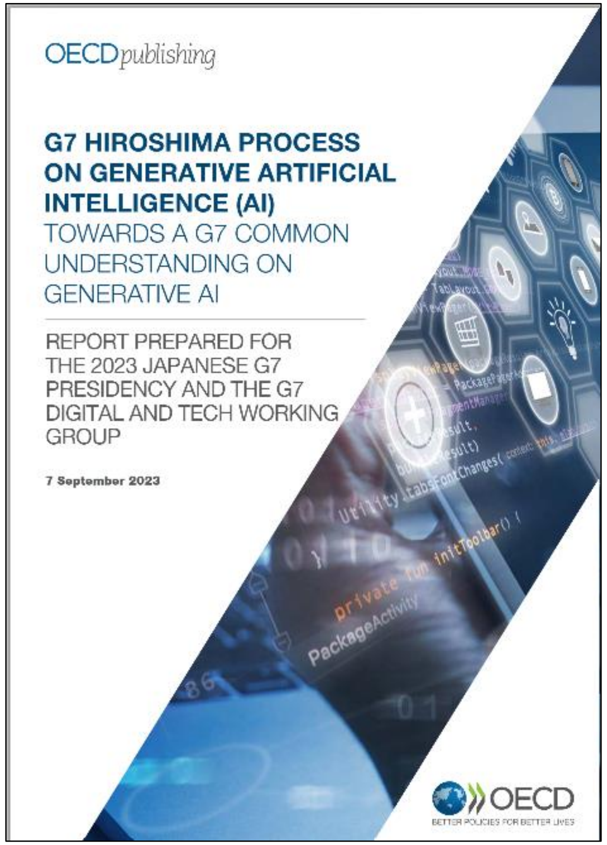
Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI System

The Hiroshima Process International Guiding Principles for Organizations Developing Advanced AI Systems aims to promote safe, secure, and trustworthy AI worldwide and will provide guidance for organizations developing and using the most advanced AI systems, including the most advanced foundation models and generative AI systems (henceforth "advanced AI systems"). Organizations may include, among others, entities from academia, civil society, the private sector, and the public sector.

This non-exhaustive list of guiding principles is discussed and elaborated as a living document to build on the existing OECD AI Principles in response to recent developments in advanced AI systems and are meant to help seize the benefits and address the risks and challenges brought by these technologies. These principles should apply to all AI actors, when and as applicable to cover the design, development, deployment and use of advanced AI systems.

We look forward to developing these principles further as part of the comprehensive policy framework, with input from other nations and wider stakeholders in academia, business and civil society.

G7 Hiroshima Process on Generative AI



OECD*publishing*

G7 HIROSHIMA PROCESS ON GENERATIVE ARTIFICIAL INTELLIGENCE (AI)
TOWARDS A G7 COMMON UNDERSTANDING ON GENERATIVE AI

REPORT PREPARED FOR THE 2023 JAPANESE G7 PRESIDENCY AND THE G7 DIGITAL AND TECH WORKING GROUP

7 September 2023

OECD
BETTER POLICIES FOR BETTER LIVES

Moving forward – G7 Italy 2024

G7 ITALIA 2024

G7 INDUSTRY, TECHNOLOGY AND DIGITAL MINISTERIAL MEETING

Verona and Trento 14-15 March 2024

Ministerial Declaration

1. We, the Industry, Tech & Digital Ministers of the G7, met on 14-15 March, 2024 in Verona and Trento under the co-chairmanship of Mr. Adolfo Urso – Minister of Enterprises and Made in Italy – and of Mr. Alessio Butti – Undersecretary of State to the Presidency of the Council of Ministers – to discuss the impact of digital innovation on economic growth and industrial competitiveness, and how to leverage the digital transformation and Artificial Intelligence (AI) to advance our economies and societies in an inclusive and sustainable way.
2. We welcomed invitations to Brazil, the Republic of Korea, Ukraine and the United Arab Emirates, the International Telecommunication Union (ITU), and the UN Secretary-General's Envoy on Technology. We are grateful for the efforts of the Organisation for Economic Co-operation and Development (OECD); United Nations Development Programme (UNDP); and the United Nations Education, Scientific and Cultural Organisation (UNESCO) as knowledge partners.
3. The future of our economies, and their sustainability and resilience, depends on our determination to promote innovation and on our ability to help translate the digital transformation into concrete benefits for our citizens and to address the related challenges. We reaffirm our unwavering commitment to the principles of the rule of law, due process,

Clause 14:
“Applauding the outcomes achieved under the Japanese G7 Presidency in 2023, we remain committed to further advancing the Hiroshima AI Process Comprehensive Policy Framework (“Framework”), in accordance with the workplan, including the implementation of the International Guiding Principles (“Principles”) and International Code of Conduct for Organisations Developing Advanced AI Systems (“Code of Conduct”). […]”

UNI
UN MONDO FATTO BENE

A contribution from UNI's CSN

A set of recommendations concerning developments and specific actions aiming at operationalizing the G7 AI Guiding Principles and Code of Conduct and making further progress towards the development of safe, secure, and trustworthy AI.

Think7 ITALY 2024 · C7 ITALIA 2024

Policy Brief

April 2024

Task Force 4: Science and Digitalization for a Better Future

Towards Safe, Secure, and Trustworthy AI: Implementing the G7 AI Hiroshima Policy Framework

Daniele Gerundino, Centro Studi per la Normazione, Italian National Standards Body (UNI); Research associate, Institut de Gouvernance de l'Environnement et Développement Territorial (GEDT), University of Geneva, Switzerland
Conor Hearn, CEO Lumolo, Director Vision Corp (Holdings)
Paul Alan McAllister, President and CEO, Global Leaders in Unity and Evolvement (GLUE), United States
Vidisha Mishra, Head of Community Engagement Programme, Global Solutions Initiative, Germany
Simona Romiti, Assistant Program Manager for Global Leaders in Unity and Evolvement (GLUE), United States
Alice Saltini, Research Coordinator, European Leadership Network (ELN)
Dennis J. Snower, President, Global Solutions Initiative, Germany
Paul Twomey, Fellow and Core Theme Leader, Global Solutions Initiative, Germany

Strategically promote standardization and conformity assessment frameworks

Promote scientific collaboration as a method for building secure and trustworthy AI and ensuring its equitable distribution

Educate AI developers to advance AI understanding and human-centred values, via whole education frameworks that address all AI actors at various levels

Empower digital citizens by promoting their fundamental right to full control, individually and collectively, of their personal data

Strategically promote standardization…

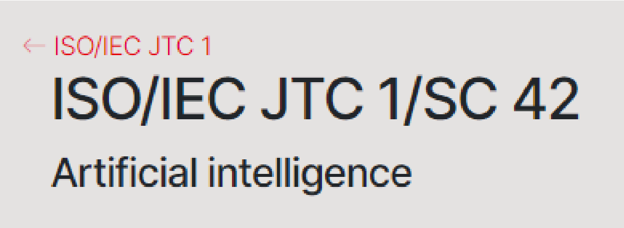
- Standardization can be extremely effective at fostering the adoption of good practices – helping to "operationalize" regulatory goals. However, it is absolutely non-trivial to ensure that the contribution of standardization is as effective and timely as required.

- *G7 countries – in a coordinated way – can provide an invaluable contribution by forwarding to standardization organizations clear input consistently with the G7 policies, to shape strategic directions and prioritization for standardization in AI*

- Standardization organizations must ensure:
  - multidisciplinary and multistakeholder input, engaging highly qualified experts
  - a transparent roadmap, with well defined priorities and milestones
  - a highly efficient process, using all the necessary means to provide timely deliverables (in prohibitively short times…)

So, we need to deliver *very good standards*

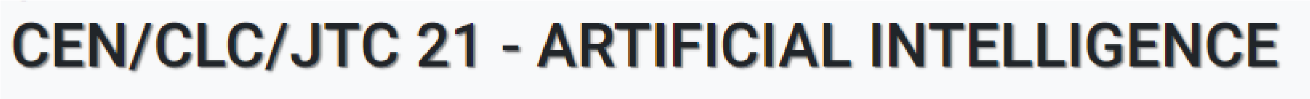
- International standardization organizations and their technical committees are actively working on the subject, notably:



← ISO/IEC JTC 1

ISO/IEC JTC 1/SC 42

Artificial intelligence

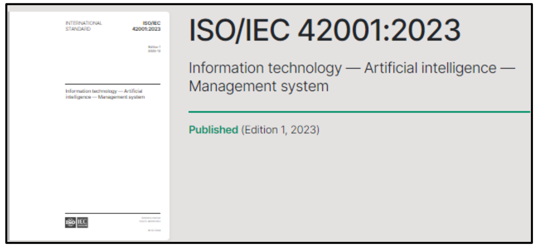


CEN/CLC/JTC 21 - ARTIFICIAL INTELLIGENCE

- They have already developed standards (especially in areas that are somehow familiar to them), such as:

AI Management systems and Risk Management

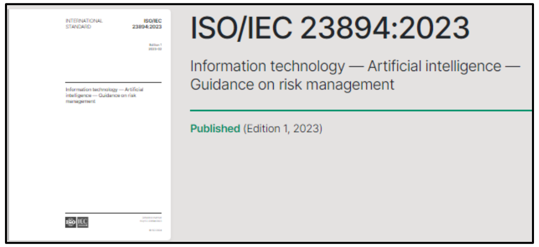
Quality requirements and quality model



ISO/IEC 42001:2023

Information technology — Artificial intelligence — Management system

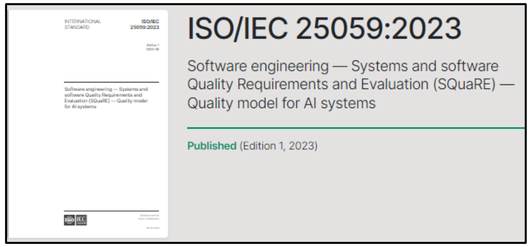
Published (Edition 1, 2023)



ISO/IEC 23894:2023

Information technology — Artificial intelligence — Guidance on risk management

Published (Edition 1, 2023)



ISO/IEC 25059:2023

Software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model for AI systems

Published (Edition 1, 2023)

The European context

The 'Blue Guide' on the implementation of EU product rules 2014

In the European context standardization and (accredited) conformity assessment already have a fundamental function in relation to product policy (*to know more, read the so called "EU Blue Guide"!*)

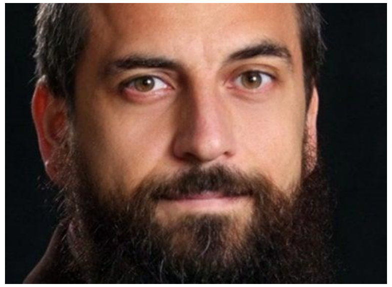
For the AI Act, a key role is played by the joint CEN/Cenelec committee JTC 21

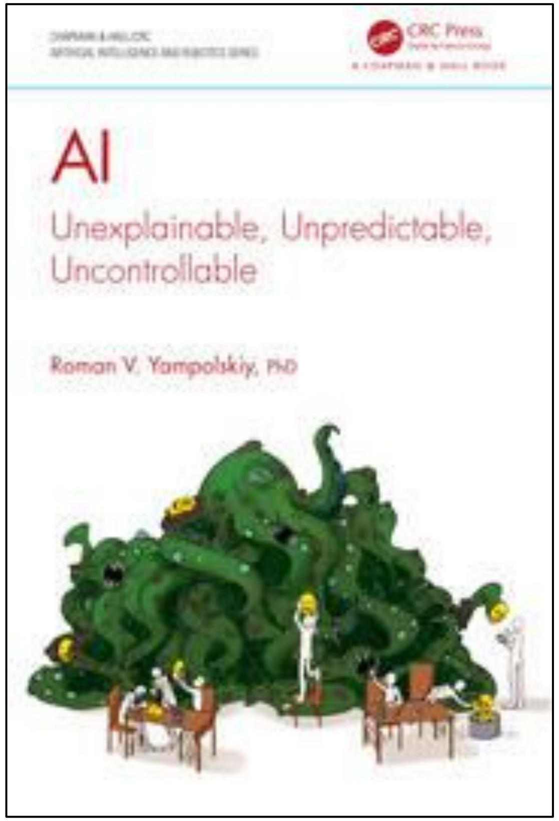
CEN/CLC/JTC 21 - ARTIFICIAL INTELLIGENCE

…but delivering *very good standards* in support of AI regulatory requirements is a formidable challenge!

On the one hand, the challenge is directly related to the intrinsic characteristics of AI and of the leading AI foundational models

Roman Yampolskiy





Yampolskiy elaborates on the concepts of **Unexpectability, Unpredictability, Incomprehensibility,** and **Non-Verifiability** as integral features of AI and the leading existing AI systems.
It highlights some key reasons why AI is very challenging to manage and control.

The discussion covers extensively the issue of AI Uncontrollability as a reflection of (lack of) human capacity to manage AI, as well as AI *Unownability* as a reference to owning the results of AI and being responsible for its consequences

…delivering *very good standards* in support of AI regulatory requirements is a formidable challenge!

On the other hand, we need to ensure that standards for AI translate broad regulatory goals into specifications that are scientific-technically sound and allow feasible implementations

- To this end, it is essential that standards for AI:
  - Incorporate knowledge from a multidisciplinary perspective (warning: many of these standards should not be developed exclusively – or predominantly – by IT professionals)
  - Involve leading subject matter experts (warning: this is not the case for many technical committees!) and maintain strong links with authoritative research and development institutions
  - Keep design approaches tailored to effectively support conformity assessment activities (warning: but not just intended "for the sake of it"…)

- …and, (if that is not enough!) for their development, meet stringent deadlines

Italy can provide a substantial contribution!

Leveraging the highly qualified experts from Italy and the body of knowledge built by them over many years…

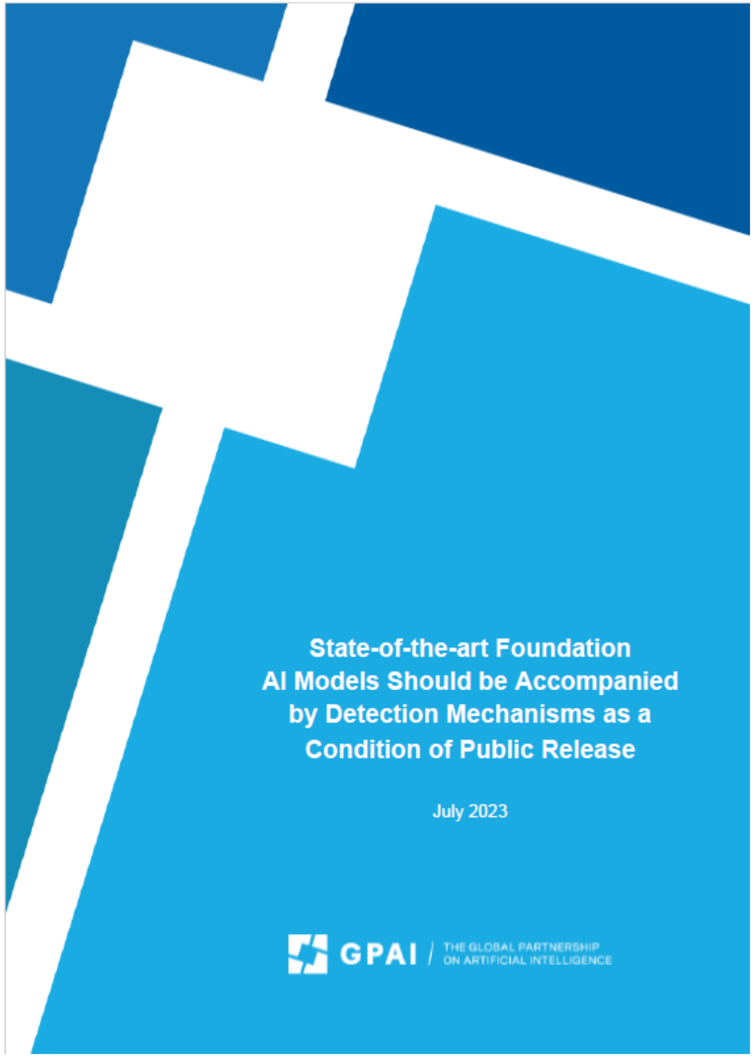
Associazione Italiana per l'Intelligenza Artificiale

500 italiani e italiane che contano
nell'Intelligenza Artificiale

Ital-IA
ITALIA INTELLIGENZA ARTIFICIALE
cini National Lab AIIS

Quarto Convegno Nazionale CINI sull'Intelligenza Artificiale, Napoli, 29 maggio 2024

Just an example of promising approaches

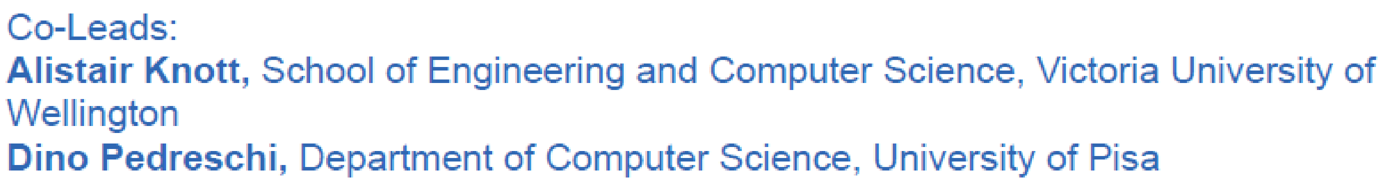


State-of-the-art Foundation AI Models Should be Accompanied by Detection Mechanisms as a Condition of Public Release

July 2023

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

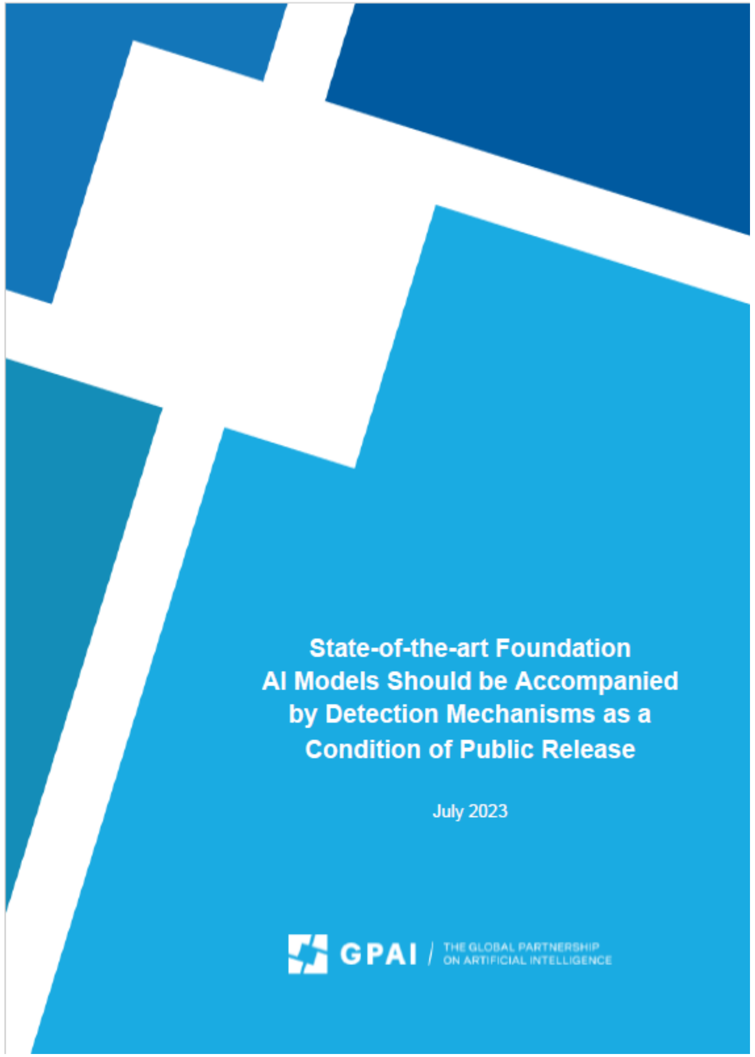
GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE



Co-Leads:
Alistair Knott, School of Engineering and Computer Science, Victoria University of Wellington
Dino Pedreschi, Department of Computer Science, University of Pisa

The report was written by: **Alistair Knott*,** School of Engineering and Computer Science, Victoria University of Wellington**; Dino Pedreschi*,** Department of Computer Science, University of Pisa**; Raja Chatila*,** Sorbonne University**; Susan Leavy*,** School of Information and Communication Studies**,** University College Dublin; **Ricardo Baeza-Yates*,**Institute for Experiential AI, Northeastern University; **Tapabrata Chakraborti†,** University of Oxford and the Alan Turing Institute**; David Eyers†,** Department of Computer Science,University of Otago**; Andrew Trotman†,** Department of Computer Science, University ofOtago; **Lama Saouma†,** GPAI's Montreal Center of Expertise - CEIMIA**; Virginia Morini†,** Istituto di Scienza e Tecnologie dell'Informazione, NIRC; **Valentina Pansanella†,** Scuola Normale Superiore, University of Pisa**; Paul D. Teal†,** School of Engineering and Computer Science, Victoria University of Wellington**; Przemyslaw Biecek*,** Warsaw University of Technology**; Ivan Bratko*,** University of Ljubljana**; Stuart Russell*,** UC Berkeley; and **Yoshua Bengio†,** Université de Montréal and Mila – Quebec AI Institute.

Example (2)



State-of-the-art Foundation AI Models Should be Accompanied by Detection Mechanisms as a Condition of Public Release

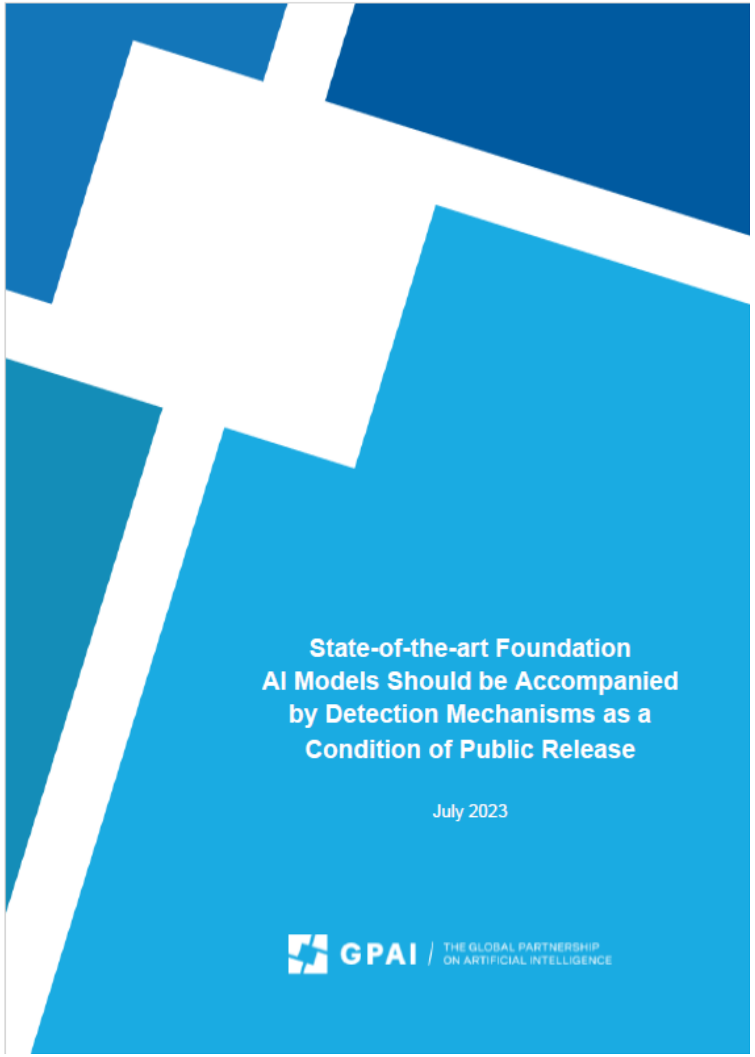
July 2023

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

"…We propose that **a central condition on release of a new state-of-the-art foundation model should be demonstration of a *detection mechanism* that can distinguish content produced by the foundation model from other content, with a high degree of reliability**.
[…]
We'll focus on textual content in the current paper. For this content, the detection mechanism could involve a **classifier**, perhaps making use of watermarks included in the generated text or image, or **methods exploiting statistical features of generated content**, or it could involve the producer company keeping **a log of all texts generated by its LLM and offering a plagiarism detector** running on this log. […]
Crucially, it would be for the organisation wishing to release a new foundation model to demonstrate a detection mechanism that is effective in the current adversarial context, and show its practicality, either unilaterally, or in collaboration with other groups."

Example (3)



State-of-the-art Foundation
AI Models Should be Accompanied
by Detection Mechanisms as a
Condition of Public Release

July 2023

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

GPAI / THE GLOBAL PARTNERSHIP ON ARTIFICIAL INTELLIGENCE

"…there are *general* reasons why mechanisms for detecting AI-generated content are essential. Foundation models are improving rapidly, and it is increasingly difficult to tell whether content is produced by a person or a machine.
LLMs can produce human-like text rapidly, and at scale: so inevitably, the world is about to be flooded with a huge amount of human-like, machine-generated text…"

[…]

"…people have a *right to know* whether the content they receive comes from a human or a machine and transparency legislation should provide this information.[…]
This is a general reason for our argument that new foundation models should not be released without evidence for a detection mechanism for the content they produce. Without a detection mechanism, the enforcement of this newly posited right would be difficult, if not impossible."

Thank you for your attention!