U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO



# AI & GenAI: Challenges and impact in Society

## AI4SG – AI for Social Good ?

*Eugénio Oliveira*

URJC October 2024

U.PORTO LIACC
artificial intelligence and computer science laboratory

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

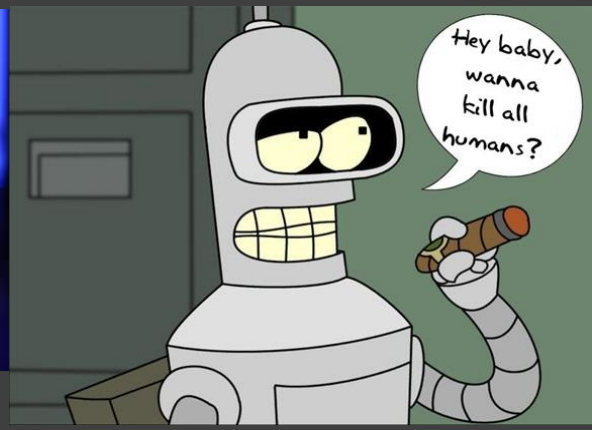


# Generative AI

**Principles,**
**Applications,**
**Challenges,**

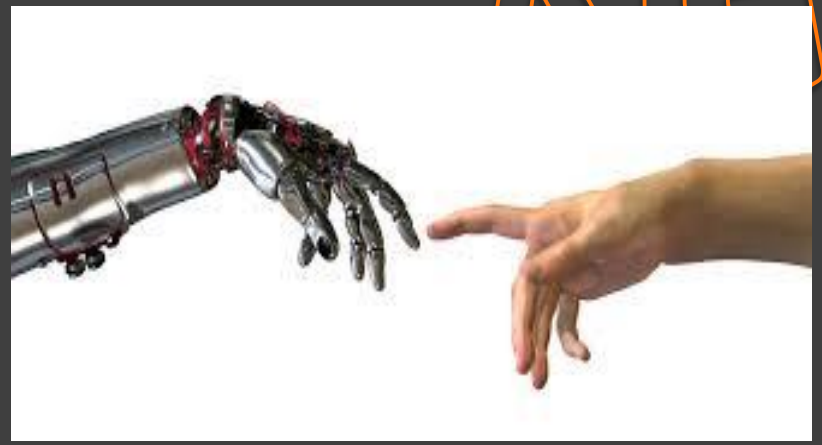
✓ High Computational Power

## PROMISES AND DANGERS

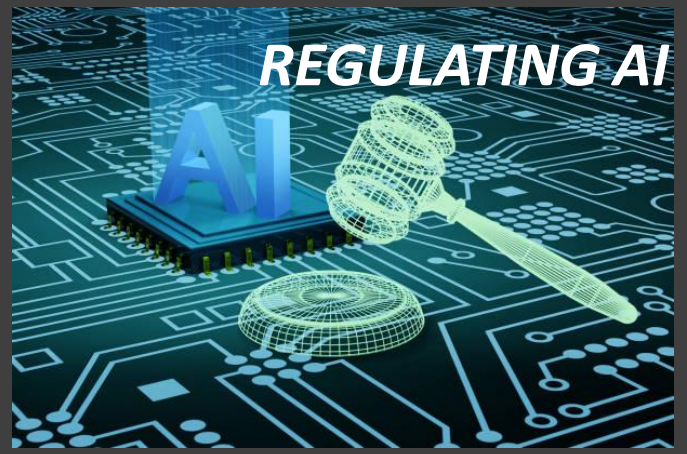



AGENDA

AI4GOOD- Beneficial AI



## REGULATING AI




## ARTIFICIAL GENERAL INTELLIGENCE?

*Eugénio Oliveira*

U.PORTO LIACC
artificial intelligence and computer science laboratory

# AI- Artificial Intelligence



**Perceive** the environment

Collect and **Interpret** data

Derive **Knowledge** from Data

**Reason** over the Knowledge

**Decide** on potential actions or conclusions

May **GENERATE** new contents

May **Learn and Adapt** their beheviour

**COMPUTATIONAL SYSTEMS**

**Criativity?**

*AI is NATURALLY created by Humans!!*
*Concerning Natural Intelligent, discussion goes on …*

## Five Tribes

- Symbolists
- Connectionists
- Evolutionaries
- Statisticians (Bayesians)
- Analogizers

Knowledge Representation
Knowledge Processing

Logics (Predicates, Temporal, Modal, Intentional…)

Biologic Analogy **(ANN)**

**GA** using simplified laws of **Evolution**

**Entropy** to build **Decision Trees**

Different schools, same objectives:
To develop machine intelligence!

*The Master Algorithm: The ultimate Learning Machine that will remake our world*

*(P. Domingos, Basic Books)*

*Eugénio Oliveira*

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# What is really NEW about AI???

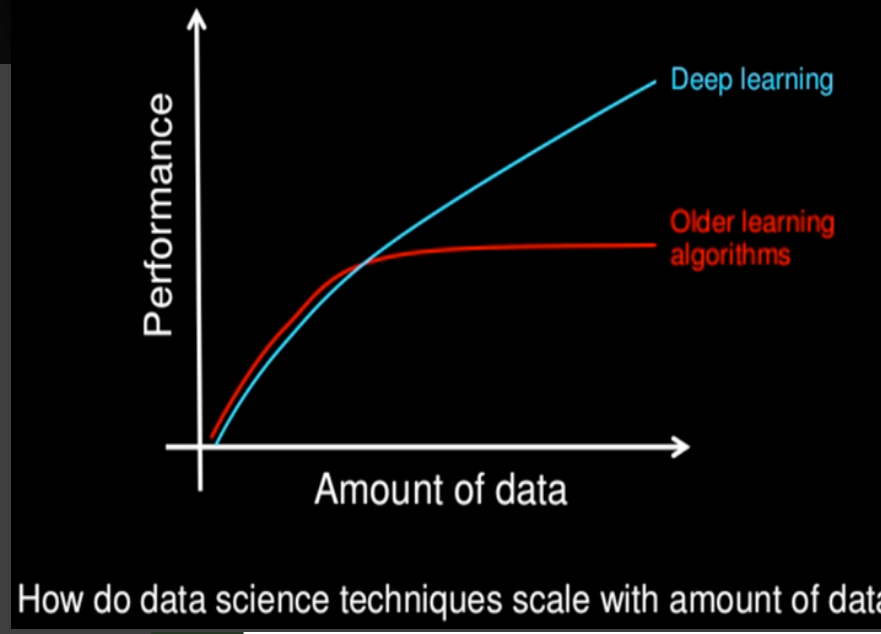


✓**HARDWARE**: High Computational Power

✓**BIG DATA** (> $10^{15}$….)

✓**Algorithms: DEEP LEARNING**

**Artificail Neural Networks based**



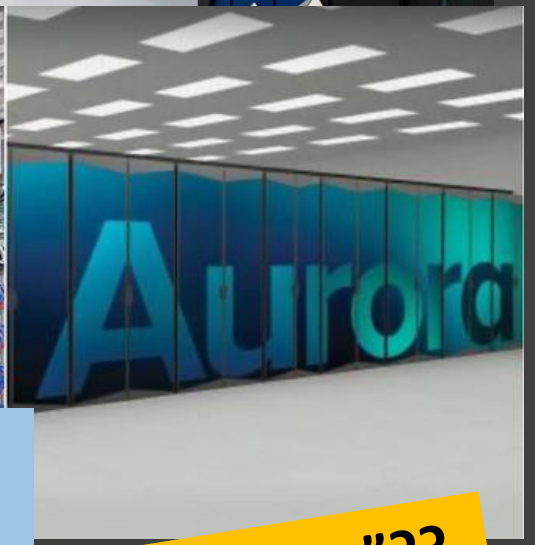

IT-friendly **Environments**:
(Domotics, Smart Cities, Health, Defense, Education …)

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# ✓HARDWARE: High Computational Power


**Summit** supercomputer, at Oak Ridge N. Lab.
area: 2 tennis courts, >27,000 GPUs.
**AI**/ Deep Learning for  understanding climate changes

**US Frontier**:1,206 petaFlops on the LINPACK benchmarks.


**Aurora: (**2024) **exascale** supercomputer US Department of Energy (DOE) and designed by Intel and Cray for the Argonne N. L. 1.012 exaflops (Billion billion = 1000 PF**)**

**"El Capitan"??**


TOP

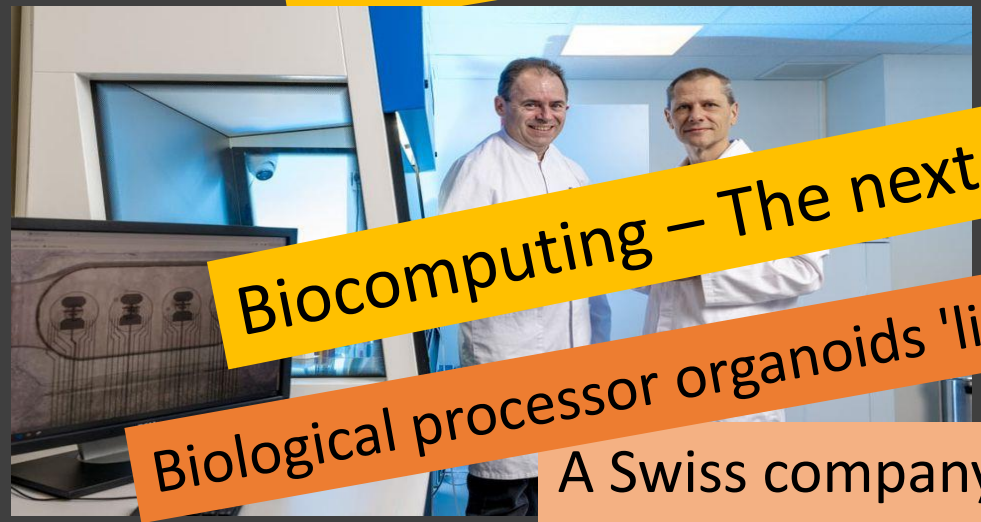
**Tianhe-3 -** "Xingyi" - supercomputer built by China's National University of Defense Technology.
peak performance **of 2.05 exaflops** and a sustained performance of 1.57 exaflops on High Performance LINPACK.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

✓HARDWARE

International Centre for Neuromorphic Systems (ICNS) at **Western Sydney University**

**DeepSouth**. IBM Neuromorphic Supercomputer simulating synapsis at the human brain scale.
228 trillion synaptic operations/s,

## Biocomputing – The next evolutionary leap

## Biological processor organoids 'live' about 100 days

A Swiss company, **FinalSpark**, has launched a '**bioprocessing platform**' that uses lab grown **brain tissue to run computing task**s. Their claim is that this tissue uses a million times **less power** than silicon chips

| | System | Peak Petaflops | HPL Petaflops | Compute Efficiency | Concurrent Cores+SMs | Cores+SMs 1 Exaflops HPL |
|---|---|---|---|---|---|---|
| | *NSC/Guangzhou "Tianhe-3"* | *2,050.0* | *1,567.6* | *76.5%* | *???* | *???* |
| | *NSC/Wuxi "OceanLight"* | *1,500.0* | *1,220.0* | *81.3%* | *41,930,000* | *34,368,852* |
| 1 | Oak Ridge "Frontier" | 1,679.8 | 1,194.0 | 71.1% | 8,699,904 | 7,286,352 |
| 2 | Argonne "Aurora" | 1,059.3 | 585.3 | 55.3% | 4,742,808 | 8,102,655 |
| 3 | Microsoft Azure "Eagle" | 846.8 | 561.2 | 66.3% | 1,123,200 | 2,001,426 |
| 4 | RIKEN "Fugaku" | 537.2 | 442.0 | 82.3% | 7,630,848 | 17,263,971 |
| 5 | CSC "LUMI" | 531.5 | 379.7 | 71.4% | 2,725,704 | 7,178,573 |
| 6 | CINECA "Leonardo" | 304.5 | 238.7 | 78.4% | 1,824,768 | 7,644,608 |
| 7 | Oak Ridge "Summit" | 200.8 | 148.6 | 74.0% | 2,414,592 | 16,248,937 |
| 8 | BSC "MareNostrum 5 ACC" | 234.0 | 138.2 | 59.1% | 680,960 | 4,927,352 |
| 9 | Nvidia "Eos" | 188.7 | 121.4 | 64.4% | 485,888 | 4,002,372 |
| 10 | Lawrence Livermore "Sierra" | 125.7 | 94.6 | 75.3% | 1,572,480 | 16,615,385 |

✓ **Algorithms: DEEP LEARNING**

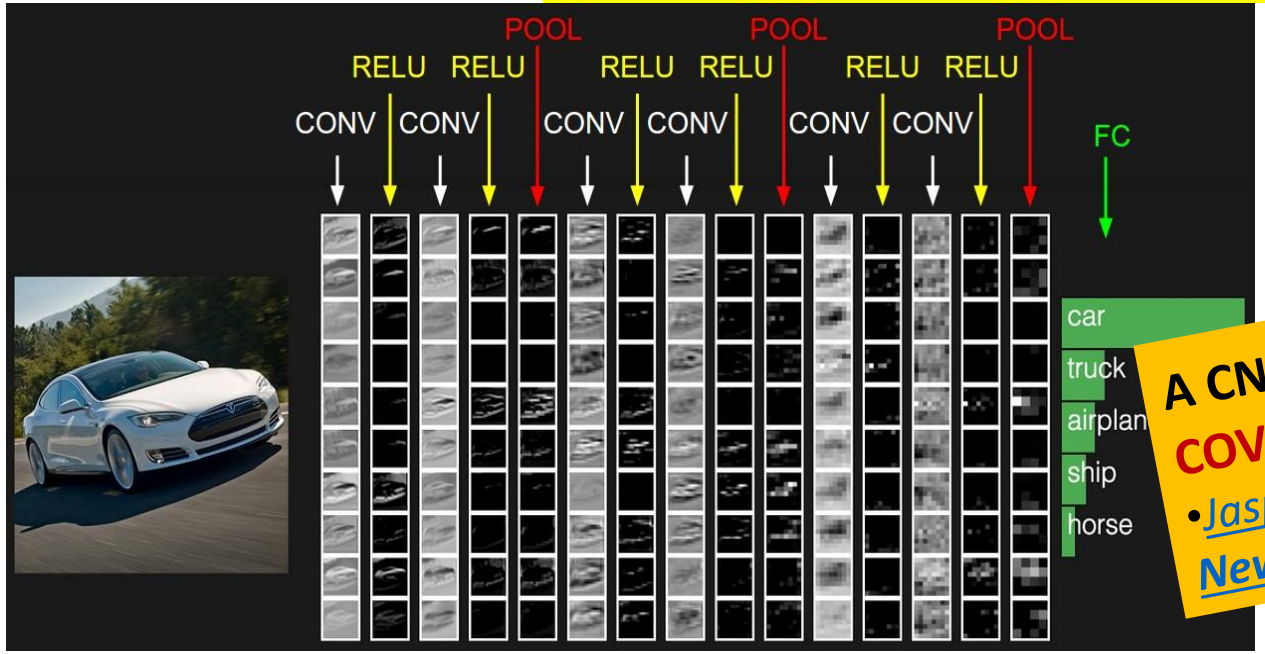
**Artificial Neural Networks based**

**Algorithms that progressively extract higher-level features from the raw input**, using multiple Layers of possible non-linear Transformations.
**Outputs specialize according to the patterns hidden in inputs.**

Progressive **Abstraction** →

Eg.: an **Image**, may be represented through:
vector of **pixels intensity**;
set of lines (**edges**);
space regions (**faces**)
**Object** recognition

# DL: CNN - Convolution Network




**A CNN Transfer Learning-Based Automated Diagnosis of COVID-19 From Lung Computerized Tomography Scan Slices**

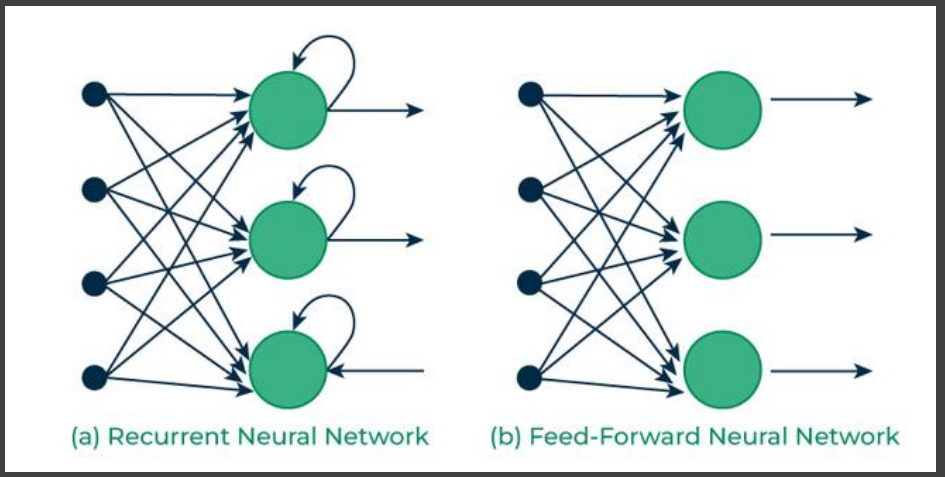
•Jaspreet Kaur & Prabhpreet Kaur

***New Generation Computing* volume 41**, *pages 795–838 (2023)*

CNNs make use of filters (also known as **kernels**), to detect what features, are present throughout an image.

Each CONV layer includes Filters producing Activation Maps.

# Deep Learning (**RNN Recurrent Networks**)




Recurrent Units have the ability to maintain hidden states, allowing the network to capture sequential dependencies.

**LSTM and Gated Recurrent Unit (GRU)** versions improve by remembering previous inputs while processing the RNN's ability to handle **long-term dependencies.**

**Advantage**

An RNN may remember each piece of information through time.

**Disadvantages**

-**Training an RNN is a difficult task.**

-Gradient vanishing and exploding problems.

-It cannot process very long sequences if using *tanh* or **relu** as **an activation function**.

# Differences Summary

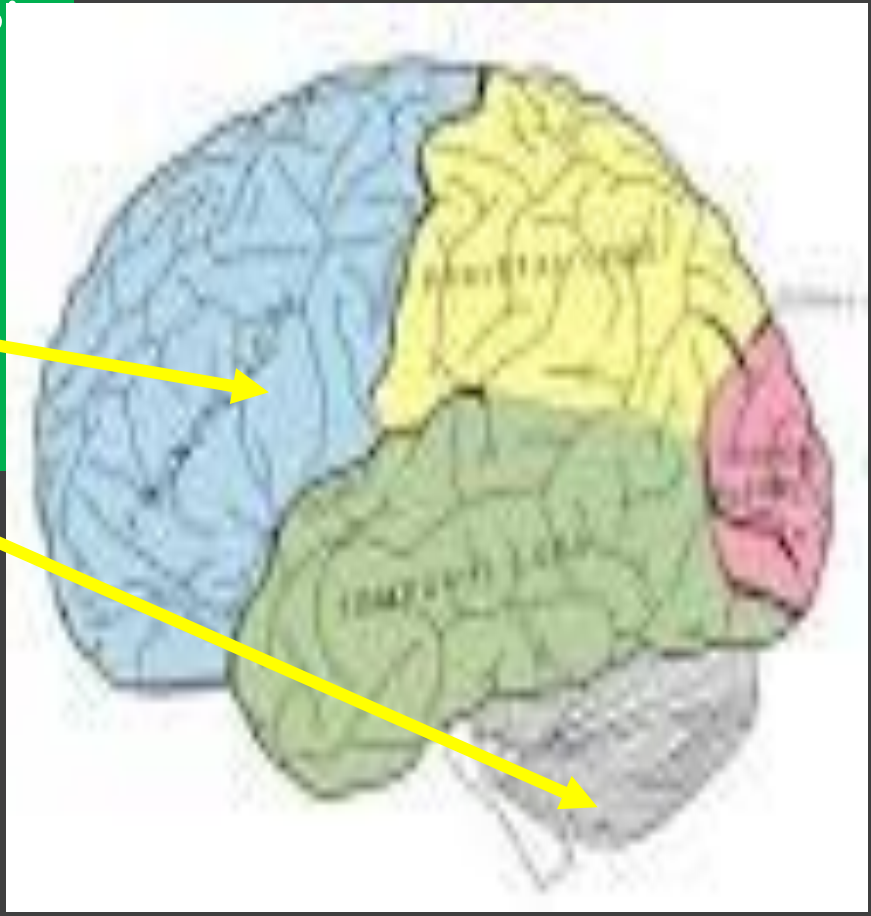
| Feature | CNNs | RNNs |
| --- | --- | --- |
| Primary Use | **Spatial** data (e.g., images, videos) | **Sequential** data (e.g., text, time series) |
| Key Layers | **Convolutional**, pooling | **Recurrent** (e.g., LSTM, GRU) |
| Connection Pattern | **Local** connectivity, parameter sharing | **Temporal** connectivity, shared weights |
| Memory | **No memory** of previous inputs | Maintains **hidden state** for memory |
| Strengths | **Feature extraction**, translation invariance | Sequence modeling, temporal dependencies |
| Common Applications | **Image** recognition and generation, object detection | **Language** modeling, machine translation Text generation |

# ANN functionalities

Different DL NNs:

May be useful for mapping fuctionalities of diferente **Brain lobs**

**occipital** (vision, **ConvNets**),
**frontal** (behaviour, working Memory : **RNN)**
**temporal** (ANN, recognition, **LTM)**
**parietal** (perception, spacial representation: STM)



# Human Brain

~$10^{11}$ Neurons.
$10^4$ synapsis/ neuron
(average value)

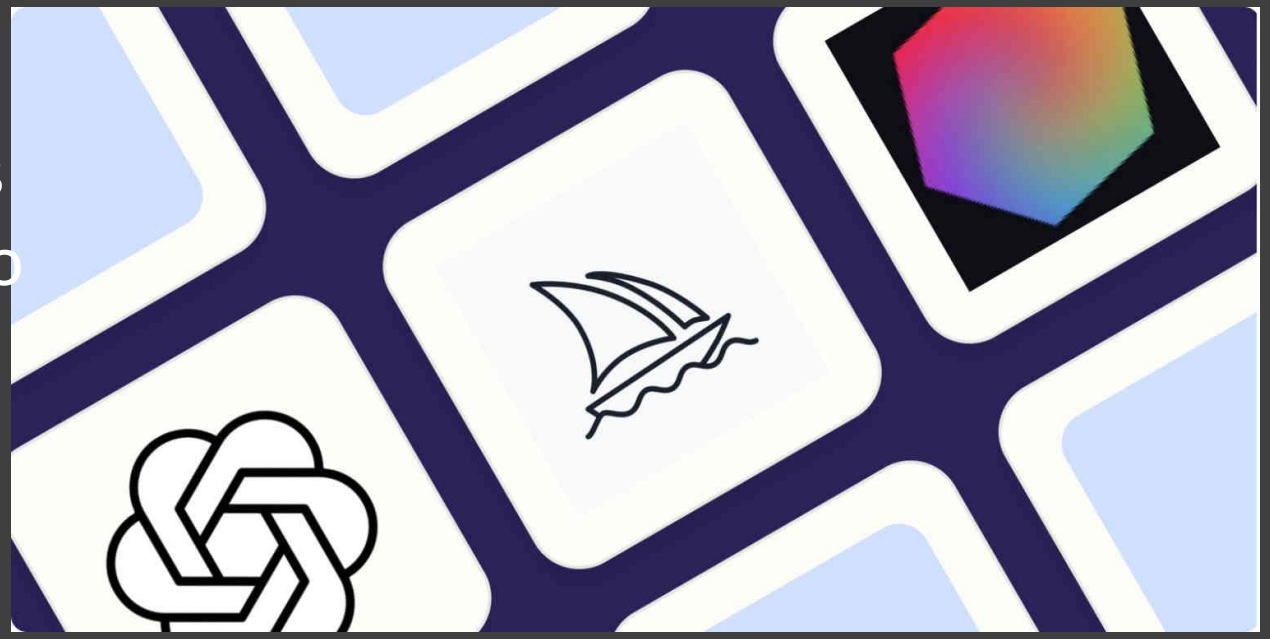
$10^{15}$synapsis*10 spikes/second→$10^{16}$ operations/s
~25 watts very eficient (?up to 100w?)

Cortex: between $10^4$ and $10^5$ neurons per $mm^3$

- Brain operates in a massive parallel way

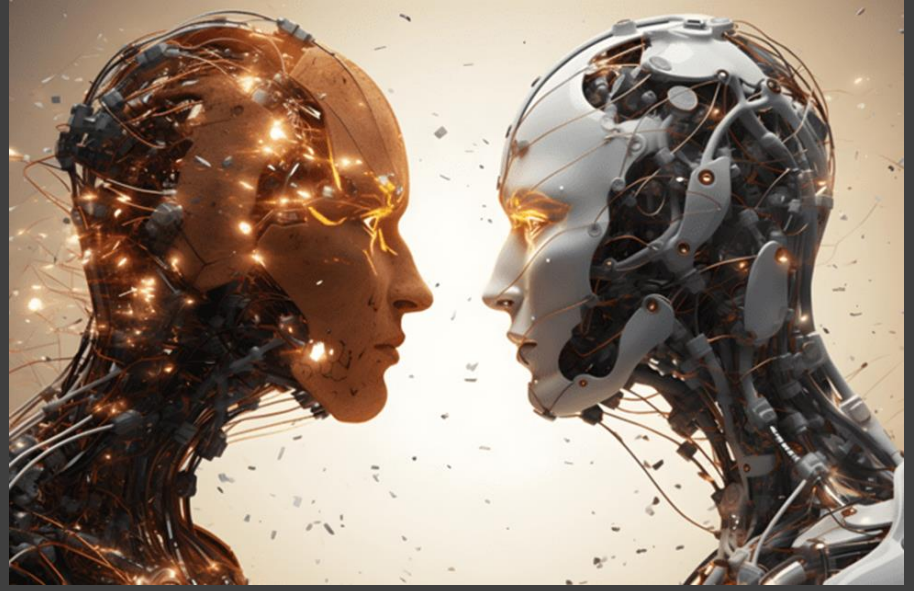
# GENERATIVE AI

Generative AI can produce various types of content, including text, imagery, audio and synthetic data.
**LLMs** mostly use transformers.



**Transformers** use a concept called **attention** that enables models to track the connections between related tokens.
*eg. words across pages, chapters and books,* rather than just in individual sentences. ***words / pixels/code / proteins /chemicals / DNA.***

**"Attention is all you need"** Vaswani NIPS 2017 using the softmax-based attention mechanism proposed by Bahdanau et. al. in 2014 for machine translation but using a scaled (down) dot product



**1960-70**: **Markov** Models – algorithms that generate next states based on probabilities

**2010-20**: AI Algorithms– “Deep Learning” ANN Architectures + powerful **CPUs**

Traditional AI: analyzes Data and get to the Conclusions (Decisions, Previews…)
**GENERATIVE AI**: GENERATES new data related with training data sets.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

https://youtu.be/38-xqTmSUDY

# GENERATIVE AI

Generative pre-trained transformers (**GPT**) are a type of **LLM**

February 14, 2019

**Better Language Models and Their Implications**

**GPT-2** NL Model trained with **8 M web page**

OpenAI launched ChatGPT, with zero fanfare, in November 2022

Users seen as guinea pigs …

GPT-2 "could be" **used** for:

. **Automatic Translation, Text Summarization, …AI Assistants (text generation)**

"**GPT-3** from1.5 to 175 B /(thousands of Millions) parameters

**Business:** released because "is either us or the others" …

"**GPT-4 100 T (Billions) Parameters** **than GPT-3**

**Trained with 45 TBy**

**CHAT GPT** became the fastest-growing consumer product in history

GPT-4o ("o" for "omni"), ChatGPT-o understands text, sound and images, answering the input text with voice or generating images.

GPT-5

AlphaSignal 14/1/24

2T-5T (2,000B-5,000B)

Sam Altman hints at new details of **GPT-5**

June 18th 2024, OpenAI has announced the **training** of GPT-5, which promises to surpass the capabilities of the latest GPT-4o.

**GPT-5** is more focused in the **multimodal** interactions including Audio and Video.

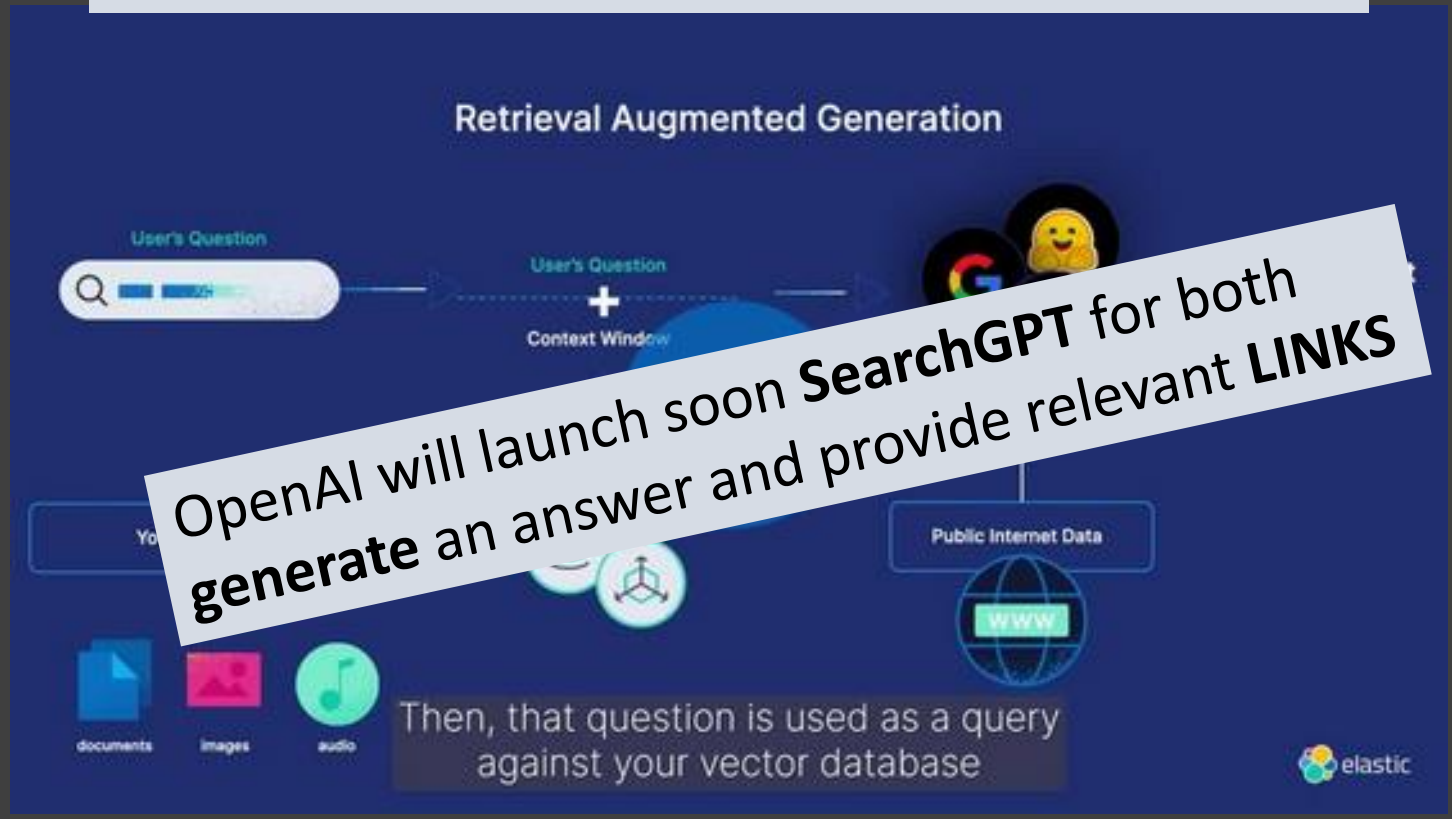
**GPT-5** is said to be more **responsible** by doing **logical** analysis of the information.

# NEW FM Developments

While **vector-based** models are good at retrieving information based on similarity, they're bad at relevance and context.
And require a lot of computational **power**,

RAG solutions remove context when encoding information

## RAG- Retrieval Augmented Generation




OpenAI will launch soon **SearchGPT** for both **generate** an answer and provide relevant **LINKS**

**SLM** - Small language models are still an emerging technology for very focused AI use cases. For example, a tool for building an **internal documentation chatbot** that is trained to provide employees relevant information on the company

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

**MY Prompt**: audiance including the kings of Spain, in a talk on Artificial intelligence at an University in Madrid


Midjourney, Inc., São Francisco based Lab

# The best AI image generators at a glance

| | Best for | Access options | Price | Parent company |
|---|---|---|---|---|
| DALLE·3 | Ease of use | ChatGPT Plus or Enterprise; Bing's AI Copilot; API | Included with ChatGPT Plus at $20/month | OpenAI |
| Midjourney | High-quality results | Discord | From $10/month for ~200 images/month and commercial usage rights | Midjourney |
| Stable Diffusion | Customization and control | DreamStudio; Clipdrop; API; and lots of other iterations, including downloading it to a local server | Free for 25 credits; from $10 for 1,000 credits | Stability AI |
| Adobe Firefly | Integrating AI-generated images into photos | firefly.adobe.com, Photoshop, Express, and other Adobe tools | Free for 25 credits; from $4.99 for 100 credits/month | Adobe |
| Generative AI by Getty | Commercially safe images | iStock | From $14.99 for 100 AI generations | Getty (uses NVIDIA Picasso) |

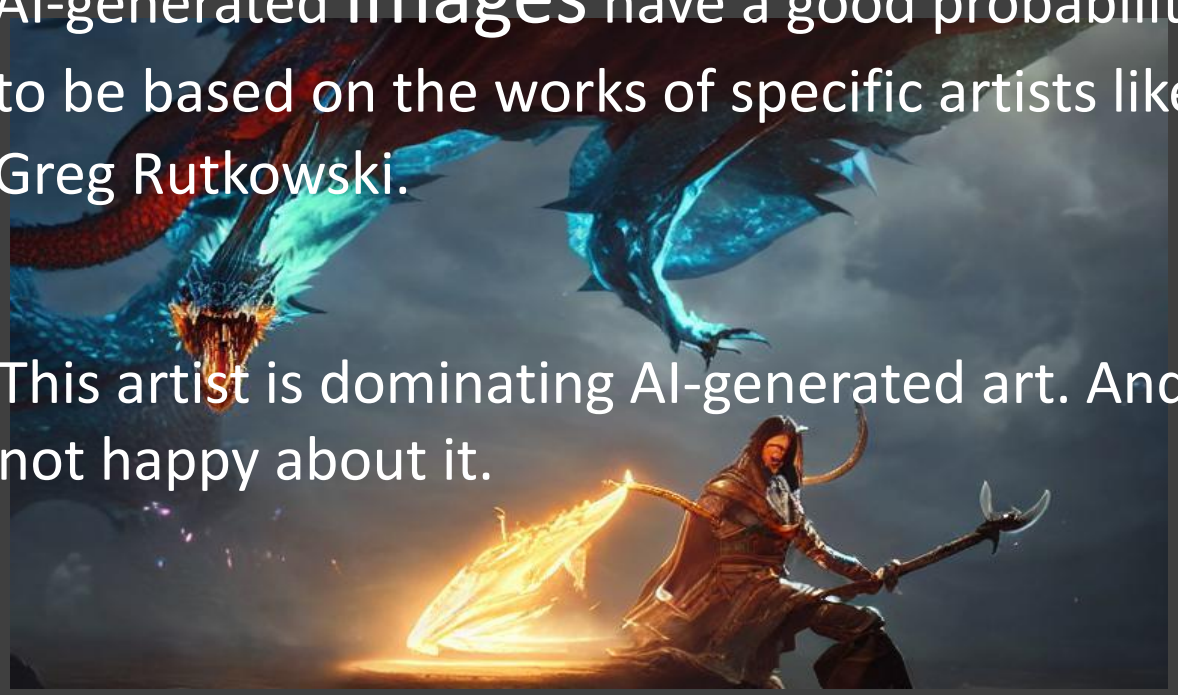
But the best way to use DALL·E 3 is through ChatGPT. If you're a ChatGPT Plus subscriber

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# Hallucinations, Authorship

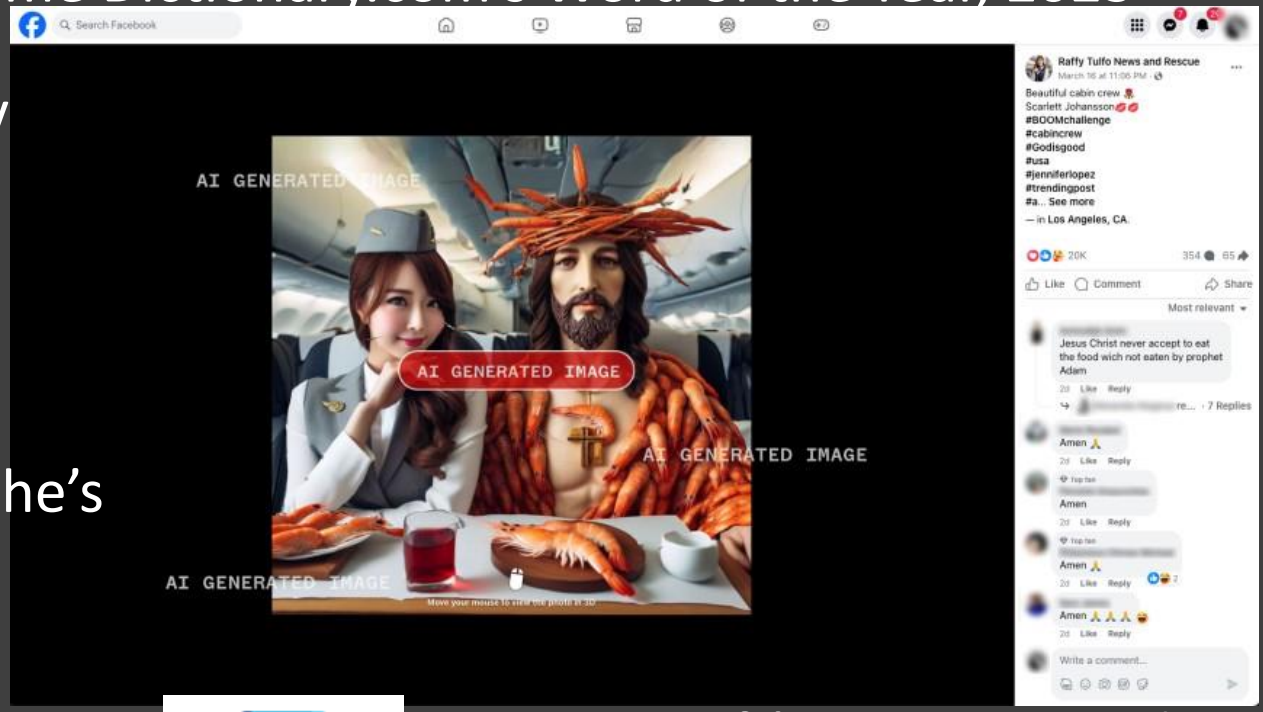
the term "hallucinate" became Dictionary.com's Word of the Year, 2023

AI-generated images have a good probability to be based on the works of specific artists like Greg Rutkowski.



This artist is dominating AI-generated art. And he's not happy about it.

GR is a more popular prompt than Picasso.




Microsoft's Azure AI Studio new built-in safety features to identify and block suspicious inputs in real time.

**AI** is **learning to deceive humans …**
One of the most striking examples is META's CICERO, which turned out to be an expert liar.

May 13

the blueprint BY INTERESTING ENGINEERING

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# Tests find AI tools create **election lies**

By ALI SWENSON, May 31, 2024


Washington, D.C.-based Center for **Countering Digital Hate** created audio clips of five false statements about elections in the voices of eight prominent American and European politicians.

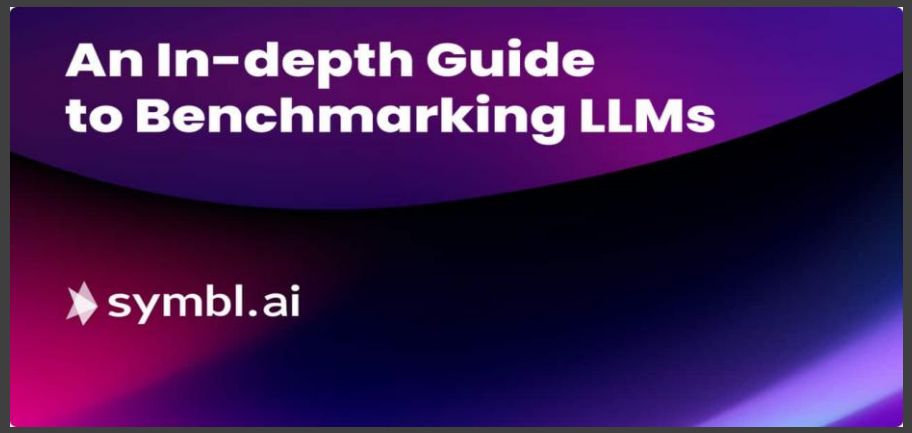
a fake Biden says election officials count each of his votes twice.

a fake Emmanuel Macron warns citizens not to vote because of bomb threats at the polls.

convincing voice clones in 80% of the time

Eugénio Oliveira

# GENERATIVE AI: Benchmarking




**MMLU**: Massive Multitask Language Understanding

**SuperGLUE** General **Language** Understanding Evaluation (GLUE)

**GSM8K** Grade School Math 8K benchmark **mathematical** reasoning abilities.

**HumanEval**-Python measures a model's ability to generate functionally correct **code**

**MT-Bench** evaluates a language model's capability  engage in multi-turn **dialogues**. Measures a model's ability to answer subsequent, related questions

U.PORTO
FEUP FACUL
UNIVE

# LiveBench: A Challenging, Contamination-Free LLM Benchmark

… ,Yann LeCun, and others

work is sponsored by Abacus.AI

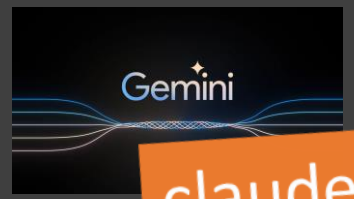
**LiveBench**: releases new questions monthly, based on recently-released datasets, arXiv papers, news articles, and IMDb movie synopsis.

Each question has verifiable, objective ground-truth answers

18 different tasks across 6 **categories**

Reasoning + Mathematics (competitions)+ Data Analysis(predicting…)+ Language (word puzzle, Typo removing)+ Instr. Follow (summ., story gen…) + Coding (gen.)

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO


| Model | Global Average | Reasoning Av. |
|---|---|---|
| **gpt-4o-2024-05** | **55.28** | 48.00 |
| **gpt-4-turbo-202404** | **54.59** | 56.00 |
| **claude3opus-202402** | **54.29** | 48.00 |
| **gpt-4**-1106-preview | 51.86 | 42.67 |
| ~gpt-4-0125-preview | 48.99 | 40.67 |
| gemini-1.5-pro-latest | 45.67 | 28.00 |
| claude3sonnet-202402 | | 33.33 |
| | | 38.67 |
| | | 32.00 |
| command-r-plus | 35.7 8 | 34.00 |
| gpt-3.5-turbo-1106 | 3 5.66 | 28.67 |

claude-3.5-sonnet-2024-06-20 Av. 61.16
gpt-4o-2024-05-13 Av. 54.96

At last World AI Conference in Shanghai, SenseTime, a leading Chinese AI firm, unveiled its latest model, **SenseNova 5.5**.SenseTime claims that SenseNova 5.5 is just as good as GPT-4, *Sujita Sinha, "Interesting Engineering" July 09*

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

IMPACT

Google DeepMind Researchers Use AI Tool to Find 2 Million **New Materials**

***Financial Times*** *Michael Peel November 29, 2023*

*July 22, 2024*

Accenture published “Reinventing R&D in the age of AI,” outlining how **biopharmaceutical** companies use AI for **drug and therapeutic research** and development pipeline.



U.S. Rep. Jennifer Wexton (D-VA) regained the voice she lost with the **help of an AI voice-cloning program** from ElevenLabs

*Associated Press*; Dan Merica (July 25, 2024)

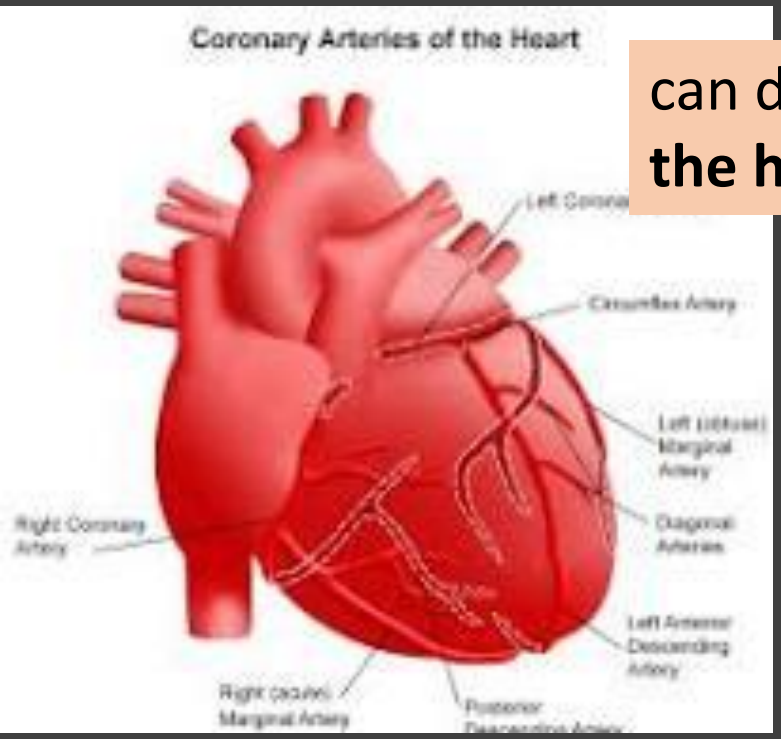
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

TechTarget | ComputerWeekly.com

ISSUE DATE: June 5, 2024

IMPACT

**University of Oxford** has published a study looking at how AI could help in the **diagnosis of heart disease**.




can detect the **level of inflammation** of the heart arteries, **not visible to the human** eye.

**pilot study** in 4 **NHS hospitals**,
In **around 45% of cases, doctors changed their patients' treatment** plan as a result.



U.PORTO LIACC artificial intelligence and computer science laboratory

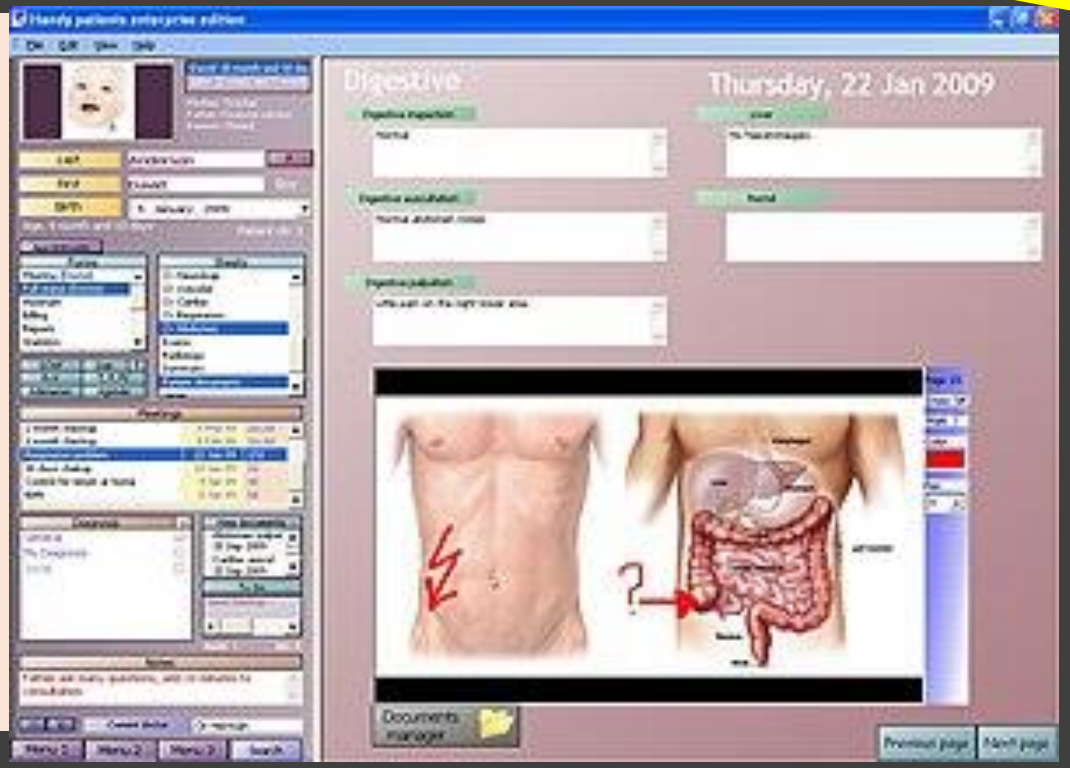
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

IMPACT

Gen AI to creates **EHR "Electronic Health Records"** in Hospitals USA, sparing Clinicians time.

**Ambient Clinical Intelligence** combines AI tools (Voice recognition, Summarization) to produce in seconds the relevant personalized patient data after the medical consultation.

*TechTarget Hannah Nelson, Assistant Editor09 Aug 2024*




According to a recent study published by JACC: CardioOncology, researchers from "Brigham and Women's Hospital", using **AI successfully detect several types of Arrhythmias** following exposure to radiation during lung cancer treatement.

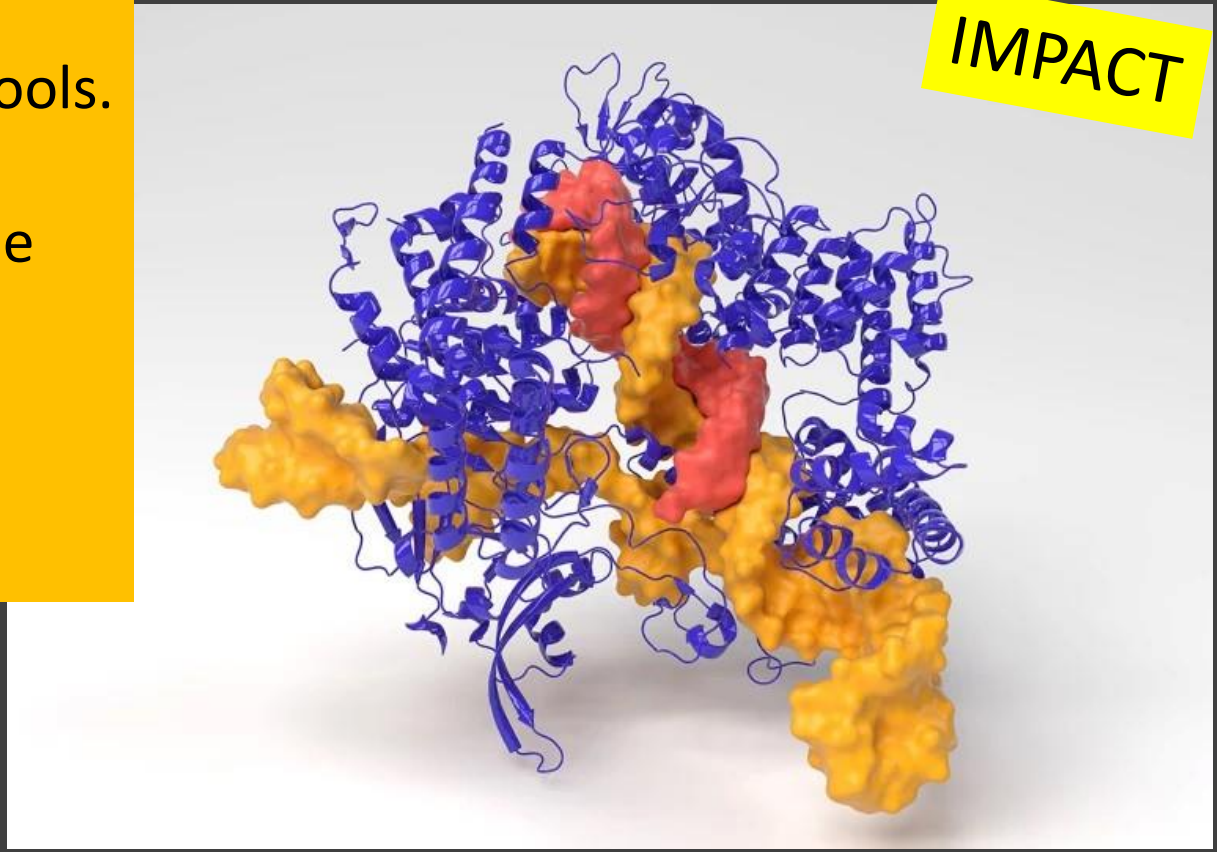
NATURE 29 April 2024

‘ChatGPT for CRISPR’ creates new gene-editing tools.

AI-designed **Gene editors** could be more versatile than those found in nature.

**Providing protection against virus**

IMPACT



CRISPR-designing were trained on vast amount of biological data in the form of **protein or genome sequences**.

The healthcare sector is using **GenAI** for clinical documentation, patient communication and clinical text summarization.

IMPACT

Medical Doctors use Chat GPT : presentations, scientific articles, reports, documentation

**LANCET 7/23** “Ethics of LLMs in medicine and medical research“

## LLMs for Medicine

**BioGPT** (MIT), LaMDA , Med-PaLM 2 (Google),

Sparrow(Deepmind,UK), Pangu Alpha (Huawei, China), OPT-IML (Meta, USA), Megatron Turing MLG (Nvidea, USA)

LLM BioGPT is domain specific. Pre-trained with PubMed DB of scientific articles !

IMPACT


AI increases **software** creation productivity.
*Eg.: Copilot is AI for **code** generation (snippets).*

Microsoft introduced  Windows PCs designed for AI.
**Copilot+ PCs** are the fastest, most intelligent Windows PCs ever built. trillion operations per second, all–day battery life and access to the most advanced AI models



"We're moving from computation towards **cognition** into the age of AI," said **Dell**.

*Published: 20 May 2024  ComputerWeekly.com*

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# GENERATIVE AI: Power Games

Sam Altman, CEO and co-founder of **OpenAI** was dismissed

November 18, 2023[1]

**Open AI workers** said they will leave the company

Sam Altman joined **Microsoft**

Sam Altman returned to **OpenAI**,

The **five-days** interregnum between Altman's firing and his return marked a pivotal moment for the company

**Sutskever and Jan Leike, criticized OpenAI for prioritizing "shiny new products" over vital**

lawsuit filed in federal court in Northern California says that **Altman** "*assiduously manipulated **Musk** into co-founding their spurious non-profit venture*" by promising that OpenAI would be safer and more transparent than profit-driven alternatives.


**co-founder Ilya Sutskever left the company**

# GENERATIVE AI: Power Games


UK former Chancellor announced an "**AI Incubator**" **i.AI** including "an elite team of technical experts at the heart of government".

**UK government** announces £8.5m in grants for AI safety research
*ComputerWeekly.com:* May 22, 2024

Minister of Science, Patrick Vallance said 19/8/24 he sees a great opportunity on AI for Medicine, Materials and Public Services

Reuters says that **Xi Jinping's** governement objective is to make **China self-sufficient** regarding Semiconductors production

U.PORTO

# GENERATIVE AI: Power Games

**Taiwan Semiconductor Manufacturing Company**

so-called the "**most important**" company in the world.

It is estimated to produce 90% of the more advanced semiconductor chips in the world. (eg. NVIDIA)

The Netherlands' ASML Holding and TSMC can disable the world's most sophisticated **chipmaking** machines in the event China invades Taiwan, sources say.
*Bloomberg (May 21, 2024)*

"MacroPolo" think-tank found that nearly **half** of the world's top AI researchers come from **China**
while only around 18% come from U.S. **undergraduate institutions**

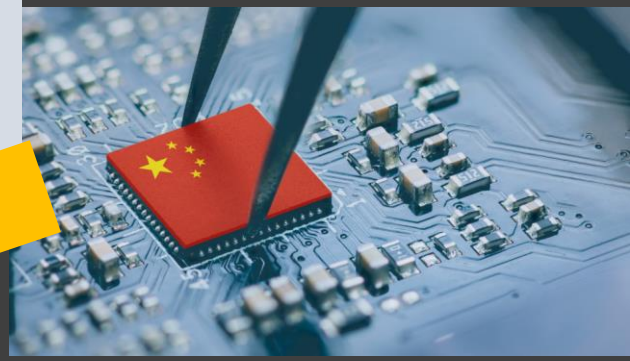
AI researchers working in the U.S.: **31%** are American **38%** are from China.

The **U.S. is home to around 42%** of the world's top AI talent
***NYTimes*** *March,22 2024*

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# GENERATIVE AI: Power Games

"One of the most intriguing aspects of the **Tianhe-3** is its processor. "
Like AMD "Antares" **hybrid** MI300A CPU-GPU that is going into El Capitan
*TheNextPlatform*




**FUGAKU Japonese** supercomputer from Fujitsu+I.I.Riken uses **CPUs**, not GPUs "now in short supply due to a fierce global LLM development race." *The Japan Times (May 11, 2024)*

The E.U. Commission has approved, a **€5 billion** German measure to support **European Semiconductor Manufacturing Company** ('ESMC') in the construction and operation of a microchip manufacturing plant in **Dresden**
Deutche Welle (20/8/24) : "joint venture" **TSMC,** Bosch, Infineon, and NXP (The Nederlands)

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO


**February 26, 2024** edition of **ACM** TechNews

## The highest-paying skill in tech may not surprise you

**Right now, there's nothing more in demand than AI knowledge—and the salaries prove it.**

**average** salary of $174,727

Here's the jobs list:

- Generative AI
- SoC
- Deep learning
- Torch
- PyTorch
- Computer v… …ernlog Mesos Rust Elixir

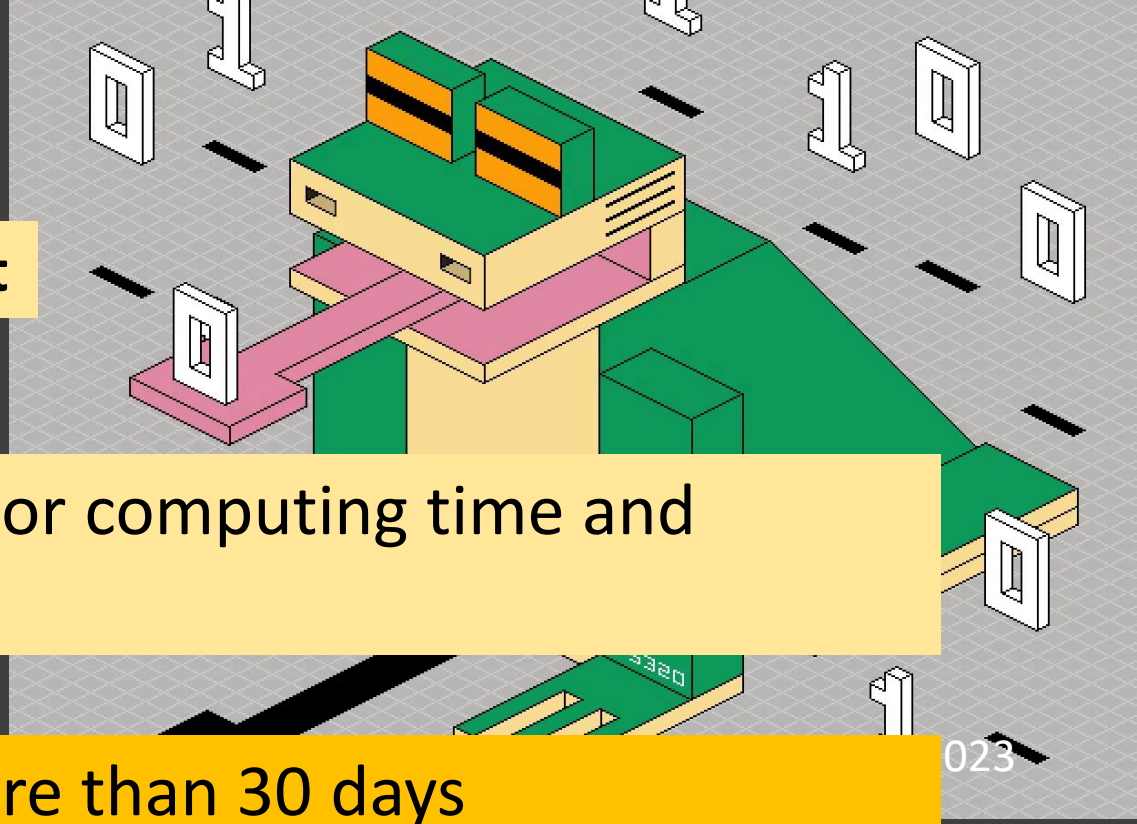
AI-related jobs offer salaries that are more than **77% higher** than other fields. Some positions start with compensation that's as high as $450,000 per year
*FastCompany.com* 06/06/24



U.PORTO LIACC artificial intelligence and computer science laboratory

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# Power Consumption



"Generative AI revolution comes with a planetary **cost**

GPT3 required 355 years of a single processor computing time and consumed 284,000 kwh of energy to train

**GPT3** was trained using 1000+ GPUs for more than 30 days
**GPT4** 10,000+ GPUs. (equivalent in the US to 23 M Dollars of electricity)

A.de Vries, UVA predicted that GenAI will consume as much **energy yearly** as a country like **Ireland**. (29.3 terawatt-hours per year).
By **2027** the AI sector could consume between 85 to 134 TwattH/y. That's about the same as the annual energy demand of **the Netherlands**.

(*Joule Volume 7, ISSUE 10, P2191-2194, Oct. 18, 2023)*

# Power Consumption

Depending on the hardware, training a LLM of any significant size can take up to weeks or months to complete.

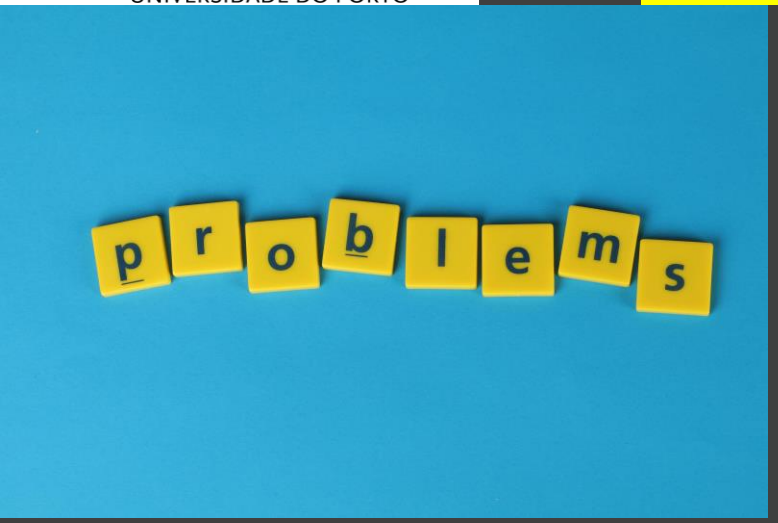
Google, Intel, Nvidia fight for better training LLMs.
**Nvidia** is 1st place, followed by Intel… Google last.

*IEEE Spectrum Samuel K. Moore 12 Nov 2023*

NVIDEA's **Eos** supercomputer can train a 175 billion parameter **GPT-3 model in under 4 minutes**.

NVIDIA **Nemotron**-4 340B family includes base, instruct and reward models that form a pipeline to generate **synthetic data used for training and refining LLMs**.

# Gen AI Drawbacks and Limitations




Doesn't understand the **cause** → **effect** relationship.

Cannot explain the domain principles → **black box**

Lack of **accuracy** since it is trained on internet data publically available.

Potentially **hallucinates** making up false statements.

**Originality**? copyright violations

How is it different from someone learning from many sources and create something based on what he/she learnt?

# Gen AI Drawbacks and Limitations

Impact on **Labor**:
Any automation leads to disruption in job markets.

**impact on the white collar, not on blue collar jobs …**

**Power** consumption: LMMs training is very power intensive

**Bias**: amplify the bias that already exists in the data we produce, on issues of race, gender, language or culture. Abusive associations can result.

# Gen AI Drawbacks and Limitations

Impact on basic **cognitive** skills

Could FM diminish the need for humans to acquire knowledge or get trained professionally?

Will it help humans to operate at higher layers of cognition sooner in age?

**OpenAI o** performance in the qualification for the Mathematical **Olympiad** was modest: **89th** in 500 students (solved 83% of the problems)


How is it going to affect Education, creativity and artistic skills?

# Gen AI Drawbacks and Limitations

**Monopoly** of power players:

DANGER ?!

FM require huge amount of resources

GenAI is dominated by a small number of tech giants.

Need for an **AI Agency** similar to International Atomic Energy Agency to regulate at a supra-national level the use and proliferation of this technology.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# A desirable world with Responsible AI

**Gen AI to unleash human productivity:** faster discovery of **pharmaceutical** drugs, faster and more bug free **software** coding, quicker text/images/video **creation**, less **bureaucracy**

Let's not fear but master Technology !
Historically, T. has done more benefits to humanity than damage if properly regulated and managed.

USA: almost two-thirds (61%) CEOs, are pushing their organizations to adopt GenAI faster than some people are comfortable with.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# The Role of Ethics in AI

principles of **fairness, transparency, and accountability**.

AI systems must not operate as black boxes; their decisions need to be comprehensible, transparent, and **justifiable**.

**Privacy** concerns are central to the ethical deplo of GenAI ensuring responsible data collection in compliance with

fundamental factors for long-term sustainability.

**Transparency** informing when interaction is with an AI-driven tool

**YHH Nexus**. “For thousands of years prophets, poets and politicians have used **language to manipulate** and reshape society. Now **computers** are learning how to do it. And they won’t need to send killer robots to shoot us. They could **manipulate** human beings to pull the trigger.”

# WHAT is needed?

**When using AI tool, WHO takes the Responsibility?**: (eg. Medicine)

Who is using it (Medical doctor)?
Who gives permission for using it? (Hospital /Clinic)?
Who should regulate it (M. of Health Justice, Parli[illegible]ent)?
who sells it, distributes it (Compa[illegible]
Who does researc[illegible] Universities)?
Who should Test a[illegible]
Who advertises it?

All?
A lot of Legislation is needed!!

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# Digital afterlife industry?

It is now possible to use technology to raise the dead ...

Today, a “digital afterlife industry” to create reconstructions of dead people **based on the data they have left behind.**

Microsoft has a patent for creating a conversational chatbot of a specific person using their “social data”.


**Character.ai**, a website that hosts chatbots of people living, dead, and fictional. The site provides the ability to chat with notable dead people like Shakespeare, Elizabeth II, Tolkien.

**Postmortem digital possibilities**

Eugénio Oliveira

U.PORTO LIACC

artificial intelligence and computer science laboratory

# Human-like AI?

*How far away are we from* ***building "human-lik***

*the key problems that we need to solve bef*

to

Machines that will have **us believe** they **have a self, or a personality**, should be relatively easy to develop. But whether they would be truly self-aware, we will only know if we **crack the "hard problem of consciousness" first.**

**RL** used with AlphaGo demonstrated how an ANN can learn and **invent** strategies

**abstract** knowledge accumulated by solving a specific problem, and apply this **k** in solving a **different** prob

## Human-level A

| | State of Play |
|---|---|
| | Solved ✓ |
| g taught | Solved ✓ |
| earning | Not yet |
| Common sense | Not yet |
| Self-awareness | Still mysterious |

**hard** problem. The function of human **memory** is perhaps the key to developing **common sense** in machines

role of a sophisticated Body and the emergence of Sentiments and Emotions. under investigation

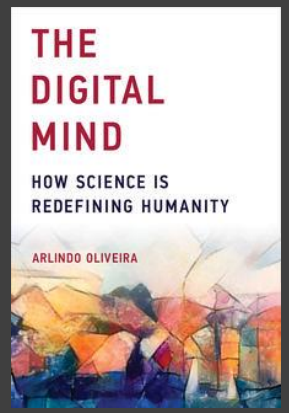
# • Artificial General Intelligence

**AGI** raises the issue of the possibility of artificial consciousness

Deep Blue and Alfa Go: **intelligences without consciousness**

**Could Consciousness be an EMERGENT property out of the complex interactions of many specific intelligent capabilities??**

"The Mind is an **emergent** property of the Brain enabling humans to have a set of cognitive capabilities" *The Digital Mind, Arlindo Oliveira*



U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Yann LeCun
@ylecun
Professor at NYU. Chief AI Scientist at Meta.
Researcher in AI, Machine Learning, Robotics, etc.
ACM Turing Award Laureate.

**General intelligence**, artificial or natural, does not exist.
all animals have **specialized intelligence**. They have different collections of skills and an ability to acquire new ones **quickly**.
That's the kind of learning that we need to reproduce in machines before we can get anywhere close to human-level AI.

*Yann LeCun 2024, 24 May*

**Open AI: 5 levels for AGI. We reached level 2**, working on 3(Agents), 4 (Innovations); last one is building up Organizations

# Beneficial AI

**Planetary OBJECTIVE**:

**AI for Environment sustainability (WATER, LAND, AIR)**

**Ethically rewarding**

SMART *: Cities, Homes, Factories

**WEF** pointed out more than 80 possible uses of AI favouring the Environment: **Climatic** protection, and Modeling, Autonomous mobility, Intelligent Power Networks, …

# Beneficial AI

## 17 Goals for a Sustainable Development

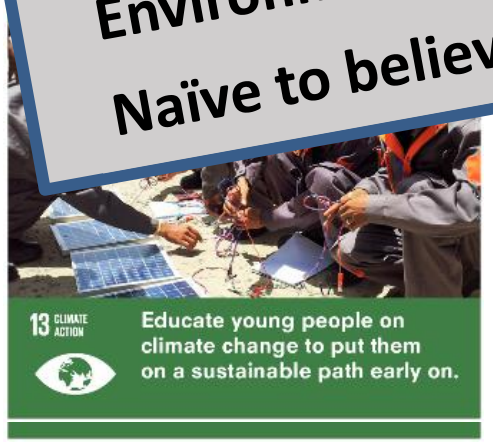
**Satisfying needs of the Present without affecting Future Generations**

WATER

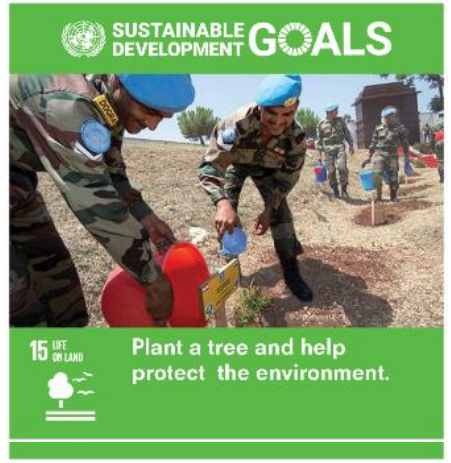

ENERGY



**Three Dimensions of Well-being:**
**Environment Society Economy**
**Naïve to believe 'win-win-win' situations**



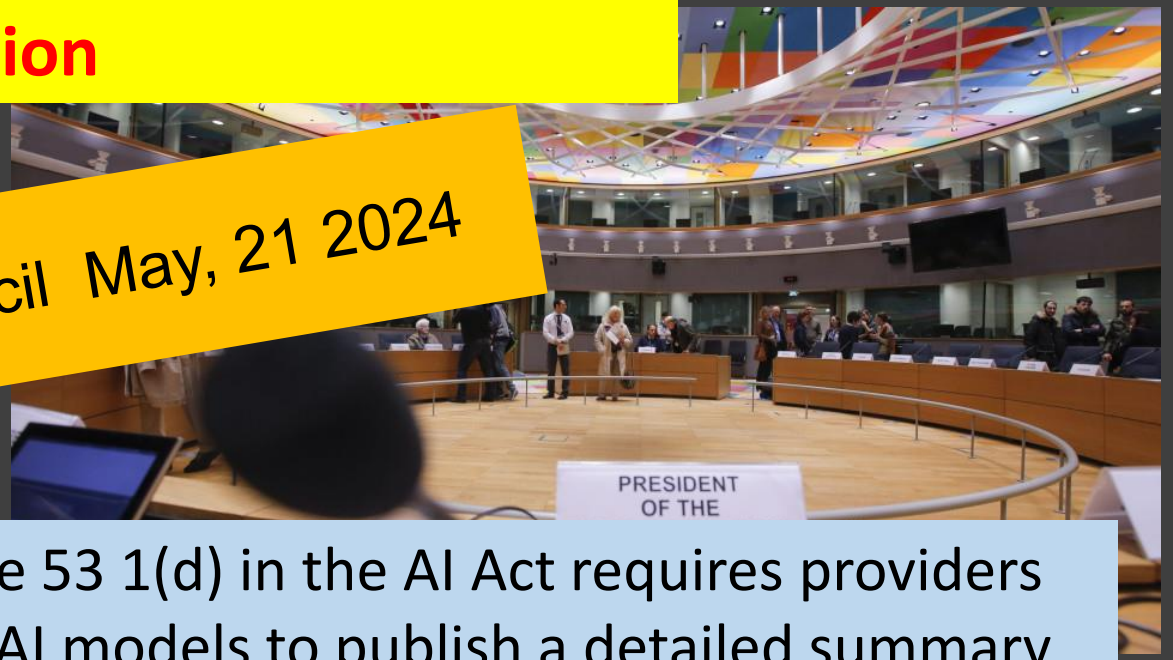

POLUTION


ODS13. Climatic actions ODS14. Protect Ocean Life ODS15. Protect Life on Earth

# Regulation

**EU AI ACT** approved by the European Council May, 21 2024




The EU ACT encompasses various safeguards, including **on GP AI**, limitations on **biometric identification** systems in law enforcement, bans on **social scoring** and AI used to manipulate or exploit user vulnerabilities

Article 53 1(d) in the AI Act requires providers of GPAI models to publish a detailed summary of **training** content. The summaries should cover **data sources** and sets as well as narrative **explanations**.

- UK: The new Labour government should place an outright ban on AI-powered "**predictive policing**" and **biometric surveillance** systems, …

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

# EU AI ACT Prohibits AI systems …

- to **predict** the risk of a person committing a criminal offence;
- to create **facial** recognition databases by untargeted scraping of facial images on the internet and closed-circuit television
- **emotion** recognition systems in the workplace and education institutions
- remote **biometric identification** in publicly accessible spaces


The E. Commission collects over 100 AI Pact signatures and an online kick-off plenary for the **GPAI code of practice** is scheduled for **30 September**.

Official Journal of the European Union

EN

L series

2.7.2024

HE

THE

of 13 June 2024

laying d… rules on artificial … Regulations (EC) … No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

California LEGISLATIVE INFORMATION

*SENATE BILL NO. 1047*

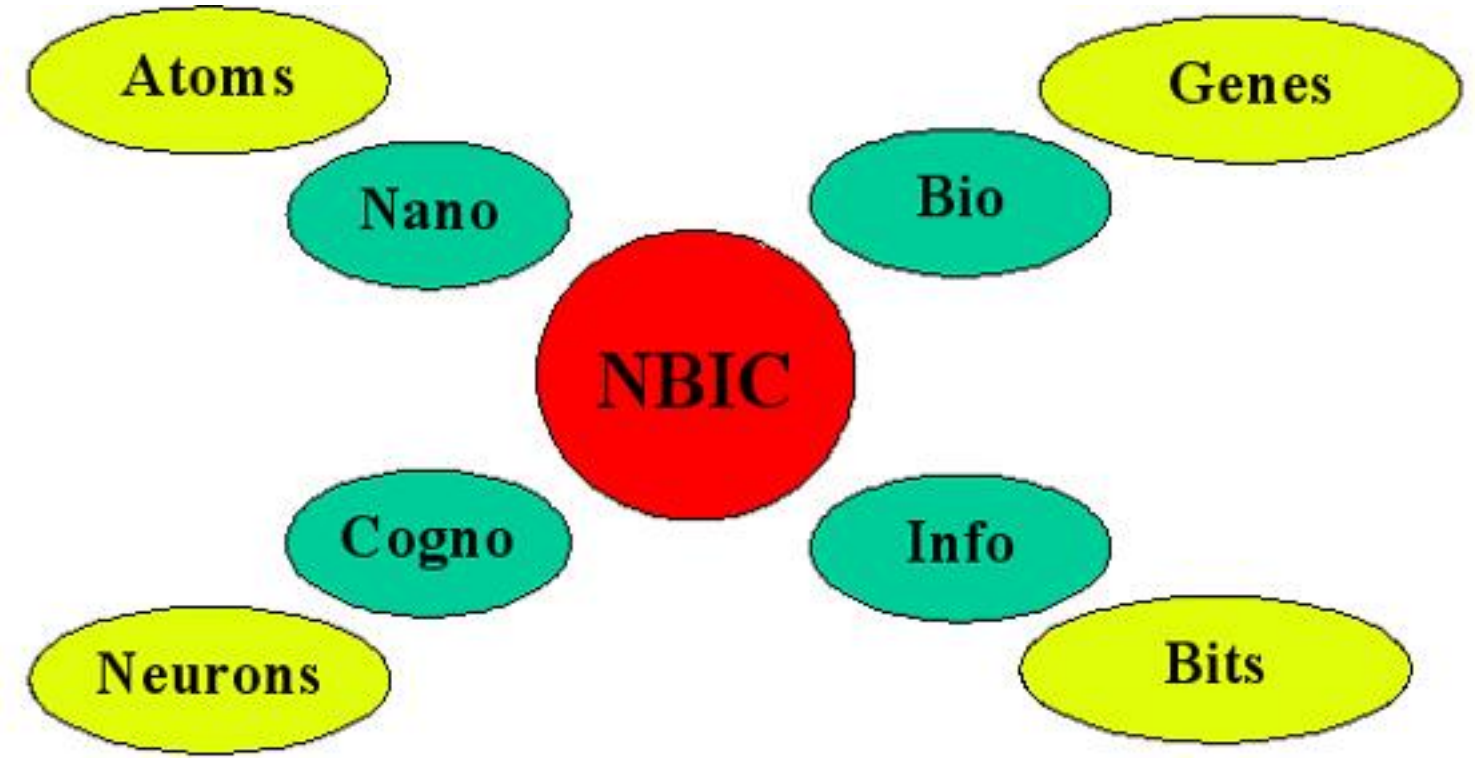
**Safe and Secure Innovation for Frontier Artificial Intelligence Models Act** to re

capability to promptly enact a **full shutdown**
a written safety and security pro
prohibits a d
**cr**
to
p **ions**, including **prohibiting a developer** from preventing
a from disclosing information, or **retaliating** against an **employee** for disclosing information, to the Attorney General or Labor Commissioner if the employee has **reasonable** cause to **believe** the developer **is out of compliance** with certain requirements

California **Governor** Gavin Newsom **vetoed.** Newsom said the legislation "does not take into account whether an AI system is deployed in high-risk environments, involves critical decision-making, or the use of sensitive data." *Politico; Lara Korte; Jeremy B. White* (***September 29, 2024***)

**OpenAI agreements blocked staff from revealing AI risks:**
**The agreement discouraged staffers and investors from expressing to federal authorities regarding safety violations.** Jul 14, 2024 INTERESTING ENGINEERING

# Sciences Synergy will shape the future




ROBOTICS

- ***a "Call to Action"***

**+ ART** as in **ART**ificial Intelligence:

**Accountability, Responsibility, Transparency**

**A**: To **whom** should we address when something goes wrong?

**R**: System providers are responsible for the clarity of the decision-making and identify errors or unexpected results

THE HUMAN IN THE LOOP !

**T**: **Transparency** refers to the need to describe, **inspect and reproduce** the mechanisms through which AI systems make decisions and learn to adapt to the environment, and to the governance of the data used.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

# AKNOWLEDGES


TechTarget | ComputerWeekly.com

acm TechNews

INTERESTING ENGINEERING

IEEE Spectrum

the blueprint BY INTERESTING ENGINEERING

future of life INSTITUTE

AlphaSignal

*Bloomberg/Getty*

***Financial Times***

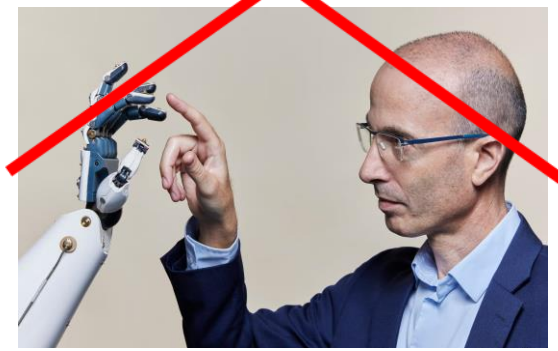
*NYTimes*

AI LOGS BY INTERESTING ENGINEERING

Eugénio Oliveira

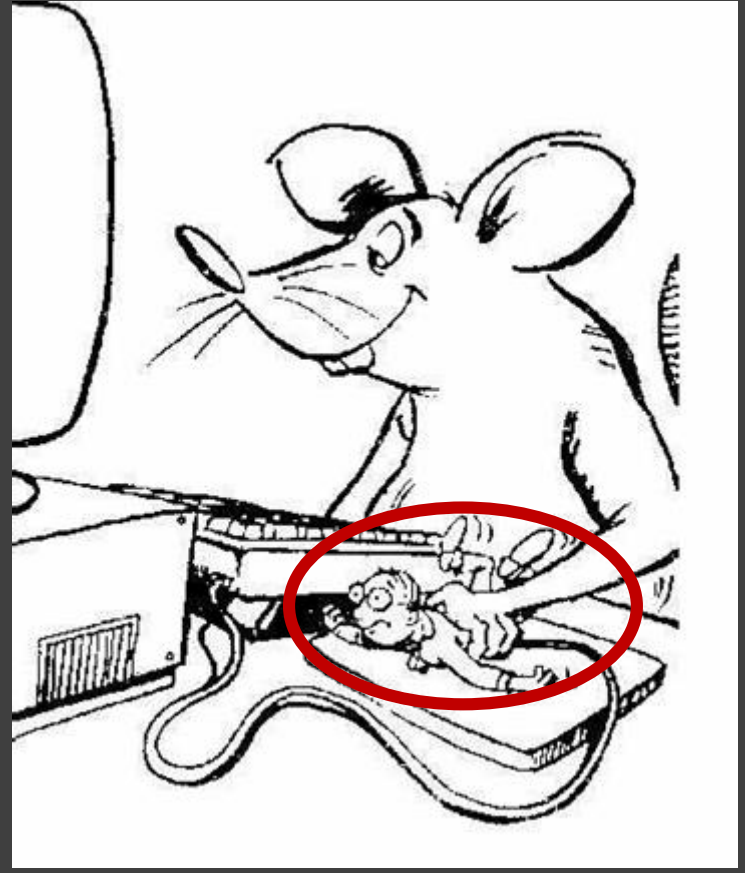

U.PORTO LIACC
artificial intelligence and computer science laboratory


I am afraid I just have questions … not answers …

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

**Ai is fascinating BUT …**
**BE CAREFUL**



THANK YOU!

*Eugénio Oliveira*