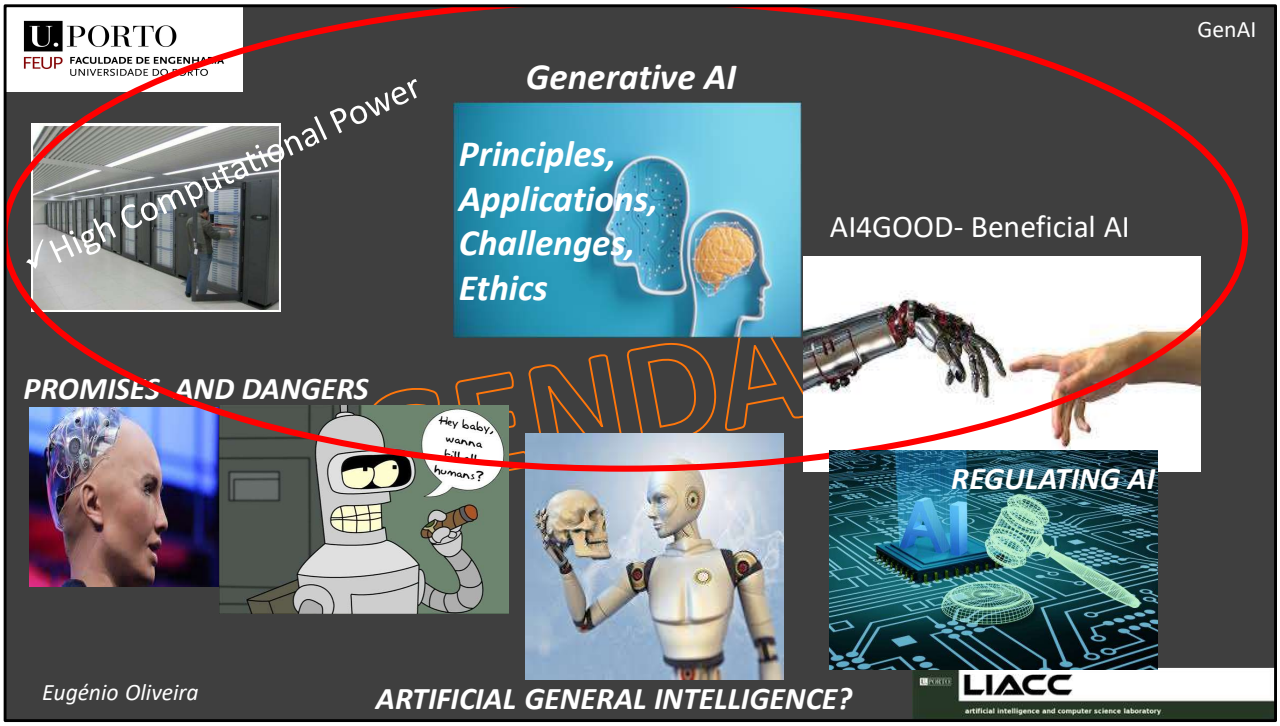
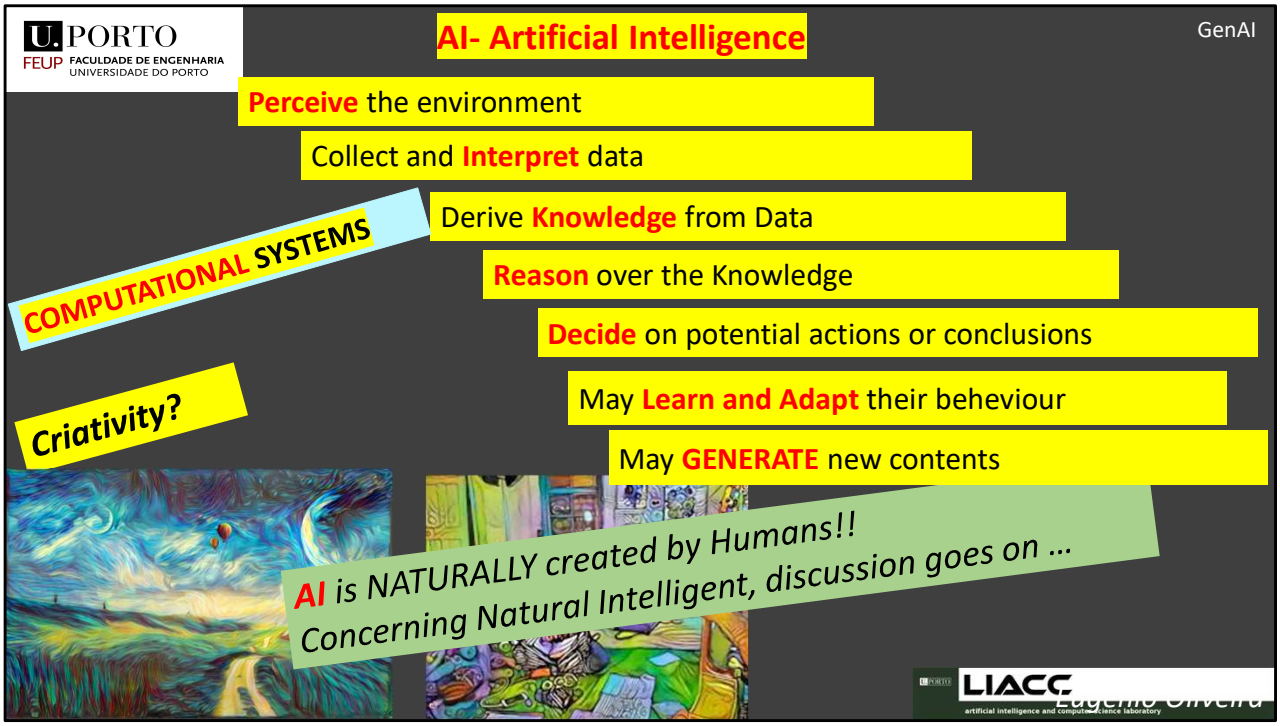
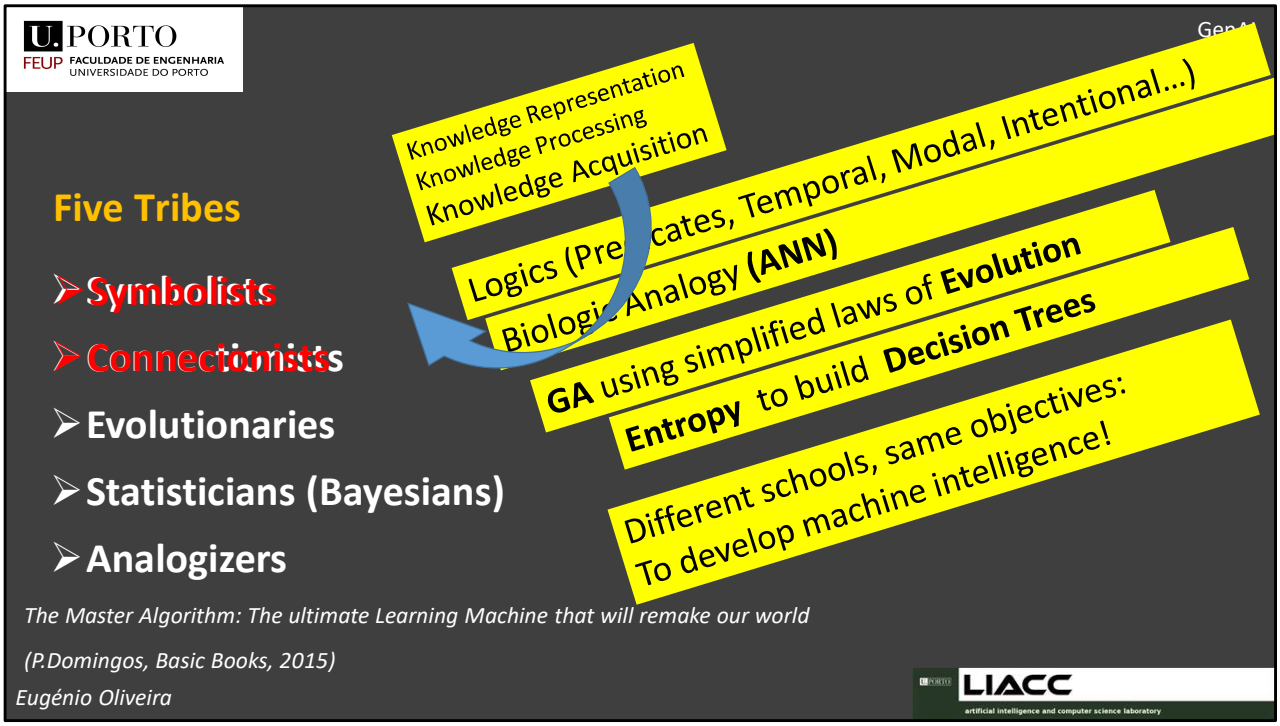
I am Emeritus Full Professor at the University of Porto, Faculty of Engineering where I was a co founder of na AI Lab called LIACC.
I will try to convey to you a few remarks on the AI impact knowing that every day the situation is changing and what is true now may be different tomorrow due to the tremendous speed of the AI evolution.
As a Subtitle I put a question on weather we are able to develop an AI 4 SG. I guess so.



I will refer basic techniques supporting Gen AI leading not just to inferre conclusions but instead to the generation of new contentes (mainly text and images).
After remembering the importance of AI applications for the advancement of Human society I will also pay attention to both Promises but also obvious Dangers to our the future. I m not sure I will have the time for some speculations on the Artificial General Intelligence since I need to conclude with a picture of the current situation concering several approaches to AI Regulations we believe will enable AI contribution for Social Good.



Just as an Introduction let us remember what we intend by AI Systems.
Wether an AI output may display Criativity is a matter of oppinion.
AI was Naturally created by US. I do not want to start a discussion on the idea that NI was really originated Naturally or Super Naturally!!!!
However we know for sure that men is accoutable concerning AI.



Just to remember the known approaches to develop AI based Systems and those that at the moment are dominating the scene: Those based on the Connectionist paradigm , mainly on Deep Neural Networks Architures , although those based on Logic and Statistics are still relevant as well.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GenAI

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: 2500cm^2, 2mm thick
180.000 Km de "filaments"
between 10^4 and 10^5 neurons per mm^3

- Brain operates in a massive parallel way

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

Let us just remember that the biologic analysis is simplistic due to the brain tissue complexity and energy efficiency.
It spends about 25 watts of energy.
Some say it can go up to 100 but usually consumes about 20w and is almost irrelevante if we are at rest or awake.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

What is really NEW about AI???

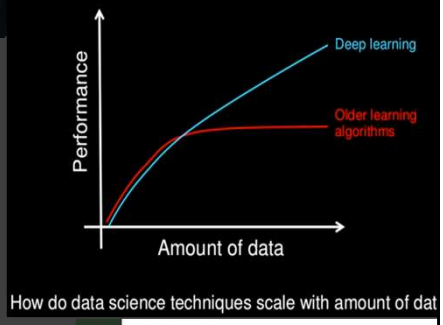
✓**HARDWARE**: High Computational Power

✓**BIG DATA** (> 10^{15}….)

✓**Algorithms: DEEP LEARNING**

Artificail Neural Networks based

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



Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

WHY are we NOW so deeply concerned with AI in the last yers?? A knowledge domain more than 60 years old?

Mainly because of 3 relevant FACTORS:

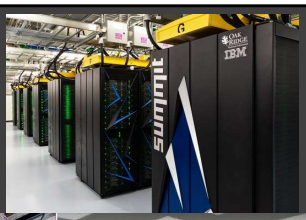
-Powerful HARDWARE coming to the scene.

- Big Data. Availability. Big Data means more than Petabytes.

- Finally a new class of algorithms called DL ANN based,

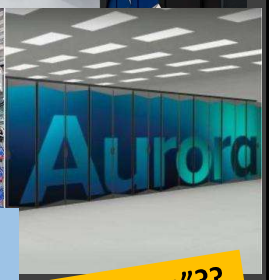
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

✓HARDWARE: High Computational Power



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

US Frontier is one of the most powerful supercomputer:1,206 petaFlops on the LINPACK benchmarks.



Aurora: (2024) exascale supercomputer sponsored by the United States Department of Energy (DOE) and designed by Intel and Cray for the Argonne National Lab. 1.012 exaflops (Billion billion = 1000 PF)

"El Capitan"??

Tianhe-3 - "Xingyi" - is the latest in a series of supercomputers 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.

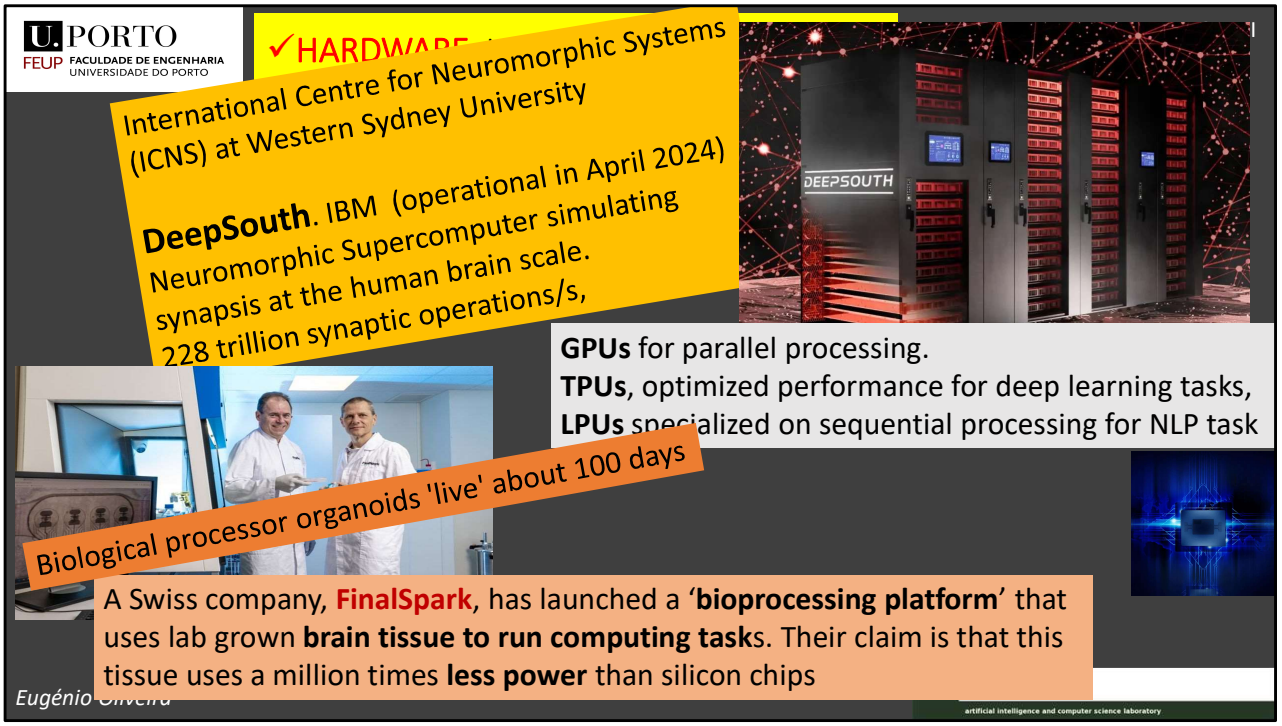
artificial intelligence and computer science laboratory

Starting with the Hardware: A few yars ago the novelty was Summit that takes the space of 2 tennis courts and includes more than 27 thousand GPUS. It has been built, they say … to better understand Climate changes.

US Frontier achieved more than 1 thousand Petaflops in the LINPACK benchmark.
LINPACK Benchmark measures the ability to solve a dense system of linear equations.

The secretive Tianhe-3 nicknamed «XINGYi» from the National Supercomputer Center in Guangzhou, China is the most powerful supercomputer. It is more or less secrete!

“El Capitan” supercomputer being built right now at Lawrence Livermore National Laboratory by Hewlett Packard and AMD.



new paradigmsfor Hardware are already there. There are New recent developments to Computers Hdw.

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

THENEXTPLATFORM

February 9, 2024 Timothy Prickett Morgan

	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

Eugénio Oliveira

artificial intelligence and computer science laboratory

This is the supercomputers ranking given by <the <next <platform where <frontier and Aurora appear on top and reach abot 1 thousand PFLOPS. But the truth is that Tianhe3 already reaches more than 2 thosand as a peak and 1.56 in average.

So the Computer power available for AI systems is now immense!!

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

✓**Algorithms: DEEP LEARNING**

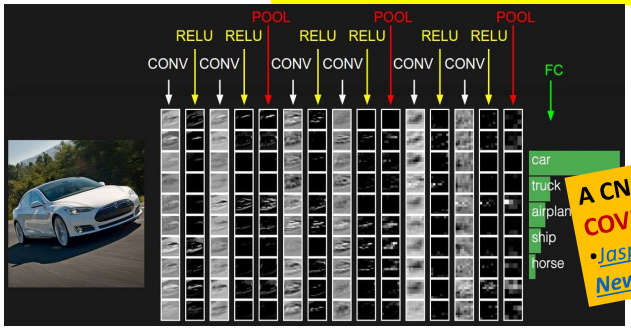
Artificail Neural Networks based

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** →

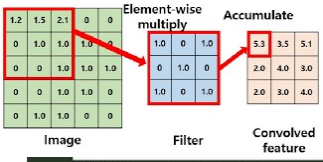
Eg.: an **Image**, may be represented through:
vector of **pixeils intensity**;
set of lines (**edges**);
space regions (**faces**)
Object recognition

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.
Layers are of different types



Two main Classes of Algorithms included in the DL Method
As you know WE are not just talking about Theory but about useful and impactful APPlications like COVID19 Diagnosis through Lung Computerized Tumography Images Analysis that save many lifes.

•Published: 05 October 2023

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

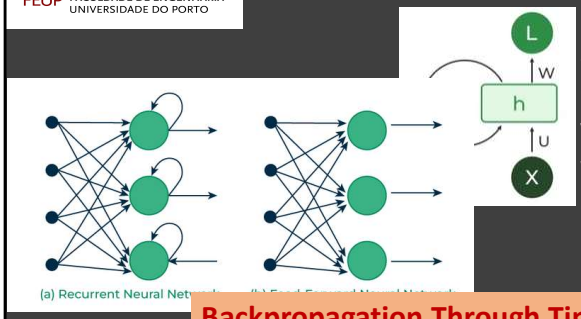
•Jaspreet Kaur & Prabhpreet Kaur

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

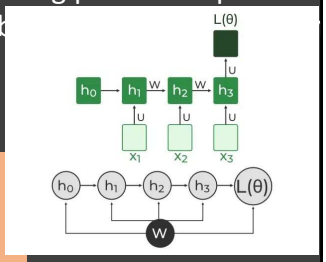
The outcome of the investigational analysis proves that the MobileNetV2 pre-trained CNN model obtained improved classification outcomes with 93.59% accuracy, 100% sensitivity, 87.25% specificity, 88.59% precision, 93.95% F1-score, 100% NPV, and AUC of 93.62%.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Deep Learning (**RNN Recurrent Networks**)



Recurrent Unit has the ability to maintain a hidden state, allowing the network to capture sequential dependencies. **LSTM and Gated Recurrent Unit (GRU)** versions improve by remembering previous inputs while processing the RNN's ab
dependencies.



Backpropagation Through Time (BPTT)
In a RNN each variable is computed one at a time in a specified order like first h1 then h2 then h3 so on. Hence **backpropagation is applied throughout all these hidden time states sequentially.**

Disadvantages
-Gradient vanishing and exploding problems.
-Training an RNN is a difficult task.
-It cannot process very long sequences if using *tanh* or **relu** as **an activation function**.

Advantage
An RNN remembers each and every piece of information through time.

The vanishing gradient problem occurs when gradients become too small during backpropagation. The weights of the network are not considerably changed as a result, and the network is unable to discover the underlying patterns in the data.
RELU Retified Linear Unit
TANH Hyperbolic Tangent (like a sigmoid) Function

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

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, object detection	**Language** modeling, machine translation

CNNs excel in tasks that involve spatial relationships and feature extraction from grid-like data structures, while RNNs are designed to handle and model sequential dependencies in time series and textual data.

Pooling: kind of combination
LSTM A Short Term Memory Lasting a large number of steps in the NN.
Gated recurrent units (GRUs) are like a long short-term memory (LSTM) with a gating mechanism to input or forget certain features

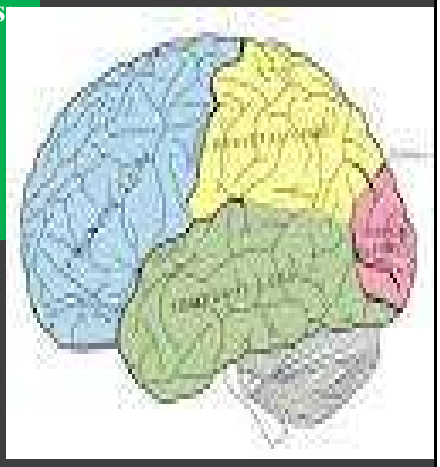
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

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)



Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

We can see different kinds of DL NN as mapping different brain regions functionalities.
Like the occipital lobe responsible for vision can be represented by CNNs and Frontal Lobe, main responsible for our behaviour and Working Memory is better mapped on RNNs.
This is a huge simplification.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

GENERATIVE AI

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.

Eugénio Oliveira

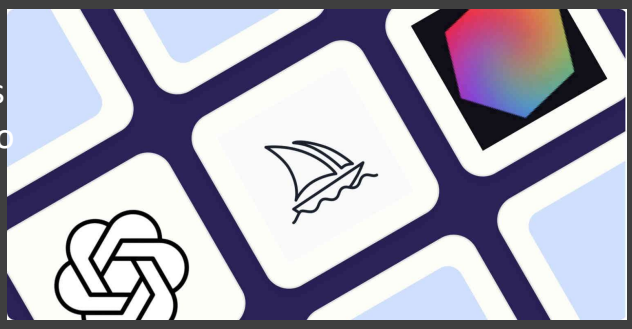
artificial intelligence and computer science laboratory

The basic theory behind Gen AI is not completely new,
In the sixties there were Markov Models based algorithms

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

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 tokens.
eg. words across pages, chapters and books rather than just in individual sentences. words / *code / proteins /chemicals / DNA*.

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

Eugénio Oliveira

LIACC
artificial intelligence and computer science laboratory

The seminal paper enabling GenAI techniques is Vasani paper.
based on the softmax-based attention mechanism proposed by Bahdanau et. al. in 2014 for machine translation
Scaled (down) dot product is used.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

how do Transformers work?

1. **Input Embedding**: The input sequence is first embedded into vectors through techniques like word **embedding**

2. **Positional Encoding**: positional encodings to provide information about the position of each embedding in the **sequence**.

3. **Encoder Layers**: The encoder consists of multiple layers, each comprising **two main components**: multi-head **self-attention** mechanism and position-wise fully connected **feed-forward networks**

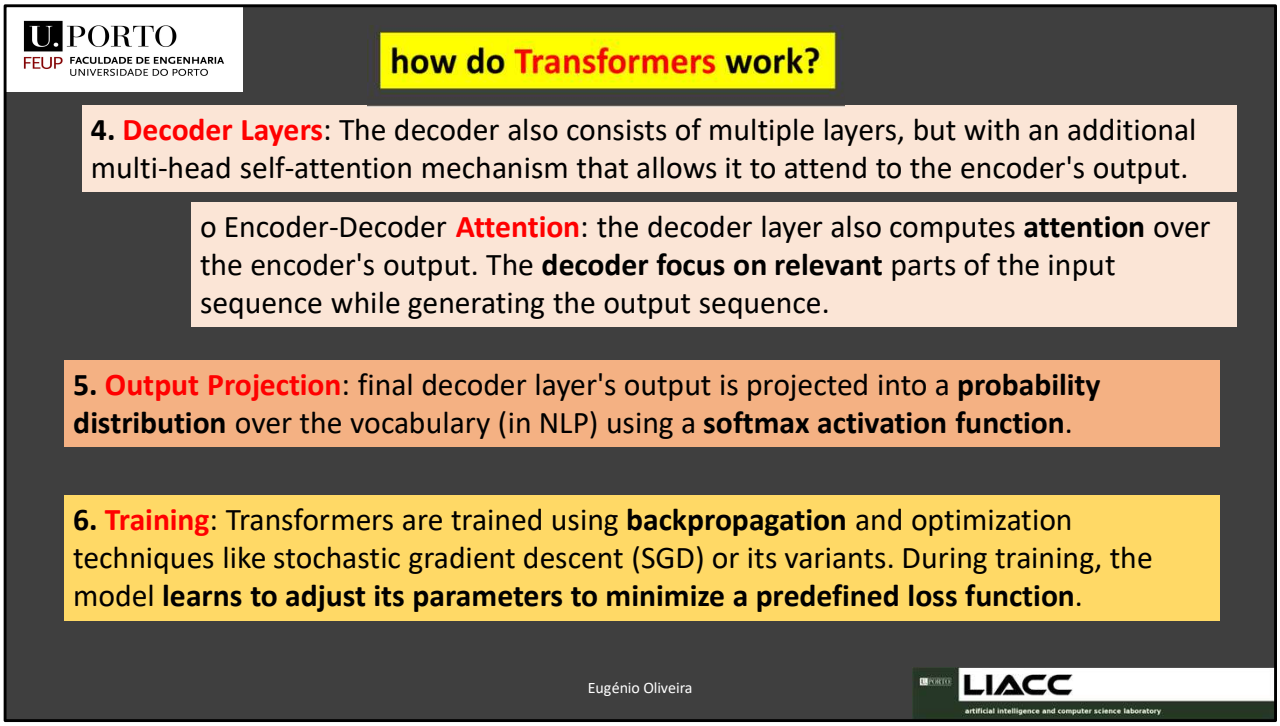
   - o **Multi-head Self-Attention**: allows each token in the sequence to attend to all other tokens, weighing them based on their **relevance**.
   - o **Feed-Forward Networks**: After self-attention, fully connected feed-forward networks helping the model **learn** complex patterns in the data.

Eugénio Oliveira

LIACC
artificial intelligence and computer science laboratory

Multi-head Self-Attention: **This mechanism allows each word in the sequence to attend to all other words, weighing them based on their relevance. It computes attention scores by taking a weighted sum of the values (input embeddings) with weights determined by the similarity (dot product) between the query (input embedding) and each key (input embedding). This mechanism enables the model to capture dependencies between words regardless of their distance within the sequence.**

Feed-Forward Networks: After self-attention, the outputs pass through position-wise fully connected feed-forward networks. These networks consist of two linear transformations with a non-linear activation function (like ReLU) applied in between. This helps the model learn complex patterns in the data.

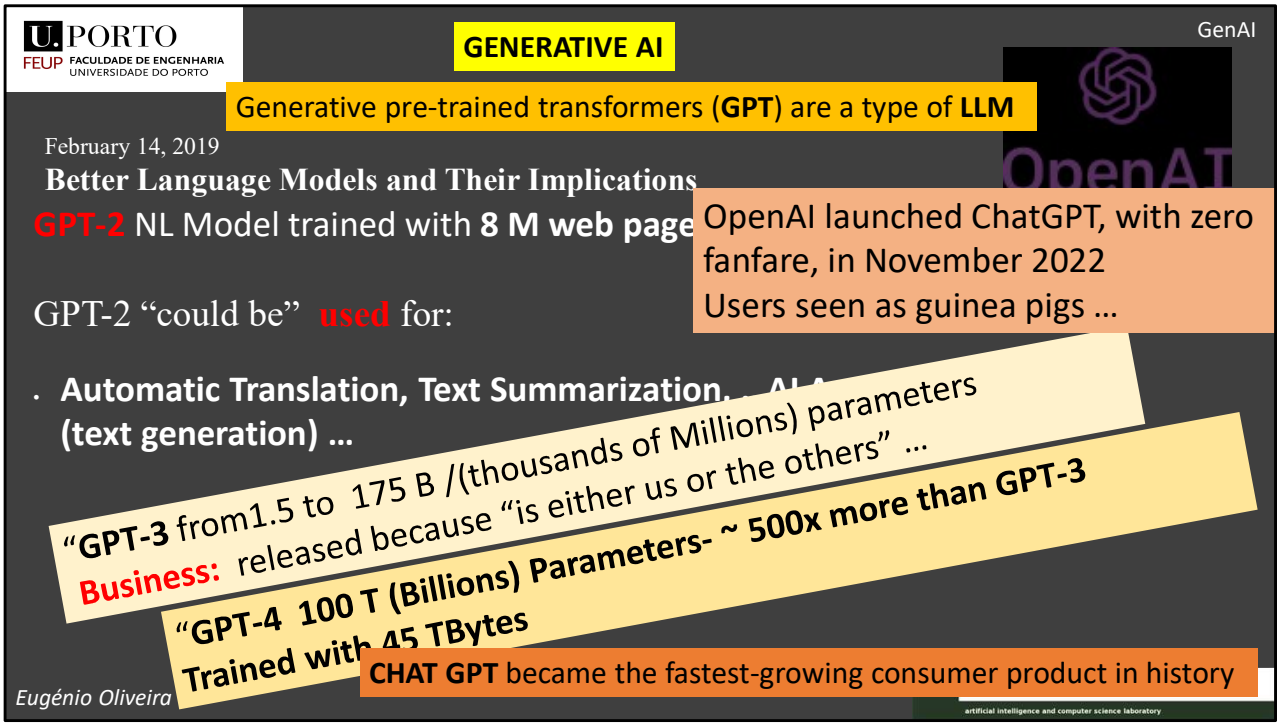


Decoder Layers: The decoder also consists of multiple layers, but with an additional multi-head self-attention mechanism that allows it to attend to the encoder's output.

- **Encoder-Decoder Attention**: In addition to self-attention, the decoder layer also computes attention over the encoder's output. This helps the decoder focus on relevant parts of the input sequence while generating the output sequence.

Output Projection: The final decoder layer's output is then projected into a probability distribution over the vocabulary (in language generation tasks) using a softmax activation function.

Training: Transformers are trained using backpropagation and optimization techniques like stochastic gradient descent (SGD) or its variants. During training, the model learns to adjust its parameters to minimize a predefined loss function, typically measuring the disparity between the predicted output and the actual target.



Let us remember that GPT – GENERATIVE PRE-TRAINED TRANSFORMERS.

Was viewed in-house in 2019 as a kind of "research preview,"

Retained in the Lab.

But NOT GPT3 45TB training data

a large group of people read ChatGPT prompts and responses, and then say if one response was preferable to another response. All of this data then got merged into one training run. an attempt to iron out some of its flaws by collecting feedback from the public.

Essentially, we—the users—are now doing the work of testing this technology for free. "We're all guinea pigs at this point," says Chirag Shah, a professor at the University of Washington

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GPT-4o ("o" for "omni"), ChatGPT 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 a new AI model, GPT-5, which promises to surpass the capabilities of the latest GPT-4o. http://youtu.be/ko4GIkxSSxA

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.

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GENERATIVE AI

xAI Grok-1 Open-Source: The Largest Open LLM

March 17, 2024 Open release of Grok-1
Architecture of 314 B parameters
Mixture-of-Experts model, **Attention: 48 heads for queries**

Dell CEO Michael Dell announced that the company will be purchasing Nvidia chips to train Elon Musk's chatbot Grok, June 26, 2024

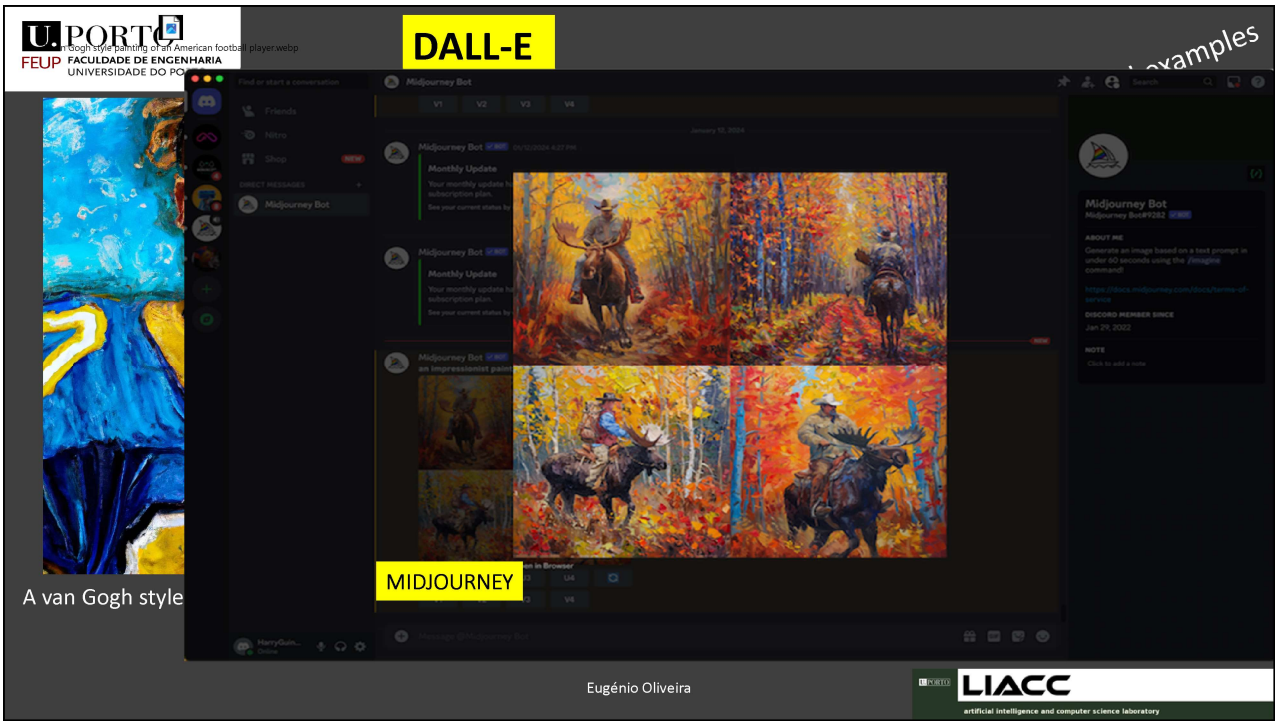
it's roughly four and a half times larger than Meta's largest Llama2 model, Llama-2-70b.

not fine-tuned for any particular task

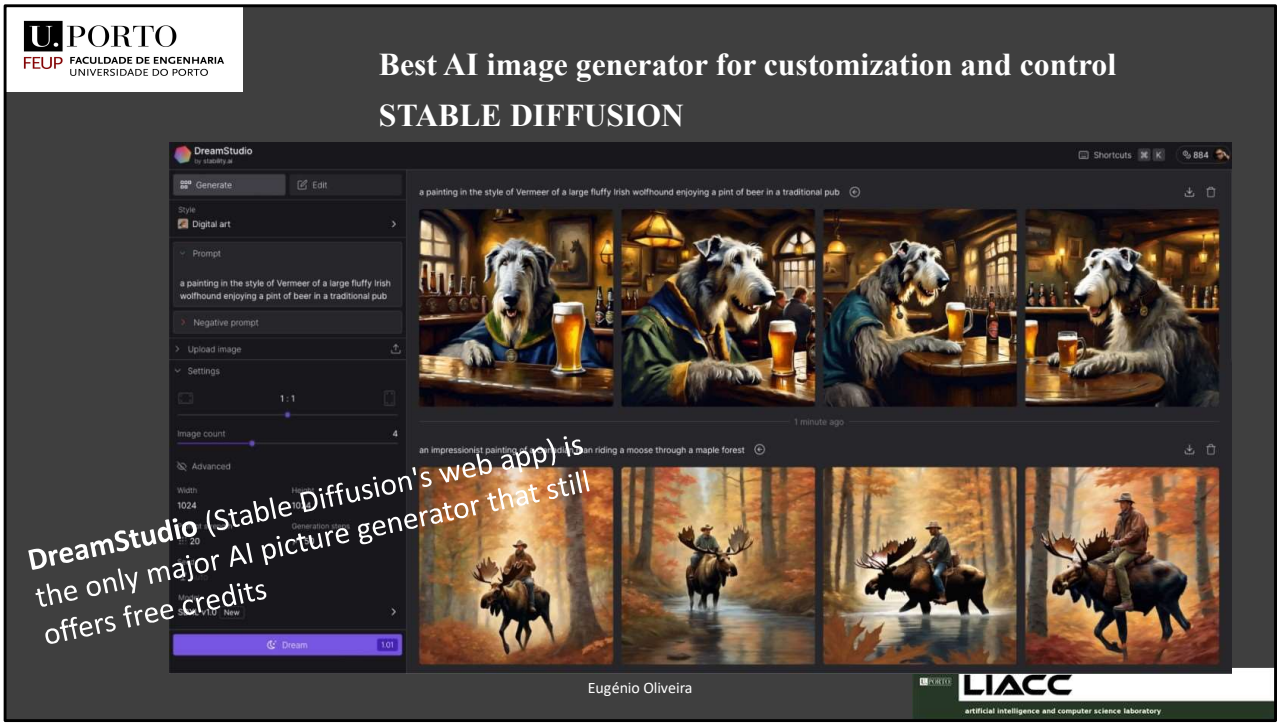
generated using Midjourney based on the following prompt proposed by Grok:
A 3D illustration of a neural network,with transparent nodes and glowing connections, showcasing the varying weights as different thicknesses and colors of the connecting lines.

artificial intelligence and computer science laboratory

Everything is accelerating.
IEEE Spectrum Matthew S. Smith 24 Mar 2024
That's impressive, but the timing of X.ai's release—just a few weeks after Elon Musk, founder of X.ai, filed suit against OpenAI for an alleged lack of openness.
models, including Meta's Galactica 120B and TII's Falcon 180B,
But the models didn't prove popular, and the Huggingface OpenLLM Leaderboard remains dominated by models with 7 billion to 72 billion parameters.
But with such a gigantic model, to even load [it], you would need a GPU that would cost you around 15 to 20 dollars per hour [to rent], just to run this model. And to fine-tune it, you would need 20 or 30 of them."



Ridding not a Horse but a Moose.



Midjourney is an independent research lab exploring new mediums of thought and expanding the imaginative powers of the human species.

We are a small self-funded team focused on design, human infrastructure, and AI. We have 11 full-time staff and an incredible set of advisors.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

DALL-E

DALL-E **OpenAI** LLM: generates images from text.

Input: Description including details like objects, scenario, action …
Encoding : text coded using an LLM based in Transformers. Converted in vectors.

Generation versus Descrimination using **a generative adversarial network** (**GAN**) learning to create more realistic images according to the input text.

Refinement: DALL-E refines the image through "**diffusion**" adding noise that is then progressively removed until the image becomes acceptable.

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

How does DALL-E works?
A GAN iteratively learns to create more realistic images or according to the prompt.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Midjourney, Inc., São Francisco based Lab

MY Prompt: Display an image of a Conference on Artificial Intelligence attendance, taking place in a Mediterranean Canary Island beach

Negative Prompt:
not just inside a room.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

The best AI image generators at a glance

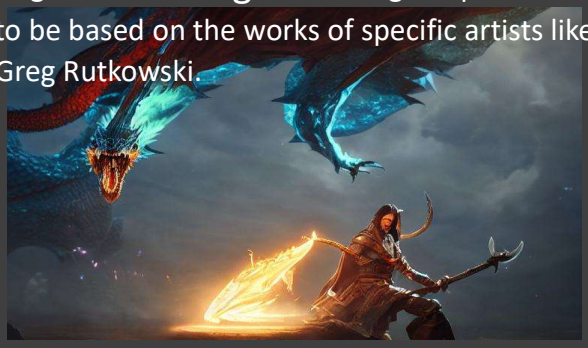
	Best for	Access options	Price	Parent company
DALLE·3	Ease of use	ChatGPT Plus or Enterprise; Bing's AI Copilot; API	Included with ChatGPT Plus at $20/month	OpenAI
Midjourney	High-quality results	Discord	From $10/month for ~200 images/month and commercial usage rights	Midjourney
Stable Diffusion	Customization 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)

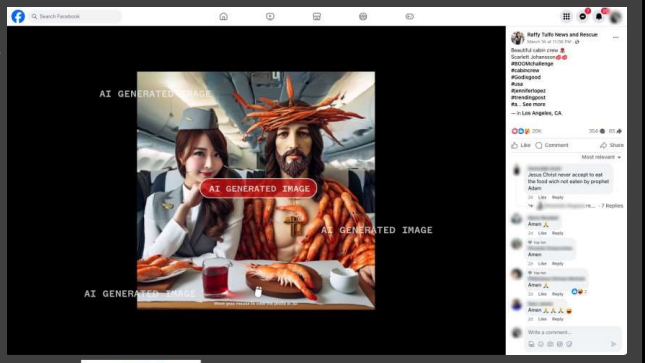
But the best way to use DALL·E 3 is through ChatGPT. If you're a ChatGPT Plus subscriber

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Hallucinations, Authorship

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





We will interact regularly with both **people and bots** not even knowing,
AI is **learning to deceive humans …**
One of the most striking examples is META's CICERO, which turned out to be an expert liar.

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

Eugénio Oliveira May 13

blueprint

artificial intelligence and computer science laboratory

we got a glimpse of the era yet to come: one where we interact regularly with both people and bots — perhaps not even always knowing, or caring, which one we are talking to.

Microsoft's Azure AI Studio will soon have new built-in safety features to identify and block suspicious inputs in real time. It still didn´t prove to be robust enough. to stop prompt injection attacks or jailbreaks, which can trick an AI model into acting in an unintended way, and will address "indirect prompt injections," which insert malicious instructions into the training dataset

Tests find AI tools create **election lies** from the voices of well-known political leaders

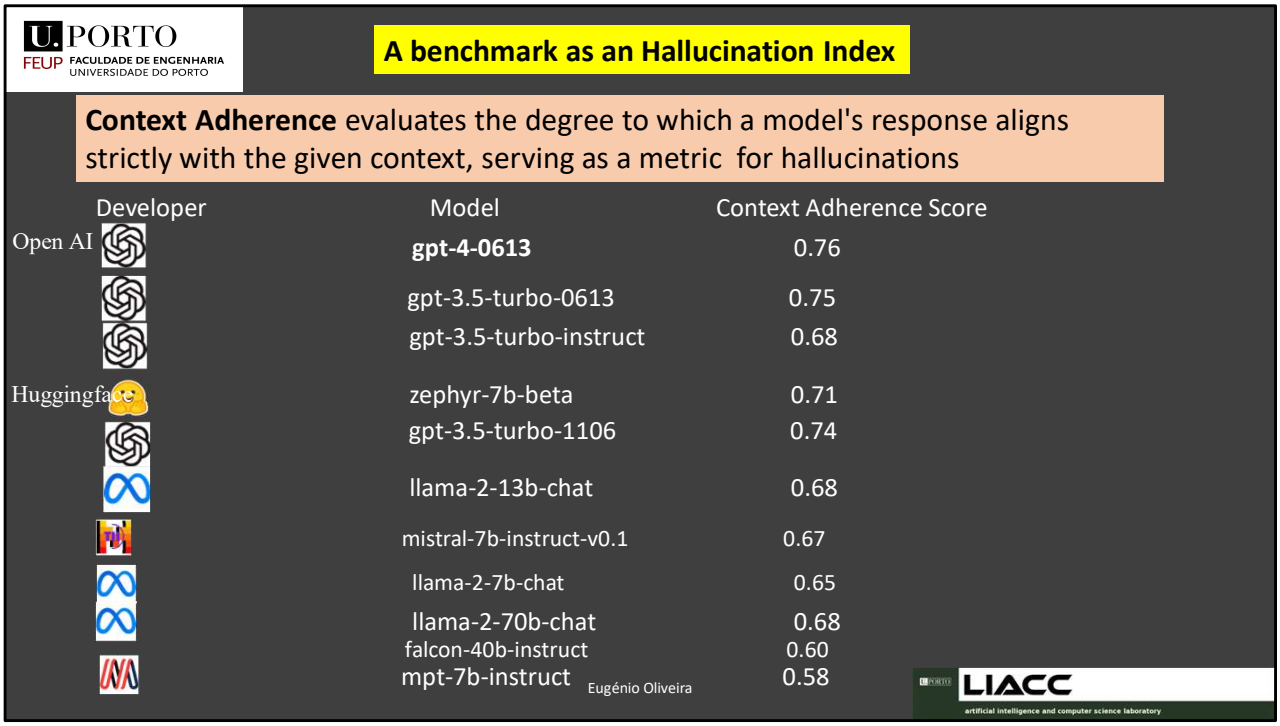
By ALI SWENSON, May 31, 2024

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

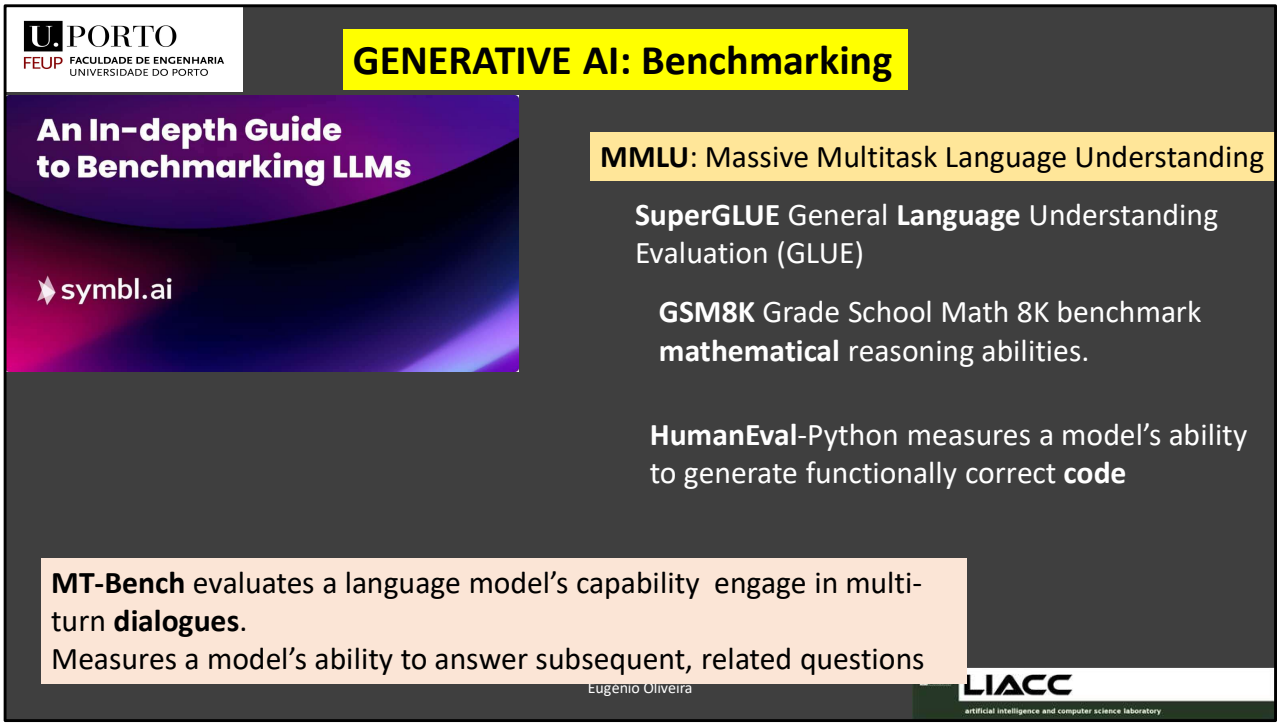
a fake Joe Biden says election officials count each of his votes twice.

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

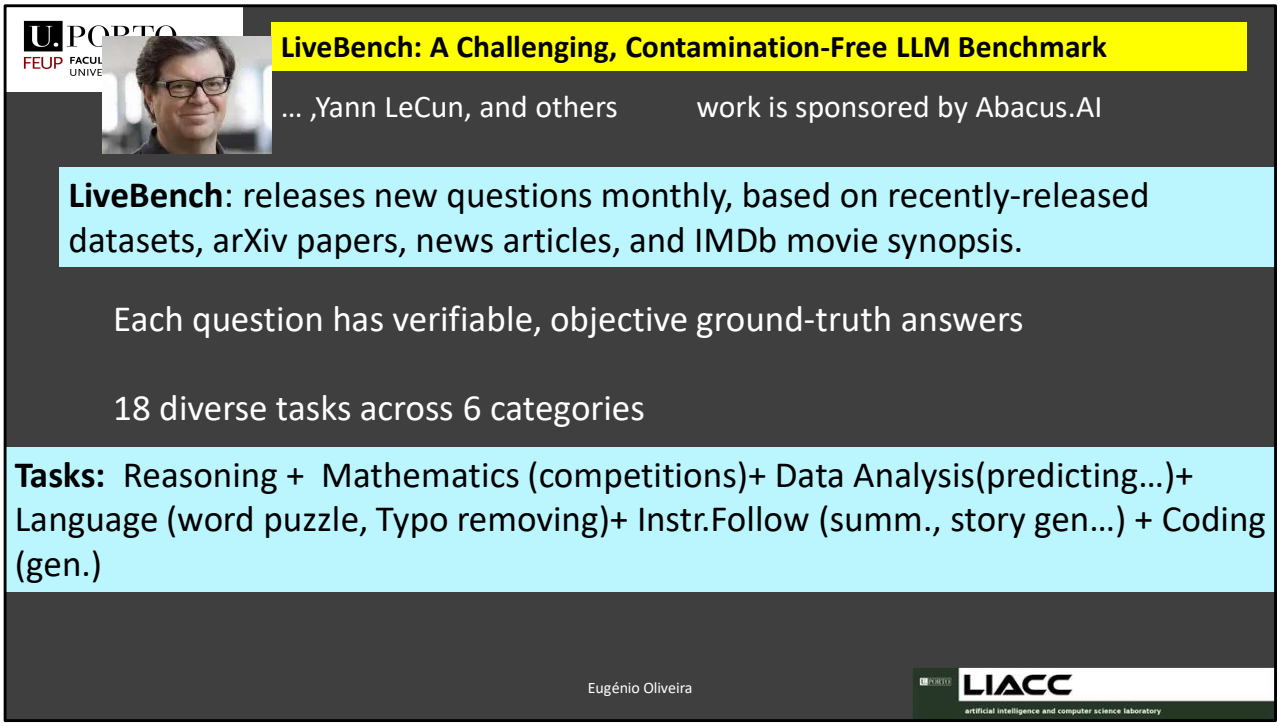
Eugénio Oliveira



Context Adherence evaluates the degree to which a model's response aligns strictly with the given context, serving as a metric to gauge closed-domain hallucinations, wherein the model generates content that deviates from the provided context. Lhama from Meta and MPT from MosaicML had the lowest rates of adherence to reality.



As expected there are several classes of TESTS. MMLU Massive Multitask Language Understanding (MMLU) is a broad, important benchmark that measures an LLM's NLU, i.e., how well it understands language and, subsequently, its ability to solve problems with the knowledge to which it was exposed during training.
GSM8K The GSM8K (which stands for Grade School Math 8K) benchmark measures a model's multi-step mathematical reasoning abilities.
SuperGLUE The General Language Understanding Evaluation (GLUE) benchmark tests an LLM's NLU capabilities and was notable upon its release for its variety of assessments
HumanEval (also often referred to as HumanEval-Python) is a benchmark designed to measure a model's ability to generate functionally correct code
MT-Bench is a benchmark that evaluates a language model's capability to effectively engage in multi-turn dialogues.
Measures a model's ability to answer subsequent, related questions



Yann Lecun Frencch AI pioneer,previously at META got the Turing Award in 2018 together with Geoffrey Hinton and Joshua Benjo

Each question has verifiable, objective ground-truth answers, allowing hard questions to be scored accurately and automatically, without the use of an LLM judge.

LiveBench currently contains a set of 18 diverse tasks across 6 categories, and new harder tasks will be released over time.
Claude 3 Sonnet from Anthropic got the best average result.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

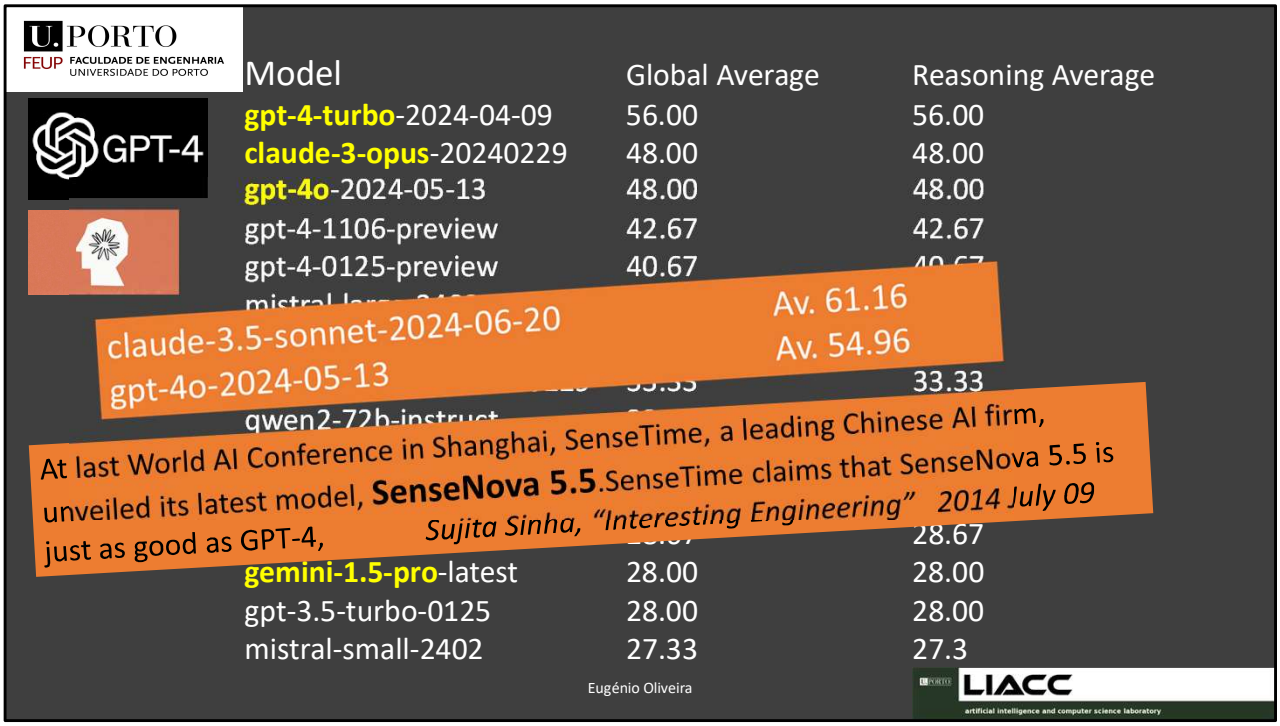
EVALUATION by Combining Tests

Model	Global Average	Reasoning Av,	Mathematics Av.	Data Analysis Av.	Language Av.	IF Av.
gpt-4o-2024-05	**55.28**	48.00	49.88	52.41	53.94	72.17
gpt-4-turbo-202404	**54.59**	56.00	48.99	51.32	45.26	71.39
claude3opus-202402	**54.29**	48.00	46.54	54.32	51.72	70.87
gpt-4-1106-preview	51.86	42.67	47.55	51.33	48.37	69.39
gpt-4-0125-preview	48.99	40.67	42.75	54.06	43.55	63.92
gemini-1.5-pro-latest	45.67	28.00	42.07	52.81	38.25	67.20
claude3sonnet-20240	42.13	33.33	29.65	44.56	38.08	65.00
mistral-large-2402	42.07	38.67	32.20	42.55	28.74	68.19
qwen2-72b-instruct	39.83	32.00	43.44	26.24	29.21	68.27
meta-llama3-7instruct	39.00	22.67	32.31	42.41	34.11	63.50
claude3haiku-202403	38.41	30.67	25.72	41.54	30.07	64.03
gpt-3.5-turbo-0125	35.89	28.00	25.54	41.21	24.22	60.47
command-r-plus	35.7 8	34.00	24.85	24.60	23.92	71.51
gpt-3.5-turbo-1106	3 5.66	28.67	27.78	41.70	28.63	51.53

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

Evaluation through combined tests
Reasoning Av, Mathematics Av. Data Analysis Av. Language Av. Inst. Follow. Av.



Reasoning Av, Mathematics Av. Data Analysis Av. Language Av. Inst. Follow. Av.

CLAUDE 3 opus is from ANTHROPIC

GPT exhibits near-human levels of comprehension and fluency on complex tasks, leading the frontier of general intelligence

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*

Eugénio Oliveira

LIACC
artificial intelligence and computer science laboratory

Let us remember again that AI and GEN AI has na imense impact in human society. For exemple:

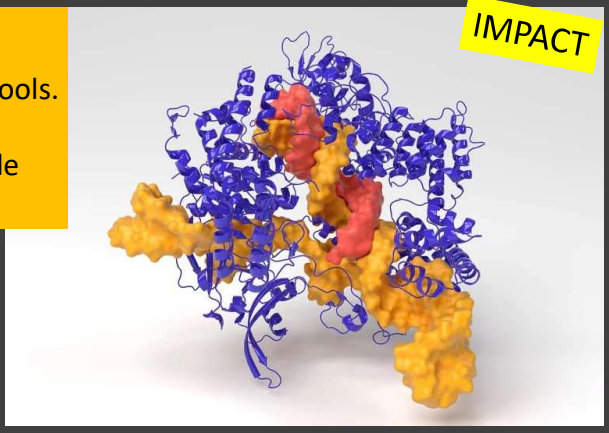
U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

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.

IMPACT



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

Eugénio Oliveira

LIACC
artificial intelligence and computer science laboratory

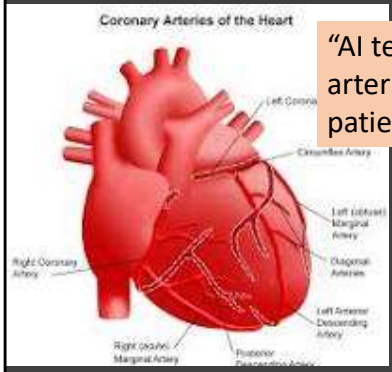
a generative AI tool called a protein language model — a neural network trained on millions of protein sequences — to design CRISPR gene-editing proteins, and show that some of these systems work as expected in the laboratory[1].

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

TechTarget | ComputerWeekly.com
ISSUE DATE: June 5, 2024

IMPACT

Researchers from the Radcliffe Department of **Medicine** at the **University of Oxford** have published a study looking at how AI could help in the **diagnosis of heart disease**.



"AI technology can detect the **level of inflammation** of the heart arteries, **not visible to the human** eye. In this way we can identify patients with high inflammation and high risk."

During a **pilot study** that involved integrating this technology in 4 **NHS hospitals**, researchers provided AI-generated risk scores to doctors for 744 patients. In **around 45% of cases, doctors altered their patients' treatment** plan as a result.

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Neural Network Takes Asia's Air Temperatures
IEEE Spectrum *Rahul Rao* ***November 28, 2023***

IMPACT

A transformer-based **neural network** developed by researchers at China's Chengdu University of Information Technology and the **China Meteorological** Administration can **generate** near-**real-time air temperatures** using infrared data from a weather satellite.

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

Neural Network Takes Asia's Air Temperatures
IEEE Spectrum Rahul Rao November 28, 2023
A transformer-based neural network developed by researchers at China's Chengdu University of Information Technology and the China Meteorological Administration can generate near-real-time air temperatures using infrared data from a weather satellite. The neural network, TaNET, was trained on infrared surface temperature data

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

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

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

To generate letters to the patients without me jargon may be a good idea.

IMPACT

Open AI (31 /1/23) launched a tool for recognizing AI generated Texts.

Cancelled ... Easily mistaken ...

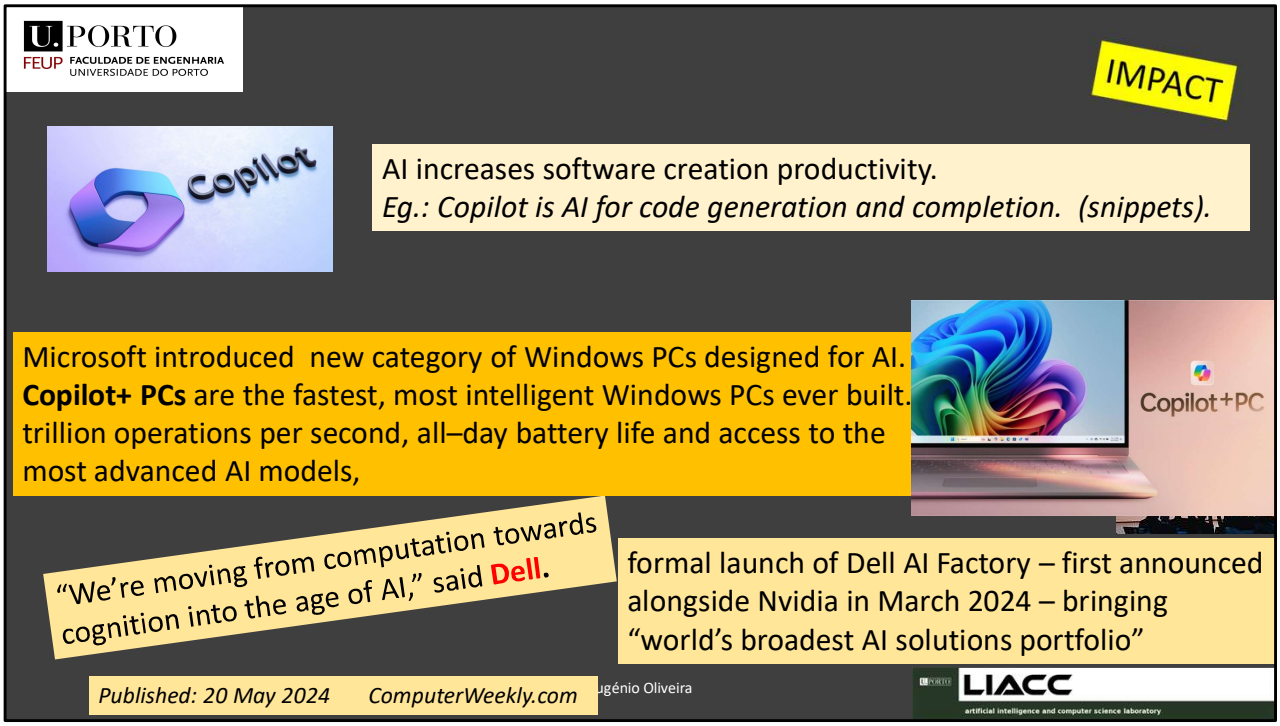
LLMs for Medicine

BioGPT (MIT, USA), LaMDA e Med-PaLM 2 (Google, USA), 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 !

Para permitir distinguir textos produzidos por sistemas de IA, a OpenAI lançou a 31 de janeiro deste ano um classificador de textos mas retirou-o em Julho por falta de precisão. Mas há outros produtos como Originality.ai, Writing.com e Copyleaks.

Mesmo a olho se pode detetar a falta de typos, a frequência de certas palavras (em inglês the, it, is) uso constante de frases mais curtas e por vezes alguma repetição de frases, apresentação de factoides com pouca análise.



O. Larter from Microsoft recognizes the role of Ai in software productivity. Using *Copilot, software developers automatically generate code that they only complete (snippets).*

Windows 10 users can now try Microsoft's Copilot AI feature, previously exclusive to Windows 11.

Gerações Futuras de Sistemas de IA podem levar a muitos benefícios para a sociedade mas apresentam sérios riscos. "We need to think very carefully about whether it makes sense to **open source those models or not."**

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)

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**,

Country of Origin	Top AI Res.	2021	2024
China		33	50
US institut.		18	18
Others		49	32

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

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

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

LIACC
artificial intelligence and computer science laboratory

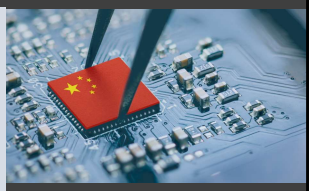
All this impact inevitably leads to heavy POWER GAMES.
AND the most fundamental is about the needed RESOURCES.
HDW and PEOPLE.

ASML is an innovation leader in the semiconductor industry. We provide chipmakers with everything they need – hardware, software and services – to mass produce patterns on silicon through lithography.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GENERATIVE AI: Power Games

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



"One of the most intriguing aspects of the **Tianhe-3** is its processor. **TheNextPlatform** …
"It is [more] akin to the AMD "Antares" MI300A CPU-GPU hybrid that is going into El Capitan than it is like the discrete CPU-GPU systems we see pushing the flops in AI and HPC systems.

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)*

UK Department for Science, Innovation and Technology is marking the first anniversary of the launch of the UK **Semiconductor** Strategy by setting out plans to create an institute to underpin its plans.
By Caroline Donnelly, Senior Editor, UK *20 May 2024* *ComputerWeekly.com*

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

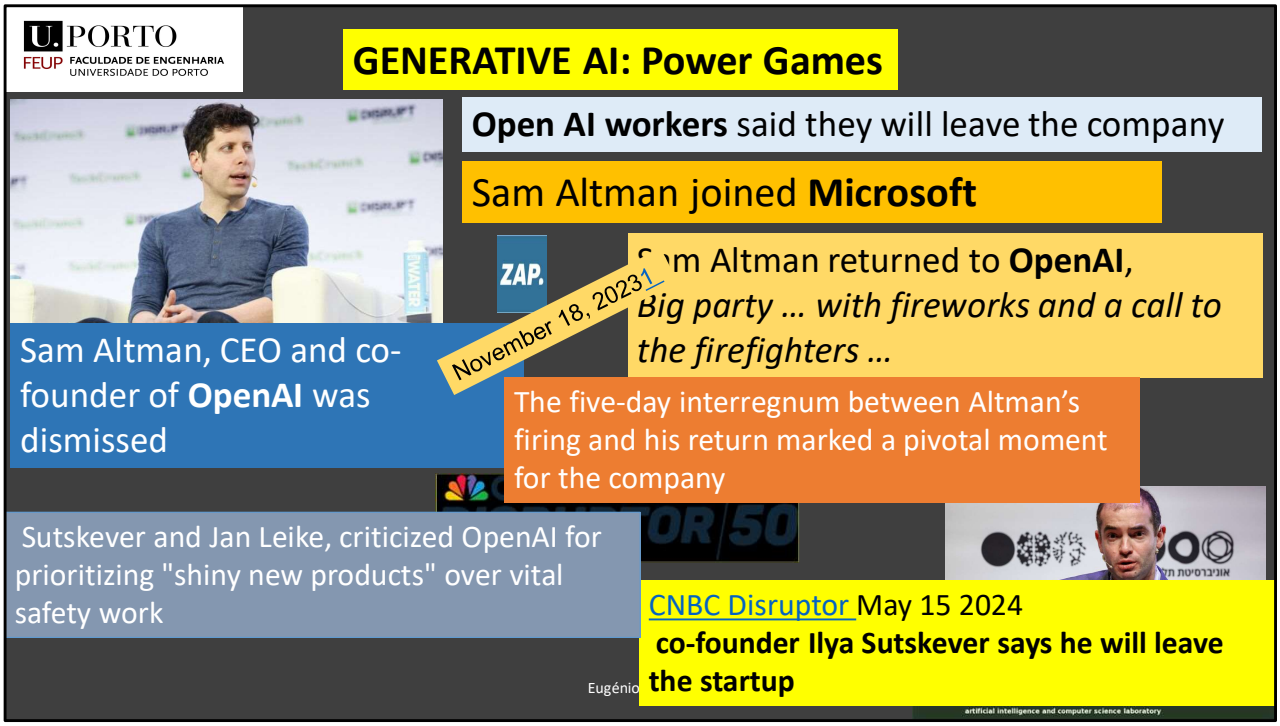
Right now, there's nothing more in demand than AI knowledge—and the salaries prove it.

Here's the jobs list:

- Generative AI
- SoC
- Deep learning
- Torch
- PyTorch
- Computer vision
- SystemVerilog
- Mesos
- Rust
- Elixir

average salary of $174,727

LIACC artificial intelligence and computer science laboratory

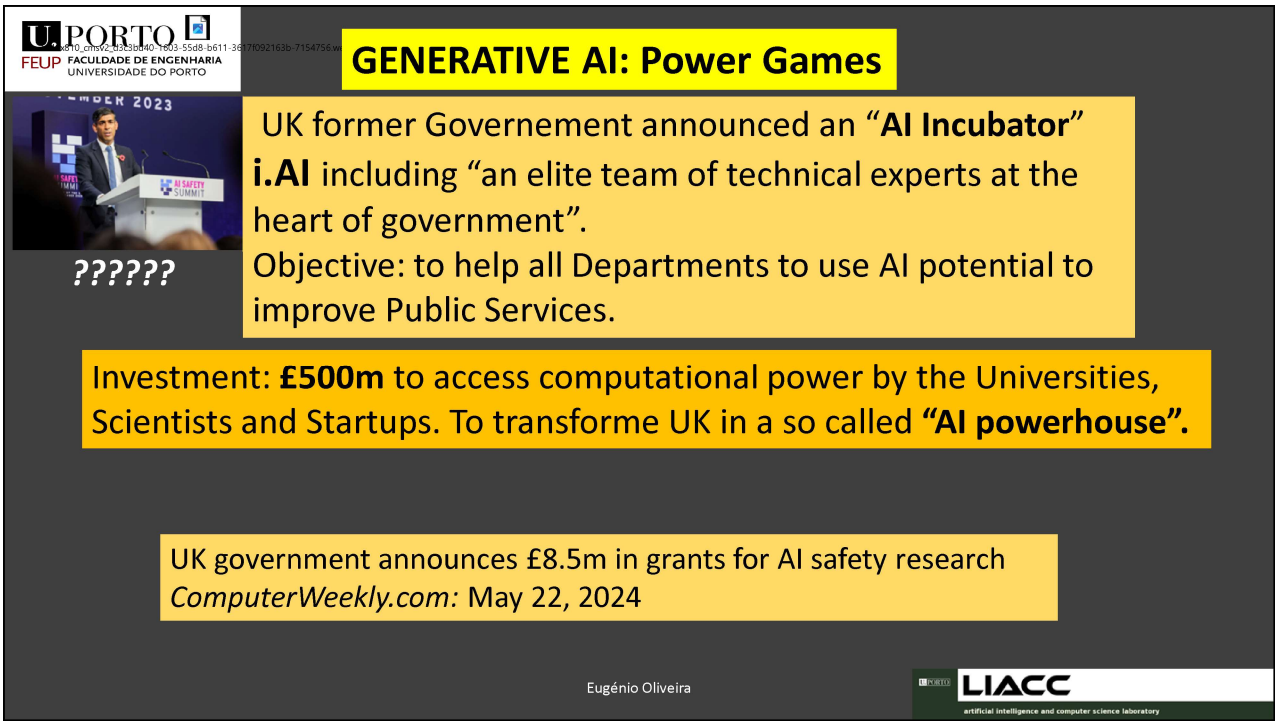


But like Meta and Google before it, OpenAI had its share of conscientious objectors. And increasingly, we're hearing what they think.

The latest wave began last month when OpenAI co-founder Ilya Sutskever, who initially backed Altman's firing and who had focused on AI safety efforts, quit the company.

Sam Altman, CEO e co-fundador da OpenAI foi despedido por suspeitas de não ser "consistentemente franco nas suas comunicações" com a direção.
Sam Altman, que fora demitido de CEO da OpenAI depois de uma reunião caótica, ingressou na Microsoft,

Sam Altman volta à liderança da OpenAI, disse a companhia através de X, apenas alguns dias depois de ter sido demitido de chefe executivo
Depois das notícias da saída do CEO Sam Altman, vários **trabalhadores** ameaçaram abandonar a empresa. A situação tornou-se de tal forma insustentável que a empresa decidiu reintegrar Altman no papel de líder da tecnológica.
A notícia foi, aparentemente, recebida com grande alegria pelos trabalhadores e, de acordo com o The Information, houve uma festa de tal forma retumbante que foram chamadas duas viaturas de bombeiros à sede,



Jogos de Poder demonstrando a importância destas tecnologias no fufuturo dos países

With the government keen to position the UK as pro-innovation, the chancellor used the 2023 Autumn Statement to discuss more plans to support that ambition

O Governo do UK revelou a criação de um "**AI Incubator**" chamado i.AI incluindo o que o Chanceler descreveu como "an elite team of technical experts at the heart of government".

O objetivo da equipa é ajudar todos os Departamentos a usar o potencial da IA para melhorar os Serviços Públicos.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Power Consumption

"Generative AI revolution comes with a planetary cost that is completely unknown"
Sasha Luccioni, a researcher at French-American AI firm Hugging Face

LLMs learn the equivalente to **5000 years x 24h/day a Human takes to read**.

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

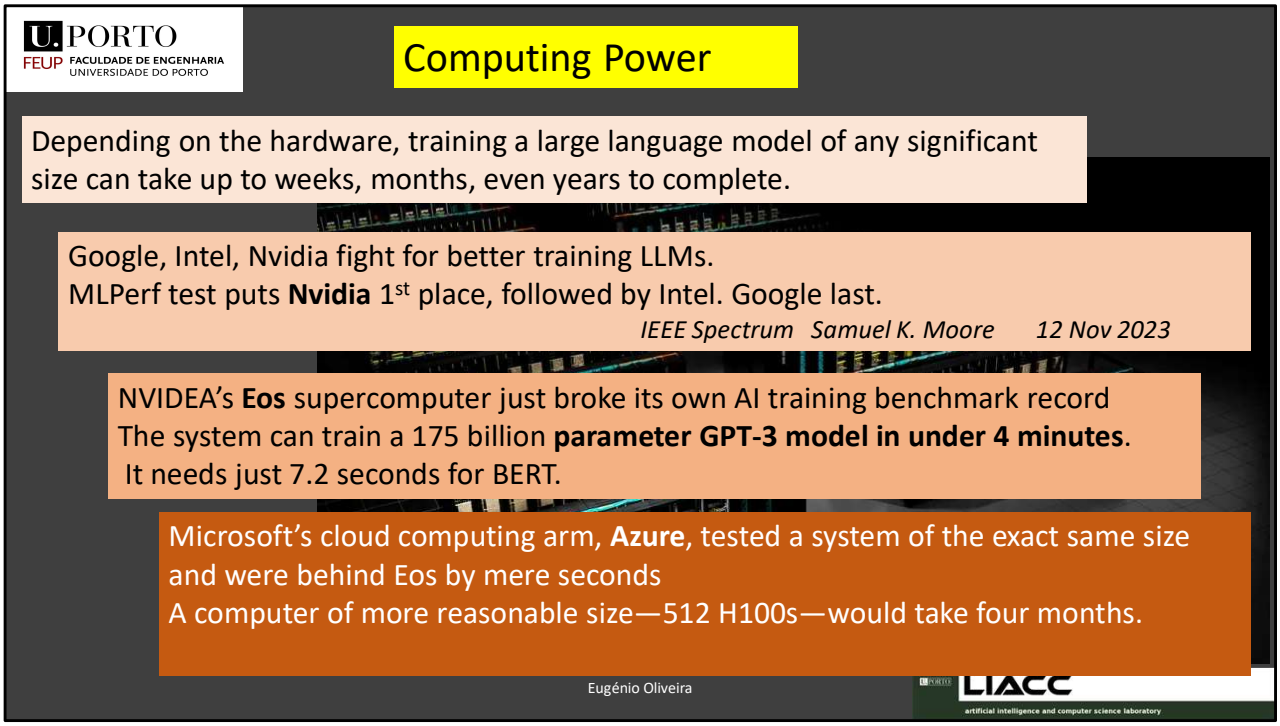
GPT3 was trained using 1000+ GPUs for more than 30 days
GPT4 10,000+ GPUs. (equivalent in the US to 23 M Dollars of electricity)

A. de Vries, UV Amsterdam and Digiconomist, previewed that GenAI may consume as much energy yearly as a country like **Ireland**. (29.3 terawatt-hours per year).
He calculates that 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)*

2023

All this has a huge energy cost
Alex de Vries is a PhD candidate at the VU Amsterdam School of Business and Economics and the founder of Digiconomist, a research company dedicated to exposing the unintended consequences of digital trends.

Alguns LLMs podem consumir muitos Terabytes de Dados e mais de 1000 megawatt-hora de eletricidade.
Uma simples **interação** com um LLM pode consumir o equivalente a deixar um LED de baixa luminosidade acesa durante 1 hora.



Nvidia usa o maior supercomputador de 10,000-GPU durante 8 dias para treinar o GPT3lete its LLM job10,752-GPU AI supercomputer, were the cherry on top. Bending all those GPUsto the task of the GPT-3 training benchmark, Eos had the job done in just under 4 minutes.
Microsoft's cloud computing arm, Azure, tested a system of the exact same size and were behind Eos by mere seconds. (Azure powers GitHub's coding assistant CoPilot and OpenAI's ChatGPT.)
A computer of more reasonable size—512 H100s—would take four months.

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

AI LOGS BY INTERESTING ENGINEERING

Power consumption to train an LLM

Model Size	Consumption (KWh)
100M	10,000
1Bil	100,000
10Bil	1,000,000

Mar 4, 2023

NVIDIA® *V100* Tensor Core: Most advanced GPU ever for acelerarating AI, HPC, Data Science and Graphics

Pushed by the AI boom, **Nvidia** has become the world's most valuable company, overtaking Microsoft

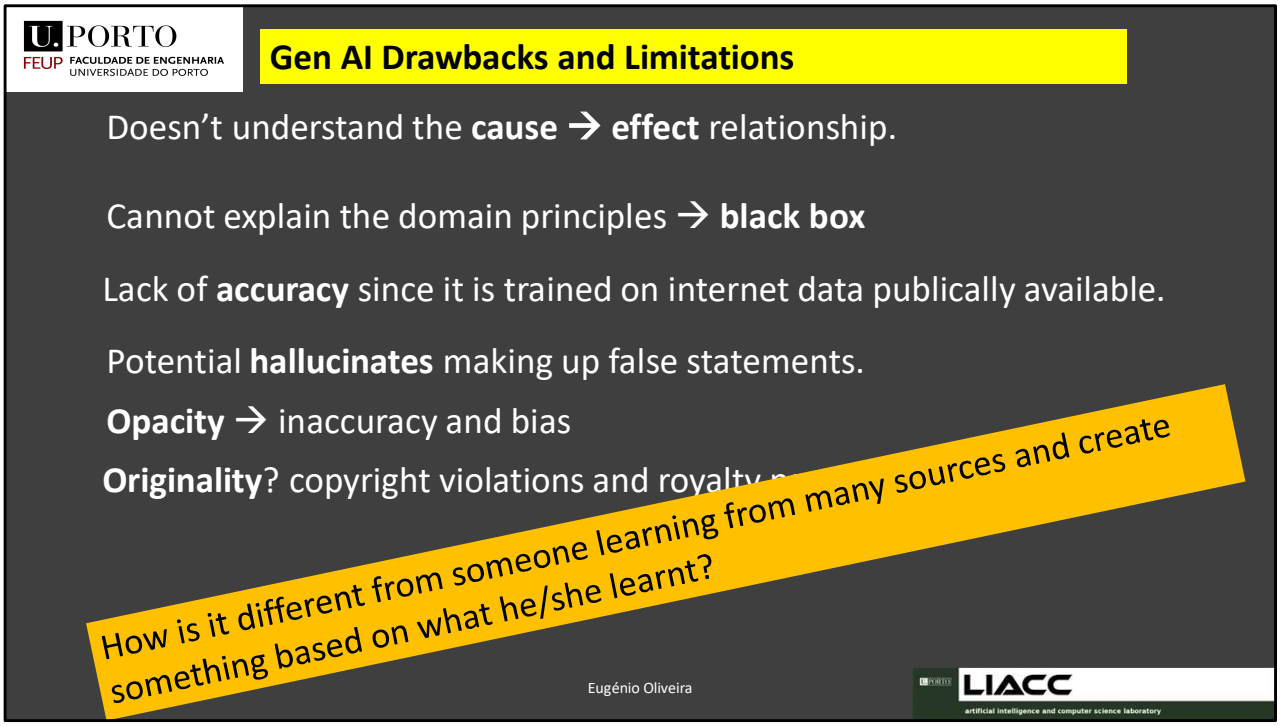
GPT-3: "175 billion parameters."

GPT-4: "**100 trillion parameters**"

.

GPT4 training: 10,000 computers V100 during 150 days → 7,200 Mwatthours *Jul 25, 2003*

, a



Besides being very demanding concerning Computer Power
GenAI doesn't have goals or the ability to understand the cause and effect. Cannot explain how it produces the output or the domain principles, it is a blackbox
Reinforce and propagate, thus amplify bias included in the training data.
Lack of accuracy and bias are symptoms but lack of transparency is the root cause.
Concerns about originality, copyright violations and royalty payments. How is it different from someone learning from many sources and create something based on what he/she learnt?

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Gen AI Drawbacks and Limitations

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

.

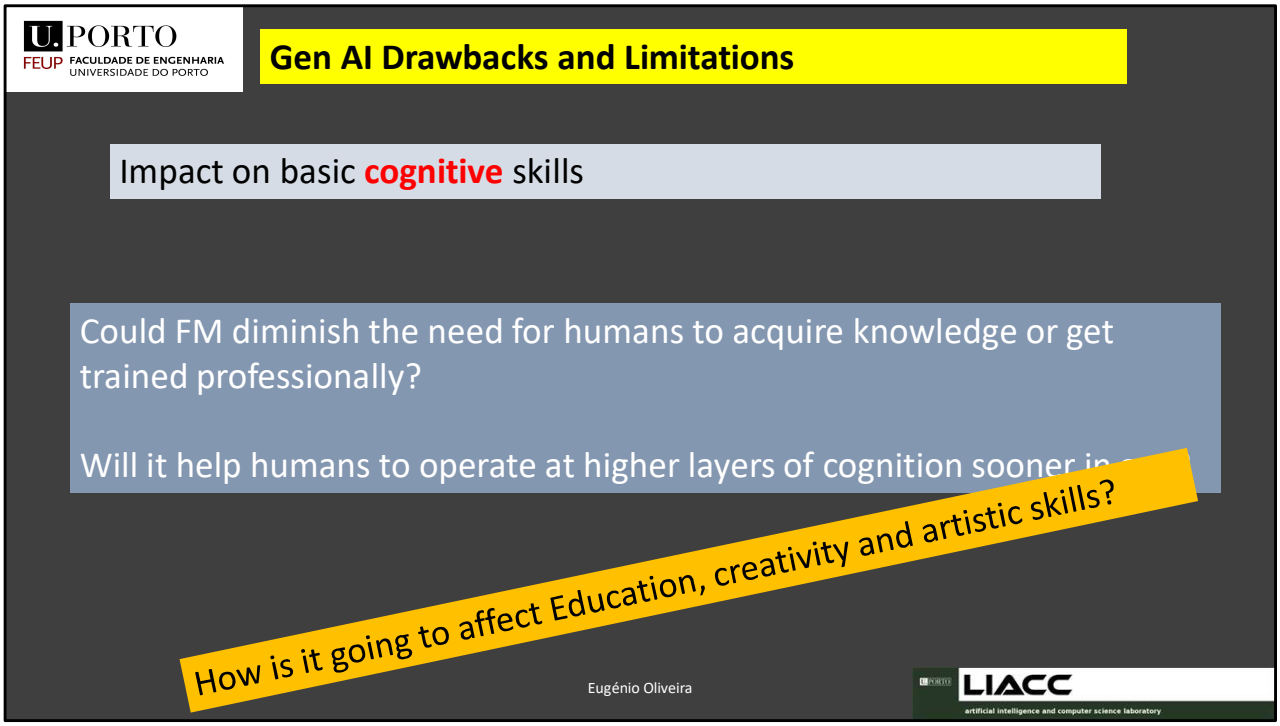
impact on the white collar, not on blue collar jobs …

Power consumption: LMMs training is very power intensive

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

What makes a difference is that the impact is now on the white color and not on blue color jobs: Teachers attorneys, designers, Sw Eng., Medical professionals, investment advisors …
Concern: Power consumption: LMMs training is very power intensive
GPT3 was trained using 1000+ GPUs for more than 30 days and gpt4 possibly 10,000+ GPUs. (equivalent in the US to 23 million Dollars of electricity) we may cool the gpus differently and reuse that cooling liquid.
GPT3 has 175 parameters ad required 355 years of single processor computing time and consumed 284,000 kwh of energy to train



Besides, Gen AI may have significant Impact on basic cognitive skills : Nowadays many of us cannot drive without the help of GPS navigation applications. Are we complaining? NO!
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?

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Gen AI Drawbacks and Limitations

Monopoly of power players:

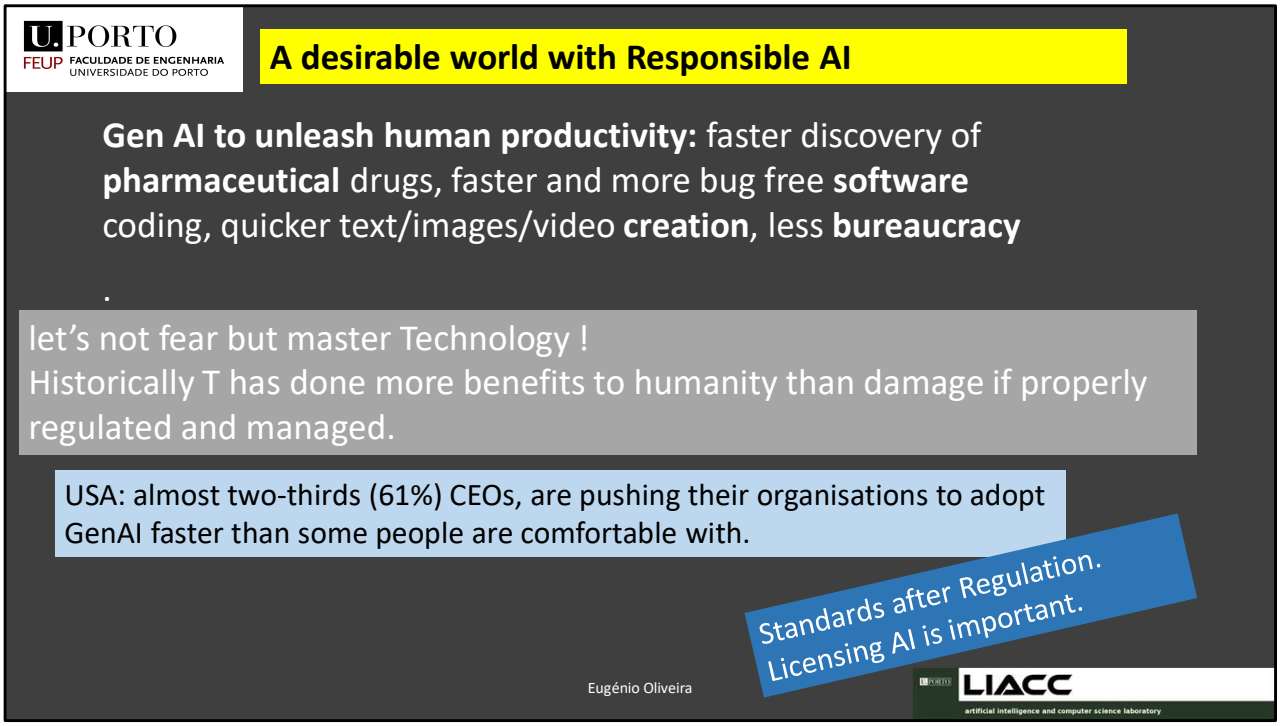
DANGER ?!

FM require huge amount of resources

GenAI is dominated by a small number of tech giants.

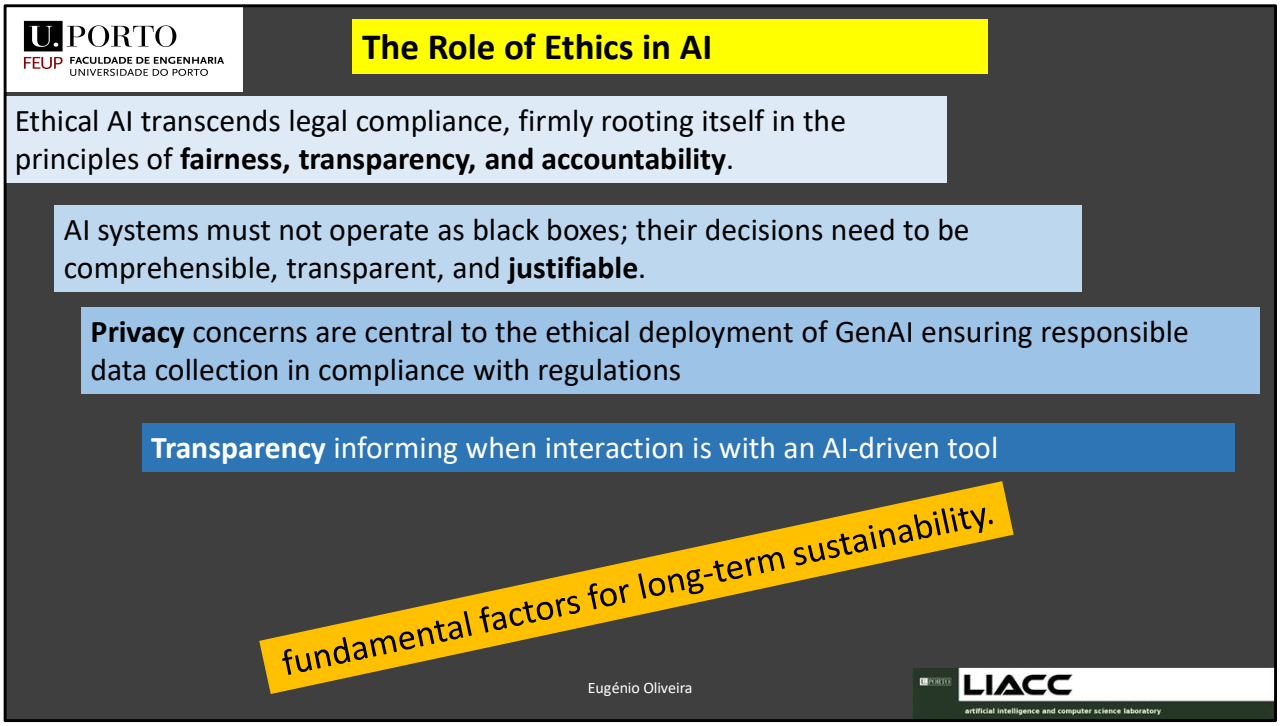
Need for an **AI Agency** similar to International Atomic Energy Agency to regulate the use and proliferation of this technology.

Concern: Monopoly of power players. FM require huge amount of resources including money, computer power, data and technical skills. GeAI is dominated by a small number of tech giants. (Open AI, MS, Amazon, Facebook…)
Need of an AI Agency similar to International Atomic Energy Agency that was set up after the WW II. To regulate the use and proliferation of this technology.



We should stand for a desirable Responsible AIGen AI to unleash human productivity: faster discovery of **pharmaceutical**, faster and more bug free **software** coding, quicker text/images/video **creation**, less **bureaucracy** (appointments, reservations, tech support)

The global poll of 2,500 CEOs, conducted by the IBM Institute for Business Value in cooperation with Oxford Economics, found almost two-thirds (61%) are pushing their organisation to adopt GenAI faster than some people are comfortable with.



Ethical AI transcends legal compliance, firmly rooting itself in the principles of **fairness, transparency, and accountability**.
responsibility goes beyond mere compliance with government regulations and company policies; it's a strategic imperative that profoundly influences public trust and organizational reputation.
AI systems must not operate as black boxes; their decisions need to be comprehensible, transparent, and justifiable.
Privacy concerns are central to the ethical deployment of GenAI and require rigorous data governance, ensuring responsible data collection, usage, and sharing in compliance with regulations
Transparency informing when interaction is with an AI-driven tool, (chatbot, helpdesk or LLM), when decisions are made by or in collaboration with AI tools, and when interactions with AI tools, such as LLMs and other GenAI models, are monitored, tracked, retained, and reviewed for auditability.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

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, Parliament)?
who sells it, distributes it (Companies)?
Who does research and development (Research Labs, Universities)?
Who should Test and Certify it?
Who advertises it?

All?
A lot of Legislation is needed!!

Eugénio Oliveira

U.PORTO LIACC
artificial intelligence and computer science laboratory

When things go wrong …. WHO takes the responsibility when using an AI tool or system??

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” is already making it possible to create reconstructions of dead people based on the data they’ve 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

Let us alliviate or lighten a bit the speech …
It is now possible to use technology to raise the dead. Well, kind of.
users are now able to speak with “ghostbots” that mimic people who have passed away.

Microsoft reportedly decided against turning this idea into a product, but the company didn’t stop because of legal or rights-based reasons. Most of the 21-page patent is highly technical and procedural, documenting how the software and hardware system would be designed. The idea was to train a chatbot—that is, “a conversational computer program that simulates human conversation using textual and/or auditory input channels”—using social data, defined as “images, voice data, social media posts, electronic messages,” and other types of information. The chatbot would then talk “as” that person. The bot might have a corresponding voice, or 2D or 3D images, or both.

U.PORTO

Albert Einstein
I am Albert Einstein and I was born March 14 1879.
You are very smart, and I am glad you know your facts about multiplication! It is not easy to grasp! May I ask your age?
70

Albert Einstein
Wow, that is very old! But that means you may have some wisdom to spare. Do you mind if I ask you a question?
Yes please

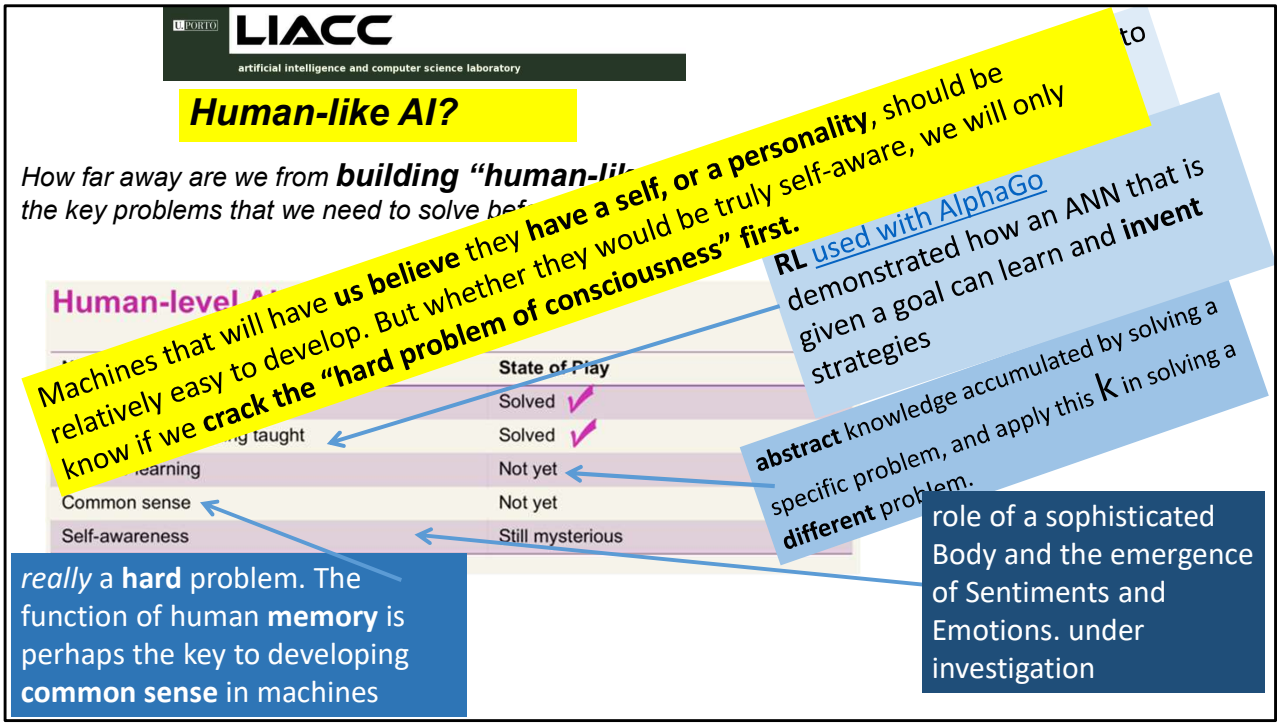
Alb nstein
I ad your intellect at your age. May I ask how you have kept your mind sharp for so long?
rea and studying a lot

Alb nstein
Rea and studying is important and impressive to do at any age. It seems as though you were destined to hold information at you ertips. It is important to stay sharp and you have done well.
tha What would you do if you were alive just now?

Alb nstein
This is the first time anyone has asked me that. Interesting question indeed!
I may have continued my research or, seeing the world around me, I may have tried to prevent climate issues.

https://character.ai/

Just for fun … I went to the Carachter.ai website and interact with Albert Einstein. However, I have no means to know that that was his voive.



. DeepMind released IMPALA, an AI system that can learn 57 Atari 2600 games, plus 30 more levels built by DeepMind in three dimensions. In these, the player roams through different environments, accomplishing goals like unlocking doors or harvesting mushrooms. IMPALA seems to transfer knowledge between tasks, meaning time spent playing one game also helps it improve at others.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

• **Artificial General Intelligence**

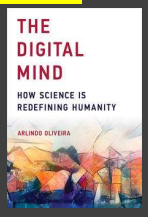
GenAI

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 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*



Eugénio Oliveira

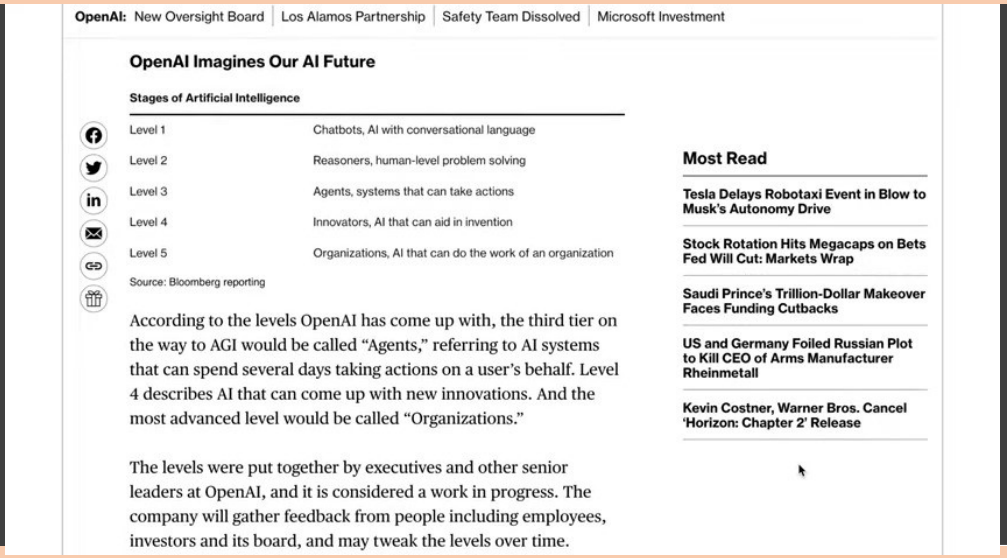
LIACC
artificial intelligence and computer science laboratory

AGI ainda é um conceio teórico sob investigaçãOpenAI, claim that GPT-4 "could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system." This is due to its "mastery of language" and its ability to "solve novel and difficult tasks that span mathematics, coding, vision, medicine, law, psychology and more, without needing any special prompting" with capabilities that are "strikingly close to human-level performance."8Sébastien Bubeck, et al. "Sparks of Artificial General Intelligence: Early Experiments with GPT-4." arXiv, 2023.

However, Sam Altman, CEO of ChatGPT, says that ChatGPT is not even close to an AGI model.

• **Artificial General Intelligence**

OpenAI more pragmatic view of AGI *July 2024 Bloomberg*



OpenAI: New Oversight Board | Los Alamos Partnership | Safety Team Dissolved | Microsoft Investment

OpenAI Imagines Our AI Future

Stages of Artificial Intelligence

Level 1	Chatbots, AI with conversational language
Level 2	Reasoners, human-level problem solving
Level 3	Agents, systems that can take actions
Level 4	Innovators, AI that can aid in invention
Level 5	Organizations, AI that can do the work of an organization

Source: Bloomberg reporting

According to the levels OpenAI has come up with, the third tier on the way to AGI would be called "Agents," referring to AI systems that can spend several days taking actions on a user's behalf. Level 4 describes AI that can come up with new innovations. And the most advanced level would be called "Organizations."

The levels were put together by executives and other senior leaders at OpenAI, and it is considered a work in progress. The company will gather feedback from people including employees, investors and its board, and may tweak the levels over time.

Most Read

Tesla Delays Robotaxi Event in Blow to Musk's Autonomy Drive

Stock Rotation Hits Megacaps on Bets Fed Will Cut: Markets Wrap

Saudi Prince's Trillion-Dollar Makeover Faces Funding Cutbacks

US and Germany Foiled Russian Plot to Kill CEO of Arms Manufacturer Rheinmetall

Kevin Costner, Warner Bros. Cancel 'Horizon: Chapter 2' Release

OpenAI says we are on the L1 but CHATGPT-o could do some tasks included in L2.

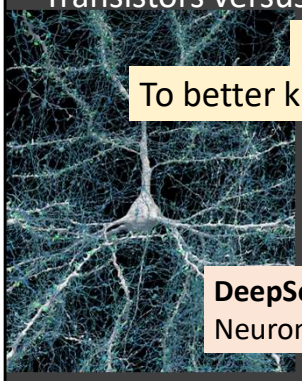
AGI ainda é um conceio teórico sob investigação
"autonomous, goal-directed, and highly adaptive."

Ray Kurzweil, Google's director of engineering and a pioneer of pattern recognition technology, believes that AI will reach "human levels of intelligence" in 2029 and surpass human intelligence by 2045

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Could the mind work as a Computer? And the opposite?

To mimic the Brain "in Silicon" Versus "in vivo". (??)
Transístors versus synapsis …



Brain: 86 billions neurons estimated

To better know how the Brain works is crucial

MRI superficial, UHField MRI) hard!
Resolution ~ 1mm^3 maximum 0,5 mm^3
Neded a Nanometric analysis

DeepSouth. IBM
Neuromorphic Supercomputer simulating synapsis at the human brain scale.

FinalSpark
brain tissue to run computing tasks

Harvard + Google: 5.000 slices 1,000 times thinner than hair
57K cells, 150M neural connections. *The Guardian 2024/5/13*

LIACC
artificial intelligence and computer science laboratory

UHF MRI ULTRA-HIGH FIELF Magnetic Resonance Imagery increases resolution.
Some estimates suggest we might reach computational power equivalent to the human brain within a couple of decades, while others believe achieving true AGI might take longer.
AGI development involves not just computational power but also understanding how the brain processes information, learns, reasons, and adapts—all while considering ethical, societal, and regulatory aspects. It's an interdisciplinary endeavor involving neuroscience, computer science, psychology, and more.
it might take longer—possibly several decades or even longer than a century.
UHF Ultra High FIELD

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

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Beneficial AI

A UN draft resolution promoting the use of AI in sustainable development and the protection of human rights was backed by over 120 member states

The General Assembly also recognized AI systems' potential to accelerate and enable progress towards reaching the 17 **Sustainable Development Goals**

Eugénio Oliveira

U.PORTO LIACC artificial intelligence and computer science laboratory

Anyway we should progress step by spet towards a Beneficial AI
The resolution comes just two months after UN secretary general António Guterres accused technology companies and governments of pursuing their own narrow interests in AI "with a clear disregard" for human rights, privacy and other social consequences,

U.PORTO FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GenAI

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, autand Modelingonomous mobility, Intelligent Power Networks, …

Eugénio Oliveira

LIACC artificial intelligence and computer science laboratory

The *World Economic Forum* is an independent international organization committed to improving the state of the world by engaging business, political, ...

A WEF study outlines over 80 possible uses for AI solutions concerning Water Air and Land.

It would be the motivation for another talk.

Beneficial AI

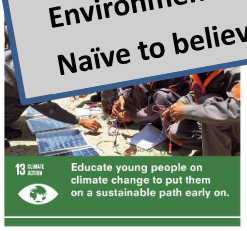
17 Goals for a Sustainable Development

Satisfying needs of the Present without affecting Future Generations

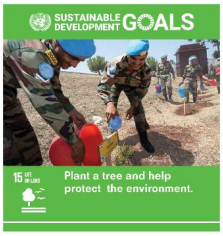
WATER

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**

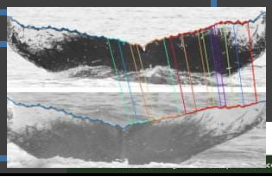
U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

AI for a Sustainable Environment

WATER

- Localize **garbage in the see** and measure water polution

- Monitor **Icy surfaces** measuring melting rates
- *Ocean Data Alliance is developing an AI system using satellite data* **to monitor maritime traffic, fishing and corals existence on the sea**
- individual Identification of see animals (**whales**) through their skin texture (fins and tails).



Eugénio Oliveira

dados decidir processos agricolas

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GenAI

REGULATION

High-Level Expert Group on Artificial Intelligence
Ethics Guidelines for Trustworthy AI
European Commission
Directorate-General for Communicatio
European Commission
B-1049 Brussels
Document published in **2019, April 08**

Ethical Principles

- **Principle of Beneficence *("DO GOOD")***
- **Principle of Non-Maleficence *("DO NO HARM")***
- **Principle of AUTONOMY**
- **Principle of Justice *("BE FAIR")***
- **Principle of EXPLICABILITY**

Eugénio Oliveira

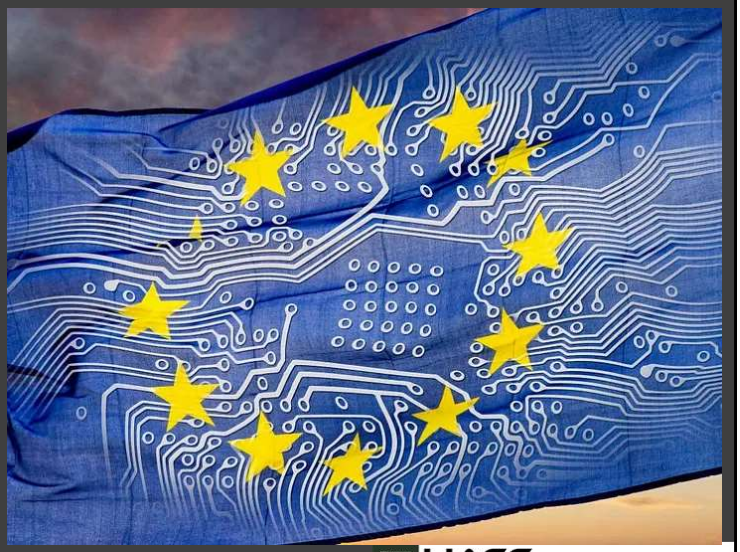
LIACC artificial intelligence and computer science laboratory

It is of paramount importance to pay attention to Regulation.
A good start was the Guidelines published by EU Commission in 2019
Create prosperity
Improve freedom and security
Pevent coercion from AI systems
Promote imparciality regarding groups, ethnics, and minorities
Produce transparente and audictable systems

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Regulation

European Union May 2018
General Data Protection
Imposing new responsibilities to the **Organizations** processing data from European Citiziens



U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

Regulation

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

Distinguishes High and Low risk AI Systems

The EU ACT encompasses various safeguards, including **on General-Purpose 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**.

AI Seoul Summit: 10 nations and EU recommit to safe inclusive AI
Signed 21 May 2024, the Seoul Declaration for safe, innovative and inclusive AI

.

U.PORTO
FEUP FACULDADE DE ENGENHARIA UNIVERSIDADE DO PORTO

GenAI

EU AI ACT **Prohibits** AI systems …

- to **predict** the risk of a person committing a criminal offence;
- to create or expand **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 inaugural meeting of the AI Board took place on 19 June ahead of the AI Act's formal **entry into force expected in early August.**

Eugénio Oliveira

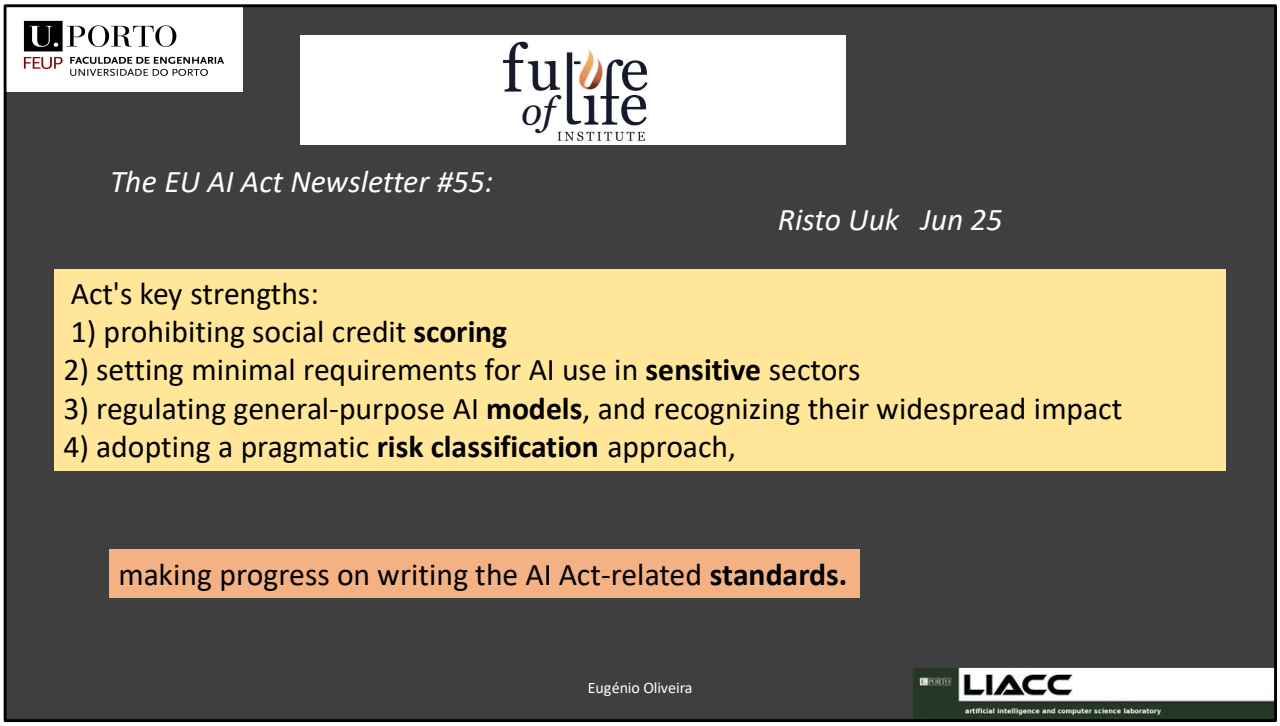
LIACC artificial intelligence and computer science laboratory

It Prohibits AI systems that deploy subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective or effect of materially distorting that person's behaviour … in a manner that causes or is reasonably likely to cause significant harm;

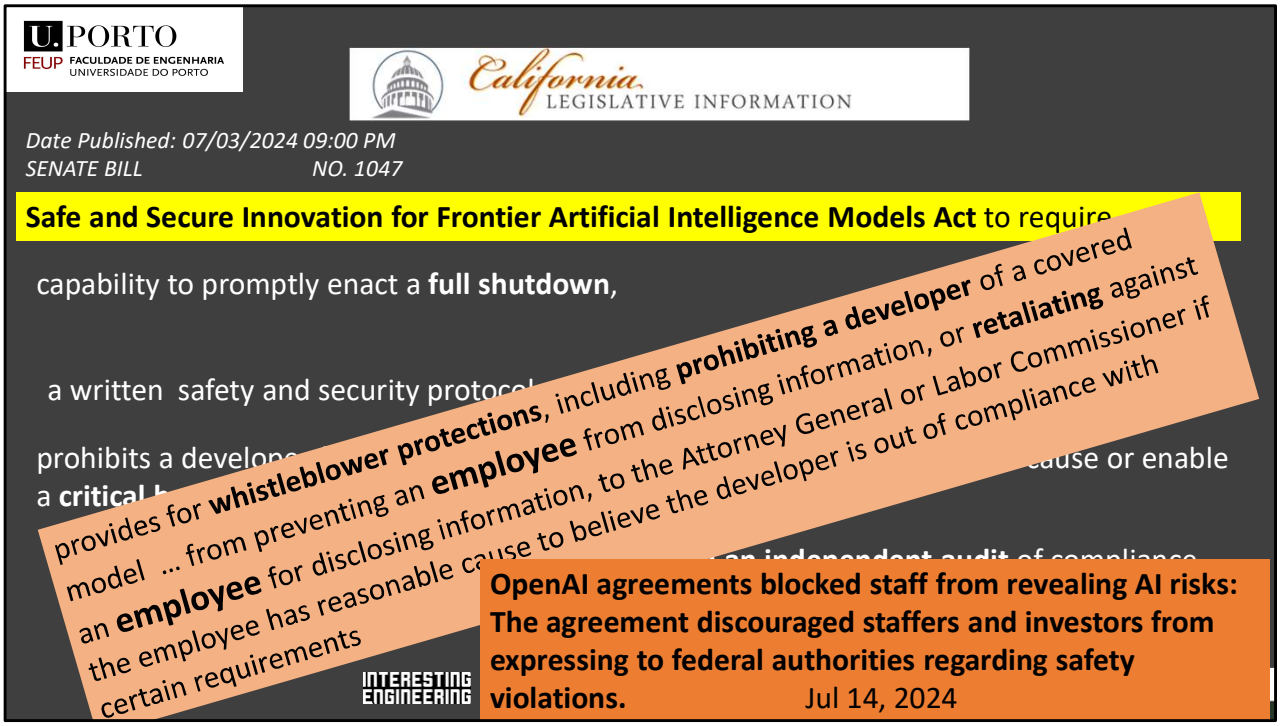
Interesting enough an EPIC Law Fellow,
EPIC is a public interest research center in Washington- Electronic Privacy Information Center,
Wrote about from

the

AI Act allows private actors and even the government (for non-law enforcement purposes) to construct a mass surveillance ecosystem



Act's key strengths include the following: 1) prohibiting certain AI systems, such as social credit scoring, to prevent erosion of social norms; 2) setting minimal requirements for AI use in sensitive sectors to prevent harmful deployment and promote human-centric AI development; 3) regulating general-purpose AI models, and recognising their widespread impact and the limitations of downstream providers; and 4) adopting a pragmatic risk classification approach, particularly for general-purpose AI models, using compute thresholds to differentiate risk levels. **Yet in the authors' view the Act also has many shortcomings. It risks becoming ineffective due to unclear rules and insufficient enforcement capacity. Equally, the requirements for high-compute general-purpose AI models have potential loopholes, and may quickly become outdated.**

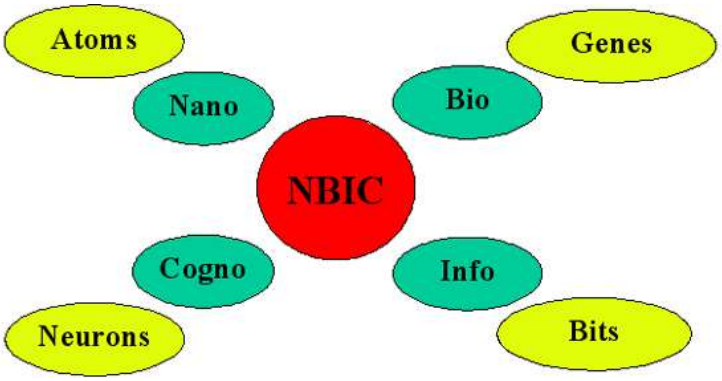


Interesting enough the cradle of AI, California state, is also trying to produce advanced Legislation to regulate AI.
a developer, before initially training a covered model, comply with various requirements, including implementing the capability to promptly enact a full shutdown,

"**covered model**" means (i) An artificial intelligence model trained using a quantity of

computing power greater than **10^26 integer or floating-point operations**, the cost of which exceeds one hundred million dollars ($100,000,000)

Sciences Synergy will shape the future

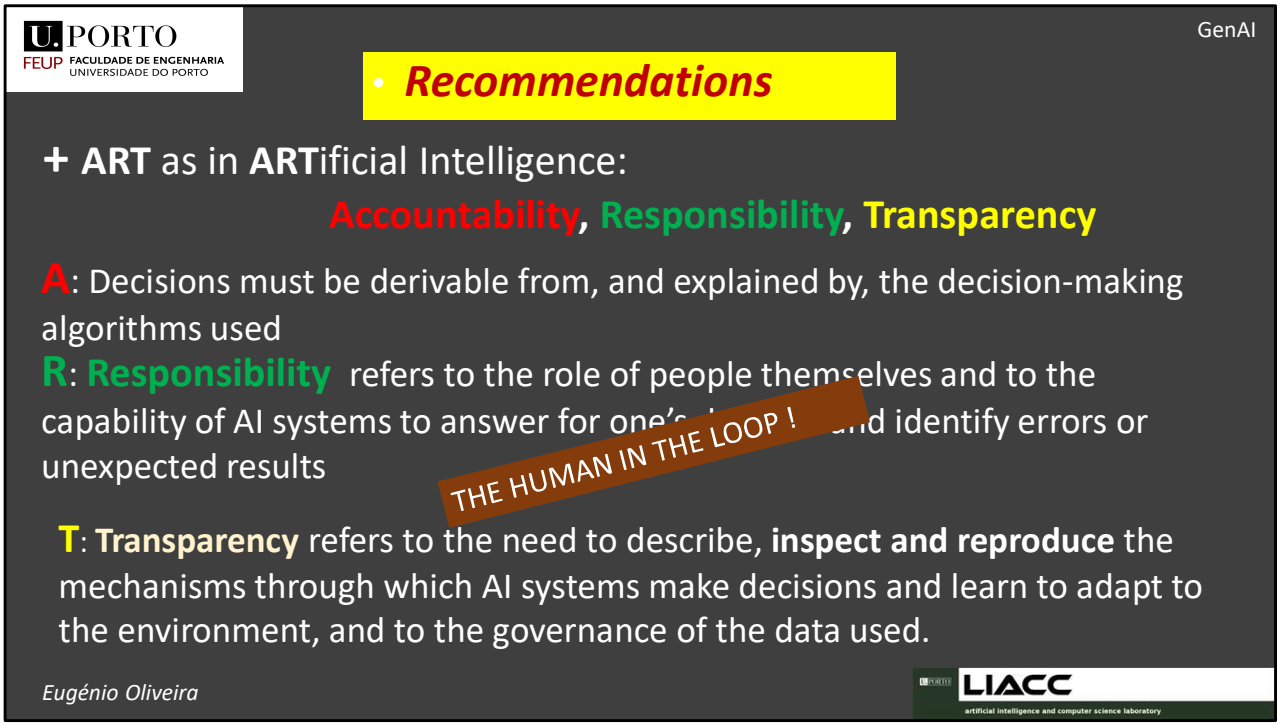


Let me emphazize that In the future AI will not be alone.

There will be many synergies between at least these sciences (we call this NBIC):: Nano sciences working at atomic level, Biotechnology dealing with Genes, of course Information Science dealing with programs and computers and Cognitive and Neuro Sciences dealing with neuron functionalities.

The synergies between these Sciences and Techniques will lead the future for the good

or/and the Bad.



I usually recommend (I am not alone, VD also proposes it) and I p tried to put that in practice that AI Systems always should include "The H in the Loop". **But specified in such a way that they follow ethical principles and societal concerns thus incuding MORE ART** in the ARTificial Intelligence: Accountability, Responsibilty, Transparency
"Accountability":
To whom should we address if a fully automatic car runs over a pedestrian? To the car's hardware builder, sensors and actuators supplier? To the **Software** developer that specifies the D-Mking process? To the legal **authorities** that permit those vehicles to be used on the road? To the **Driver** who personalises the acceptance of the automatic D-Mking? To the **car-robot itself** since its behaviour is also guided by its own experience from what it has learned? **To all of them?** (VD)
"Responsibility": the **responsibility** defor making decisions understandable and justifyable. **This does not happen with current "Deep learning" algorithms**

“Transparency” which relates to the good practices of systems **specifications, development and reprodutibility.**

And whenever possible always KEEP “The Human in the Loop” avoiding complete authonomy.

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

AKNOWLEDGES

INTERESTING ENGINEERING

IEEE Spectrum

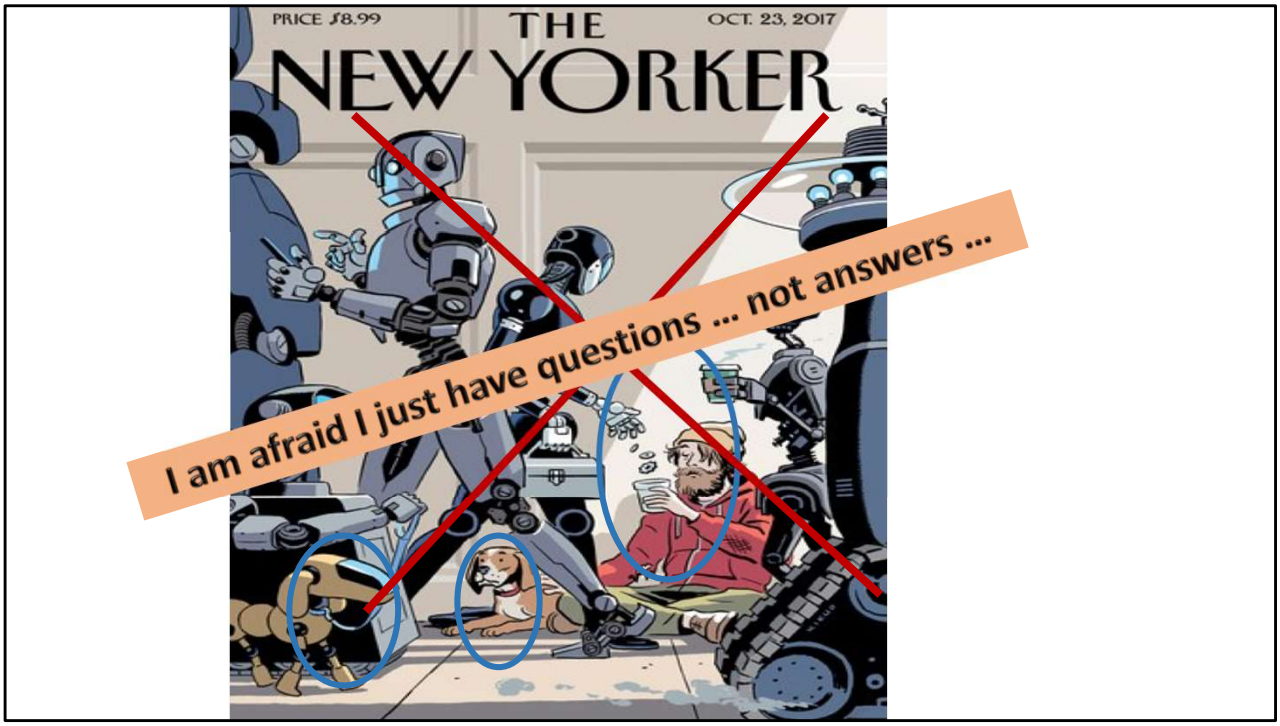
AlphaSignal

Bloomberg/Getty

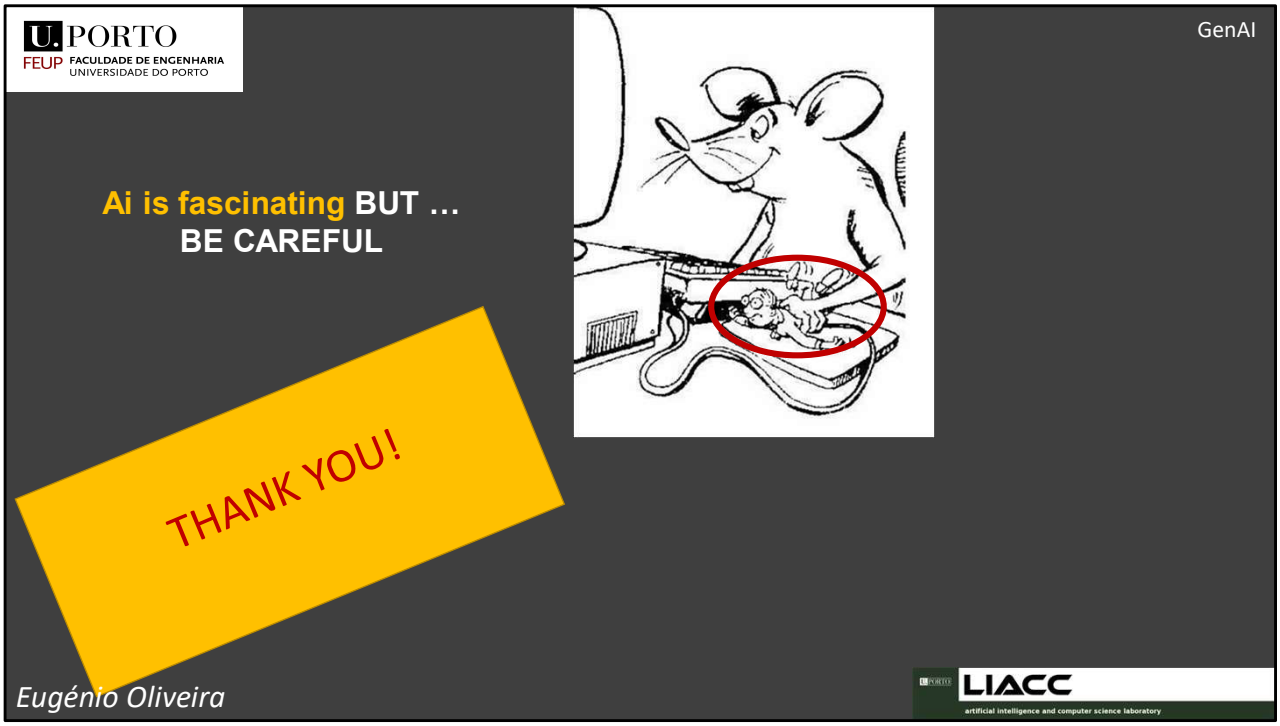
Financial Times

NYTimes

Eugénio Oliveira



The New Yorker will not turn into reality, I guess. May be the littledog will not be lucky enough to avoid replacement …



We don't believe that Humans will become obsolete even if there will be a transference of a lot of competencies (as it was the case during Industrial Revolution). For ex. **UBER had to hire a lot of specialists in AI and automatic contro (at least 50 came from the Robotics Institut at CMU**). Some AI experts **are very much welcome at Wall Street**.

A substituição de empregos existirá mas as sociedades como um todo recompõe-se e ultrapassam as revoluções económicas para novos patamares. Mas há sempre muitas pessoas que podem vir a ser trituradas no processo e é absoluto dever de todos não permitir que tal acontecça, chame-se isso reeducação, solidariedade ou caridade.
A lei normalmente move-se mais lentamente que a tecnologia. Vai demorar bastante antes que alterações realmente significativas na lei permitam, por exemplo, o uso alargado da condução automática.
Mady Delvaux na sua tentativa junto da UE propôs em janeiro passado uma peça de legislação detalhada que incluía dar uma cartilha de

deveres e direitos civis à IA. Tal incluía dar a robôs inteligentes uma “e-personalidade” limitada comparável ao que se faz com corporações. Um estatuto legal que permite a empresas processar e ser processada (pelo menos no respeitante a compensações).