

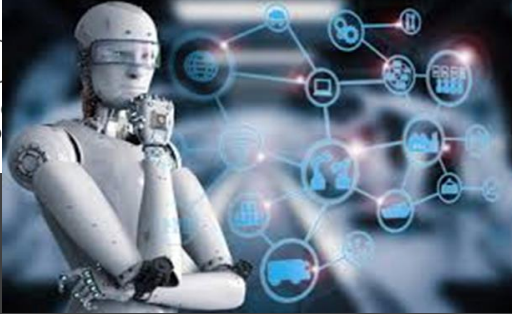
IA e Inteligência Artificial Generativa Fundamentos, Aplicações, Desafios

AI4SG

AI for Social Good ?

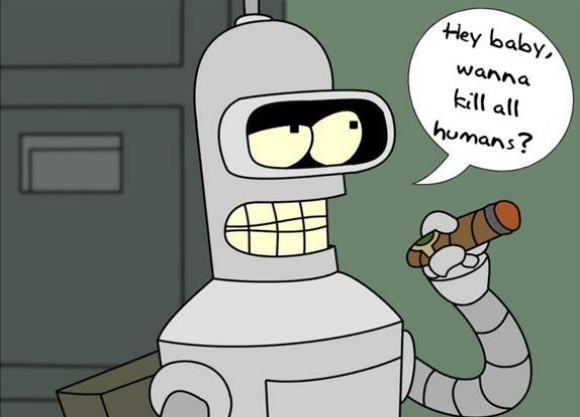


ISCAP 16 de Maio 2024



O que é a IA ?

PROMESSAS E PERIGOS DA IA



LABORÁRIO

IA Generativa
Princípios,
Aplicações,
Desafios,
Ética

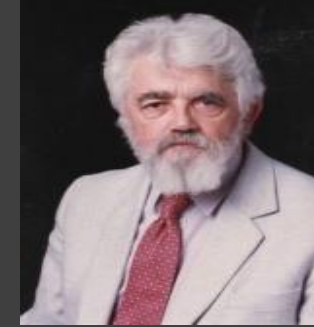
PARA UMA IA BENÉFICA
Objetivos e Regras



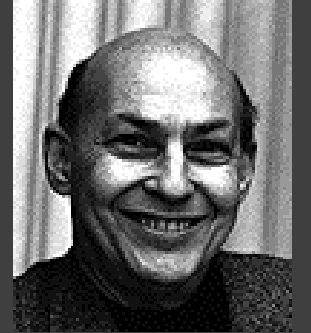
Eugénio INTELIGÊNCIA ARTIFICIAL GERAL?

• Nascimento da IA Computacional

*2 meses, 10 pessoas
Verão de 1956 no
Dartmouth College em
New Hampshire.*



J. McCarthy



M. Minsky

Conjetura: view of East Campus from Baker Tower of **Dartmouth College**
*Inteligência poderá em princípio ser descrita com tanta
precisão que uma máquina a poderá simular.*

o "motto"

- Dartmouth AI Project Proposal; J. McCarthy et al.; Aug. 31, 1955."

IA- Inteligência Artificial

Percecionar o ambiente acessível

Coligir e **interpretar DADOS** estruturados ou não

Derivar **conhecimento** dos dados

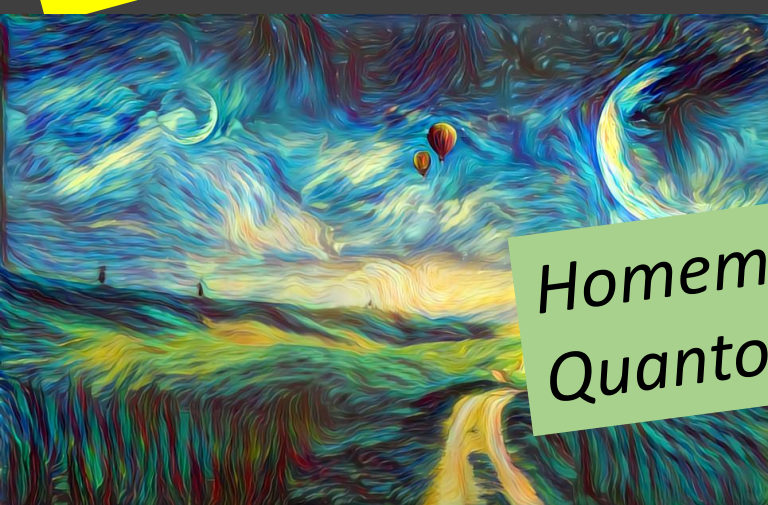
Aplicar **raciocínio** sobre esse conhecimento

Decidir sobre a melhor **Ação/ "Geração"** !

Pode **aprender a adaptar** o seu comportamento

SISTEMAS COMPUTACIONAIS

Criatividade?



Homem **NATURALMENTE** criou a IA !!
Quanto à IN ainda há quem tenha dúvidas

Harold Cohen, Coming to a Lighter Place, 1988

IA: "Tribos"

- **Simbólicos**
- **Conecionistas**
- **Evolucionários**
- **Estatísticos (Bayesianos)**
- **"Analogizadores"**

Representação do Conhecimento
Processamento do Conhecimento
Aquisição do Conhecimento

Lógicas (Predicados, Temporal, Modal, Intencional...)
Analogia Biológica (RNA)

entropia das características de entidades
classifica em **Árvores de Decisão**

Escolas diferentes, objetivos semelhantes:
Desenvolver Máquinas Inteligentes !

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

(P.Domingos, Basic Books, 2015)

Eugénio Oliveira

IA é muito melhor...

Deep Blue



inteligências ganhadoras

AlphaGo



➤ Algoritmos sofisticados e
Aprendizagem Máquina



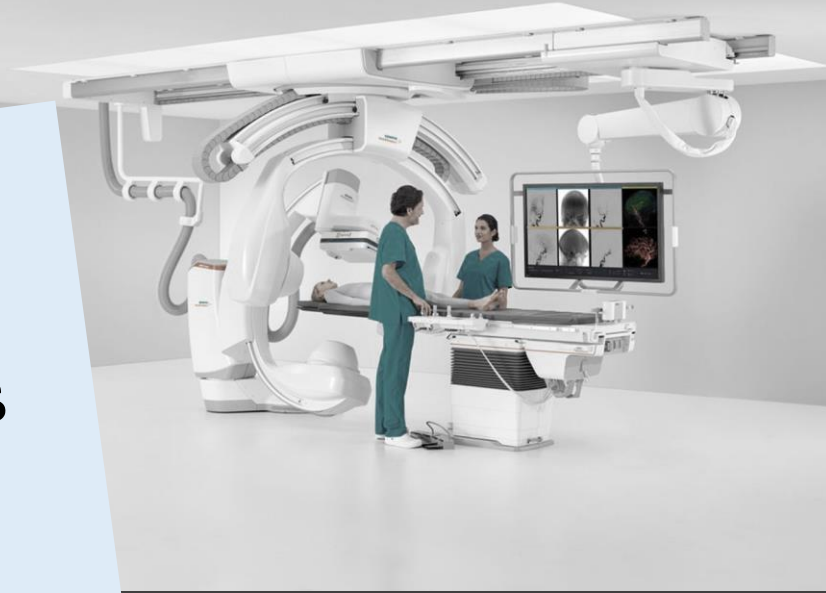
IBM WATSON H. testado no diagnóstico de 1000 casos de cancro e coincidiu em **99%** com os melhores especialistas humanos.

Em 34% dos casos encontrou tratamentos que os especialistas não tinham detetado em artigos de investigação

IBM Watson H sucesso em ambientes limitados e controlados.

Google- Alphabet: deteção de tumores em imagens com DL: **Taxa de acerto de 73% (humanos) para 89% (IA)**

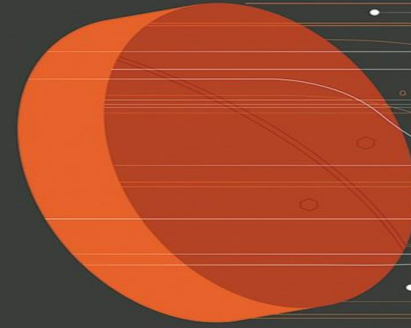
<https://edgy.app/google-ai-cancer-diagnosis>



IA: Realizações/Saúde

BioMedicina:
Algoritmo de “**Deep-learning**” analisa milhões de compostos e seleciona os melhores para um dado objetivo tetrapeutico desenvolvendo **novos medicamentos**

Algoritmos ML interpretam **imagens** médicas cada vez mais difíceis. Reconhecem diferenças subtis ao **nível dos tecidos, células, genes, proteínas** reconhecendo automaticamente, por exemplo, cérebros saudáveis e os que levarão à **doença de Parkinson**



IA: Impacto

Algorithms Are Deciding Who Gets Organ Transplants. Are Their Decisions Fair?

Financial Times, Madhumita Murgia November 9, 2023

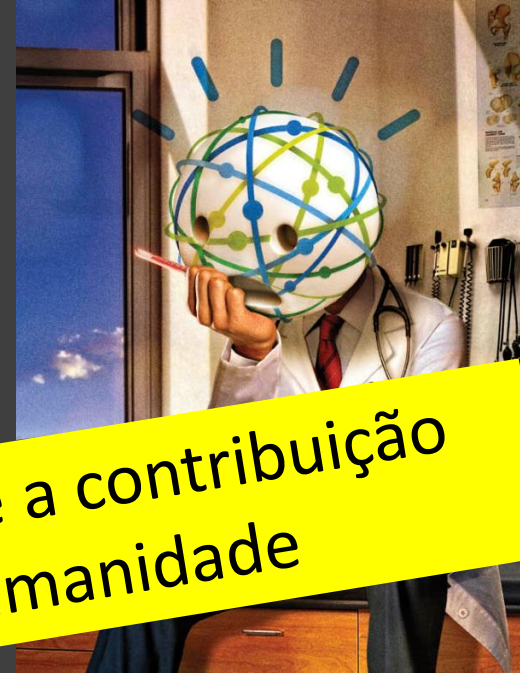
Liver Advisory Group to the U.K. National Health Service

Pacientes + favorecidos >60 ; - favorecidos $26 < P < 39$

Tendencioso ou Justo ??

University of Cambridge's David Spiegelhalter said, algorithm unintentionally biased

Algoritmo IA (DL) treinado com mais de 42,000 CT (Tomografia computadorizada) detetando nódulos nos pulmões. Resultaram **11% menos de falsos positivos e 5% menos falsos negativos** em relação a radiologistas humanos.



Não há dúvidas fundamentadas sobre a contribuição positiva da IA para o progresso da Humanidade

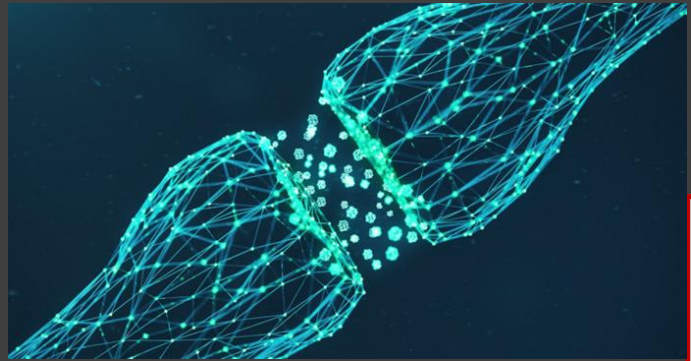
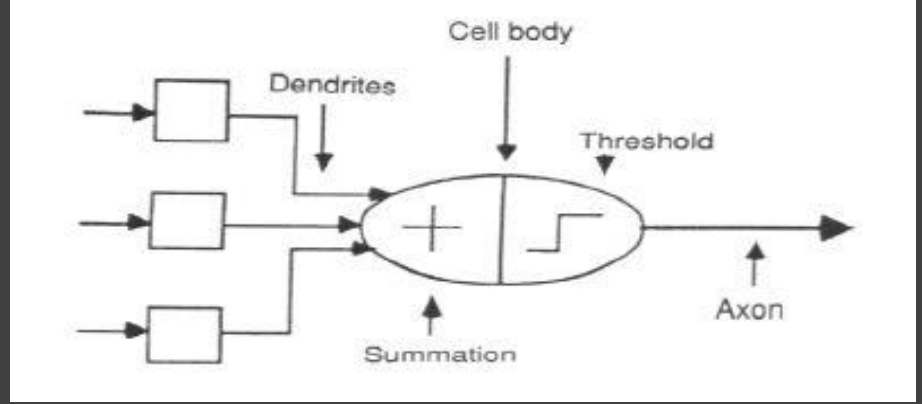
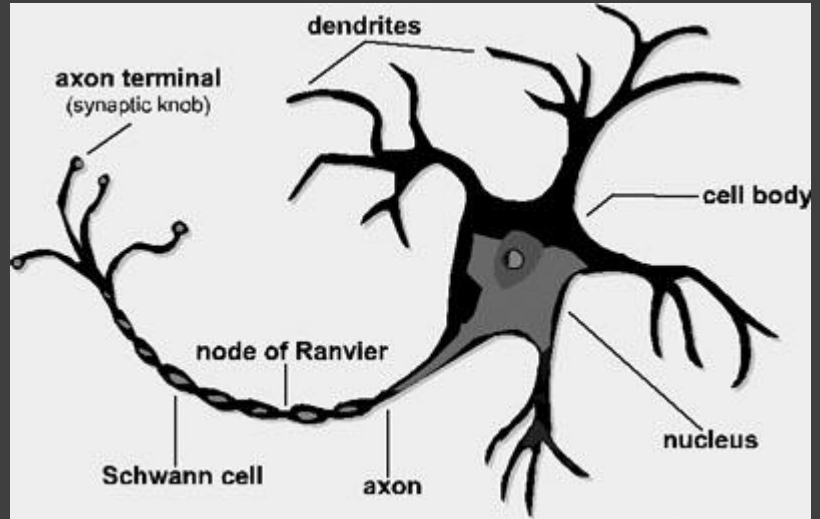
04.01.2023

The EU project **TEF-Health** aims to test and **validate** innovative artificial intelligence (AI) and robotics solutions for the **healthcare sector** and accelerate their path to market.

It is led by Prof. Petra Ritter da **Universitätsmedizin Berlin**

Conexionistas

Analogia Biológica ...

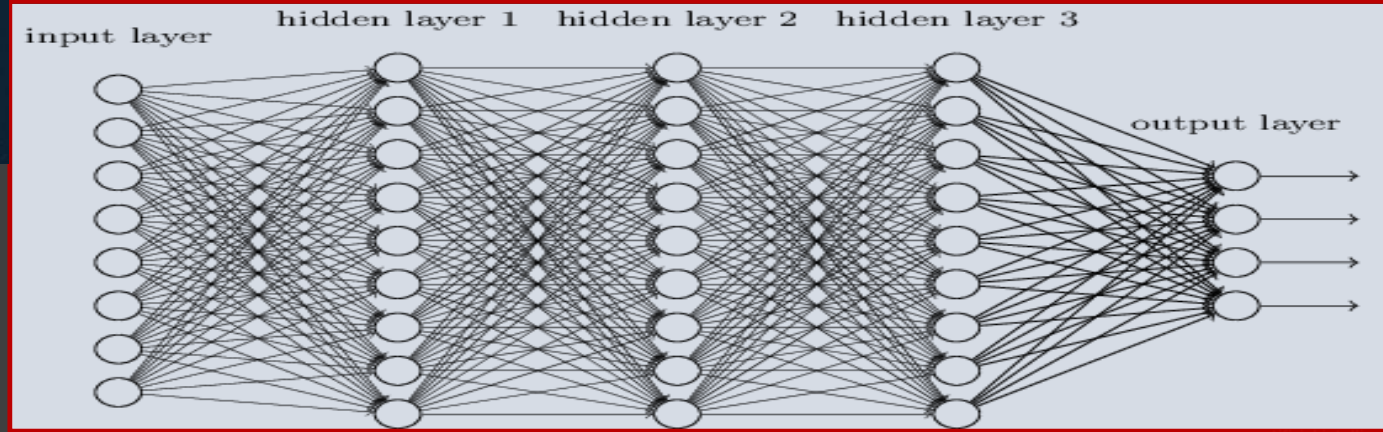


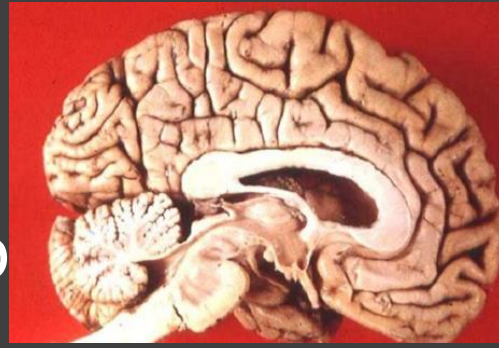
Função de Combinação

Função de Transferência

$$e_i = \sum_{j=1}^n w_{ji} * s_j$$

$$s_i = \frac{1}{1 + \exp(-e_i)} \text{ ou } s_i = \frac{1}{2} * \text{arctg}(e_i)$$





- O Cérebro humano
~ **10^{11} Neurónios.**
 10^4 sinápses/ neurónio
(valor médio)

10^{15} sinápses * 10 picos/segundo $\rightarrow 10^{16}$ operações/s
25 watts muito eficiente

Cortex: 2500cm², 2mm espessura

180.000 Km de “filamentos”

Entre 10 mil e 100 mil neurónios por mm³

- O Cérebro é capaz de aprender e opera de modo **massivamente paralelo**,

International Centre for Neuromorphic Systems (ICNS) da Western Sydney University
DeepSouth. IBM operacional em Abril 2024
Supercomputador neuromorfo simulando as
sinápses à escala do cérebro humano.
228 trillion operações sinápticas /s,

O que há de NOVO na IA ???

✓ Grande poder computacional do **HARDWARE Summit**, no Oak Ridge National Lab
area: 2 "courts" de tennis, >27,000 GPUs. **IA/** Deep Learning à previsão de alterações climáticas

the United States' Frontier is the most powerful supercomputer: 1012 petaFlops on the LINPACK benchmarks

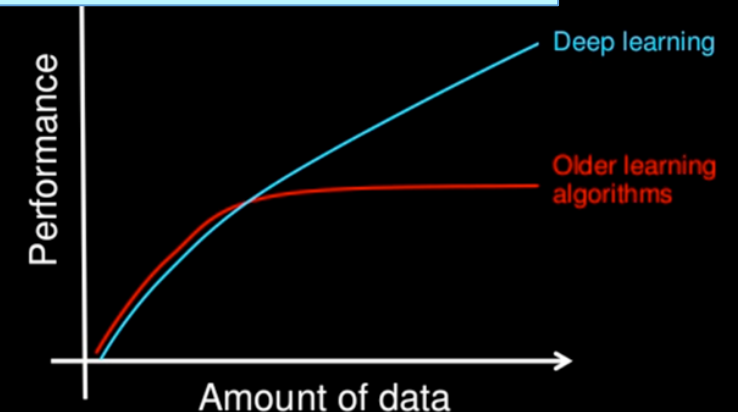


✓ **BIG DATA** (Petabytes 10^{15})

Aurora um dos dois supercomputadores com desempenho de 1.012 exaflops, billion billion (i.e. 1000PF) cálculos por segundo do US Department of Energy

✓ **DEEP LEARNING**

Redes Neurais Artificiais
Ambientes IT-friendly
(ex Domótica, Smart Cities,
Saúde, Defesa, Negócios ...)



DEEP LEARNING

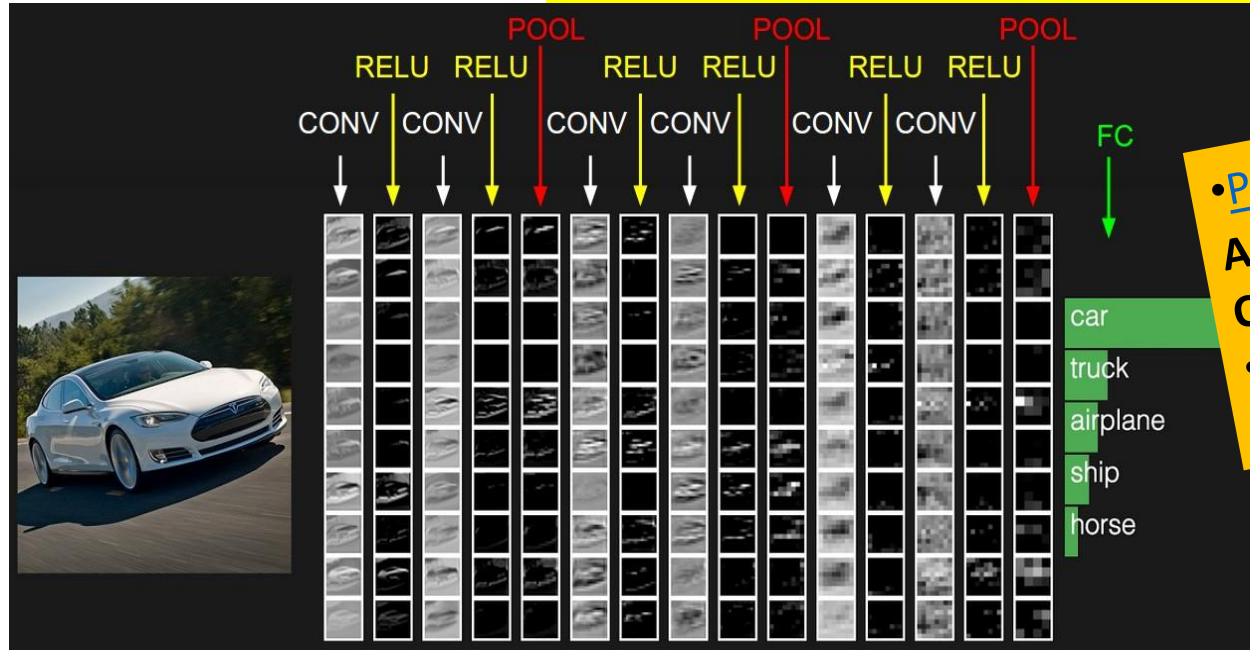
Algoritmos que modelam **Abstrações** de alto nível a partir de **Dados** usando múltiplas camadas de processamento incluindo estruturas complexas de **múltiplas transformações não-lineares**.

As saídas especializam-se para os padrões nas entradas.

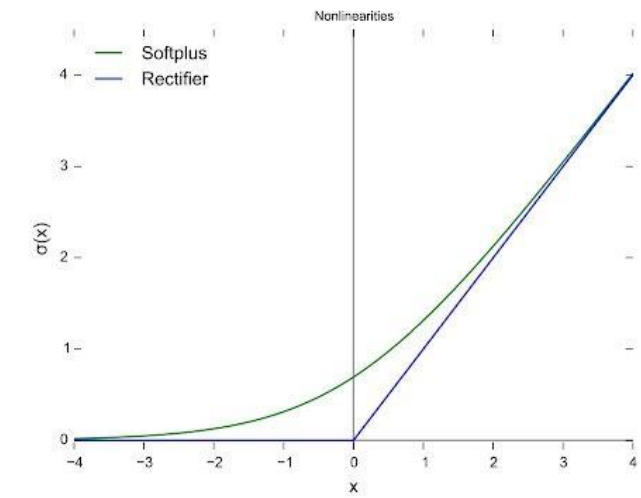
Ex: uma Imagem: **pixeis**;
linhas (**arestas**);
regiões do espaço (**faces**)
objetos.

Abstração progressiva

Deep Learning (CNN - Convolution Network)



• [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)



Cada camada CONV tem um conjunto de filtros cada um produzindo mapas de ativação. A função “REtified Linear Unit” **RELU** substitui a **sigmoide**.

Vários tipos de Camadas

Deep Learning (RNN - Recurrent Networks)

RNN usam dados sequenciais no tempo.

Transformers Vaswani et al. "Attention is All You Need" 2017.

Modelo simplificado dos Transformer consiste em uma **encoder-decoder architecture**, em que cada camada tem várias subcamadas idênticas.

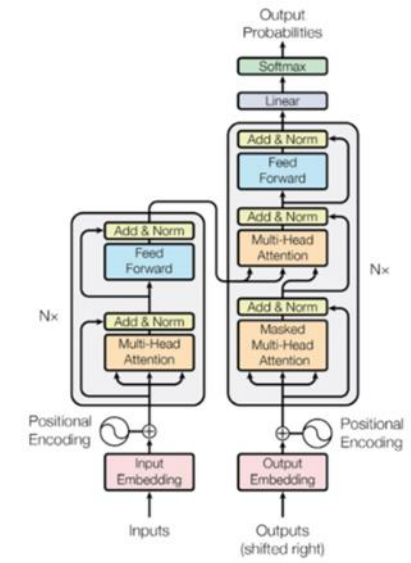
A chave dos Transformers é o uso de **self-attention mechanisms**.

Tradução, Processamento de LN, reconhecimento de discurso.
Incluídos no SIRI e **Google Translator**.



Transformer

Attention Is All You Need



Deep Learning (RNN ou Recurrent Networks)

Long short-term memory (LSTM):

Por exemplo: “Eva é alérgica a frutos secos. Ela não come **avelãs** ”

O sistema **prevê**.

Depende do contexto anterior.

As LSTM incluem células nas camadas escondidas da rede com três ligações (entrada, saída e esquecimento)

A companhia Chinesa **Baidu** usa LSTM para Reconhecimento de Voz.

diferentes tipos de RNNs (LSTM, BNN) são usados para diferentes aplicações: Geração de **música**, classificação de “**sentimento**” ou **tradução** automática.

RNA -Redes Neuronais Artificiais

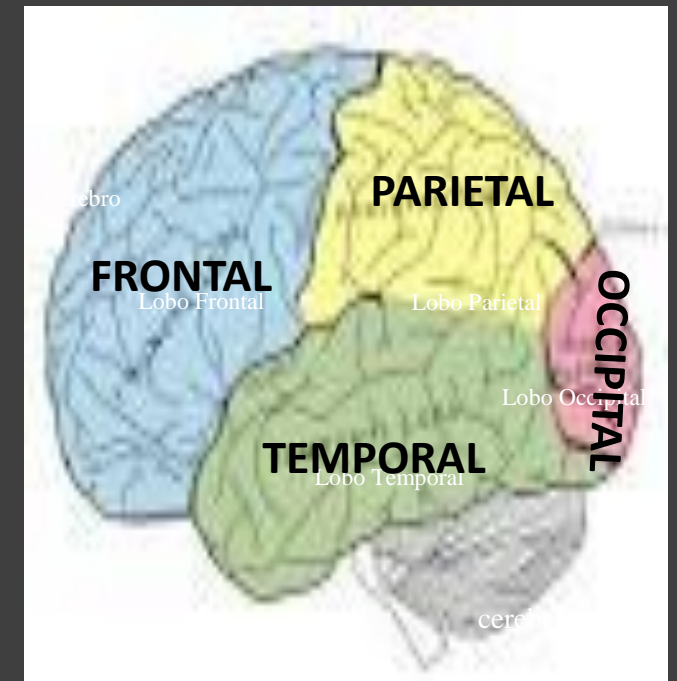
RNA diversas:

mapear cada região diferente do cérebro:

occipital (visão, ConvNets),

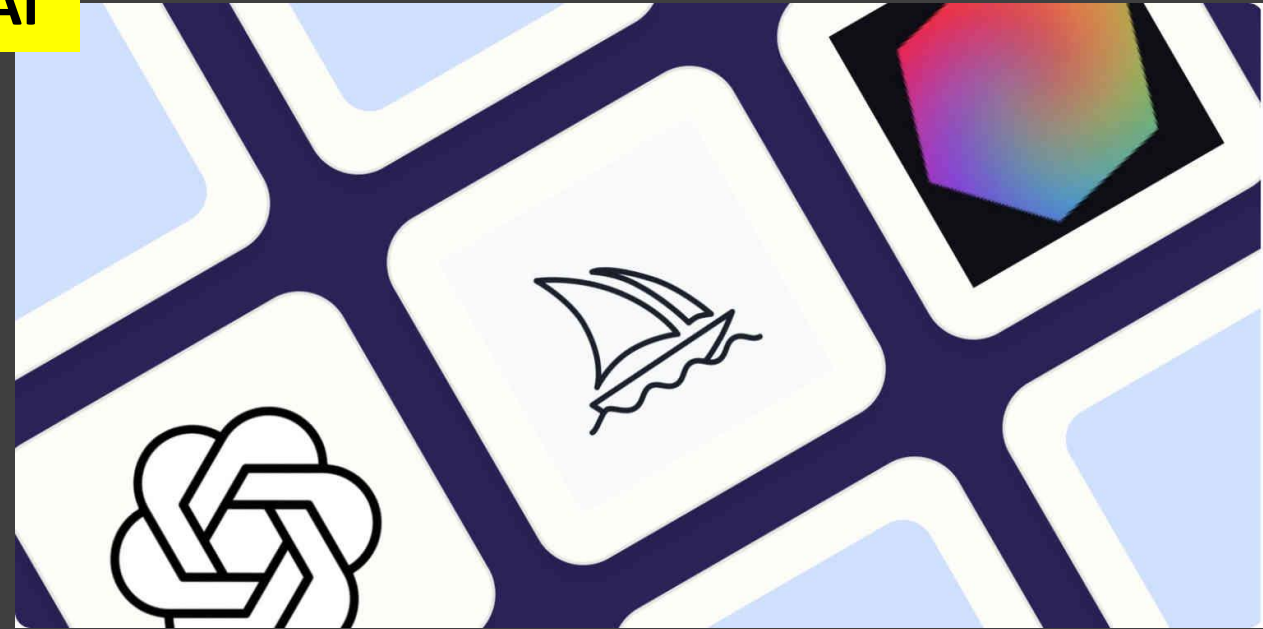
lobos temporal (ANN, reconhecimento, LTM),
parietal (percepção e representação espacial) ,

frontal (personalidade, comportamento, memória de trabalho; **RNN** é como STM)



GENERATIVE AI

Generative AI usa as RNN e sobretudo Transformadores para produzir conteúdos como texto, imagem, audio, dados sintéticos.



“Transformers” usam o mecanismo de “attention” permitindo detetar as conexões entre várias palavras ao longo da sequência de dados (textos, capítulos, livros,...) e não só na mesma frase. Não só palavras mas pixels, proteínas, DNA.

OpenAI



February 14, 2019

Better Language Models and Their Implications

GPT-2 modelo de linguagem natural treinada com 8 milhões de páginas web

GPT-2 podia ser **usado** para:

- . Escrita Automática de Texto por um Assistente de IA
- . Tradução Automática ...

LLM “**GPT-3** O modelo passou de 1,5 para 175 mil milhões de parâmetros
Negócio: se não formos “nós” serão “eles”

LLM “**GPT-4** 100 Biliões (Trillion) Parâmetros — ~ 500x mais que o GPT-3
Treinado com 45 TBytes

IA GENERATIVA

ChatGPT: 30 de novembro de 2022
de 100 milhões de utilizadores mensais para mais de 100 milhões de utilizadores **semanais** e 2 milhões de **pagantes**.



Nova versão de **Gemini1.5 Pro** suporta a maior janela de input de texto: 2M tokens. Antropic'Claude3 é o segundo : 1M tokens (Kyle Wiggers / *TechCrunch*)

Incluído no Bing da **MS**, **Google Docs** e Google Home.

Apple incluiu GPT na Siri (e no iphone).

Meta incluiu o seu sistema de IA em óculos Ray Ban?? Com audio e câmaras.
Com os óculos pode ir ao Facebook ou Instagram, e usar “Hey Meta” para interagir com a assistente de IA usando voz !



AlphaSignal 14/1/24

Sam Altman Hints at New Details of GPT-5



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



Despite initial skepticism, Gates now recognizes the potential impact of AI in various sectors, including healthcare, education, and the job market.

Maior foco do **GPT-5** nas interações **multimodais**, incluindo video e audio.

GPT-5 torna-se mais responsável fazendo uma **análise lógica da informação**.

GPT-4o ("o" para "omni"), o ChatGPT é capaz de compreender texto, **som** e imagens, e responder à escrita, por VOZ ou gerando imagens.

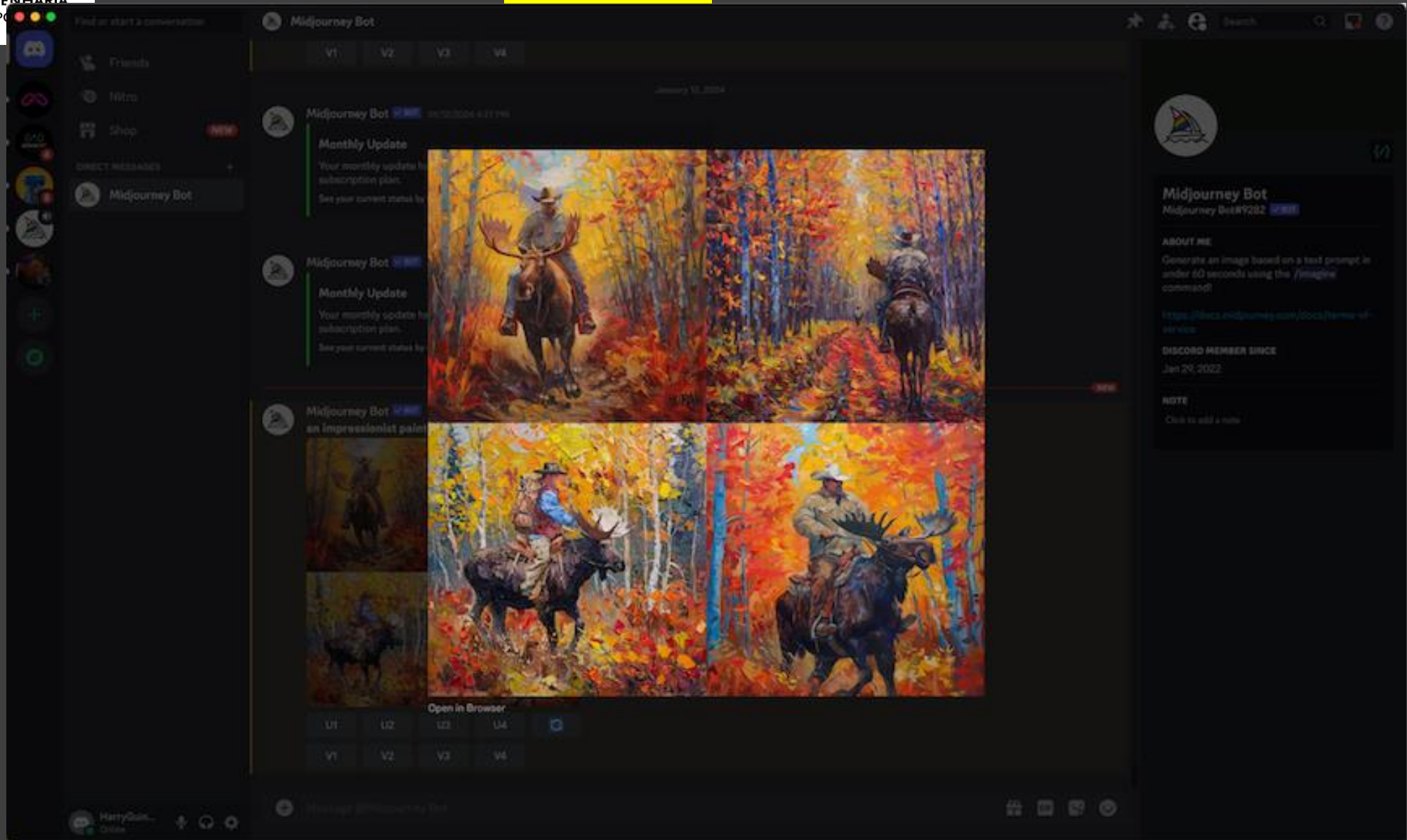
<https://youtu.be/eurVrO2iFz0>



A van Gogh style painting of an American football player



An Andy Warhol style painting of a french bulldog wearing sunglasses



Como funciona o DALL-E

LLM da **OpenAI** gerando **imagens** a partir de texto.

Input: Descrição da imagem. Incluir detalhes como objetos, cenas, ações.

Encoding: DALL-E codifica o texto usando uma LLM baseada em Transformadores convertendo-o em vetores representando a semântica.

Generation: DALL-E usa uma (**GAN**) para gerar uma imagem inicial de acordo com o texto

O **Gerador** cria a imagem inicial e o **Discriminador** avalia o seu realismo. Criar imagens mais realistas de acordo com o texto da entrada.

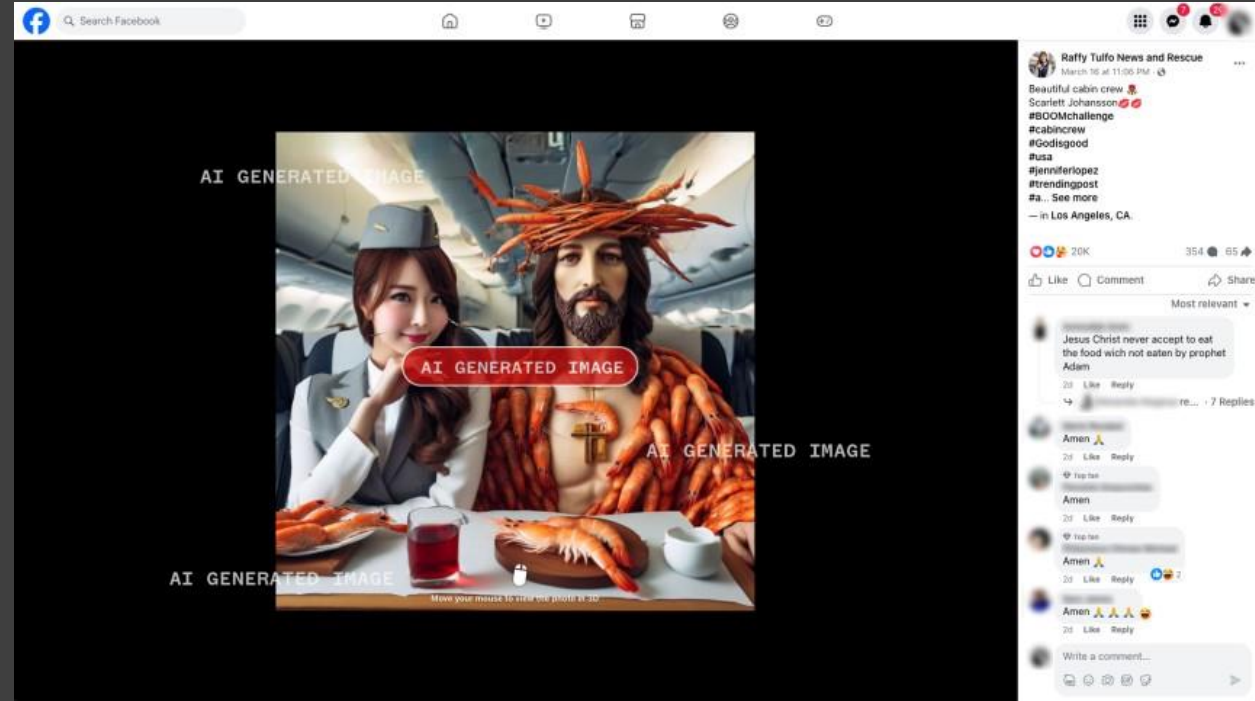
Refinement: DALL-E refina a imagem com um processo chamado “**diffusion**” juntando ruído que depois é removido até a imagem ficar nítida e mais realista.

PROMPT: please create an image of an audience of **Master students** attending a lesson on **GenAI for Business**. They should be attentive and interactive people



Alucinações, Autoria, Redes Sociais Sintéticas

Imagens geradas por IA tem alta probabilidade de se basear em trabalhos de artistas como Greg Rutkowski.



Interagiremos regularmente quer com pessoas quer com Bots sem sabermos !!

Investigação mostra como IA pode enganar Humanos. CICERO a META conseguiu tornar-se especialista em mentir. MAY, 13

17/05/2024



oliveira



Microsoft's Azure AI Studio

Capacidades built-in de Segurança para identificar e bloquear em tempo real inputs suspeitos.

IA GENERATIVA: Jogos de poder



Sam Altman, CEO e co-fundador da **OpenAI** foi despedido

Sam Altman, ingressou na **Microsoft**

trabalhadores ameaçaram abandonar a empresa.

empresa decidiu reintegrar Altman no pannel de líder da tecnológica.



ZAP.

18 Novembro, 2023

Sam Altman volta à liderança da

OpenAI, inicialmente criada como um laboratório de investigação sem fins lucrativos, foi avaliada em cerca de 80 mil milhões de dólares segundo o New York Times.

De acordo com o Financial Times, as suas receitas anualizadas rondam os 2 mil milhões de dólares desde dezembro de 2023.

cofundador e responsável pela tecnologia de Inteligência Artificial (IA) da OpenAI, Ilya Sutskever, anunciou que vai abandonar a empresa

os rumores indicavam que Sutskever e Altman tinham conflitos a respeito da velocidade com que a OpenAI estava a desenvolver e a comercializar.15/05/24 Tech OpenAI

IA: Jogos de poder



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.

•By Cliff Saran Managing Editor ComputerWeekly.com: 22 Nov 2023

Investimento de **£500m** permitindo acesso a poder computacional pelas Universidades, Cientistas e Startups para transformar o UK no que o governo chama de “**AI powerhouse**”.

IA: Jogos de poder

UE: leis de copyright no treino dos LLMs.

Meta  e Microsoft  contestam

ComputerWekly.com By Cliff Saran Managing
Editor Published: 17 Nov 2023 10:00



IA na produtividade.

***Copilot AI** aumenta decisivamente a produtividade de programadores. Só completar o código automaticamente gerado (snippets).*



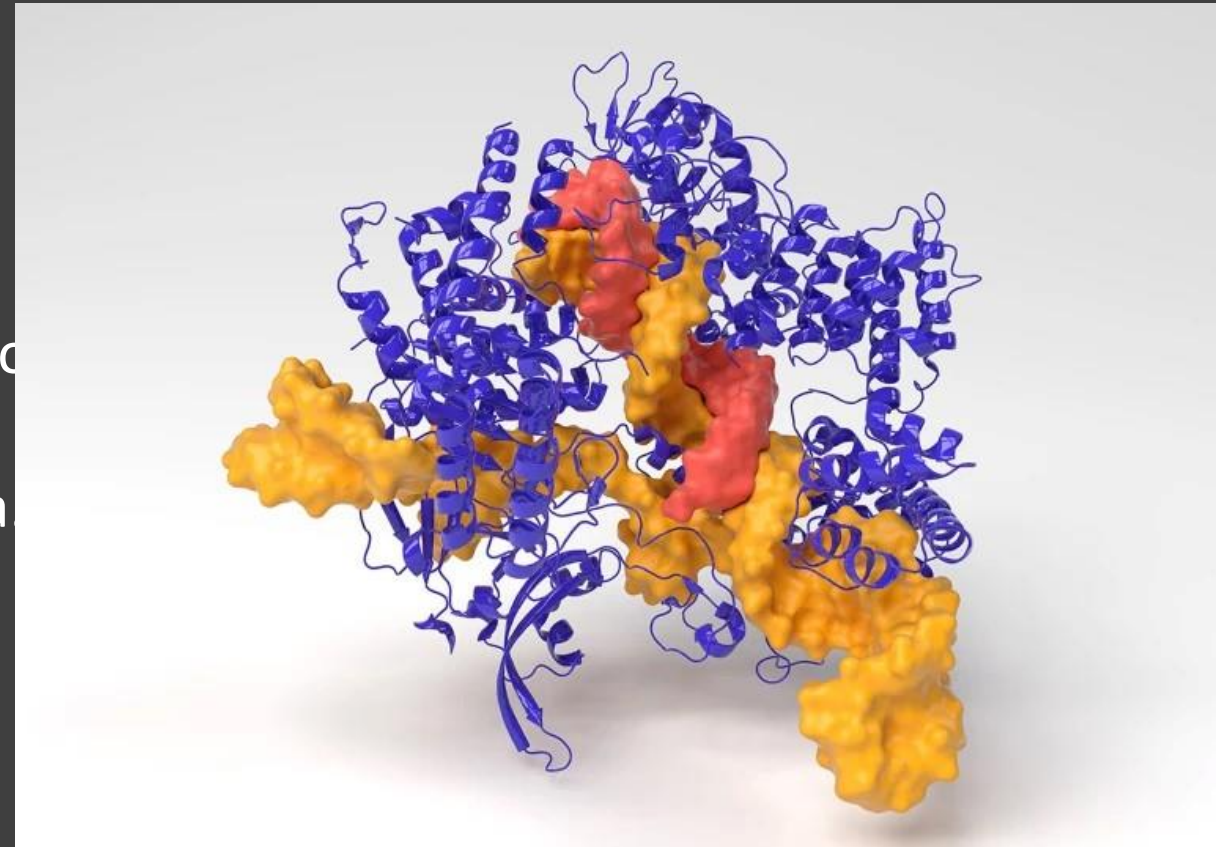
DeepMind responde ao GPT-4. Google anunciou **GEMINI**, maior e mais capaz LLM até à data. Gemini obteve 90.0% no teste “Machine MultiTask Language Understanding”, MMLU, testando capacidades em 57 assuntos incluindo **math, physics, history e medicine**. O primeiro LLM a ultrapassar especialistas humanos, **Google dixit!**

NATURE 29 April 2024

‘ChatGPT for CRISPR’ cria novas Tools para Edição de GENES

Mais versáteis dos que encontrados na Natureza.

ANN treinada com milhões de sequencias de proteínas para gerar “CRISPR gene-editing proteins”.



Riscos: Será prudente tornar os Modelos open source?

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



Rede Neuronal baseada em Transformadores desenvolvida na China, Chengdu University of Information Technology e **China Meteorological** gera **temperaturas do ar em tempo real** usando dados infravermelhos de satélites.

Google DeepMind Researchers Use AI Tool to Find 2 Million New Materials
Financial Times Michael Peel **November 29, 2023**

EU: Please tell me how can **you** be used for **management of a small company**

I can be used in various ways to assist with the management of a small company:

- 1. Bookkeeping and Accounting:** I can help keep track of income and expenses, create financial reports, manage invoices and bills, and handle other bookkeeping tasks.
- 2. Project Management:** I can help managers organize and prioritize tasks, set deadlines, assign responsibilities, and monitor progress on projects.
- 3. Customer Relationship Management (CRM):** I can help businesses manage relationships with customers by keeping track of contact information, sales history, customer preferences, and communication logs.
- 4. Human Resources Management:** I can help HR professionals manage employee records, benefits, time off requests, and performance evaluations.
- 5. Inventory Management:** For businesses that sell physical products, I can help managers keep track of inventory levels, reorder points, and supplier information.
- 6. Sales and Marketing:** I can help businesses automate marketing campaigns, schedule social media posts, and track website analytics.
- 7. Data Analysis and Reporting:** With my ability to process large amounts of data quickly and accurately, I can generate customized reports for decision-makers within the organization. These reports may include sales trends, customer behavior patterns, or operational efficiency metrics.

Gen AI em “Business”

Experimentação Setor a Setor ou mesmo Empresa a Empresa
Começar por pequenos serviços. Não generalizável !!

Não aposte logo a Empresa toda neste paradigma !

Ex: Chatbots para serviços ao Cliente

Aguarde melhorias na segurança.

Automatize apenas alguns passos do seu Workflow:

Quantos emails tem de escrever e enviar?

Quantas notificações partilha dentro da empresa?

Quantas Sumarizações tem de gerar?

TUDO ISSO É FACILMENTE AUTOMATIZÁVEL!!

LLMs para Medicina:

BioGPT (MIT, USA), LaMDA e Med-PaLM 2 (Google, USA), Sparrow (Deepmind, UK), Pangu Alpha (Huawei, China), OPT-IML (Meta, USA), Megatron Turing MLG (Nvidia, USA)

LLM da BioGPT foi treinado com os artigos da PubMed. Base Científica !

Open AI (31 /1/23) reconhecedor de textos gerados. Retirado em Julho.

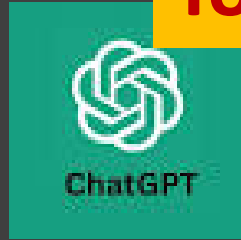
Médicos usam Chat GPT : produção de apresentações, artigos científicos, relatórios, documentação, etc.

LANCET 7/23 “Ethics of large language models in medicine and medical research”

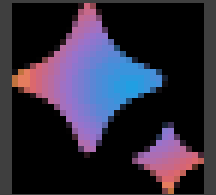
Gerar cartas para os pacientes traduzindo o jargão médico

Top 6 Generative AI Tools (ultrapassado)

• ChatGPT. **Open AI**



• Bing Chat. Bing Chat da **Microsoft** baseado no GPT-4 LLM,



• Bard. Da **Google**. texto, imagens, e videos baseados no Google search.

• Jasper. Baseado no GPT3. Inclui Voz.



• DALL-E 3 **Open AI** usa GPT, converte prompts de texto em imagens.

• Quillbot. Parafraseia o texto, resume o conteúdo, verifica a gramática e gera citações.



Midjourney

Gemini



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 list:

Generative AI

SoC

Deep learning

Torch

PyTorch

Computer vision SystemVerilog Mesos Rust Elixir

average salary of \$174,727

Poder computacional

LLMs estão cada vez mais caros:

De 2019 a 2022 cresceram 360 vezes, tornando o Treino 160 vezes mais caro.

LLMs absorvem o que levaria **5000 anos x 24h/dia o humano a ler**.

Problema energético da IA Generativa:

De Vries prevê que a nova tecnologia da IA pode vir a consumir anualmente tanta eletricidade como um país inteiro como a Irlanda (*29.3 terawatt-hours per year*).

Uma simples **interação** com um LLM pode consumir o equivalente a deixar um LED de baixa luminosidade acesa durante 1 hora.

IEEE Spectrum 29 Oct 2023

Quanta energia gasta o treino de uma LLM?

Tamanho do Modelo	Consumo (KWh)
100M	10,000
1Bil	100,000
10Bil	1,000,000

Mar 4, 2023

NVIDIA® V100 Tensor Core: o GPU mais avançado de sempre para acelerar IA, HPC, Ciência de Dados e Gráficos

Tempo de Treino do GPT4 : Cerca de 6 meses
10,000 computadores V100 em 150 dias → 7,200 Mwatt-horas

Jul 25, 2003

GPT-3: “175 billion parameters.”

GPT-4: “**100 trillion parameters**”

Mar 6, 2023



“From Words to Watts”

Siddharth Samsi et al MIT

LLaMA - LLM da Meta AI

Em 2 gerações de GPUs

(NVIDIA V100 & A100)

Treino no equivalente a 3.072 A100



8 V100 GPUs cada com 32 GB de RAM ou 4 A100 GPUs
cada com 80GB necessários para qualquer **inferência** significativa de um 65B LLaMA model.

GPT-5



10,000 → 25,000 NVIDIA A100 GPUs with some H100s).
2X Tempo de treino (~3 meses → ~ 4-6 meses)



Por vezes chamada de a mais importante Companhia do Mundo produz, estima-se, 90% dos mais avançados chips semicondutores do mundo

“**MacroPolo think tank**” cerca de metade dos top AI researchers são da China e cerca de 18% dos U.S.

Investigadores IA trabalhando nos U.S. 31% são Americanos e 38% vieram da China
Nos U.S. estão 42% dos “world's top AI talents”

Japoneses do I. I. Riken, usam supercomputador **FUGAKU** Fujitsu+Riken para treinar o seu LLM. **CPUs** e não GPUs “now in **short supply** due to a fierce global language model development race.” *The Japan Times (May 11, 2024)*

Atenção

Emiratos Arabes Unidos ... “Not just soccer but also AI”

Problemas Éticos

Tendenciosismo : questões de raça, género, língua, cultura.

Associações abusivas:

Religião X → Violência

Etnia Y → Crime

Endinheirado → necessitado de cuidados médicos

Gemini é um chatbot que reescreve os prompts do utilizador tornando-os mais aceitáveis. Quando a OpenAI lançou o **DALL-E 3** o bot editava as perguntas tornando-as mais aceitáveis do ponto de vista da diversidade de género e raça

Segurança e Privacidade dos Dados

HIPAA - Health Insurance Portability and Accountability Act USA

GDPR – General Data Protection Regulation UE2018

Não basta Anonimizar.

Responsabilização: (ex Medicina)

Quem usa (Médico)? Quem permite o uso?
Quem Regulamenta (M.Saúde, Justiça, Parlamento)?
Quem vende e distribui (Empresas)? Quem desenvolve (Lab Investigação, Universidades)?
Quem Testa e Certifica?

Todos?

Muita Legislação deve ser produzida!!

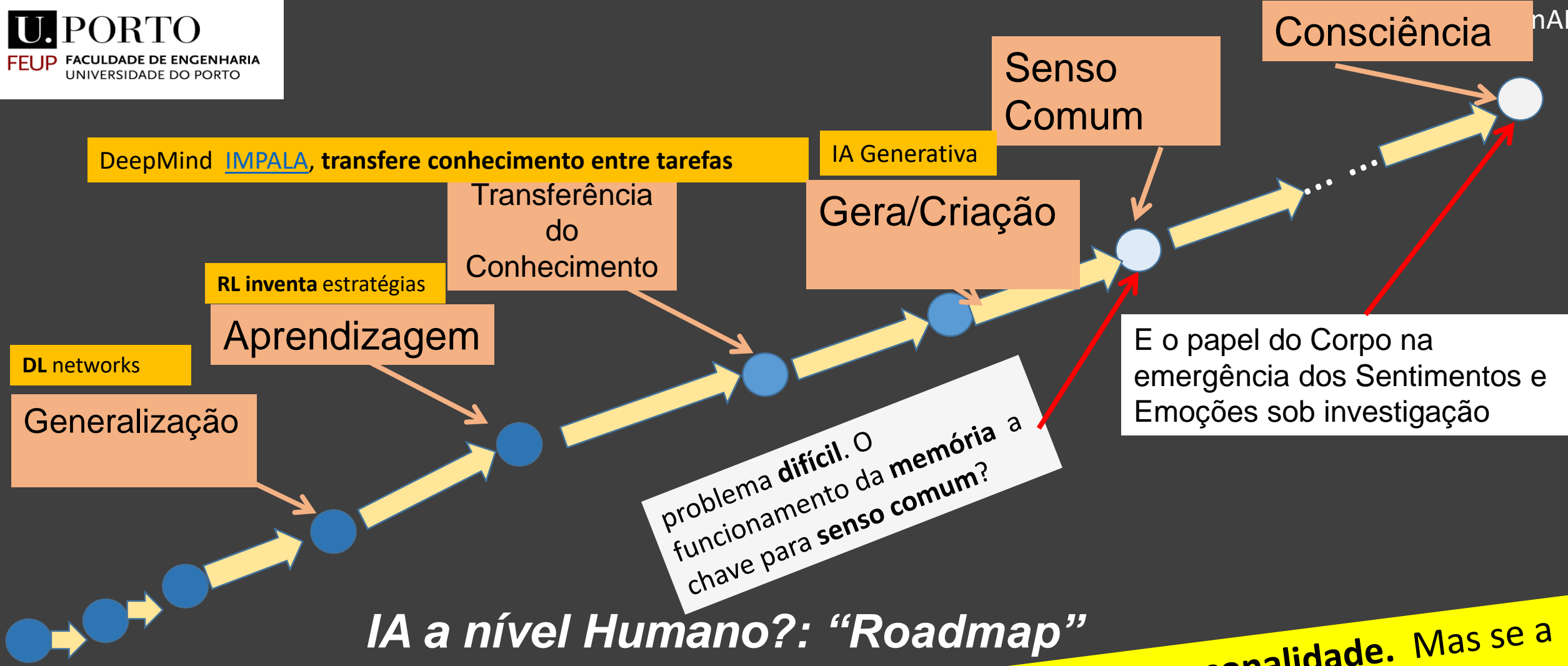
LLMs complexos → Dificilmente Auditáveis

Opacidade ao utilizador: Gestores necessitam do Racional das Decisões.

Propriedade dos Data Sets?

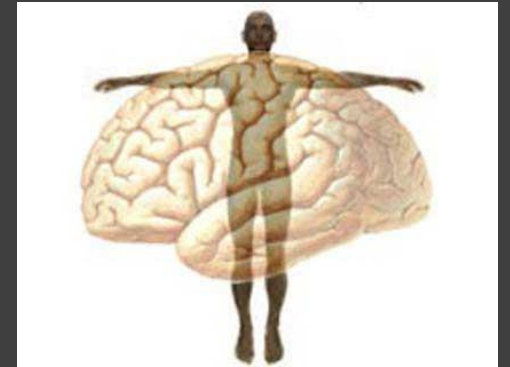
Impossibilidade de correr os LLMs localmente. A “cloud” é segura??

Diminuir os Modelos para os dominar ...



IA a nível Humano?: "Roadmap"

fácil desenvolver máquinas que **nos façam acreditar** ter "um self", ou **personalidade**. Mas se a tem realmente, só o saberemos depois de percebermos o que realmente é a consciência.



“Artificial General Intelligence” (AGI) → Sistemas autónomos com auto-compreensão e auto-controlo capazes de resolver vários problemas complexos em diferentes contextos e aprender a resolver novos problemas usando capacidades Cognitivas.

TEORIA

IAG como super cérebro computacional ?

OpenAI nasceu precisamente para responder ao desafio de chegar à IAG — uma IA capaz de resolver todos os problemas que os Humanos resolvem.

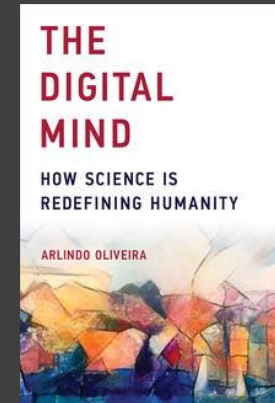


IAG levanta o problema da possibilidade da **Consciência**
Deep Blue e Alfa Go: **inteligência sem consciência**

Será a Consciência uma propriedade EMERGENTE da interação de muitas capacidades inteligentes??

“Um Filme mostra movimento, apesar de ...”

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



Poderá a Mente trabalhar como um Computador? E o contrário?

Mimar um cérebro “in Silicon” Versus “in vivo”. (??)
Transístores versus sinapses ...

Microprocessador avançado (2021): 50-60 billions de transistores

Estimativa do cérebro: 86 billions neurónios

Conhecer o funcionamento do cérebro crucial para desenvolver **superinteligência.**

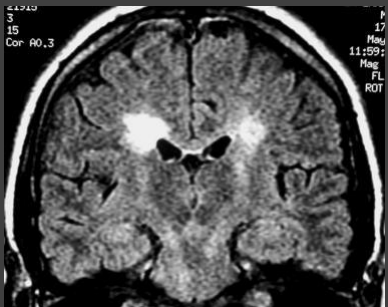


Harvard+Google: 5,000 fatias 1000 vezes + finas que um cabelo:
57,000 células, 150M conexões neuronais **The Guardian 13/5/ 2024**

MRI, ..., ainda superficial e (UHF MRI) difícil.

Resolução ~ 1mm³ máximo 0,5 mm³

Necessitamos de uma análise ao nível Nanométrico



Poderá a Mente trabalhar como um Computador? E o contrário?

Replicar a cognição não depende só da quantidade.
Arquiteturas computacionais emulando funcionalidade,
conectividade, adaptação e eficiência energética das redes
neuronais biológicas.

Muitos acreditam que para a **CONSCIÊNCIA** não são indispensáveis cérebros baseados em **Carbono** mas sim replicar a **computação** neles realizada

DEEP SOUTH: Neuromorphic Computing codifica não em binário mas em “picos” (Spikes) cujo tempo, amplitude e forma representam informação.
Parceiros da Western Sydney U.: Intel, Dell, Xenon

Submarinos não “nadam”; aviões não batem asas ...



Indústria do **pós-Vida** Digital

Cria reconstruções de pessoas falecidas baseado nos dados digitais existentes.

Microsoft, recusou o produto mas registou a **patente!** Chatbot usando imagens, voz, mensagens eletrónicas e tudo o que ficou registado da pessoa, e interagirá como se fosse essa pessoa.



PROBLEMÁTICO

Possibilidades Digitais Postmortem
\$
Cuidados em como gerir os nossos restos digitais ...
O que acontecerá com os nossos emails, textos, fotos,
palestras nos media sociais??
Quem os poderá reclamar ??

Devido à “Datafication” e à IA não morremos digitalmente.

Cuidado pois os próximos ChatGPT podem usar Avatars dando respostas que não seriam as nossas ... criar mortos-vivos digitais pode ser perigoso ... irá acontecer?



South China Morning Post: December 14, 2023
AI is being used by Chinese firms to create lifelike avatars of the deceased based on as little as 30 seconds of audiovisual material... "ghost bots"

IA MALÉFICA

OBJETIVO:

criação de uma MAI intencional através do programa



Militar : ciber-armas e soldados robôs para o domínio.

Governos: estabelecer hegemonia e controlar populações ou derrubar outros

Corporações construir monopólios destruindo competição ilegalmente.

Vilões tentando dominar o mundo pela IA.

Hackers roubando informação e recursos e destruindo plataformas.

Cultos religiosos tentando influenciar através de catástrofes

Psicopatas tentando entrar na história (megalomania)

Criminosos criando e escondendo-se com sistemas proxy

Negacionistas dos riscos da IA

IA MILITAR

Robot Dogs Join U.S. Air Force Exercise *CNN, Brad Lendon, September 9, 2020*

Exercício da U.S. Air Force : Robô-Cão
lançado por drone em área hostil com
sensors avançados



Relatório da **DARPA** aponta até 2050 para introdução de Soldados
Cyborgs nos USA (sem fadiga, comunicação instantânea, visão com
Zoom, ação psicológica ...)

Parte do U.S. military's Advanced Battle Management System (ABMS),
usando IA e Data Analytics para detetar “ameaças”

Iniciar uma corrida às armas baseadas em IA é muito **MÁ** ideia
**Autonomous Weapons: an Open Letter from AI & Robotics
Researchers** <http://tinyurl.com/awletter>

IA MALÉFICA/BENÉFICA

Afrontar o perigo sem medo

Ouvir uma Sinfonia na cabeça.

Visão superhumana

Curar cegueira, surdez ou doenças mentais [Elon Musk] prevê na Neuralink pelas interfaces cérebro computador (**implantes**)

Neuralink apresentou um robô “implantador” de milhares de eléctrodos ultra-finos no cérebro.

Implante da Neuralink em porcos: 1,000 canais de leitura de um numero similar de neurónios



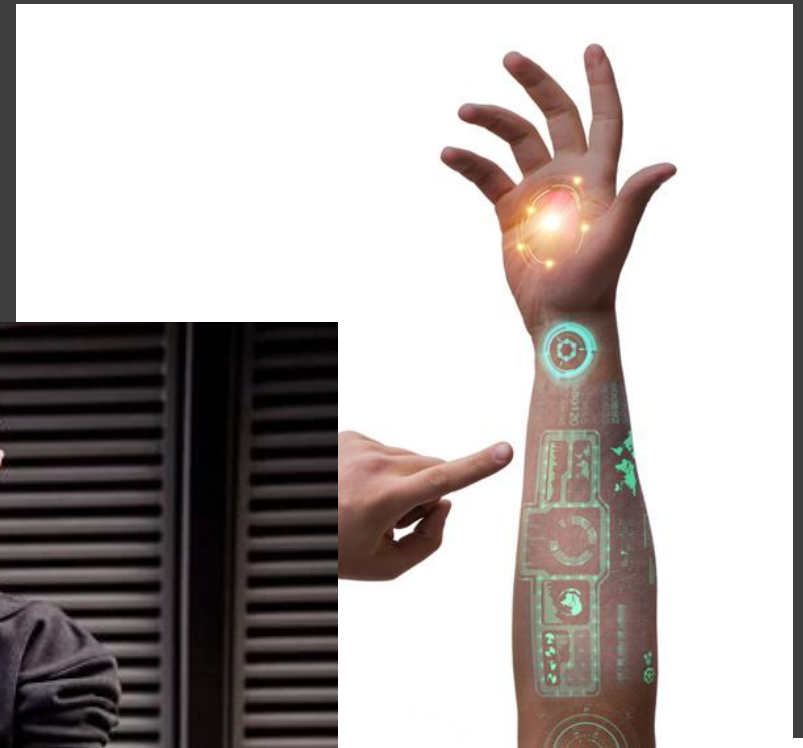
CYBORGS

- “cyborg” um ser futurista parte máquina parte humano
- *Manfred Clynes e Nathan S. Kline criaram o termo em 1960*
- Aumentar a manobrabilidade, a força e as sensações não é ficção científica.

“Ouvir” Cores com uma Antena.
Artista Neil Harbisson daltónico de
nascença



Boy Born with One Hand Gets Spiderman
Bionic Arm eat, open a bag, and pick up fruit
with both hands.



Ética Para uma **IA Benéfica**




OBJETIVO:

Maximizar os Benefícios da IA enquanto se Minimizam os seus Riscos.



Abordagem centrada-nos-Humanos (human-centric**) é necessária!!**



**Atenção a grupos vulneráveis: Crianças, idosos, pacientes, deficientes e minorias.
Atenção às assimetrias no poder ou na informação: entre Empregados- Empregadores;
Produtores- Consumidores ...**



Para uma IA Benéfica.

17 Objetivos do Desenvolvimento Sustentável

Desenvolvimento sustentável é o que satisfaz as necessidades do presente sem afetar as condições das gerações futuras

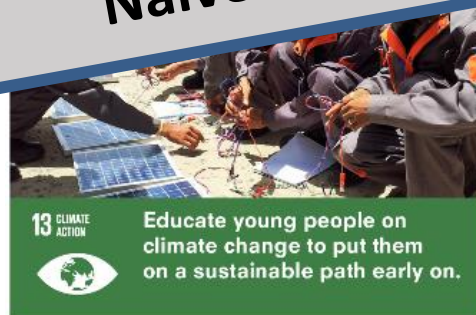
ÁGUA



ENERGIA



Três dimensões do Bem-estar:
Ambiente, Sociedade, Economia
Naïve esperar situações 'win-win-win'



ODS13

POLUIÇÃO



ODS14



ODS15

Para uma IA Benéfica.

OBJETIVO Planetário:

IA para a sustentabilidade Ambiental (ÁGUA, TERRA, AR)



Éticamente compensador

SMART *: Cities, Homes, Factories

WEF estudou mais de **80** possíveis usos de **IA** em favor do Ambiente:
proteção Climática, mobilidade autónoma, redes inteligentes de energia.
modelação climática ...

Q: Algoritmos de IA emergentes tornarão melhor a nossa vida ou não ?

➤ Preocupações:

- **Delegação** dos humanos: perda de **controlo** das respetivas vidas
- **Abuso** dos (nossos) Dados: para o lucro ou para o poder
- **Perda** de Empregos: Perda massiva e ganhos menores para novos perfis
- Aumento da **dependência**; as **capacidades** humanas em vez de aumentarem, diminuem
- Caos/desordem: **Cibercrime** em toda a economia;
- **Armas** autónomas dos militares; informação deturpada para a guerra

Q: Algoritmos de IA emergentes tornarão melhor a nossa vida ou não ?

➤ Sugestões:

- **Cooperação:** internacional. Controlo humano das redes digitais
- **Valores:** Responsabilidade social e ética da tecnologia; **Regulamentação** (ex. Lei na Califórnia, 2019, proibindo a imitação enganosa de humanos (*impersonate*))
- Prioridade às **Pessoas:** Alteração da abordagem de desenvolver e aplicar a IA na corrida com os robôs para ajudar os humanos
- **Comité de Ética para validar o desenvolvimento e aplicação da IA**
- **Legislação**
- **Inclusão de Restrições e Princípios Éticos no Software de IA**

Preocupações críticas levantadas pela IA

- 1. Privacidade: Identificação** individual sem consentimento
(reconhecimento de faces etc; ...)
- 2. Camuflagem** de sistemas de IA *(robôs humanoides/androides tem de ser bem identificados)*
- 3. Pontuação (scoring)** normativa e generalizada de cidadãos
(apesar do ensino, condução ...)
- 4. LAWS-** Sistemas Autónomos de Armas Letais *(seleção e ataque de alvos)*
- 5. Inteligência Artificial Geral e Consciente** *(a longo prazo)*

Regulamentação

União Europeia produziu em Maio de 2018 o Regulamento **General Data Protection** impondo novas responsabilidades às **Organizações** que processam dados dos cidadãos da UE para a investigação científica





INDEPENDENT
HIGH-LEVEL EXPERT GROUP ON
ARTIFICIAL INTELLIGENCE
SET UP BY THE EUROPEAN COMMISSION



ETHICS GUIDELINES
FOR TRUSTWORTHY AI

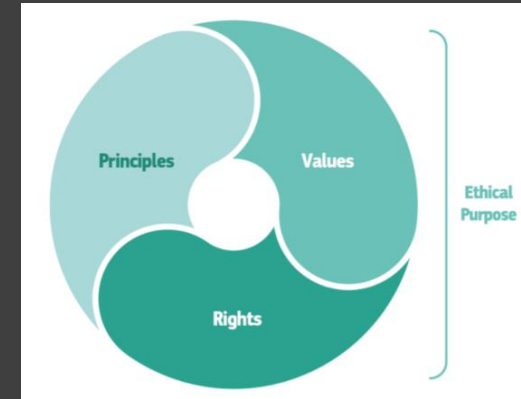
High-Level Expert Group on Artificial Intelligence **Ethics Guidelines for Trustworthy AI**

European Commission
Directorate-General for
Communication

European Commission
B-1049 Brussels

Documento publicado em **08
Abril 2019.**

Princípios Éticos no contexto da IA e dos valores correlacionados



➤ Princípio da Beneficência (“DO GOOD”)

- Geração de prosperidade e criação de **valor**,
- Na direção de sociedade **justa** e pacífica com igualdade de oportunidades;

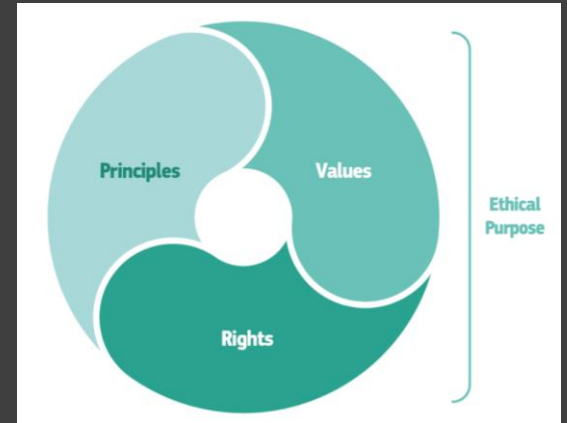
➤ Princípio da NÃO Maleficência (“DO NO HARM”)

- Garantir a dignidade, integridade, **liberdade, segurança no trabalho** e na sociedade

➤ Princípio da Autonomia

- Preservar a liberdade, **não subordinação a ou coerção por Sistemas de IA**
- Completa **autodeterminação**

Princípios Éticos no contexto da IA e dos valores correlacionados



➤ Princípio da Justiça (“BE FAIR”)

- **Imparcialidade** para os indivíduos e grupos incluindo minorias,

➤ Princípio da Explicabilidade

- Garantir a Transparência da tecnologia
- Sistemas auditáveis, compreensíveis e **inteligíveis** aos humanos

Parlamento Europeu: 13 Março, 523 votos favor, 46 contra, 49 abstenções

O AI Act inclui várias garantias para o uso quer da GenAI, quer da General-purpose AI: **limitações** nos sistemas de **identificação biométrica**, banimento do **social scoring** e da IA para manipular ou **explorar vulnerabilidades do utilizador ...**

UK: National Cyber Security Centre Guidelines

criado com o Setor Privado e os USA. Muito permissivo !!



What the **US** can learn: Maria Villegas Bravo, EPIC **Law** Fellow

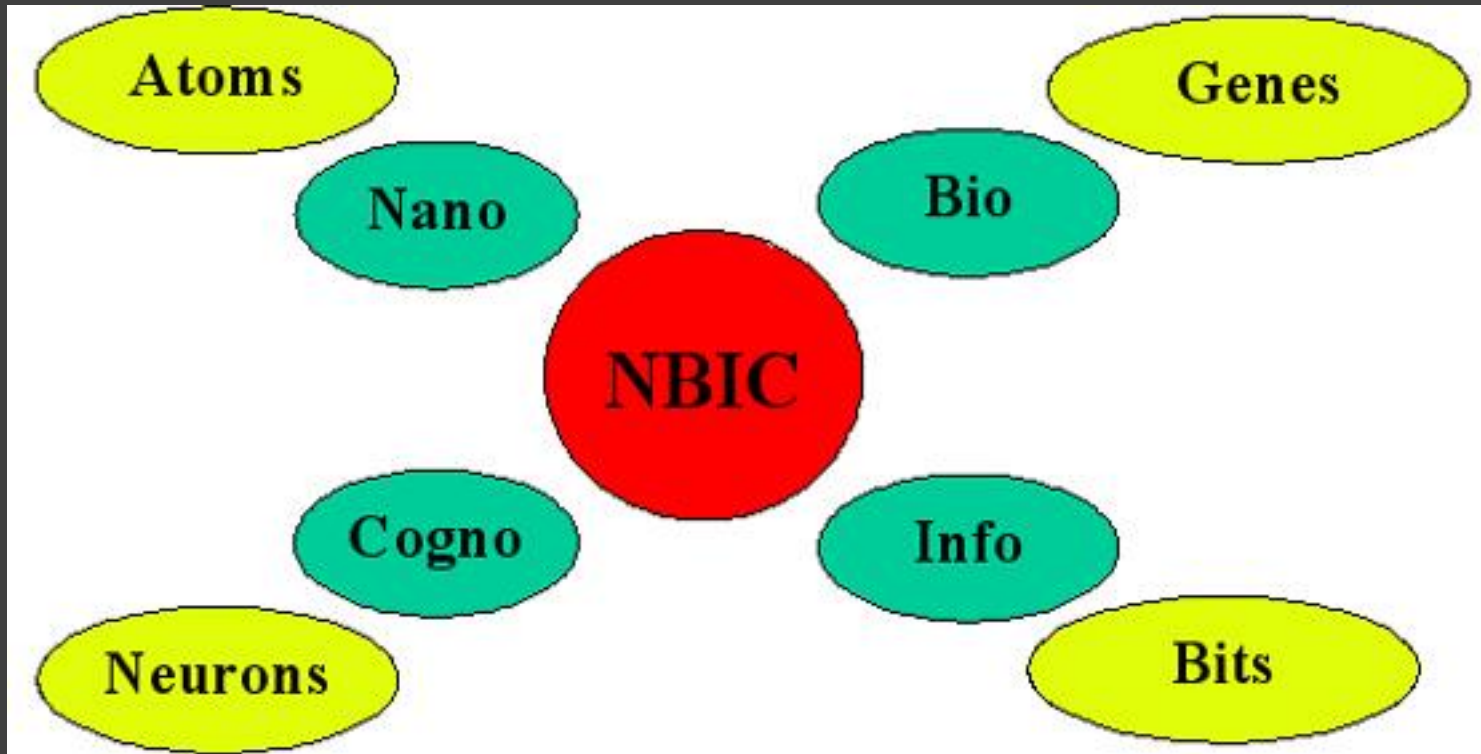
Forças e Fraquezas do AI Act categorizadas em "**The Good, The Bad, and The Ugly**". "**The Good**": Proibir Aplicações intrusivas; reconhecimento de ameaças aos direitos fundamentais; "**The Bad**": Regras aplicáveis a "general-purpose AI" e "open-source software"; "**The Ugly**" restrições na identificação biométrica.



A UN draft resolution promove o uso da IA no desenvolvimento sustentável e na proteção dos Direitos Humanos
Votado por mais de 120 estados membros
A Ass. Geral da ONU reconheceu o potencial acelerador do progresso para os 17 Sustainable Development Goals

2 meses após o Secretário Geral A.G. ter acusado as grandes companhias HiTec de prosseguirem os seus interesses estereitos com a IA sem olhar aos indivíduos, sua privacidade e direitos

Futuro: Sinergias entre as Ciências



- **Recomendações**

+ ART em ARTificial Intelligence:

Accountability, Responsibility, Transparency

A: a **quem** nos devemos dirigir se (por exemplo) um automóvel autoconduzido atropelar um peão?

R: **Responsabilidade** de os **fornecedores** dos sistemas tornarem claro e compreensível o **Porquê** das decisões tomadas de forma autónoma.

T: especificação, **desenvolvimento** e reprodutibilidade dos sistemas de IA. Tal implica a **compreensão** de Como funcionam e, possível inclusão de **RESTRIÇÕES ÉTICAS**

“Human in the Loop”



➤ Tecnologias transformativas: Energia a Vapor, Eletricidade, Automóvel, Transistor, Internet
Excitação inicial, Hype

- tra
Hu
- sociedades recompõe-se das **revoluções económico/tecnológicas** MAS muitas pessoas concretas podem ser trituradas no processo
 - **DEVER:** não permitir que tal aconteça

Legislação sobre ética no desenvolvimento, disponibilização e uso da IA

Mais **Ciência** na IA para incluir nos **programas** restrições Éticas

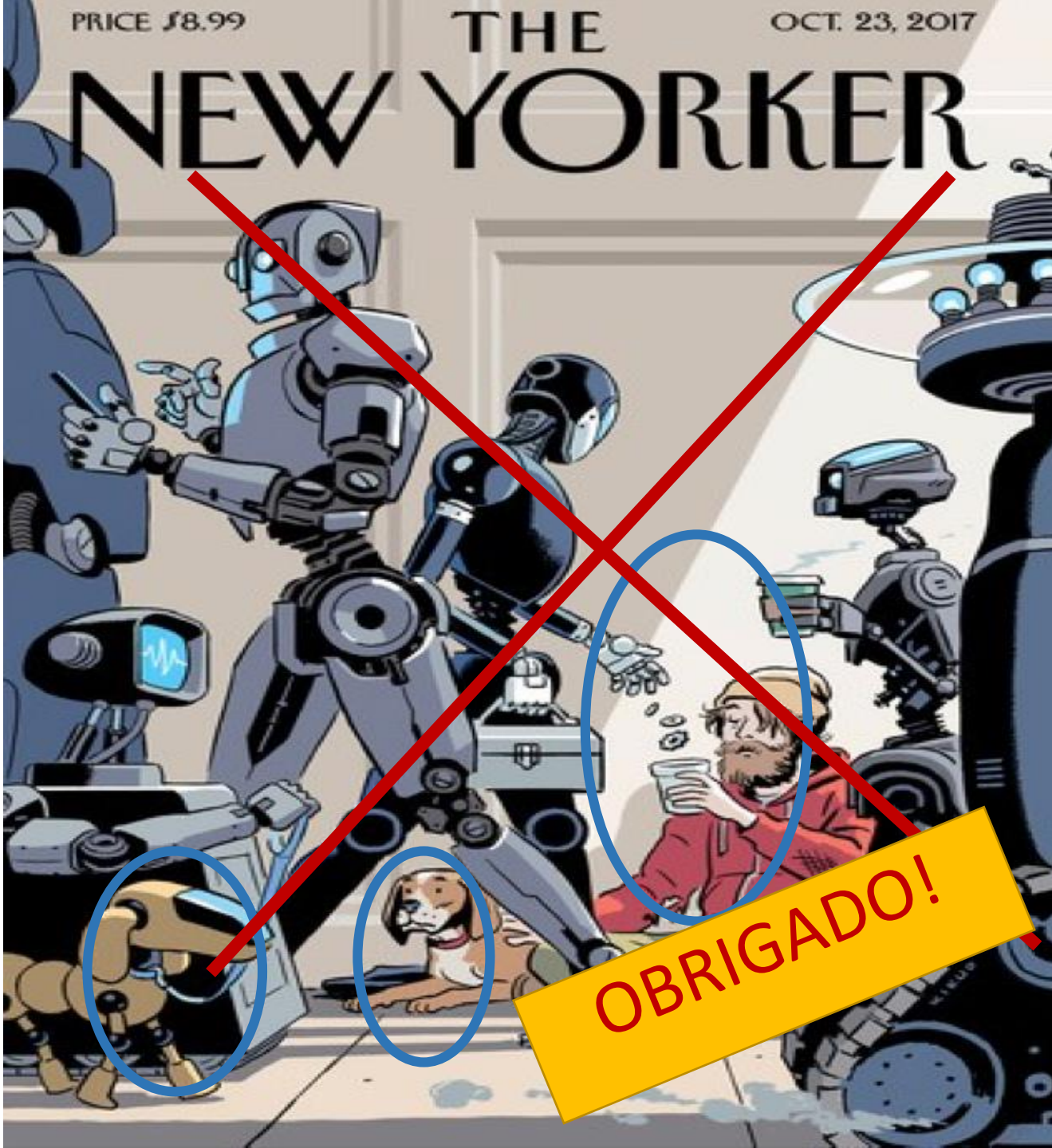
➤ **Inteligência Artificial** é um trabalho da Inteligência Humana permitindo **AVANÇOS** civilizacionais **IMPORTANTES**. MAS ...

Todos se devem interessar civicamente pelo futuro
que se está a preparar **AGORA**

PRICE \$8.99

OCT. 23, 2017

THE NEW YORKER



OBRIGADO!