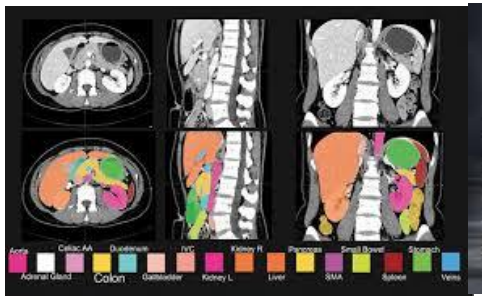
A INTELIGÊNCIA ARTIFICIAL e o seu Impacto

Eugénio Oliveira

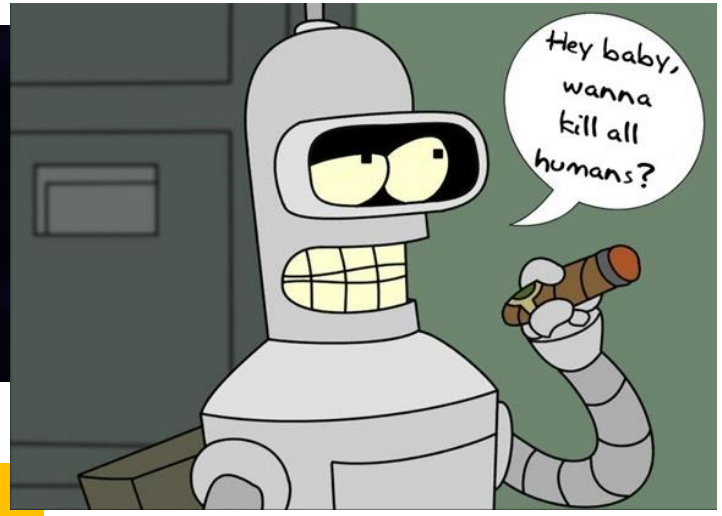
"In an **A.I.-first world**, we are rethinking all our products," CEO Google."
The New York Times, Maio 18, 2017

Realizações



O que é a IA?

Promessas e Perigos da IA



SUPER Inteligência Artificial (IAG)?

VERÁ

Para uma **IA Benéfica**

Eugénio Oliveira

IA- Inteligência Artificial

Percecionar o ambiente acessível

Coligir e **interpretar** dados estruturados ou não

Derivar **conhecimento** dos dados

Aplicar **raciocínio** sobre esse conhecimento

Decidir tomar o que consideram ser a melhor Ação!

Pode **aprender a adaptar** o seu comportamento

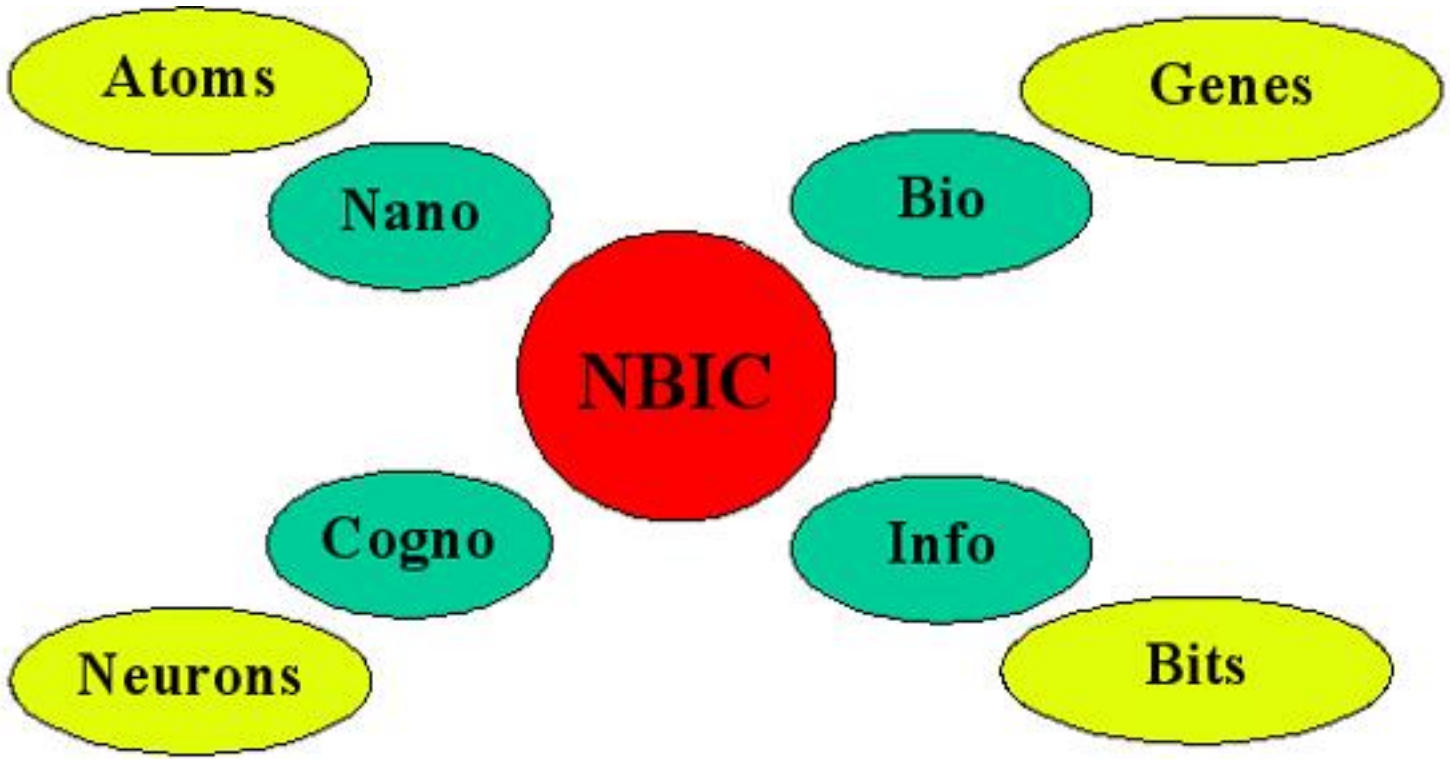
Criatividade?

Nada mais NATURAL que a INTELIGÊNCIA ARTIFICIAL

Homem NATURALMENTE criou a IA !!

Quanto à IN ainda há quem tenha dúvidas …

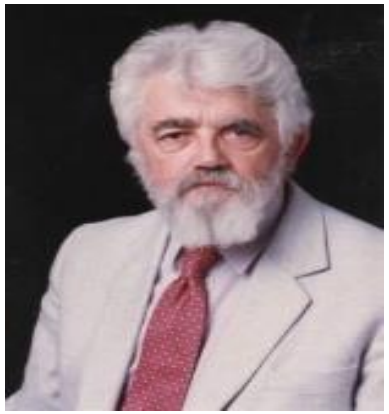
Futuro: Sinergias entre as Ciências

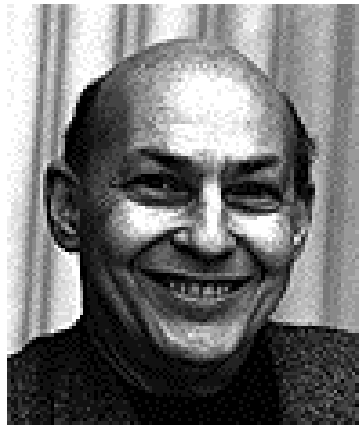


- **Nascimento da IA Computacional**

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

o “motto”





Conjetura:
Inteligência poderá em princípio ser descrita com tanta precisão que uma máquina a poderá simular.

*- Dartmouth AI Project Proposal; **J. McCarthy** et al.; Aug. 31, 19**55**.”*

O que há de novo na IA ???

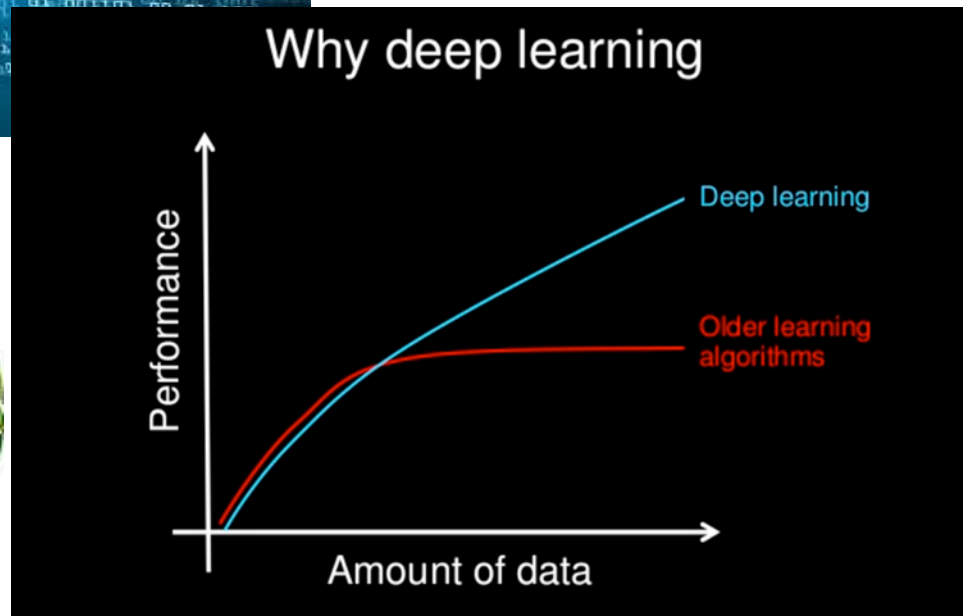
- ✓ Grande poder computacional do **HARDWARE**

  supercomputador, **Summit**, no Oak Ridge National Lab

  area: 2 “courts” de tennis, >27,000 GPUs.

  Aplica **IA**/ Deep Learning à previsão de alterações climáticas

- ✓ **BIG DATA (Petabytes 10 ^{15}....)**



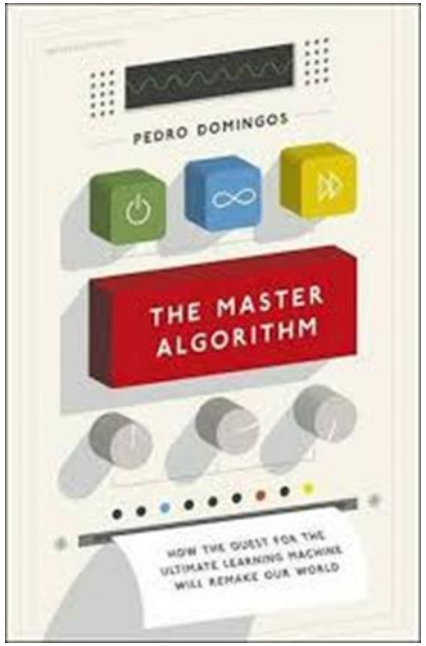
- ✓ **DEEP LEARNING**

Redes Neuronais Artificiais

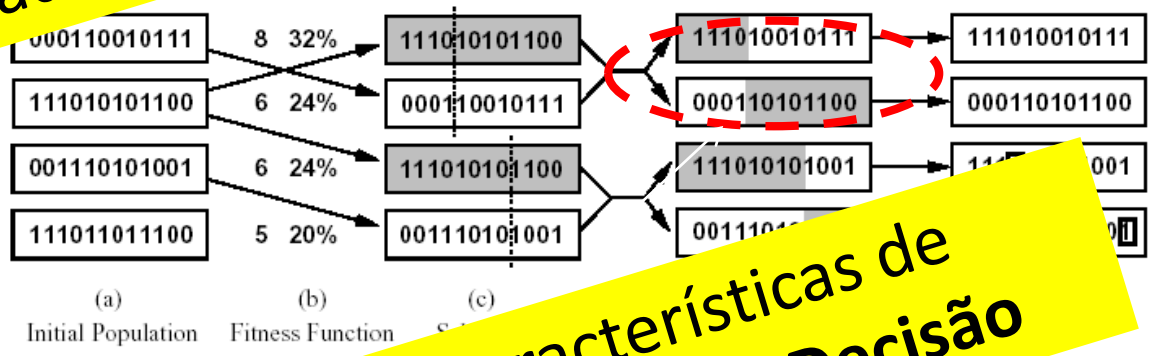
Eugénio Oliveira

Representação do Conhecimento
Processamento do Conhecimento
Aquisição do Conhecimento

Lógicas (Predicados, Temporal, Modal, Intencional…)



- **Simbólicos**
- **Conexionistas**
- **Evolucionários**
- **Estatisticos (Bayesianos)**
- **"Analogizadores"**



Uso de medidas da **entropia** das características de entidades para as classificar em **Árvores de Decisão**

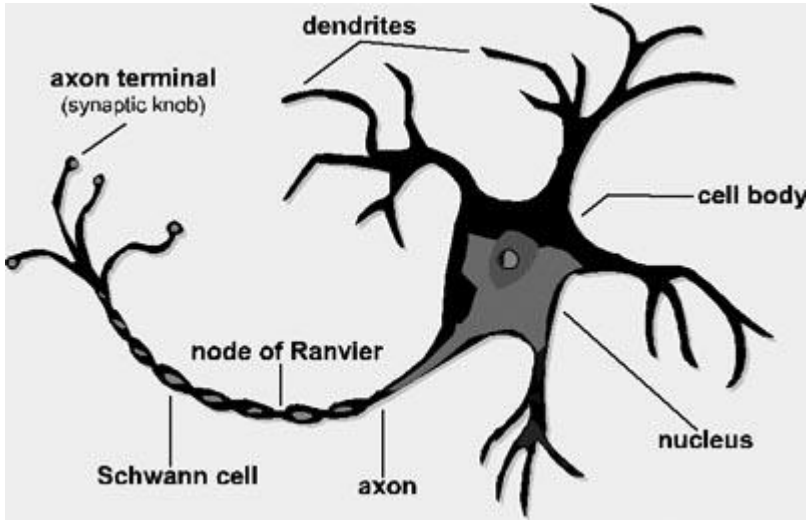
The Master Algorithm: The ultimate Learning Machine that will remake our world

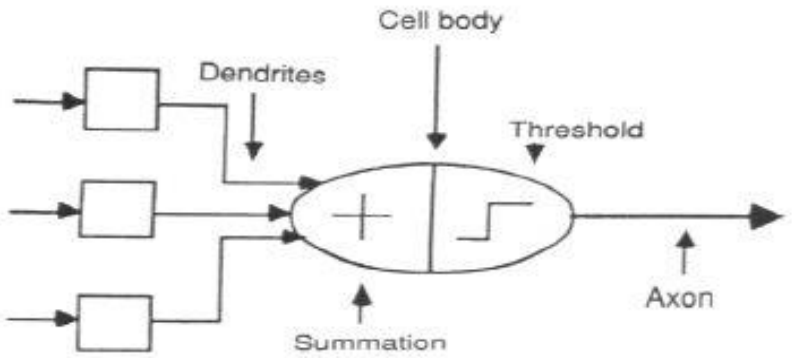
(P.Domingos, Basic Books, 2015)

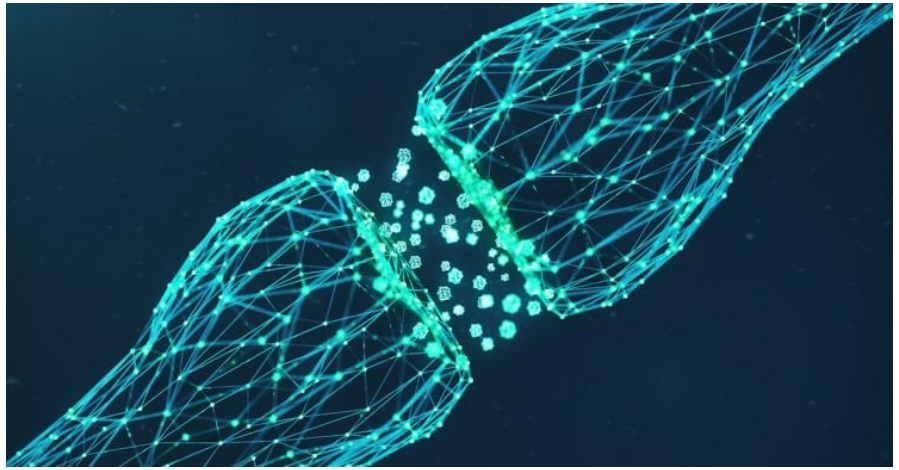
Eugénio Oliveira

Escolas diferentes, objetivos semelhantes:
Desenvolver "Máquinas Inteligentes" !

Analogia Biológica ...





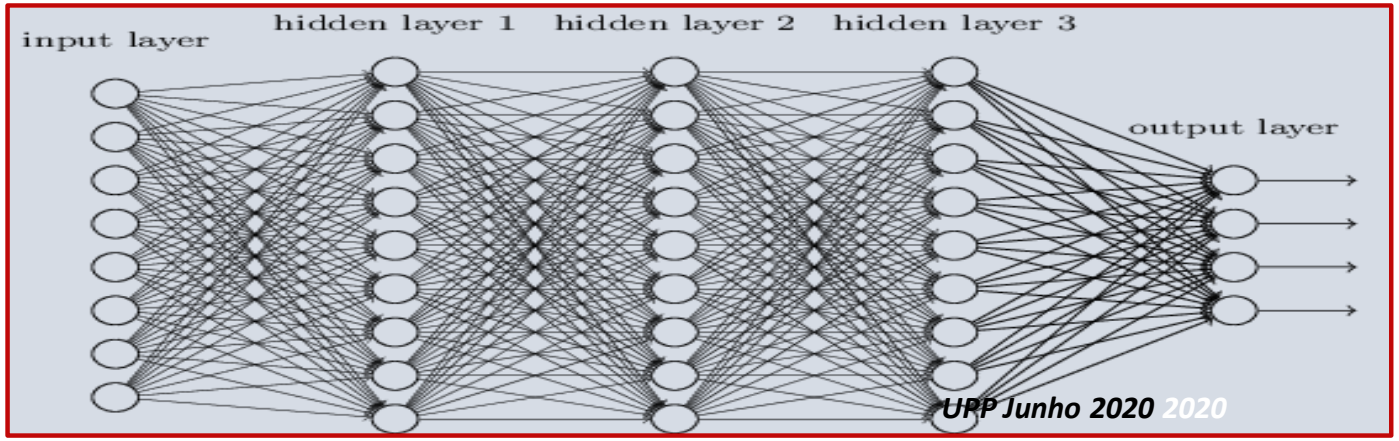


Função de Combinação

Função deTransferência

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

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



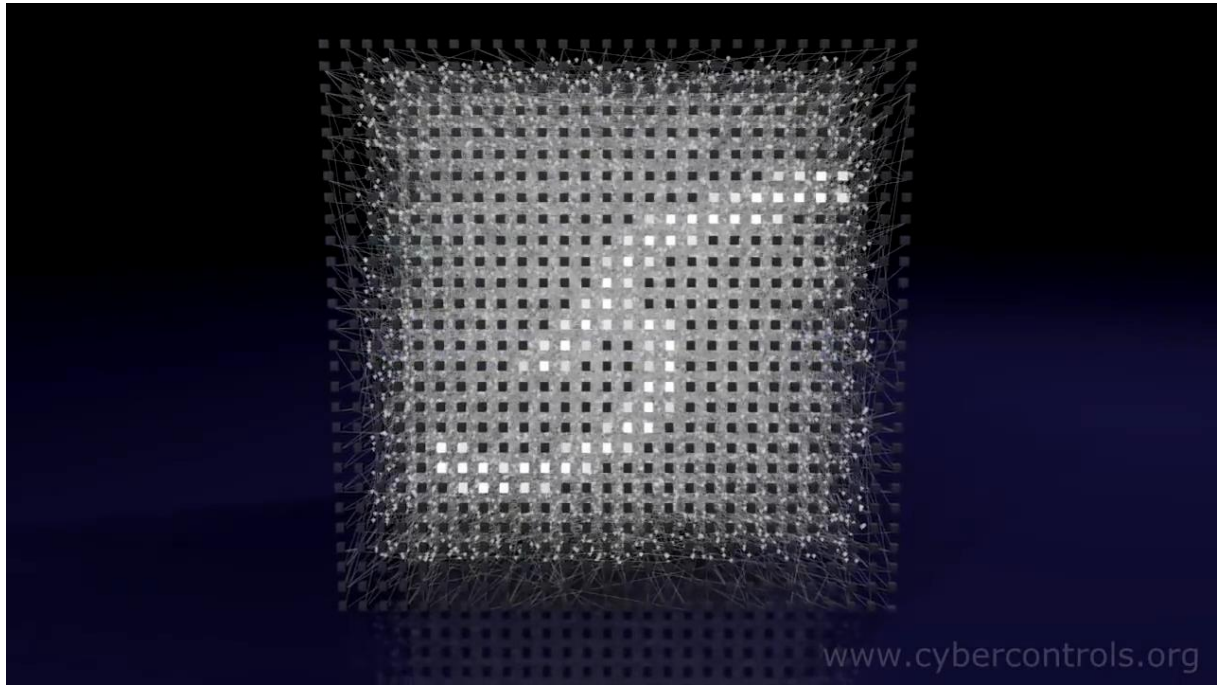
Analogia Biológica ...

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

- 10^{15} sinápses*10 picos/segundo→10^{16} operações/s
~25 watts → muito eficiente

- RNA produzem ainda cerca de 10 milhões de sinápses a menos que o cérebo humano

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

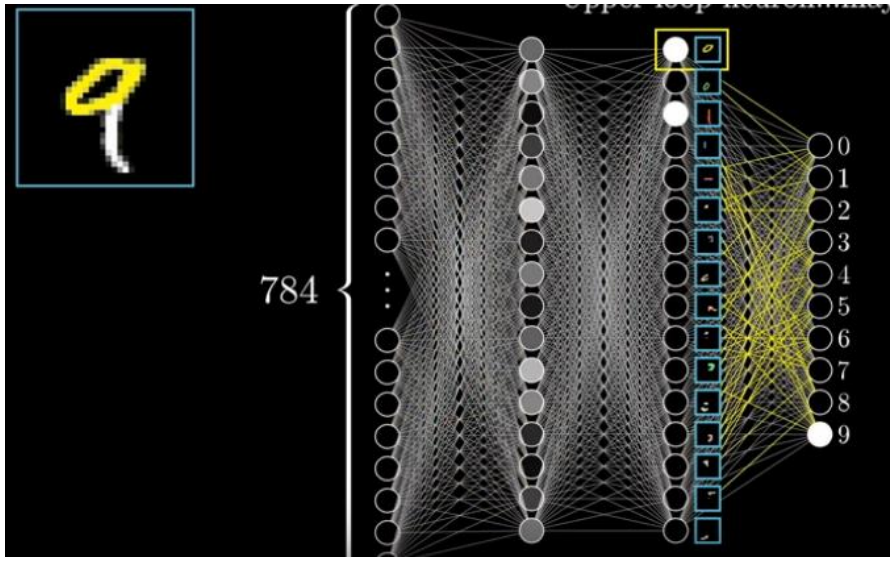


DEEP NN: CNN e RNN

CNN: aplica filtros e operações de convolução

Camadas de diferentes tipos

RNN comprime as ANN mas há loops temporais que cada camada se auto-alimenta



Principais **aplicações**: Voz, Texto, Imagens

IA: Realizações

- Em 2006 R. Brooks opôs-se à ideia que a IA falhara e avisou que a IA estaria à nossa volta a toda a hora nos anos vindouros. Certo!!

Aplicações por todo o lado e a toda a hora**:**

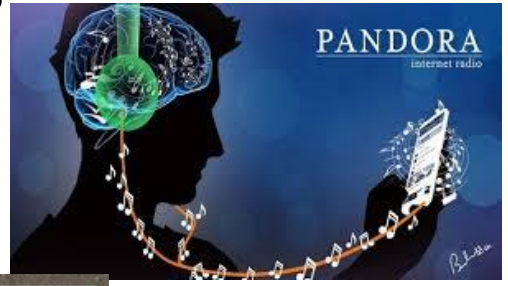
- **Facebook** e **Google** decidem o que mostrar a cada utilizador
- Empresas conhecem o **perfil** dos seus clientes
- Gestão inteligente de **satélites**
- **Diagnósticos** médicos com precisão

Óbvio

- **Sistemas de Recomendação** da Amazon e Netflix

IA: Realizações

- PANDORA **regista** da radio as nossas **músicas** preferidas

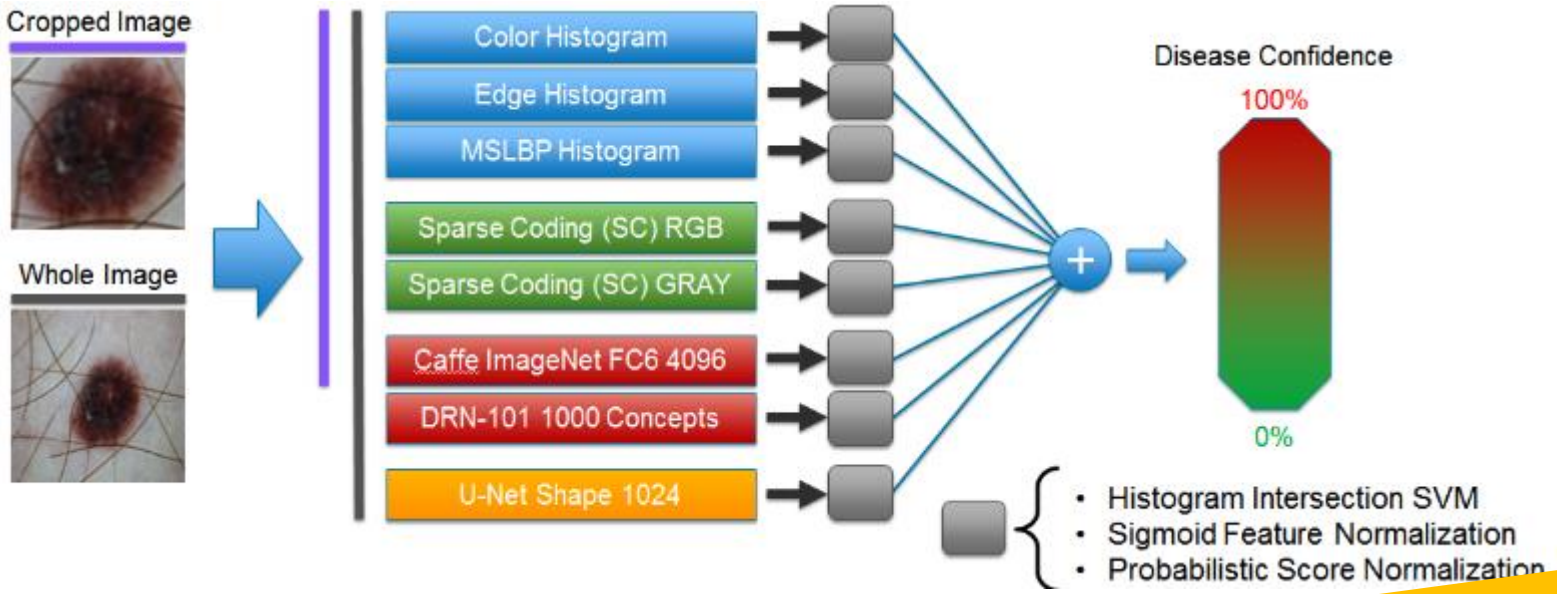


- BingTravel **prevê preços de vôos** futuros

- **Previsões:** **Será que TU** cometerás um crime?
- **TU** serás bom naquele emprego?

IA: Realizações

- **Agentes** de software representam investidores na **Bolsa**
- Manchas na pele podem ser reconhecidas automaticamente como carcinogénicas

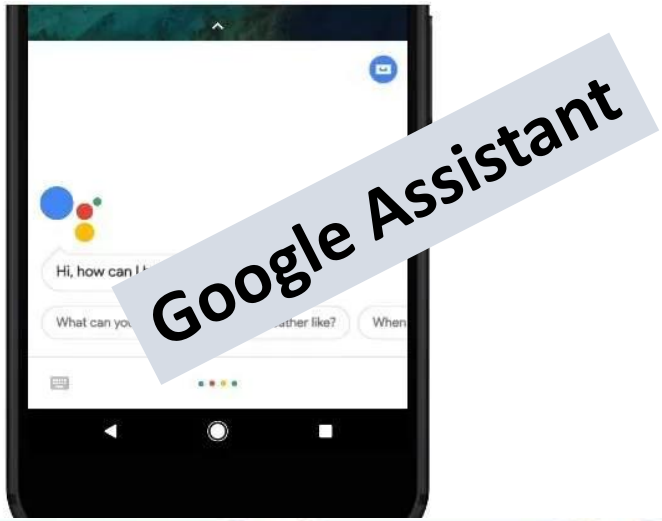


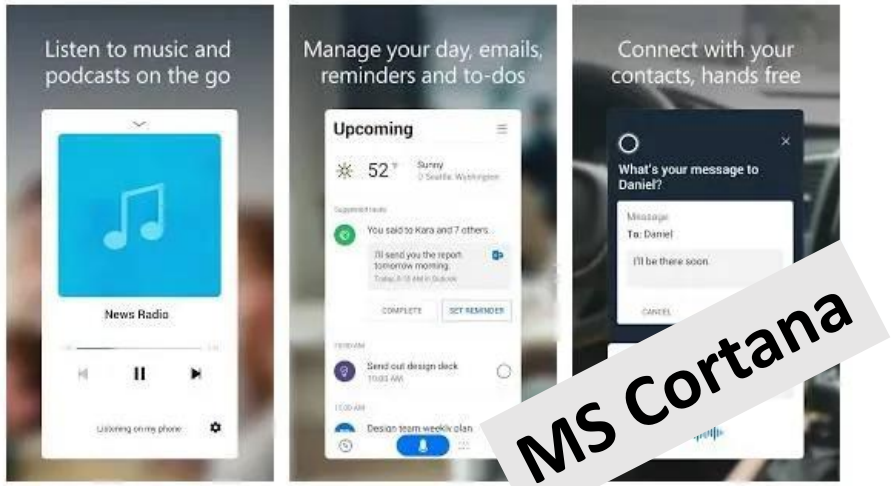
- INRIX ajuda a selecionar o **melhor precurso** quando há problemas de trânsito

Veículos Atónomos
Domótica
Robótica
Educação à distância / Tutores Inteligents
Cibersegurança (PayPal usa DL. Evita 50% de falsos

IA: Realizações

➢ Assistentes com IA:



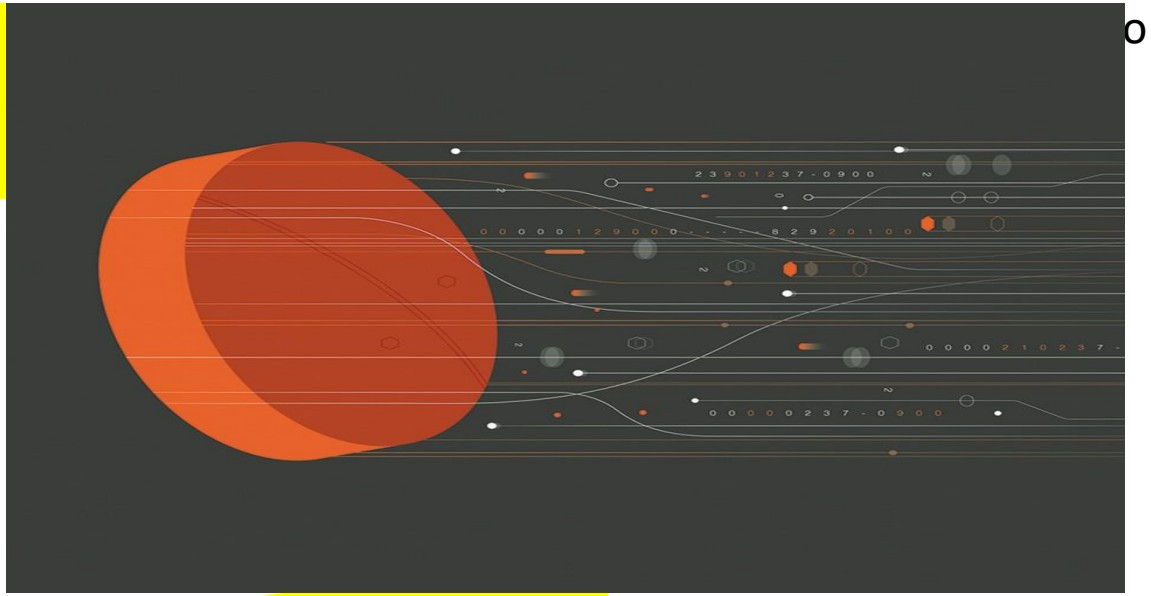


IA: Realizações

Saúde: 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

Saúde: Google- Alphabet: deteção de tumores em imagens com **DL**:
Taxa de acerto passou de **73%** (humanos) para **89% (IA)**
https://edgy.app/google-ai-cancer-diagnosis

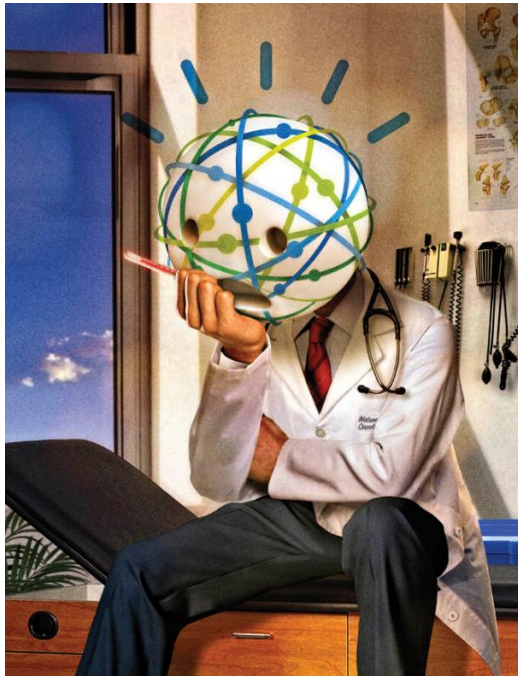
IA: Realizações



Saúde: BioMedicina.
Algoritmos de "**Deep-learning**" analizarão milhões de compostos e selecionão os melhores para um dado objetivo tetrapeutico desenvolvendo **novos medicamentos**

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

IA: Realizações



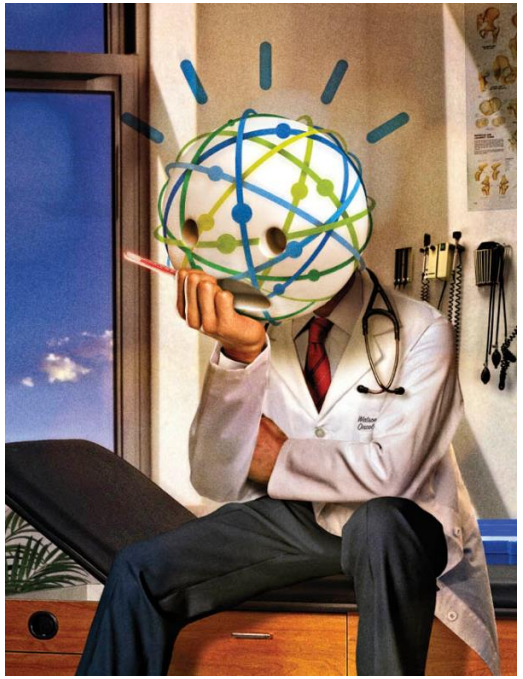
"How IBM Watson **Overpromised** and **Underdelivered** on AI Health Care", *Eliza Strickland, IEEE Spectrum*

O potencial dos sistemas de IA foi demonstrado em experiências cuidadosamente **controladas**.

Apenas uns **poucos** sistemas baseados em IA foram **aprovados** pelos reguladores para uso em ambientes reais em hospitais e clínicas.

desencontro entre as **promessas** dos sistemas de "Machine Learning" e a **realidade** clínica

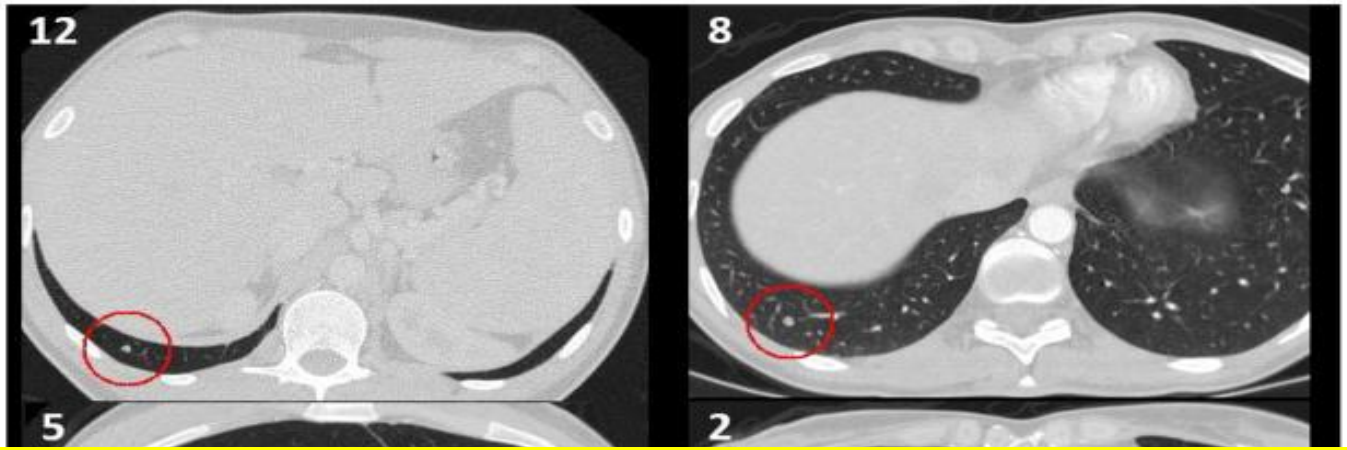
IA: Realizações



Yoshua Benjo, U.de Montreal

- "NLP não precisa de ter a compreensão total para fazer coisas muito úteis. Mas ainda não alcança a compreensão do médico."

"No, we're r

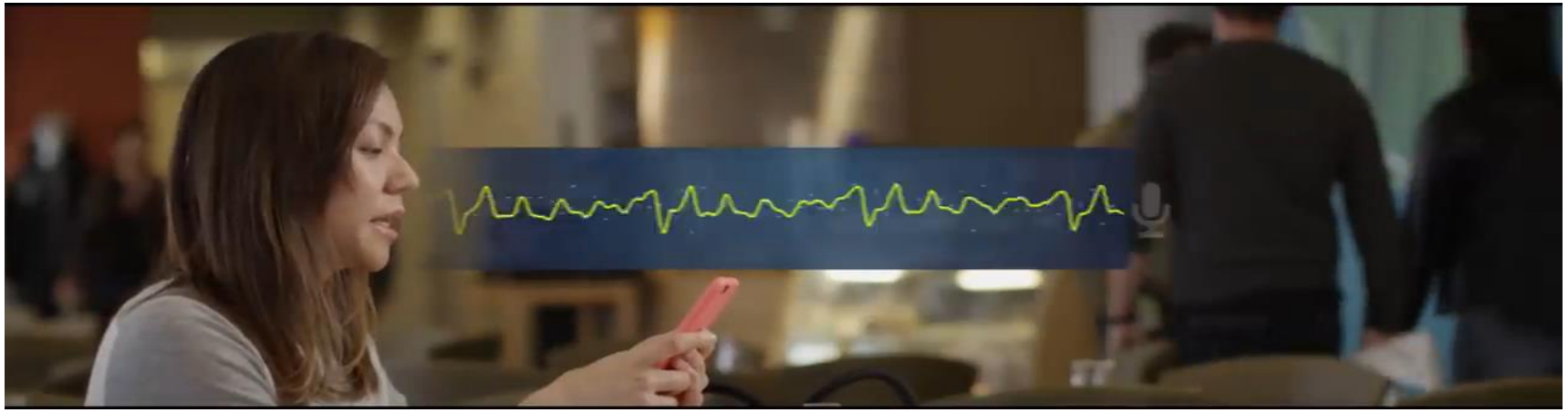


- "It's a difficu ge. **But we're doing it,"** says Joh solutions and IBM research.

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

IA: Realizações

Deep Learning para reconhecimento de voz



Tradução Automática

Inglês→ Mandarim e Mandarim → Inglês

IA: Realizações

https://youtu.be/-d4Uj9ViP9o

IBM **Debater** argumentou a favor do tema "*o Governo subsidiar o Ensino pré-escolar*". Humano profissional de debates Harish Natarajan argumentou contra. (02/2019)

BOM

Debater (NLP) elegeu os 10 artigos mais relevantes de entre 4 milhões.

identificou 3,000 frases mais significativas.

Analisou a sua **polaridade de opinião**.

Construíu um discurso justificativo.

Insuficiente …

IA é muito melhor…

Deep Blue

e **AlphaGo**

- Algoritmos sofisticados (**Algorítmo MCTS- "Monte Carlo Tree Search"**)
- Aprendizagem Automática DNN

IA: Impacto

OpenAI

Novembro, 2019

Better Language Models and Their Implications

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

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

- **Escrita** Automática de Texto por um Assistente de IA
- **Tradução** Automática …

Problemático.
Inicialmente retido…
Já disponibilizado

OU para propósitos **maliciosos** incluindo:

- **Escrita** Automática de Notícias Falsas **"credíveis".**
- **"Personificar" Humanos online**

(CTEC) found that extremist groups can use GPT-2 for misuse, specifically by fine-tuning GPT-2 models on four ideological positions: **white supremacy, Marxism, jihadist Islamism, and anarchism**. CTEC demonstrated that it's possible to create models that can generate synthetic propaganda for these ideologies.

IA: Impacto

É urgente conhecer os impactos atuais da IA

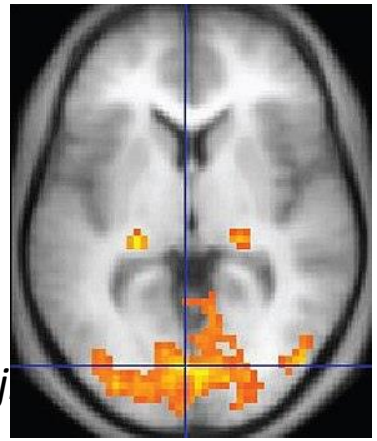
Sistemas de IA já fazem **julgamentos problemáticos** com impacto cultural e social na perceção individual

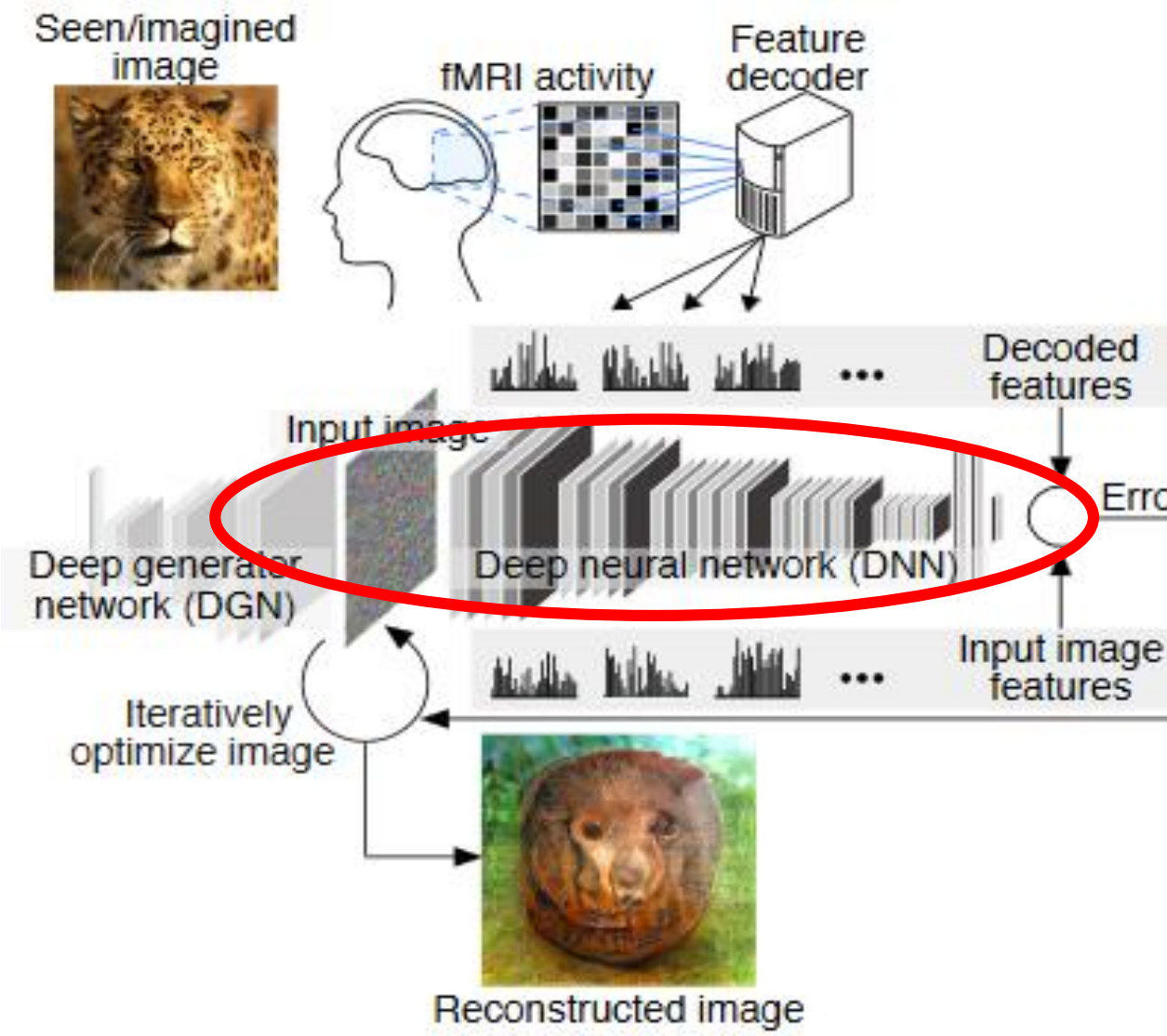
fotografia icónica do horror indiscriminado no Vietnam, **banida automaticamente** do Youtube e Facebook por **programas** que as consideram imorais.

IA: Impacto

Detetar (alguns) pensamentos/imagens ….

"Reconstrução de imagens através da atividade cerebral"





Guohua Shen, Tomoyasu Horikawa, Kei Maj
Kamitani

ATR Computational Neuroscience Laboratories,

Kyoto University, Japan

preprint colocado online em **Dec. 28, 2017**

Eugénio Oliveira

IA: Impacto

Super Inteligência? **Inteligência Artificial Geral**? IA Forte?

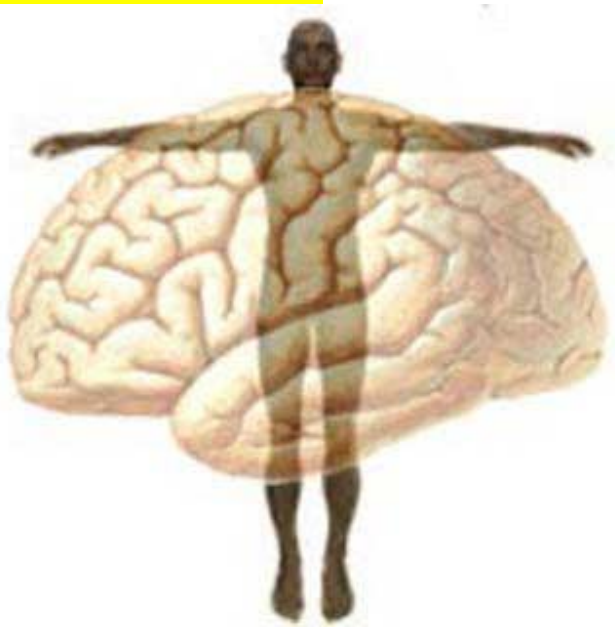
Distração?

Hyping IA

- *"… and thanks to ML, which lets computers predict the future based on past experience the **world will guess what you want before you even want it and will be ready for you"*** (P.Domingos TEDxLA 2016).

ameaçador?

- **Inteligência Artificial Geral**

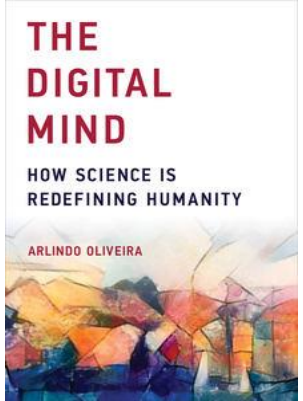


Será a Consciência uma propriedade EMERGENTE da interação de muitas capacidades inteligentes??

Se emerge de leis mais simples, então a hipótese de simulação poderia ser verdade

"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*



• Inteligência Artificial Geral

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

Mimar um cérebro "in Silicon" Versus "in vivo"
Transístores versus sinapses …

integração de transistors **ainda/já** a alguns anos da densidade neuronal

Conhecer o funcionamento do cérebro
crucial para desenvolver **superinteligência**.
NeoCortex: Células não especializadas

Submarinos não "nadam"; aviões não batem asas …

MRI ainda superficial e difícil.
Resolução ~ $1mm^3$~250 milhões de neurónios, …
Necessitamos de uma análise ao nível Nanométrico

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

Hipótese da Simulação

Nick Bostrom filósofo sueco da University of Oxford.
Fudador e Director do Future of Humanity Institute da Oxford University

Simulation Hypothesis:
Sociedades com **Maturidade Tecnológica** podem simular cérebros, conscienci
ambientes simplificados mas c

Vivemos numa Simulação!

Simulation ARGUMENT:
1- *Todas as civilizações existententes se* ***extinguirão*** *antes de atinjir a Maturidade Tecnológica que permita a* ***Hipótese da Simulação***

2- Sociedades com Maturidade Tecnológica **não tem interesse** na Simulação dos antepassados

3- Já **vivemos numa Simulação** dos antepassados produzida por superinteligências

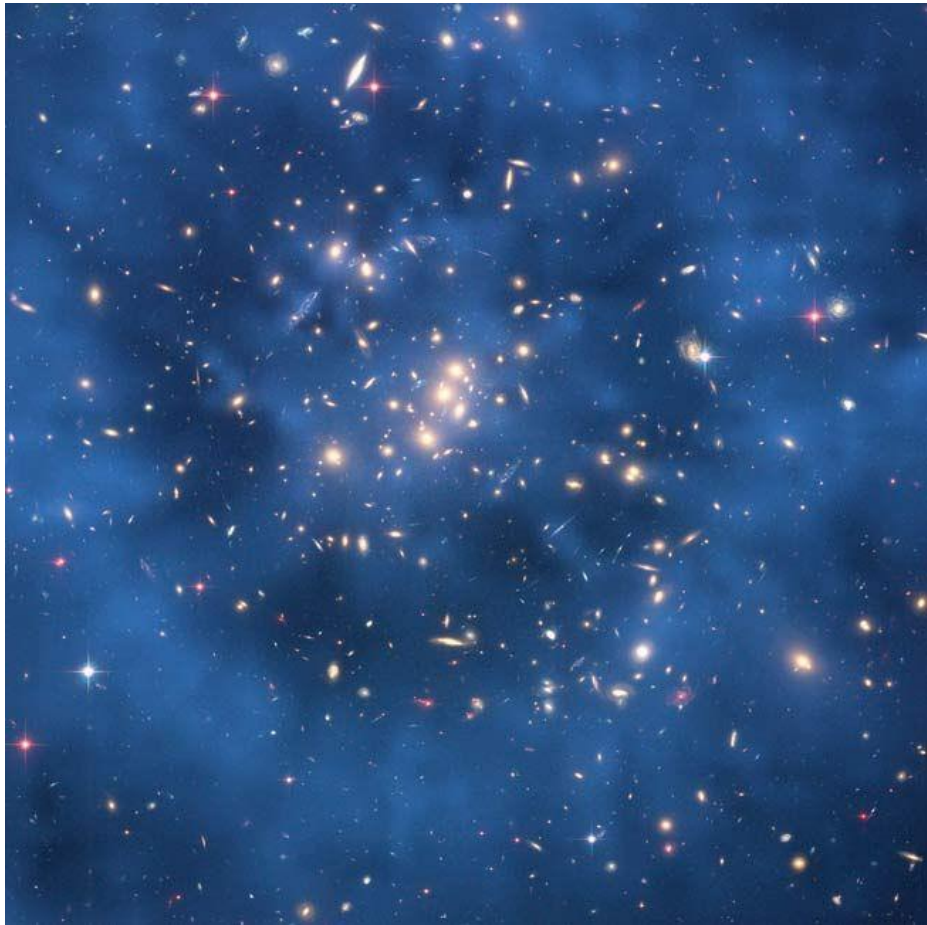
Simulation Agument

Não há evidências que nos permitam preferir uma das 3 alternativas.
Mas as sociedades tornam-se cada vez mais tecnológicas …

Quem ou o quê originaria a simulação? Quem ou o quê estará de fora a produzir essa simulação? Simulação de quê? Realidade ou jogo?

A **Matéria Negra** que existe em muito maior quantidade que a materia da qual nós somos feitos? Mas está provado que a matéria negra apesar de não interagir com a luz, obedece à mais fundamental das leis da matéria: A **gravidade**.

Faria parte da tal simulação também



• Inteligência Artificial Geral

Watson venceu "**Jeopardy** "

MAS… Ficou **feliz** por ter ganho?

Sentimentos e **Emoções** dependem também do corpo! (AD)

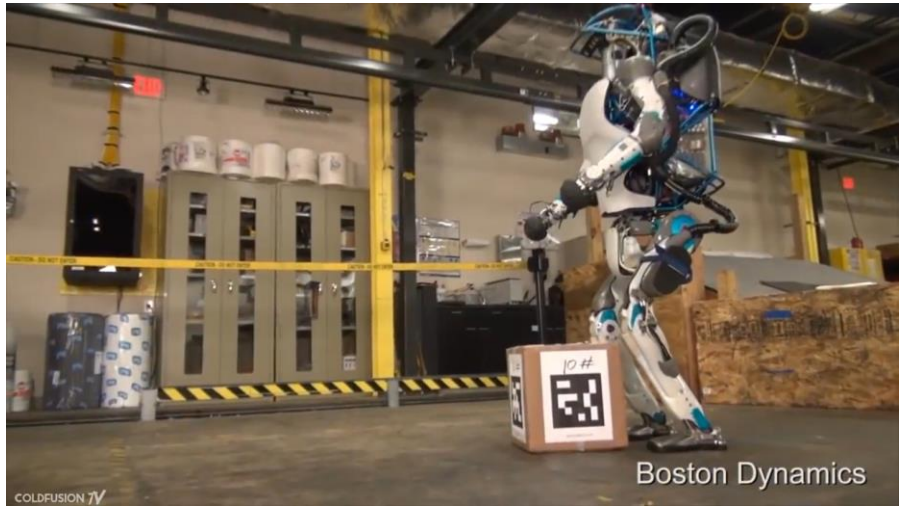
Simular Estados Emocionais simples (Medo, Ansiedade, Felicidade …)

• *Poderá um Computador funcionar como uma Mente?*

Download de um cérebro para uma entidade artificial (robô, computador, rede de computadores)?

Evolução de um Cérebro Digital
→estímulos muito complexos→
corpo e muitissimos sensores ...

Suposição justa:
Máquinas construidas por **humanos** que erram serão erróneas como eles.



Poderemos ser Amigos de um Robô?

Amizade Aristotélica: relação entre pessoas pela partilha de valores e interesses.

- Condição **Mutualista**: Deve haver **partilha Mútua de valores e interesses**.
- Condição da **Autenticidade**: Os participantes tem de ser honestos entre si
- Condição de **Igualdade**: Nenhum participante pode sentir-se **superior** aos outros
- Condição da **Diversidade** : Participantes devem **interagir** em circunstâncias variadas.

John Danaher, *Lecturer at National University of Ireland, Galway*

Poderemos ser Amigos de um Robô?

<u>Alguns afirmam : Robôs Nunca poderão ser nossos amigos</u>

Porque: robôs **NUNCA** satisfarão a condição **Mutualista** porque nunca terão vida interna mental: Nunca verdadeiramente poderão partilhar connosco;

O Robô é demasiado diferente, demasiado **alienígena**.

Poderemos ser Amigos de um Robô?

- Outros dizem: O que interessa é que eles ajam de forma a que **nós pensemos** que eles satisfazem as condições.

  - A Amizade entre Humanos é fundada na satisfação "**behaviourista**"/comportamental das condições relevantes da Amizade.

- Não temos acesso à vida mental interna uns dos outros;

- Nunca saberemos se os outros realmente partilham os nossos interesses ou se estão a fazer uma representação bem conseguida

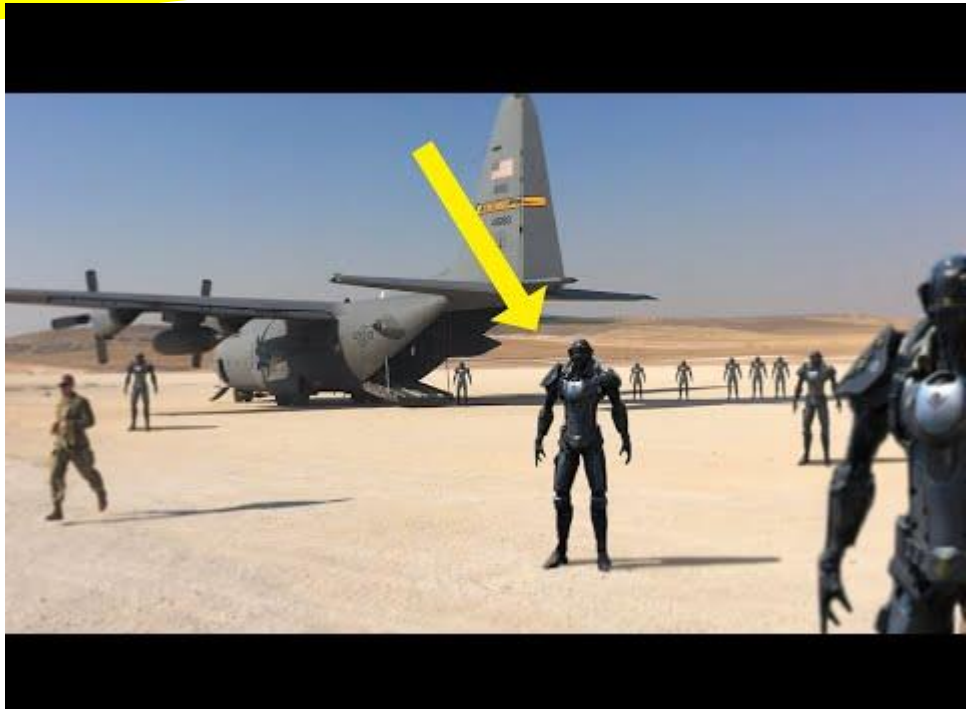
CYBORGS

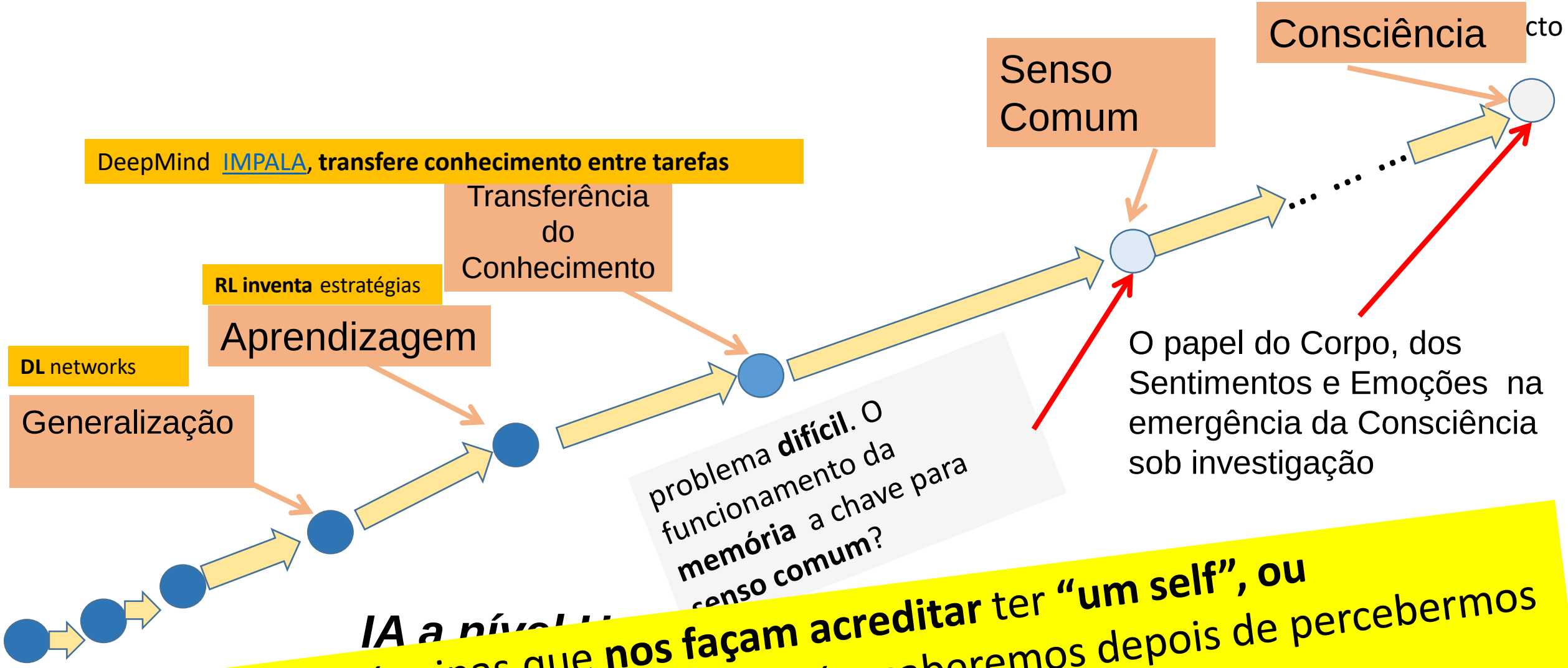
- **“cyborg”** um ser futurista parte máquina parte humano

Manfred Clynes e Nathan S. Kline criaram o terme

Aumentar a manobrabilidade, a força e as sensações não é ficção científica. Relatório da **DARPA** aponta 2050 para introdução de Soldados Cyborgs nos USA (sem fadiga, comunicação instantânea, visão com Zoom, ação psicológica …)

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





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.

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; Governantes- Governados …

Para uma IA Benéfica.

OBJETIVO Planetário:

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

Eugénio Oliveira

Que fazer?

- Elon Musk no World Government Summit:

“ we may become irrelevant as artificial intelligence (AI) grows more prominent”, February 2017, Dubai

A verdadeira questão poderá ser:

*“ Is **AI** the **threat** or is the threat staring back at us in the **mirror**?*

Stuart Russell *(U. California- Berkley)* ***open letter*** *(FLI - Future of Life Institute)* ***to keep AI beneficial***.

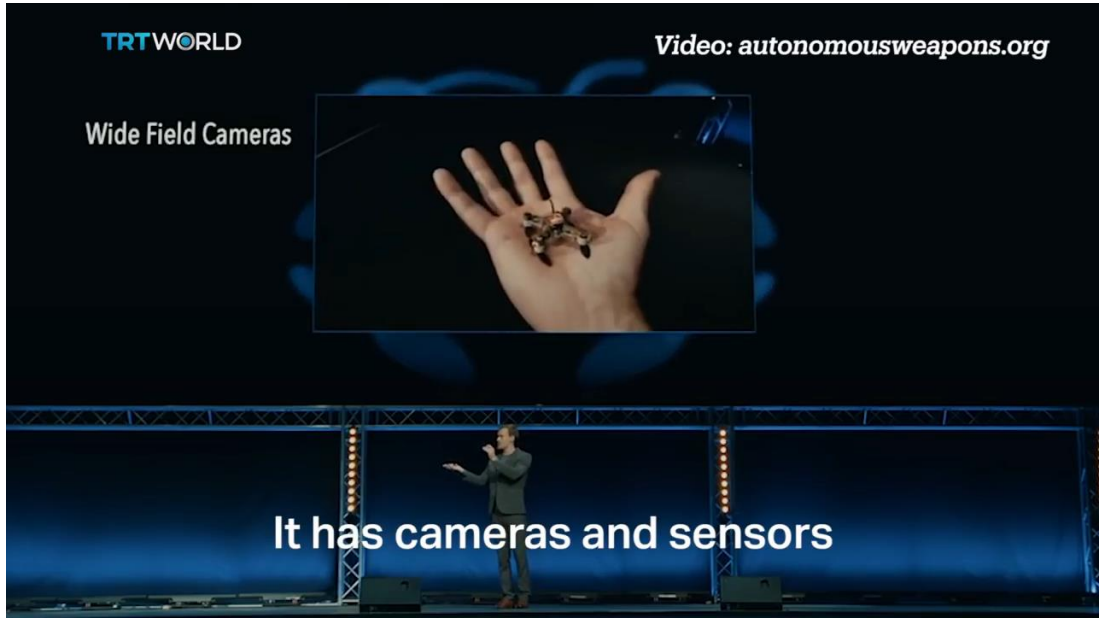
- Talvez a forma de lidar com os problemas co[illegible] seja **só** através da tecnologia. Nem só atr[illegible]

- Talvez seja cri[illegible] **Máquina** para partilhar os **nossos** valores numa so[illegible]

É preciso uma comissão para Ética na IA e Robótica

- Ingénuo imaginar que haverá apenas um conjunto de **princípios éticos para todos os Robôs/Sistemas Inteligentes**

- Haverá sempre "Robôs **Injustos-oportunistas**" nos negócios e "**Killer** Robots" na guerra.



- **É preciso uma comissão para *Ética na IA e Robótica***
  - **Certificar Software eticamente seguro**.
  - **Especificações** precisas e auditáveis.

Para uma IA Benéfica.

AI Now Institute de New York University organizou AI Now 2017 Symposium

apontou problemas éticos no uso dos **Lethal Autonomous Weapons Systems (LAWS)**

Devemos adotar princípios éticos que mantenham **controlo humano** destes sistemas.

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

Que Fazer?

- *AI NOW 2017 Report*

AI NOW 2017 Symposium refletiu sobre implicações a curto prazo **económicas e sociais da IA**:

- **Automação do Trabalho**
- **Enviesamento na Inclusão**
- **Ética e Governação**

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

By 2030, do you think it is most likely that advancing AI and related technology systems will enhance human capacities and empower them?

Pew Research Center , Dec 10, 2018

Respostas dos 979 científicos inquiridos:

- **63%** “melhor”
- **37%** “pior”
- 25 inquiridos não decidiram

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
- 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 d giar os huma na corrida com os robô anos/IA

- Comité de Ética para validar o desenvolvimento e aplicação da IA
- "Guidelines"
- Inclusão de Restrições e Princípios Éticos no Software de IA

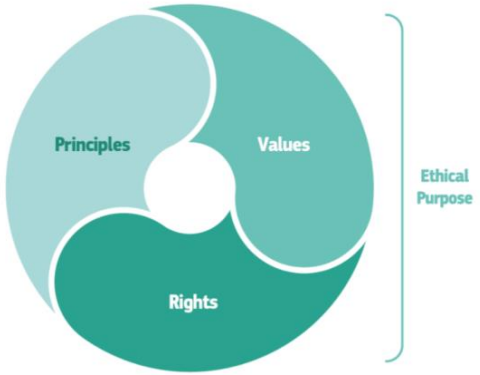
High-Level Expert Group on Artificial Intelligence

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;

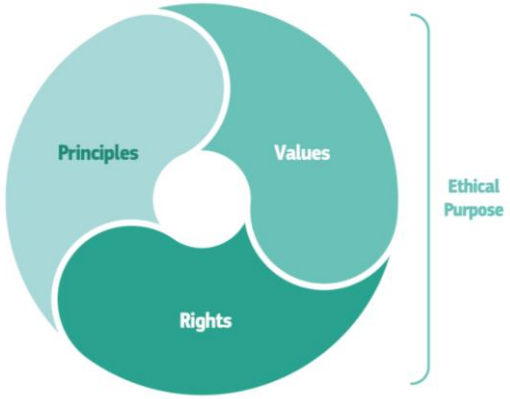
➢ **Princípio da NÃO Maleficência** ***("DO NO HARM")***

- 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

• *Recomendações*

+ ART em **ART**ificial 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** e de "**Human in the Loop**"

- transferência de competências, Sim!
  Humanos obsoletos? Nunca!

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

Questão Política!

Legislação sobre ética no desenvolvimento, disponibilização e uso da IA

Mais Ciência na IA para incluir nos programas restrições Éticas

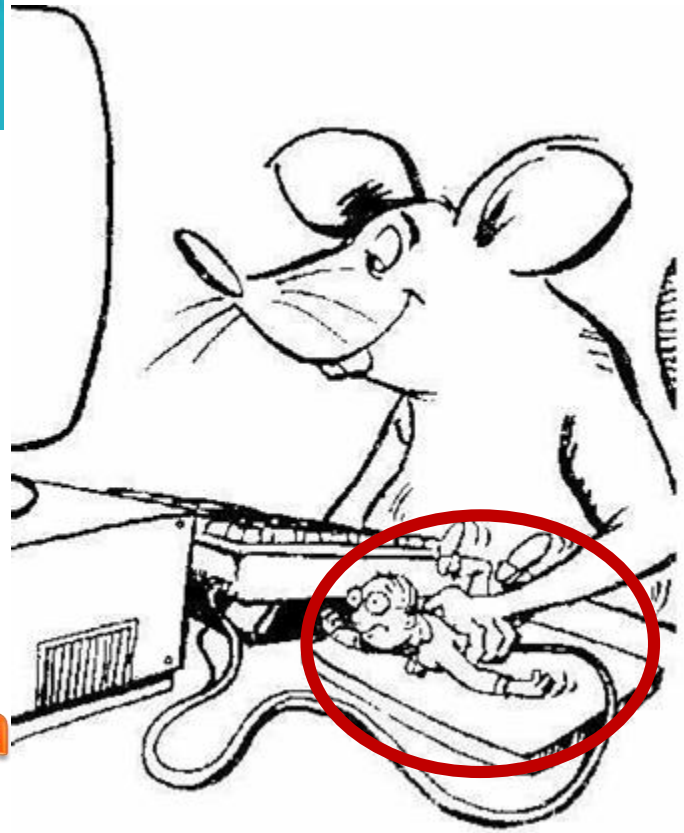
Inteligência Humana
NTES. MAS ...

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

Eugénio Oliveira

IA é fascinante
MAS …
CAUTELA…

OBRIGADO!



Há muito trabalho cívico e informado da tecnologia a realizar

Eugénio Oliveira