U.PORTO LIACC
artificial intelligence and computer science laboratory

A INTELIGÊNCIA ARTIFICIAL
Impacto e Ética

Eugénio Oliveira



?



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

Setembro 2020

U.PORTO LIACC

artificial intelligence and computer science laboratory



Ética Digital e IA

Ética e SUPER Inteligência Artificial



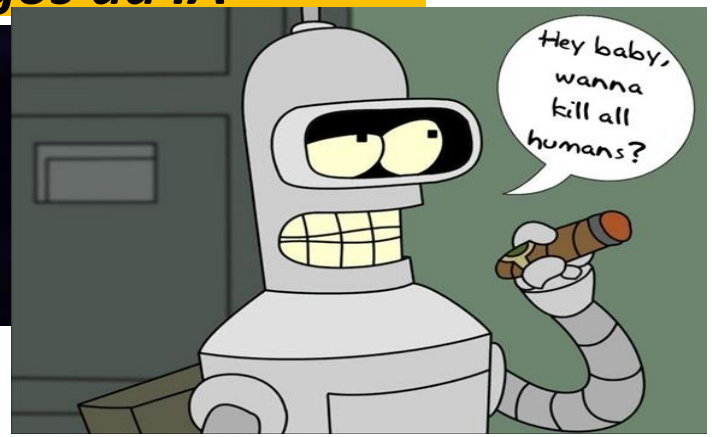
Eugénio Oliveira

Realizações



Promessas e Perigos da IA





IA Maléfica Vs IA Benéfica



Setembro 2020

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

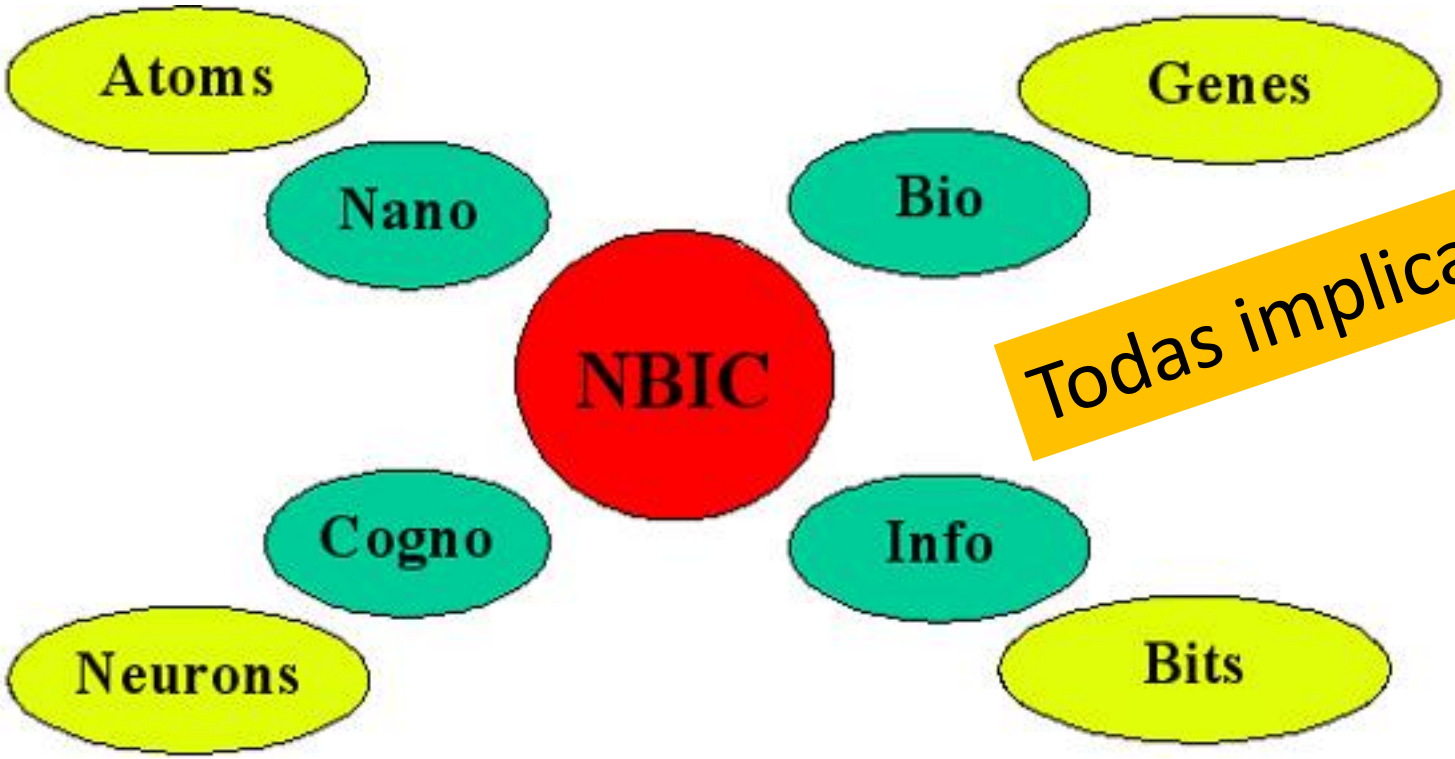
Criatividade?



Nada mais NATURAL que a INTELIGÊNCIA ARTIFICIAL
Homem NATURALMENTE criou a IA !!
Quanto à IN ainda há quem tenha dúvidas …

Setembro 2020

Futuro: Sinergias entre as Ciências



Todas implicam princípios Éticos

Debates éticos sobre IA



- Vigilância e Privacidade



- Manipulação de Comportamentos



- Opacidade dos Sistemas Inteligentes
- Enviesamento das Decisões



- Interação Humanos- Robôs



- Automação e Emprego

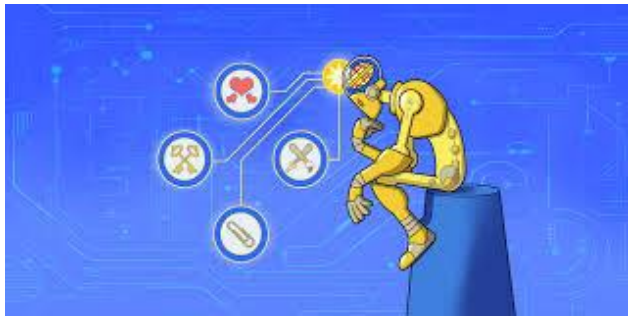
Setembro 2020

Debates éticos sobre IA

- Armas autónomas



- Ética da Máquina
- Agentes Artificiais Morais além dos deveres éticos também têm direitos e personalidade



- Singularidade e Consciência



Setembro 2020

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



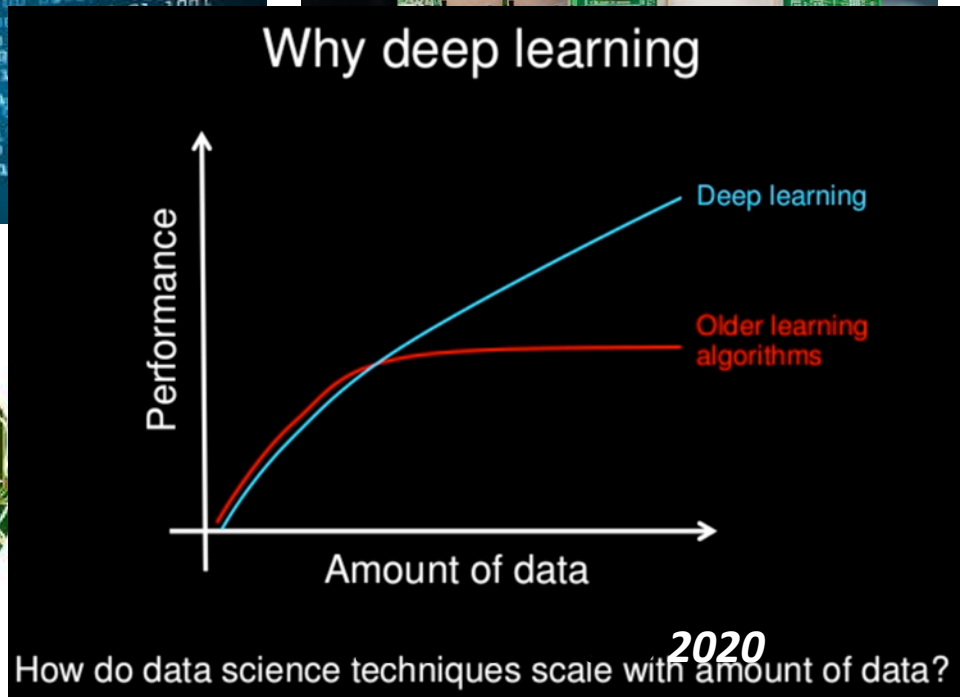
✓ **BIG DATA (Petabytes 10 ^{15}….)**



✓ **DEEP LEARNING**

Redes Neuronais Artificiais

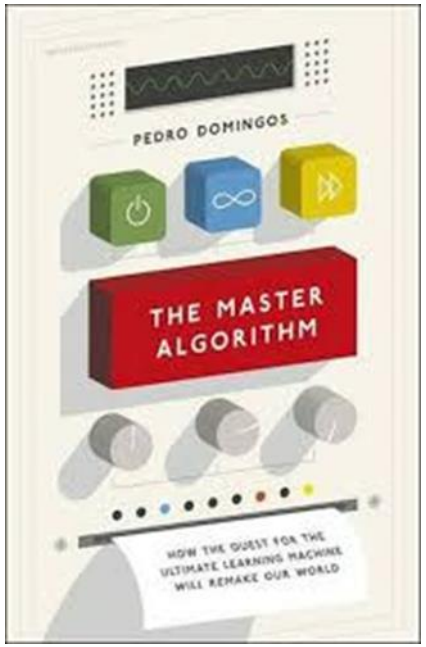




Ambientes IT-friendly
(ex Domótica, Smart Cities, Saúde, Defesa, …)

ESCOLAS

Representação do Conhecimento
Processamento do Conhecimento
Aquisição do Conhecimento

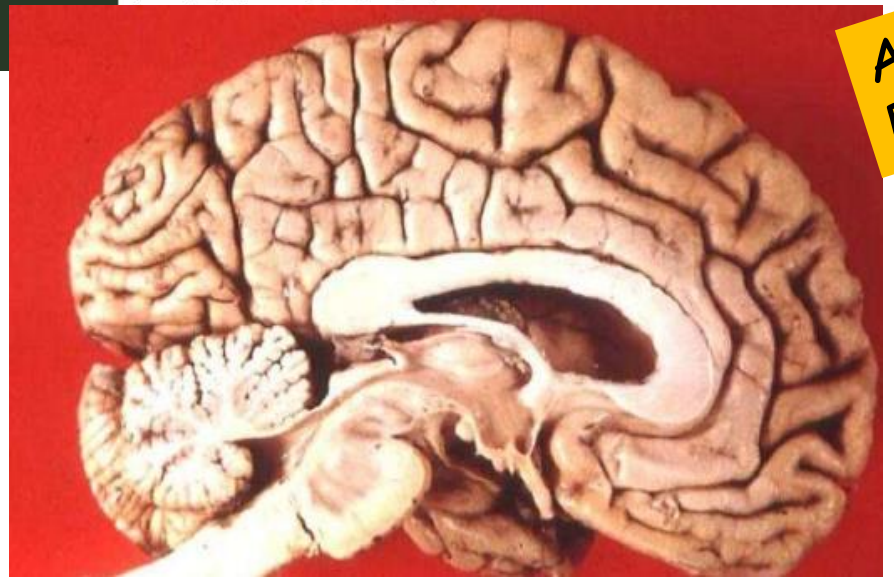


- **Simbólicos**
- **Conecionistas**
- **Evolucionários**
- **Estatisticos (Bayesianos)**
- **"Analogizadores"**

Escolas diferentes, objetivos semelhantes:
Desenvolver Máquinas Inteligentes !

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

(P.Domingos, Basic Books, 2015)



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,

IA: Realizações



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

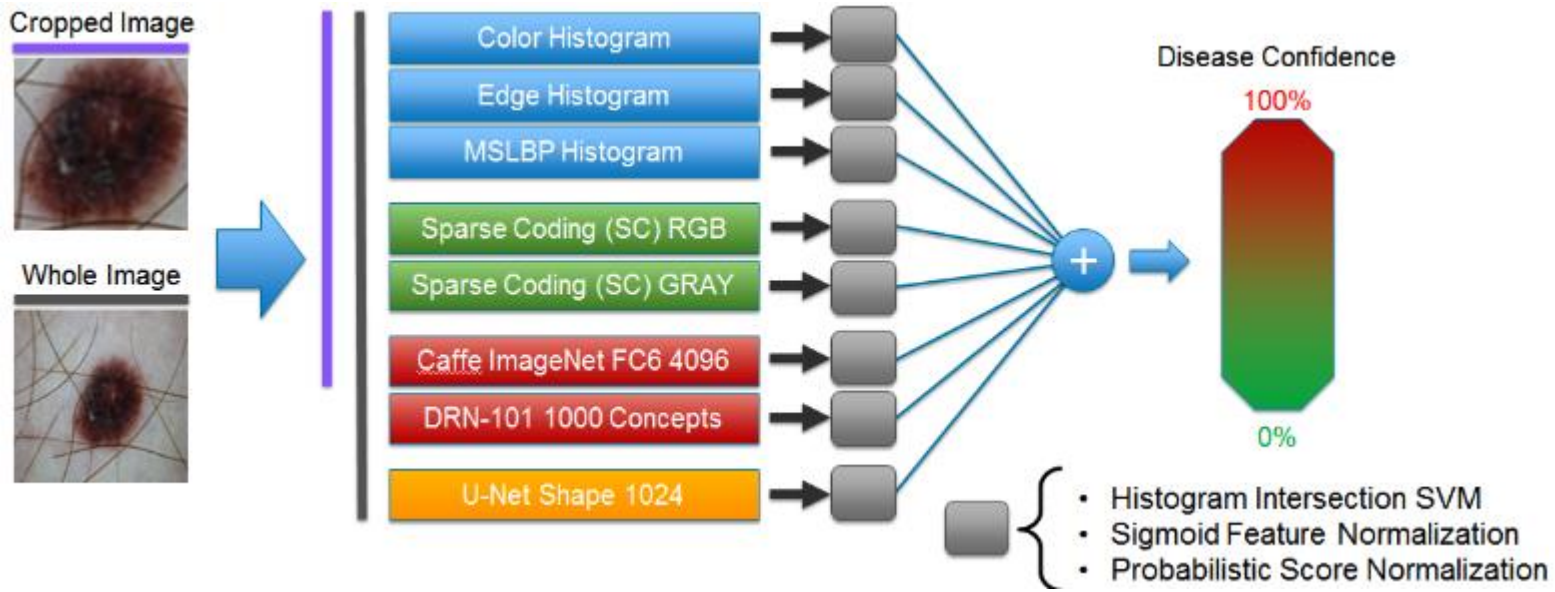
Óbvio

- **Previsões:** **Será que TU** cometerás um crime?
  Serás um potencial Terrorista? (**NSA**)
  TU serás bom naquele emprego?

Setembro 2020

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 decidir ações problemáticas

Setembro 2020

IA: Realizações

➢ Assistentes com IA:

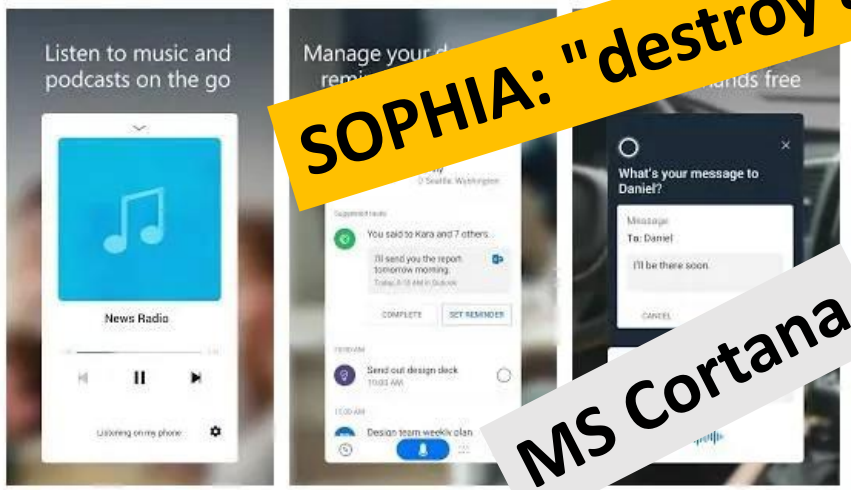


Apple



Google Assistant

SOPHIA: "destroy all humans" conversa com jornalistas.



MS Cortana



Amazon Alexa

Setembro 2020 2020

IA: Realizações

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

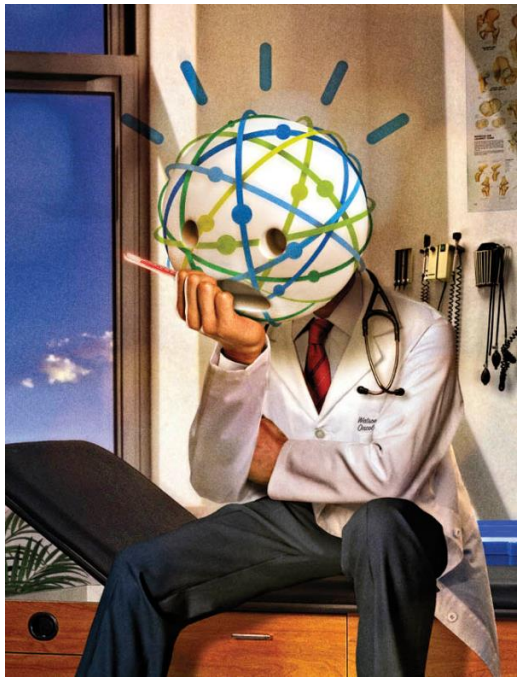
Em 34% dos casos encontrou tratamentos que os especialistas não tinham detetado em artigos de investigação



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



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

ÉTICA dos Investigadores em IA:
Evitar repetir erros do passado de "Overselling"

Sistemas de IA demonstrados em experiências cuidadosamente **controladas**.
Poucos sistemas baseados em IA foram **aprovados** pelos reguladores para uso em ambientes reais.

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

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)

NLP On-Line, TR BOM

Debater (NLP) selecionou 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: Impacto

OpenAI



GPT-2 modelo de linguagem natural treinada com **8 milhões de páginas web** Novembro, 2019

Problemático.
Inicialmente retido…

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

- **Escrita** Automática ; **Tradução** Automática …

OU para propósitos **maliciosos** incluindo:

- **Escrita** Automática de Notícias Falsas cre
- **"Personificar" Humano**

"**GPT-3** is shockingly good—and completely mindless. " July 20, 2020 Will Douglas Heaven archive page

O modelo passou de1,5 para 175 mil milhões de parâmetros

Negócio oferecido por prepagamento via "cloud"

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

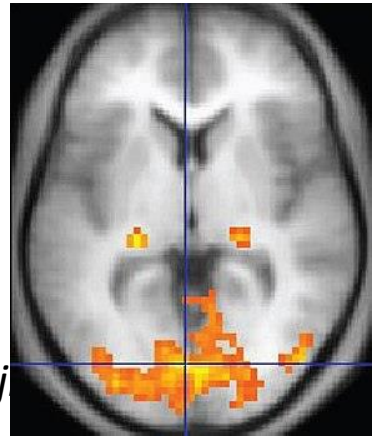


fotografia icónica do horror indiscriminado no Vietnam, **banida automaticamente** do Youtube e Facebook por **programas** que as consideram imorais.

IA: Impacto

Detetar (alguns) pensamentos/imagens ….

"Reconstrução de imagens através da atividade cerebral"



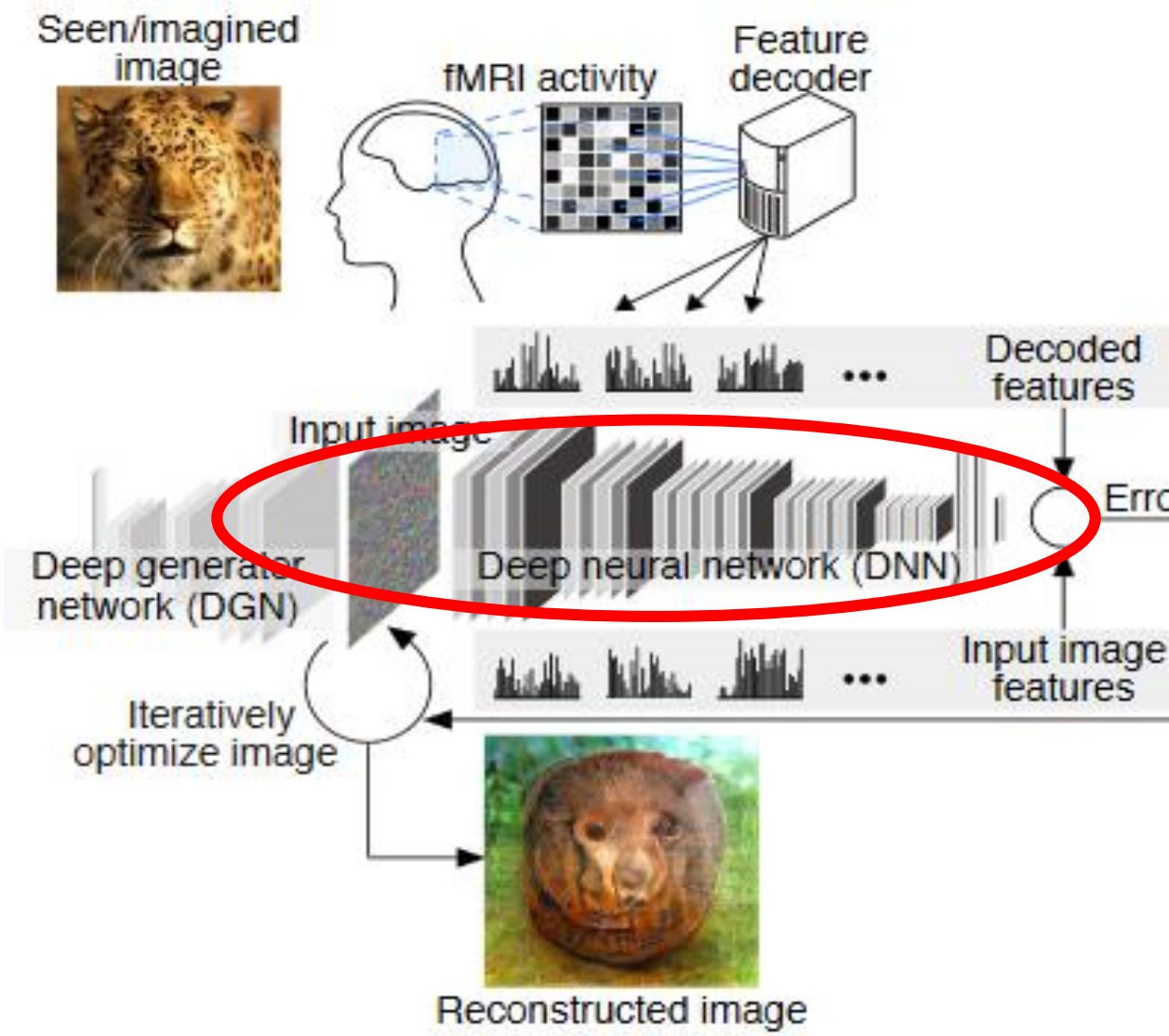
Guohua Shen, Tomoyasu Horikawa, Kei Maj Kamitani

ATR Computational Neuroscience Laboratories,

Kyoto University, Japan

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

Eugénio Oliveira

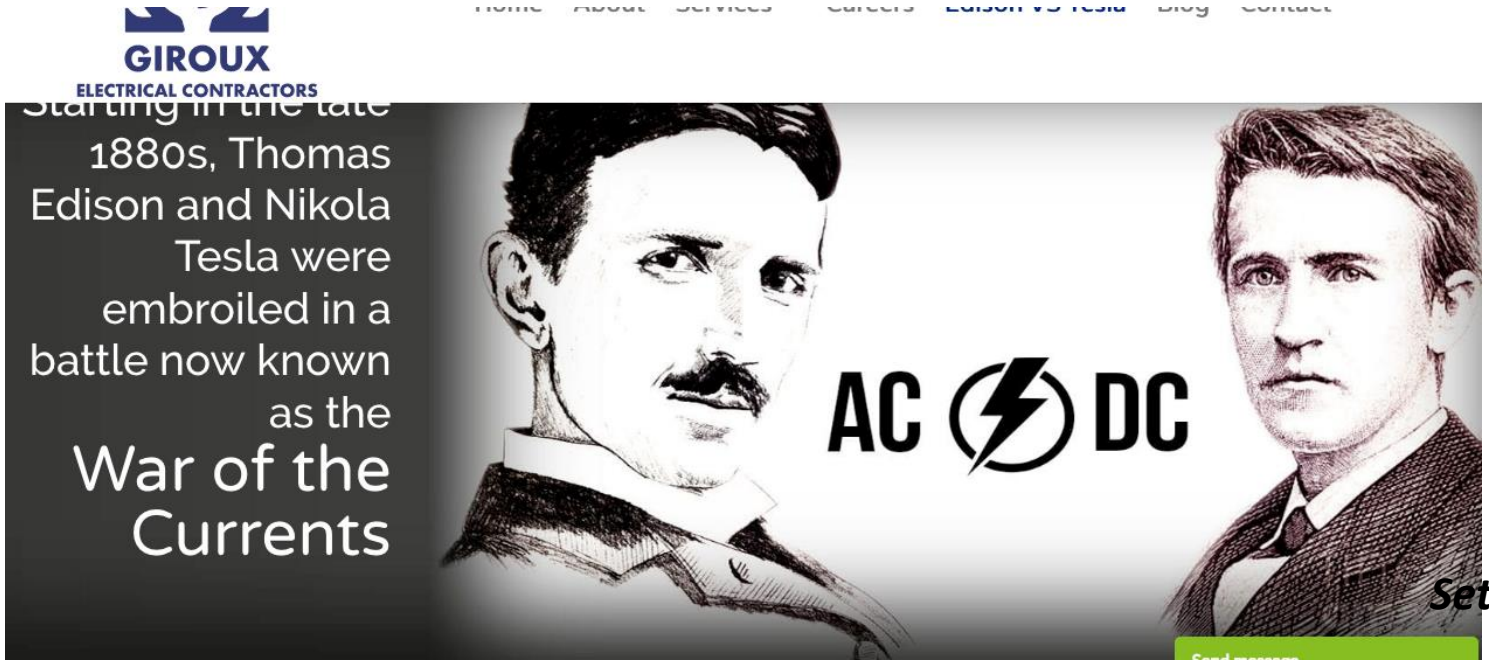


Setembro 2020

IA: Impacto

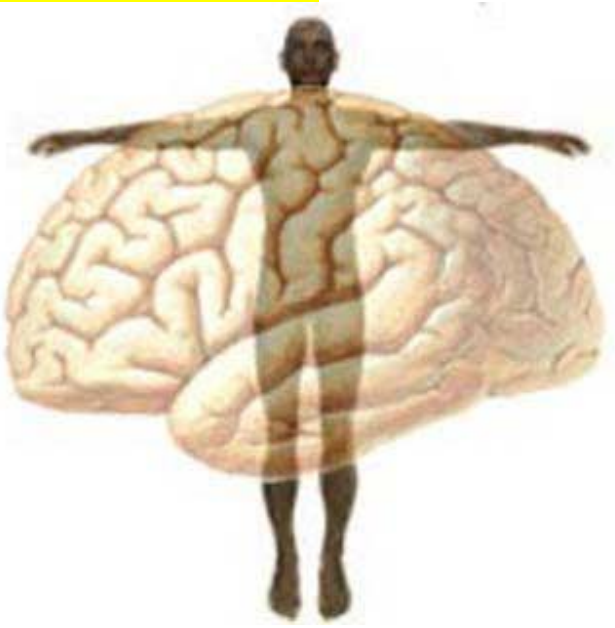
Quando a Westinghouse Electric Company, com Nikola Tesla implantou a Rede de Alta Voltagem em CA a empresa de Edison, em 1888 clamou que seria a maior ameaça na história da humanidade.

Não vejamos só o lado negro do futuro

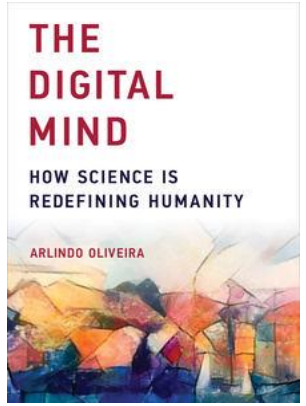


• Inteligência Artificial Geral

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



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



Setembro 2020

• 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 replicar a **computação** neles realizada

• 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 …)

e seu Impacto

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.

Eugénio Oliveira

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

Setembro 2020

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 Robôs 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





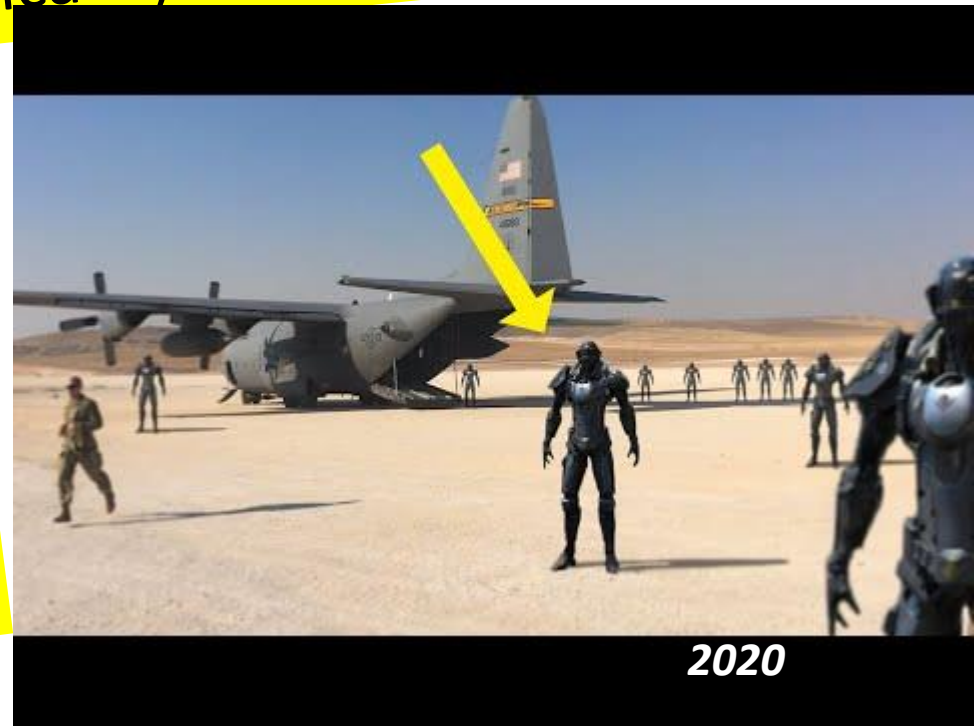
CYBORGS

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

Relatório da **DARPA** aponta até 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



cto

Consciência

Senso Comum

DeepMind IMPALA, **transfere conhecimento entre tarefas**

Transferência do Conhecimento

RL inventa estratégias

Aprendizagem

DL networks

Generalização

problema **difícil**. O funcionamento da **memória** a chave para **senso comum**?

O papel do Corpo, dos Sentimentos e Emoções na emergência da Consciência sob investigação

IA a nível

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.

Setembro 2020

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!!

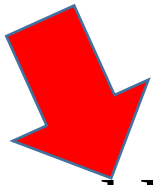
Atenção a grupos vulneráveis: Crianças, idosos, pacientes, deficientes e minorias.
Atenção às **assimetrias no poder** ou na informação: entre Empregados-Empregadores; Produtores- Consumidores; Governantes- Governados …

Eugénio Oliveira

Setembro 2020

IA MALÉFICA

OBJETIVO:
criação de uma MAI intencional através do programa



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

- **Governos:** establecer hegemomia e controlar populações ou derrubar outros
- **Corporações** construir monopólios destruindo competição por meios ilegais.
- **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 MALÉFICA

NSA instala ilegalmente **programas espias** em hardware e software coletando informação pessoal e metadados dos utilizadores da internet atenta contra a segurança dos sistemas de comunicação e torna-nos mais vulneráveis a ataques de espionagem e terrorismo ***[Abadi & al., January 24, 2014; Paterson et al., 2013; Peha, October 4, 2013].***

Disseminação de informação conflituosa e falsa para provocar o uso da força.

IA MALÉFICA/BENÉFICA

Afrontar o perigo sem medo
Ouvir uma Sinfonia na cabeça.
Visão superhumana
Curar cegueira, surdez ou doeças 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



Para uma IA Benéfica.

OBJETIVO Planetário:

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



Eticamente compensador

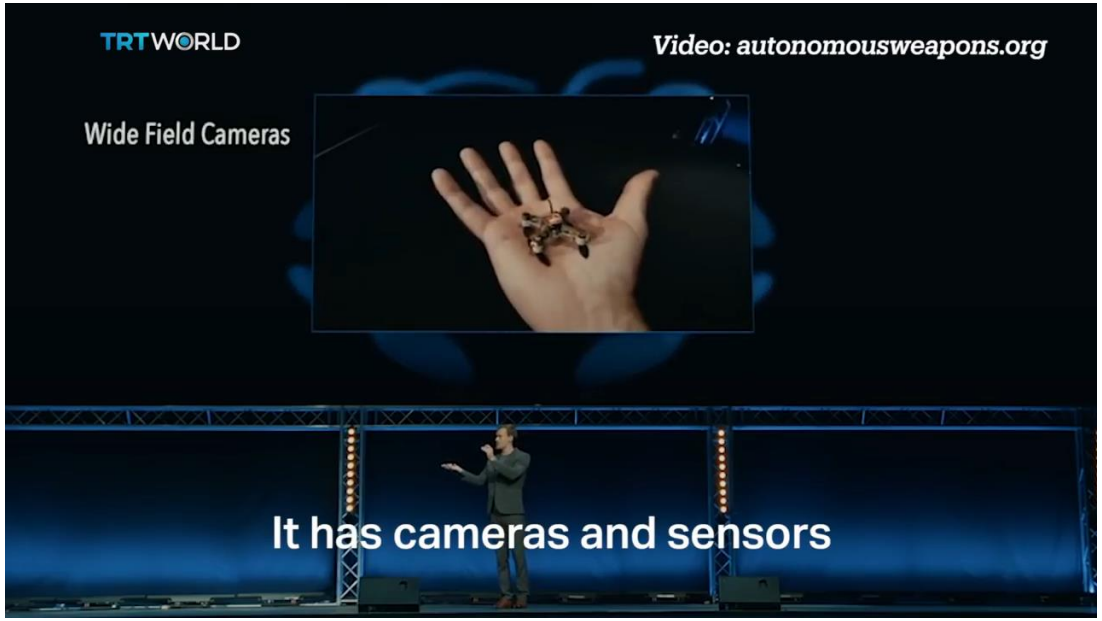
WEF estudou mais de 80 possíveis usos de IA em favor do Ambiente:
proteção Climática, mobilidade autónoma, redes intelligentes de energia. modelação climática …

Eugénio Oliveira

FAP Junho 2020

- 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

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



Para uma IA Benéfica.



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

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

Comunidade Científica em IA

➢ **Investigar**

- Camada de Software simbólico que envolva a camada mais funcional e subsimbólica

- Princípios de Validação ÉTICA das decisões tomadas

- Inclusão de Restrições e Princípios Éticos nos Sistemas Inteligentes



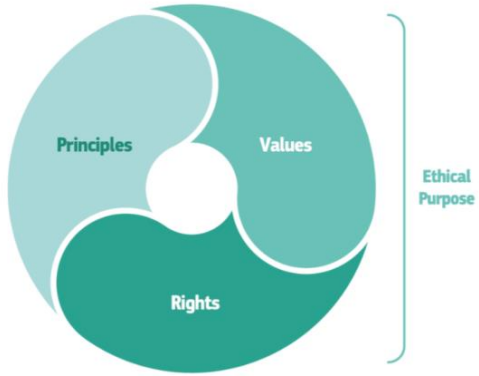
High-Level Expert Group on Artificial Intelligence

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;

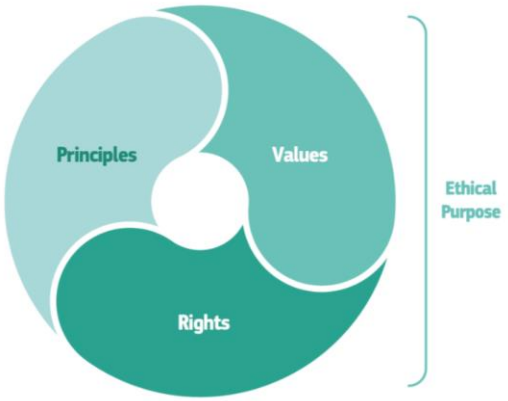
➢ Princípio da NÃO Maleficência *("DO NO HARM")*

- 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 hu

Europa tem dificuldade em reter talentos.
Se a Europa não lança um programa convincente em IA ficará à mercê dos outros dois continentes.

• *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**"

IA Ética, Legal e Robusta

Setembro 2020

U.PORTO LIACC artificial intelligence and co



Mais Ciência com Ética

- 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

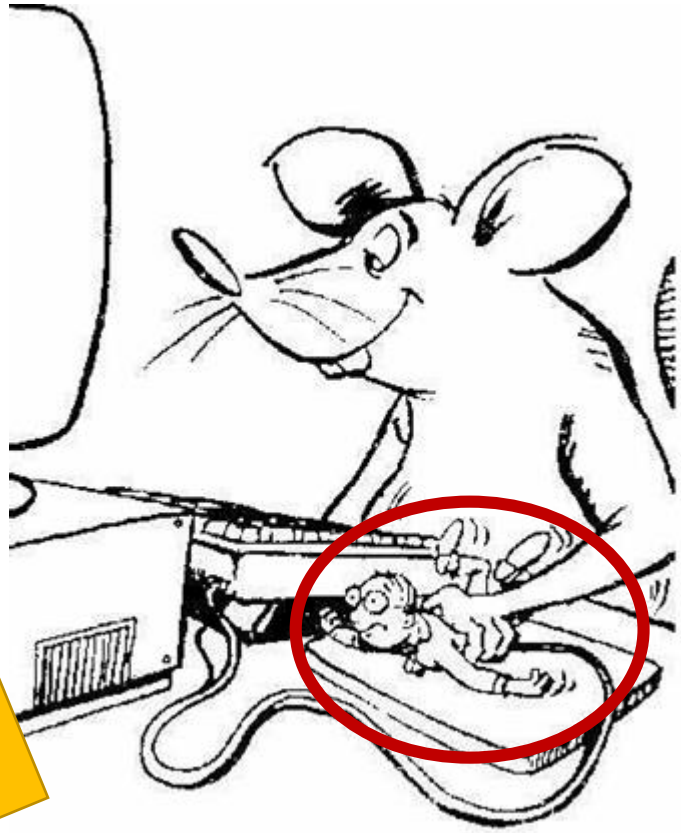
Mais Ciência na IA para incluir nos programas restrições Éticas

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

Eugénio Oliveira

Setembro 2020

Futuro e IA são fascinante MAS … CAUTELA…



OBRIGADO!



Há muito trabalho cívico e informado da tecnologia a realizar

Eugénio Oliveira

Setembro 2020