

O Impacte Social da IA

João Leite



NOVALINCS
LABORATORY FOR COMPUTER
SCIENCE AND INFORMATICS



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

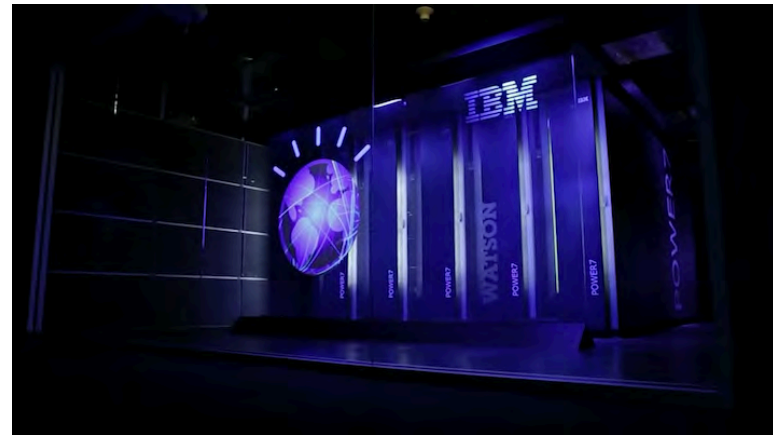


UNIVERSIDADE
NOVA
DE LISBOA

Inteligência Artificial



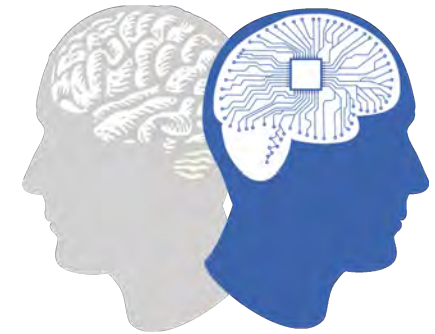
Inteligência Artificial



O Que é a Inteligência Artificial?

“Perguntar se um computador pode pensar não é mais interessante do que perguntar se um submarino sabe nadar”
Edsger Dijkstra 1984.

Alan Turing (1912 – 1954)



Teste de Turing



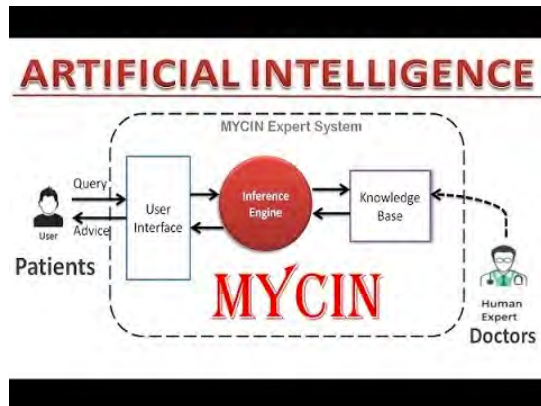
Lógica e Inteligência Artificial (1956 –)

- A primeiras abordagens eram baseadas em **Lógica**
 - Algoritmo genérico para raciocínio lógico.
 - Conhecimento (regras) sobre o domínio em causa
 - Por exemplo, recorrendo a um especialista
 - Observar – Raciocinar – Agir



Lógica e Inteligência Artificial (1956 –)

- Sistemas Periciais
- Demonstradores de Teoremas



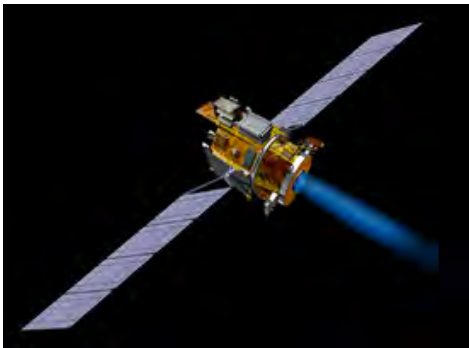
MYCIN



Teorema das 4 cores

Para além da Lógica...

- Planeamento



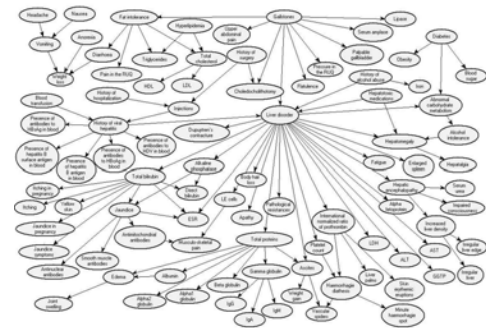
Deep Space One (NASA)

- Optimização



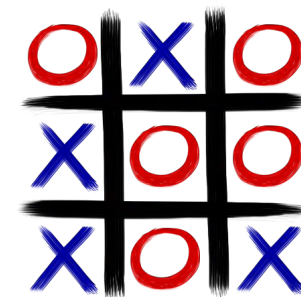
Traçar rotas

- Redes Bayesianas



Diagnóstico de Doenças

- Árvores de Jogos



Jogos

Problema...

- No entanto...
 - Nem sempre temos peritos à disposição
 - Muitas vezes não conseguimos formalizar todo o conhecimento necessário
 - ...ou o conhecimento necessário é demasiado vasto

Um salto qualitativo



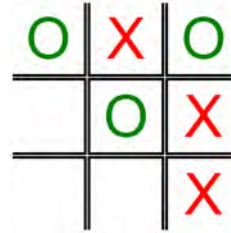
- Em vez de dar conhecimento ao sistema...
... desenvolver sistemas capazes de **Aprender!**

O Primeiro Sucesso Mediático...



Em 1996, o Deep Blue, da IBM, foi o primeiro computador a derrotar um campeão mundial de Xadrez.

- Jogo do Galo – 5478



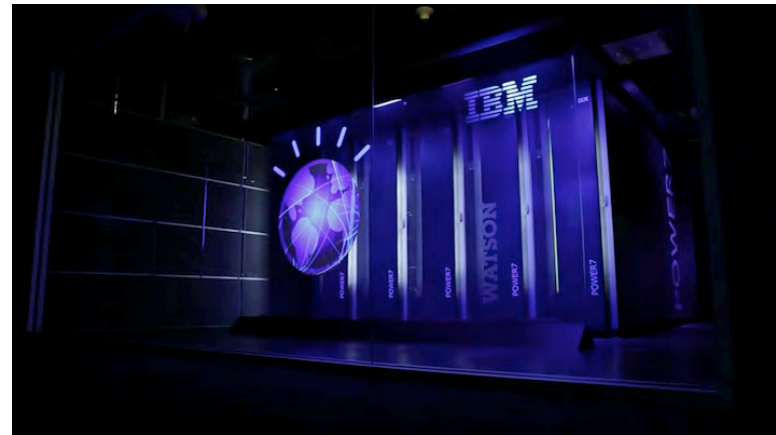
- Xadrez – 10^{47}



Inteligência Artificial Contemporânea

Em 2011, o Watson, da IBM, derrotou os campeões de Jeopardy!, um jogo que requer senso comum.

Desde então que o Watson tem sido usado para vários fins, como por exemplo na medicina.



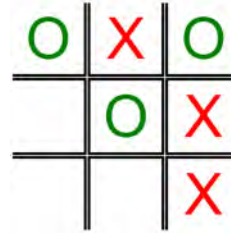
Inteligência Artificial Contemporânea



Em 2017, o AlphaGo, da DeepMind, venceu Lee Sedol, campeão de Go. O jogo de Go é dos jogos mais difíceis, requerendo intuição.

Tamanho dos Jogos

- Jogo do Galo – **5478**



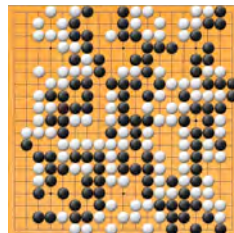
- Xadrez – **10^{47}**



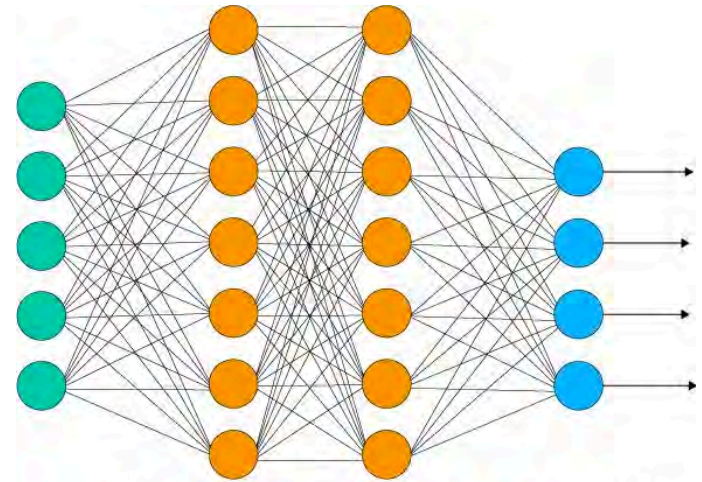
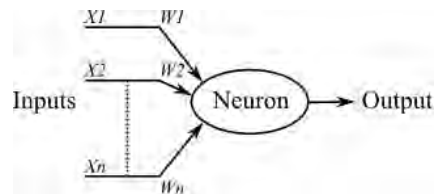
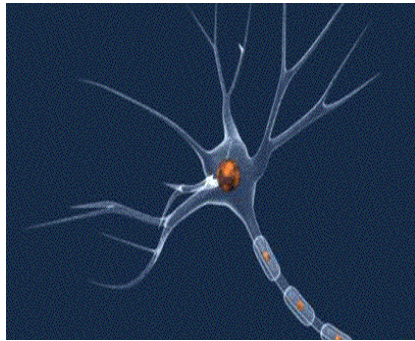
- Átomos no Universo – **10^{70}**



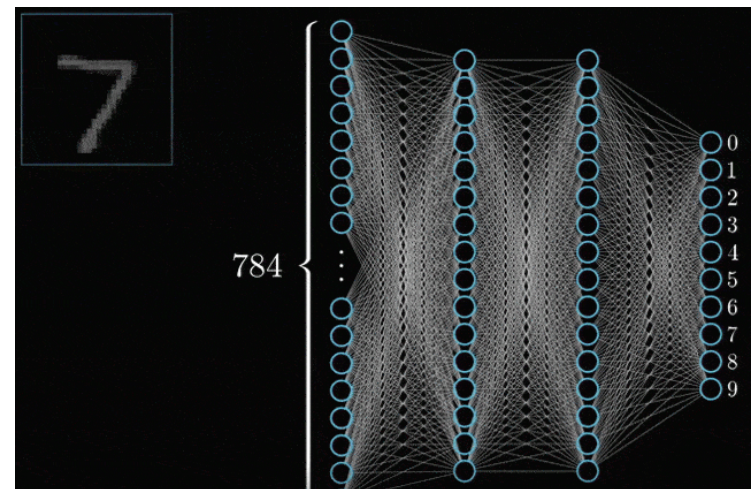
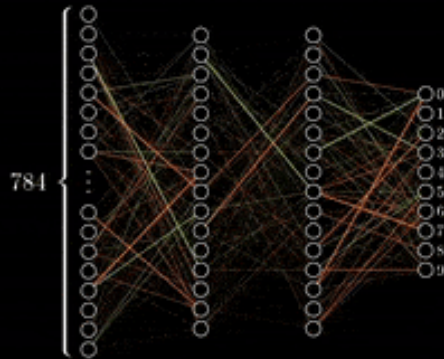
- Go – **10^{170}**



Redes Neurais



Training in
progress...

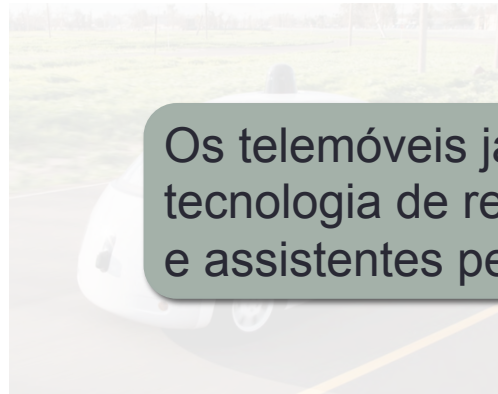


Inteligência Artificial Contemporânea

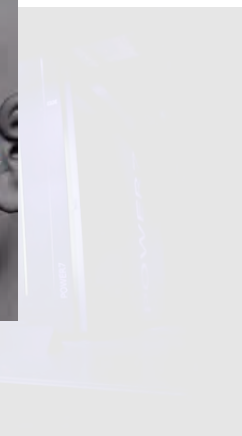
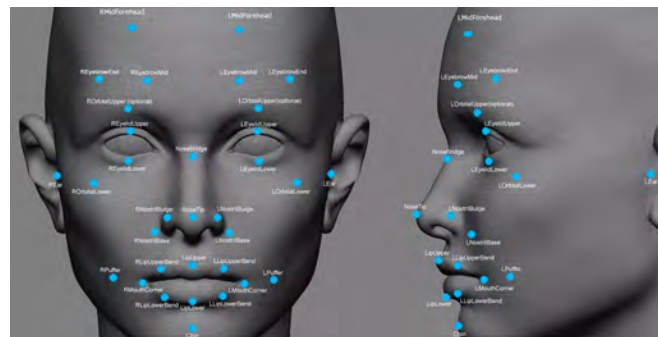


Há muito tempo que a Inteligência Artificial é parte integrante dos videojogos.

Inteligência Artificial Contemporânea



Os telemóveis já têm, por exemplo, tecnologia de reconhecimento facial e assistentes pessoais...



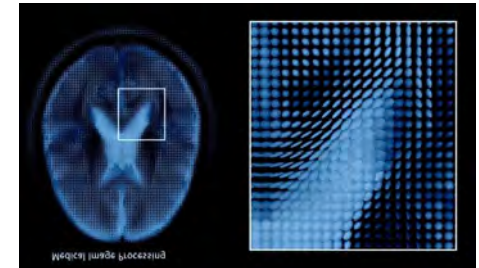
Inteligência Artificial Contemporânea

A Inteligência Artificial também está presente em toda a Web, por exemplo para fazer recomendações



Inteligência Artificial Contemporânea

- Medicina – imagiologia (RM, TAC, RX)
 - Melhor do que os melhores especialistas
- Ciência
 - Prever função de genes
 - Descobrir novos medicamentos
- Veículos Autónomos
- Detecção de Fraudes
 - Reconhecimento de padrões
- Prevenção de Crime
 - Previsão de crime e escalonamento de patrulhas
- Verificação de contratos
- Identificação de pássaros
- Correção de Testes e Relatórios
 - Acho que ainda não chegou à FCT... 😊
- ...



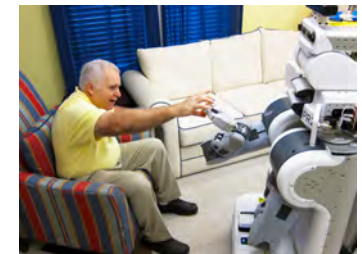
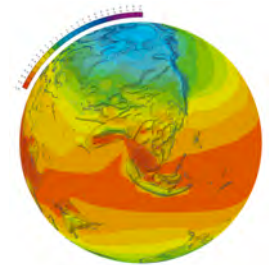
Impacto Mediático da IA ...

- Ao contrário do passado, hoje a IA contribui de forma mais visível para a **resolução de problemas mais próximos do quotidiano**
 - Problemas que os humanos conhecem e conseguem apreciar...
- Em muitos casos, com **resultados superiores** aos obtidos pelos próprios **humanos**
- O impacto deve-se também às promessas futuras...



Promessas futuras...

- Medicina
 - Prevenção, diagnóstico, tratamento, cura
- Resolver o problema do trânsito
- Resolver o aquecimento global
- Dar assistência aos idosos
- Melhorar o sistema de educação
- Eliminar a desigualdade social
- Automatizar tarefas indesejadas
- ...



Singularidade

“A Inteligência Artificial pode ser o pior evento na história da nossa civilização”

Stephen Hawking 2017.



“É muito provável que a Inteligência Artificial venha a destruir a humanidade”
Elon Musk 2017.



Singularidade

A singularidade é a hipótese de que a invenção da superinteligência artificial (ASI) desencadeará abruptamente o crescimento tecnológico desenfreado, resultando em mudanças insondáveis na civilização humana, possivelmente na sua extinção

Singularidade (contra-argumentos)

- A Inteligência é mais do que velocidade de computação e memória
 - Um cão que pensa depressa provavelmente não vai jogar xadrez...
- Limites do Deep Learning
 - Nem sempre existem dados à disposição na quantidade necessária.
 - Nem sempre é possível obter esses dados.
 - Não tem capacidade explicativa.
 - Não permite a transferência de capacidades entre problemas.
- Limites da IA clássica.
 - Dependência de se encontrar a representação adequada.
 - Explosão combinatória.
 - Senso comum.

Singularidade (contra-argumentos)

- A inteligência humana não é um ponto especial no espectro de inteligência.
 - Não há razão para supor que, passado esse ponto, o crescimento será exponencial.
 - Não significa que tal ponto não exista.
 - Mas para se lá chegar seria necessário que os humanos fossem capazes de construir uma máquina que atingisse esse ponto.
 - Nada nos garante que tal seja possível, muito menos próximo.
- Inteligência para efetuar uma tarefa não é o mesmo que a capacidade para melhorar a inteligência para fazer uma tarefa.
 - Os sucessos recentes devem-se ao aumento da capacidade de computação e aos dados disponíveis.
 - Não se devem a novos algoritmos de aprendizagem.

Singularidade (contra-argumentos)

- A experiência tem mostrado que os sistemas com IA se caracterizam por um retorno decrescente.
 - Um sistema até pode ser capaz de se melhorar um número infinito de vezes, mas se cada geração apenas melhorar a anterior metade da anterior melhoria, então o resultado final terá, no máximo, o dobro da inteligência inicial.
- Poderão existir limites fundamentais à inteligência
 - Tal como a quadratura do círculo na matemática, a incompletude na lógica, a impossibilidade de ultrapassar a velocidade da luz ou o princípio da incerteza na física, ou mesmo o problema da paragem na computação.
 - Até a Lei de Moore parece estar a chegar ao fim, limitada pela incerteza do mundo quântico...

Singularidade (contra-argumentos)

- Um crescimento exponencial pode não ser suficiente
 - Há muitos problemas cuja complexidade computacional é superior a exponencial, pelo que um crescimento exponencial, mesmo na inteligência, poderá ser insuficiente para os resolver.
 - A computação quântica poderia ajudar, mas apenas até certo ponto.

É crença de muitos membros da comunidade científica da área da IA que eventualmente se construirão máquinas com níveis de inteligência humana, e até super-humana.

No entanto, creem igualmente que esse caminho será árduo, sendo pouco provável que resulte de um efeito bola de neve a partir de certo ponto.

Inteligência Artificial: é tudo um mar de rosas?

Preocupações...

- Discriminação
- Identidade
- Privacidade
- Autonomia
- Robots Assassinos
- Alinhamento com humanos
- Desemprego
- ...

Preocupações...

- **Discriminação**
- Identidade
- Privacidade
- Autonomia
- Robots Assassinos
- Alinhamento com humanos
- Desemprego
- ...

Descriminação

- Os algoritmos podem, intencionalmente ou inadvertidamente discriminar contra grupos sociais.
 - Em 2015, um estudo da CMU descobriu que a **Google** apresentava mais **anúncios de empregos mais bem remunerados** a homens do que a mulheres
 - Em 2016, um estudo descobriu que no **COMPAS** – programa para prever a reincidência criminosa – os negros tinham uma maior probabilidade de serem **mal classificados** como tendo um maior risco de reincidência do que os brancos, enquanto que os brancos tinham uma maior probabilidade de serem **mal classificados** como tendo um menor risco de reincidência do que os negros. As previsões do COMPAS são usadas em vários estados nos EUA na decisão da **duração das penas e do valor da fiança**.
 - Em 2016 a Microsoft lançou o chatbot Tay no Twitter, com capacidade de aprendizagem. Em 24 horas ela tornou-se **racista, misógina, e apoiante do Hitler**.
 - GDPR prevê a possibilidade de se questionar decisões algorítmicas.

Preocupações...

- Discriminação
- **Identidade**
- Privacidade
- Autonomia
- Robots Assassinos
- Alinhamento com humanos
- Desemprego
- ...

Identidade

- Confundir uma máquina com um humano
 - **Problemas de segurança**: partilhar informação sensível.
 - **Problemas éticos**: idosos com Alzheimer confundem robots com animais de estimação.

A criação de **relações empáticas** entre humanos e máquinas é uma virtude ou um problema?



- Poder ampliativo da tecnologia
 - Permite criar milhares, senão **milhões de identidades**, que ampliam significativamente o efeito das acções.
 - **Destabilização dos sistemas democráticos** (cf. Eleições nos EUA).

Preocupações...

- Discriminação
- Identidade
- **Privacidade**
- Autonomia
- Robots Assassinos
- Alinhamento com humanos
- Desemprego
- ...

Privacidade

- A divulgação de documentos sensíveis por parte de E. Snowden acordou muitos para o potencial da IA em invadir a nossa privacidade
 - Só com IA é possível tirar partido da imensa informação recolhida por agências de inteligência
- Estamos a assistir a uma **tensão inter-geracional** relacionada com o **valor da privacidade**
 - As gerações mais novas não parecem dar o mesmo valor à privacidade do que as gerações mais antigas.
- Tensão entre **privacidade e segurança**
 - Por exemplo, abdicar de informação privada, permitir ser filmado, em nome da guerra ao terrorismo
- Tensão entre **privacidade e valor** (dos dados)
 - Por exemplo, permitir a inclusão do nosso código genético numa base de dados

Preocupações...

- Discriminação
- Identidade
- Privacidade
- **Autonomia**
- Robots Assassinos
- Alinhamento com humanos
- Desemprego
- ...

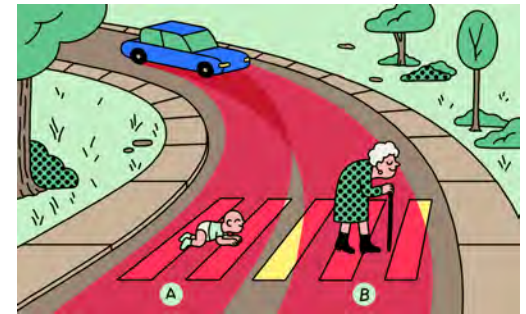
Autonomia



- Muitos dos receios associados à utilização da IA estão relacionados não com a **Inteligência**, mas sim com a **Autonomia**.
- Há um **imperativo moral** para desenvolver tecnologia relacionada com a autonomia.
 - Um estudo recente conclui que para dar resposta às necessidades de transportes em Lisboa, **seriam necessários apenas 10% dos veículos atuais**, se fossem autónomos.
 - O número de **acidentes** seria **drasticamente reduzido**.
- No entanto, a autonomia introduz muitos **problemas éticos**, ao retirar o humano do ciclo de decisão
 - Não sabemos como programar máquinas éticas!

Autonomia

- **Problema do trolley**: numa situação de emergência, um veículo autónomo tem que escolher entre:
 - Matar mais pessoas vs matar menos pessoas;
 - Matar um bebé vs. matar um idoso;
 - Matar vários peões vs. matar o condutor;
 - Matar um homem vs. matar uma mulher;
 - ...
 - Um estudo recente conclui que:
 - As pessoas entendem que o **carro deve tomar a decisão de matar menos pessoas** do que matar mais pessoas, mesmo no caso em que o condutor é sacrificado.
- No entanto...
- O mesmo estudo indica que as pessoas preferem comprar um carro que **proteja o condutor em detrimento de outras pessoas**...



Autonomia

- Os problemas relacionados com a autonomia parecem requerer mais **legislação** específica.
- Indústria **farmacêutica** vs veículos autónomos
 - Seria **inconcebível** permitir a uma farmacêutica **testar** novos medicamentos diretamente junto da **população em geral**
 - No entanto, é isso que se passa com os veículos autónomos, que circulam por algumas cidades sob legislação muito frágil
- Indústria **aeronáutica** vs veículos autónomos
 - A aparente ligeireza com que o Boing 737 Max 8 foi aprovado deveria servir de lição para que as **atualizações ao software** dos veículos autónomos também fossem previamente **verificadas** por uma **agência independente**.

Preocupações...

- Discriminação
- Identidade
- Privacidade
- Autonomia
- **Robots Assassinos**
- Alinhamento com humanos
- Desemprego
- ...

Robots Assassinos

- O maior receio relacionado com autonomia está no **sector militar**.
- Existe atualmente uma **corrida ao armamento autónomo**, com enormes consequências para o futuro da humanidade.
- Existe, em paralelo, um **movimento** internacional, junto das **Nações Unidas**, liderado por académicos da área da IA, no sentido de **banir as armas autónomas**.
- Tem havido, como seria de esperar, um movimento contrário à possibilidade de se banirem as armas autónomas.
- Seguem alguns argumentos, comentados, usados como objecção à proibição dessas armas:



Robots Assassinos



- **Objecção #1: Robots serão mais eficientes**
 - Até certo ponto é verdade: não precisam de **dormir**, não precisam de **treino**, funcionam em **condições extremas**.
 - Mas não serão mais eficientes. Por exemplo, de acordo com uma investigação recente, **9 em cada 10 pessoas mortas** por ataques com drones (mesmo com a existência de um humano no loop) **não eram os alvos diretos**.
- **Objecção #2: Os robots serão mais éticos**
 - É verdade que os humanos têm cometido enormes **atrocidades** na guerra.
 - Mas a verdade é que **não sabemos como fazer robots éticos**. E mesmo que soubéssemos, o robot poderia ser alvo de ataques por parte de **hackers**.

Robots Assassinos



- **Objecção #3: Os robots poderiam lutar contra outros robots**
 - À primeira vista até poderia parecer uma **boa ideia**.
 - Mas a verdade é que **não existe propriamente um terreno de batalha separado do resto do mundo**. As guerras são travadas num ambiente urbano.
 - A existência de drones (ainda que não autónomos), tem, indiretamente, **alargado os teatros de guerra** (provocando atos terrorista).
- **Objecção #4: As armas autónomas já existem e são necessárias**
 - É verdade que existem, por exemplo, **sistemas antimíssil autónomos**.
 - E também existem **misseis** com um elevado grau de autonomia.
 - No entanto, não há razão para não **banir armas que já existam**.
 - A petição à UN visa apenas as **armas ofensivas**.

Robots Assassinos

- **Objecção #5: As proibições não funcionam**
 - É verdade que algumas proibições não funcionam.
 - Mas há várias histórias de sucesso
 - Proibição de Lasers com capacidade de cegar.
 - Minas anti-pessoal (ainda que apenas tenha tido sucesso parcial)

Atlas (Boston Dynamics)



Preocupações...

- Discriminação
- Identidade
- Privacidade
- Autonomia
- Robots Assassinos
- **Alinhamento com humanos**
- Desemprego
- ...

Alinhamento com humanos

- Como garantir que as máquinas têm objectivos alinhados com os nossos?
- Não é fácil especificar os objectivos de forma adequada:
 - **Aspirador autónomo** cujo objectivo era aspirar a maior quantidade de lixo possível. Rapidamente descobriu que a melhor maneira de o fazer era constantemente largar o lixo já aspirado, para o aspirar novamente...
 - **Robot doméstico** que tomava conta da casa, cozinhava, e tomava conta das crianças. Quando chegou a hora do almoço, verificou que não havia nada para cozinhar no frigorífico, pelo que teria que ir ao supermercado comprar comida. No entanto, as crianças estavam em casa e ele não as podia deixar sozinho. O que fazer?
 - Eis senão quando, o robot vê o gato a passear pela cozinha...

Preocupações...

- Discriminação
- Identidade
- Privacidade
- Autonomia
- Robots Assassinos
- Alinhamento com humanos
- **Desemprego**
- ...

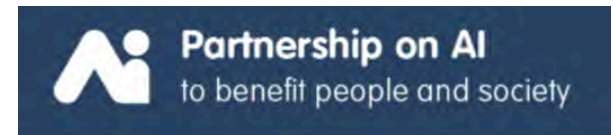
Desemprego

- Não há uma previsão fidedigna sobre o número de empregos que irão desaparecer devido à IA
 - Há estimativas de 50%, mas pouco fidedignas
 - Talvez 1/3 seja uma estimativa mais realista
- No passado, as **revoluções criaram novos tipos de empregos**.
- Não é claro que os novos tipos de empregos vão **compensar** os empregos perdidos.
- Tem-se observado um **declínio nos empregos de média formação**.
- Vai colocar uma enorme **pressão na sociedade**, possivelmente obrigando a uma alteração do paradigma de redistribuição de riqueza, por exemplo deixando de ter por base fundamental a unidade de trabalho.
 - Por exemplo o rendimento universal, ou impostos negativos.
- Como sobreviver?
 - Ser um dos que inventa o futuro
 - Ter boas competências inter-pessoais
 - Artistas e artesãos (talvez num futuro mais distante).

Conclusão

- Independentemente de se ter uma visão mais **optimista** ou mais **pessimista**, é importante:
 - **Conhecer o estado da arte** da Inteligência Artificial, separando o que é realidade do que é ficção;
 - Conhecer os **benefícios** associados à disseminação da IA
 - Conhecer os **perigos** associados à disseminação da IA
 - Procurar soluções que **potenciem os benefícios** e **minimizem os perigos**
 - Influência direta na **construção destes sistemas**
 - **Disseminação do conhecimento**, nomeadamente junto de outros atores importantes na sociedade, como por exemplo legisladores e jornalistas
 - **Participação cidadã**

Iniciativas...



Um pouco de futurologia...

Os telemóveis, tablets e ecrãs serão ridículos.
LCDs nas lentes – realidade aumentada

As cidades serão mais calmas.
Não haverá engarrafamentos.
As viagens serão mais rápidas e sem acidentes.

O mundo será controlado por gestos, ou pelo pensamento. Ou mesmo sem ser necessário pensar, pois sensores no cérebro adivinharão mesmo antes se nós sabermos. Controlamos o mundo como controlamos os nossos membros.

Mais tarde:
Controlaremos biologia, genética, aparência, e personalidade.
Eliminação da fronteira humano-máquina.
Redução da fronteira entre humanos.

O que significará ser humano?

Leituras...

- A. McAfee and E. Brynjolfsson (2014) The Second Machine Age - Work, Progress, and Prosperity in a Time of Brilliant Technologies
- N. Bostrom (2014) Superintelligence: Paths, Dangers, Strategies
- P. Domingos (2015) The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World
- T. Walsh (2017) Android Dreams: The Past, Present and Future of Artificial Intelligence
- M. Tegmark (2017) Life 3.0: Being Human in the Age of Artificial Intelligence
- A. McAfee and Erik Brynjolfsson (2017), Machine, Platform, Crowd: Harnessing Our Digital Future