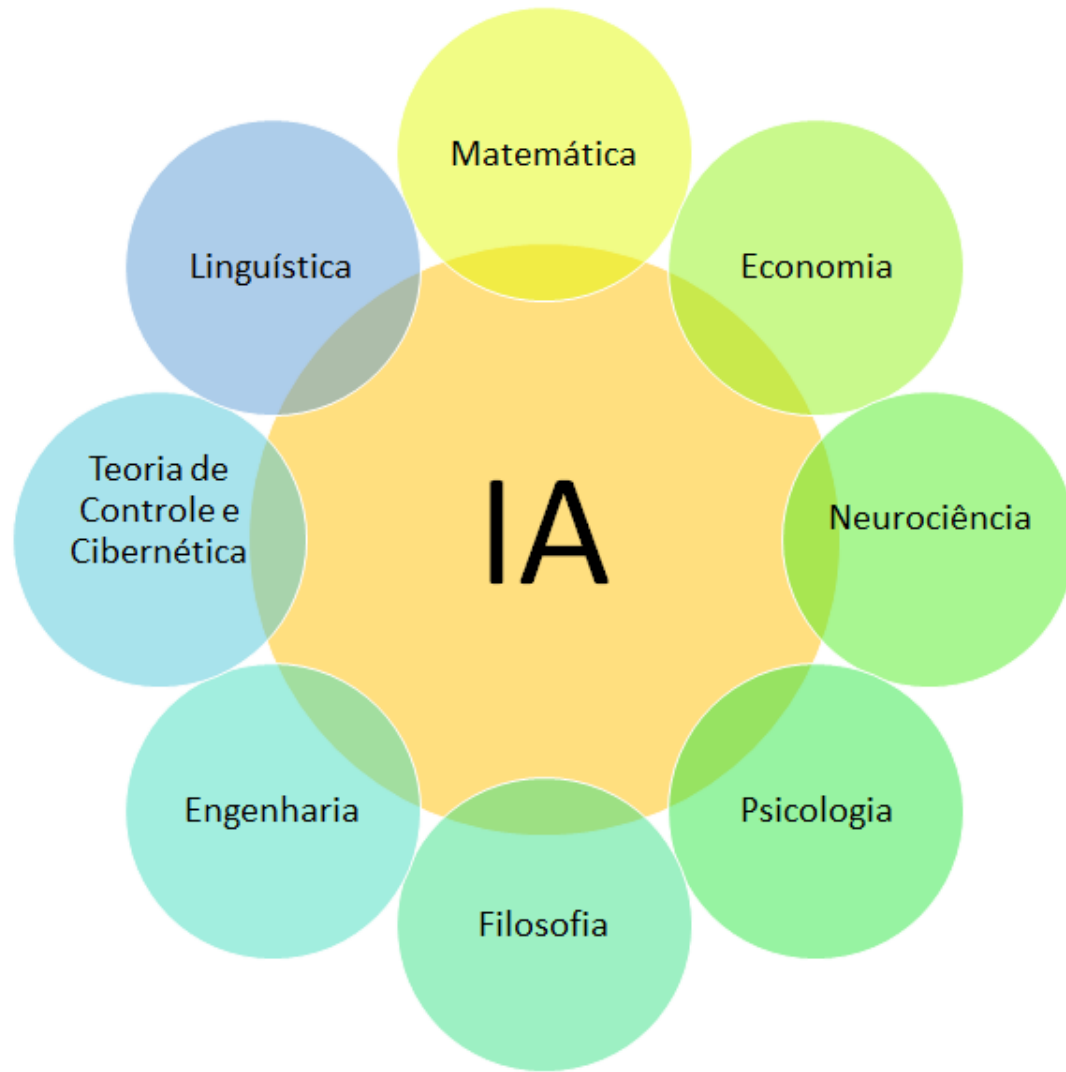




# DEUS EX MACHINA

Problemas éticos na criação da Inteligência Artificial

Marcos Bazmandegan



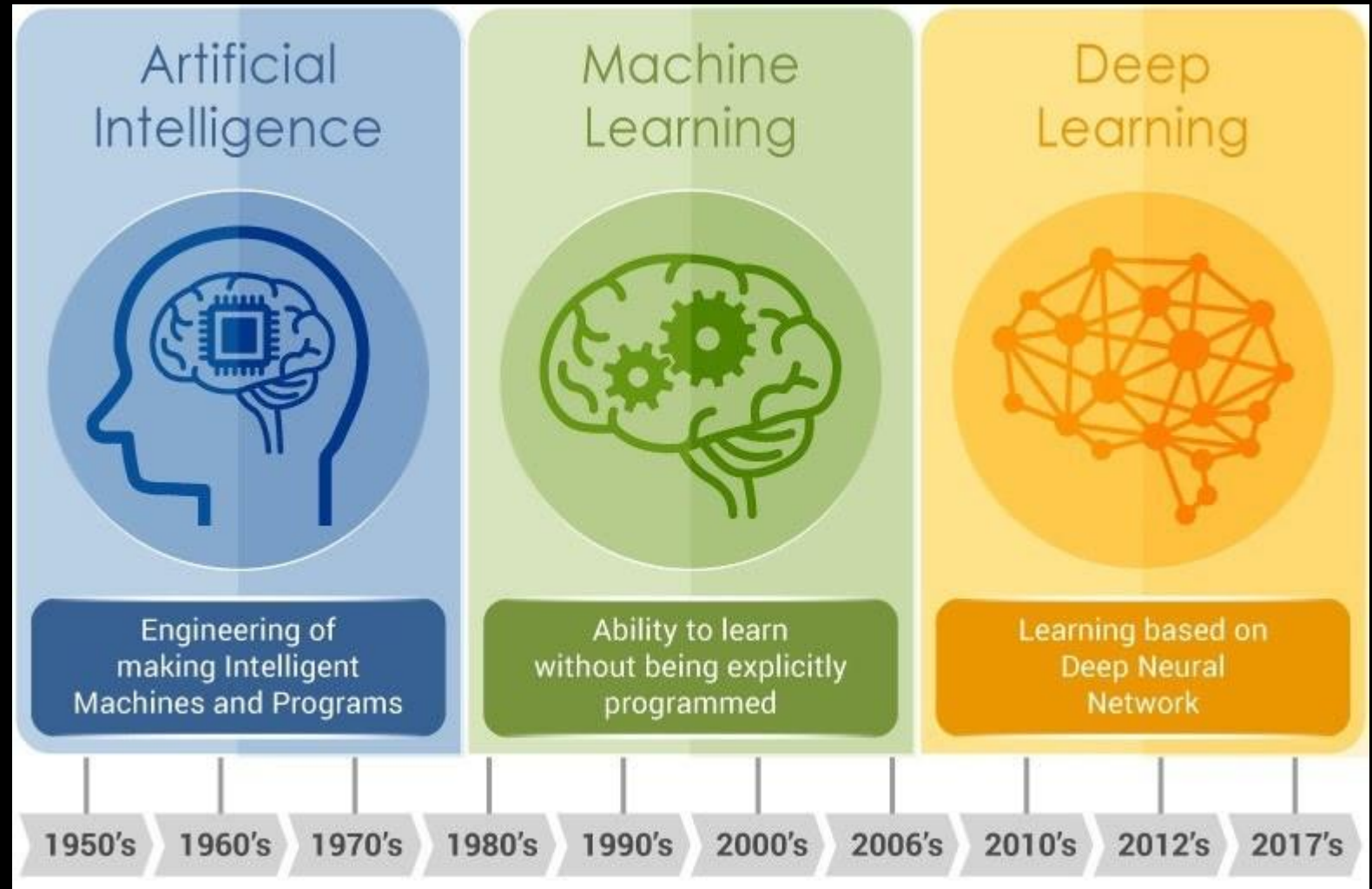
A Inteligência Artificial (IA) não é uma disciplina, mas um campo de estudo multidisciplinar

# Breve História da Inteligência Artificial



Alan Turing  
(1912-1954)

Allen Newell e  
Herbert Simon

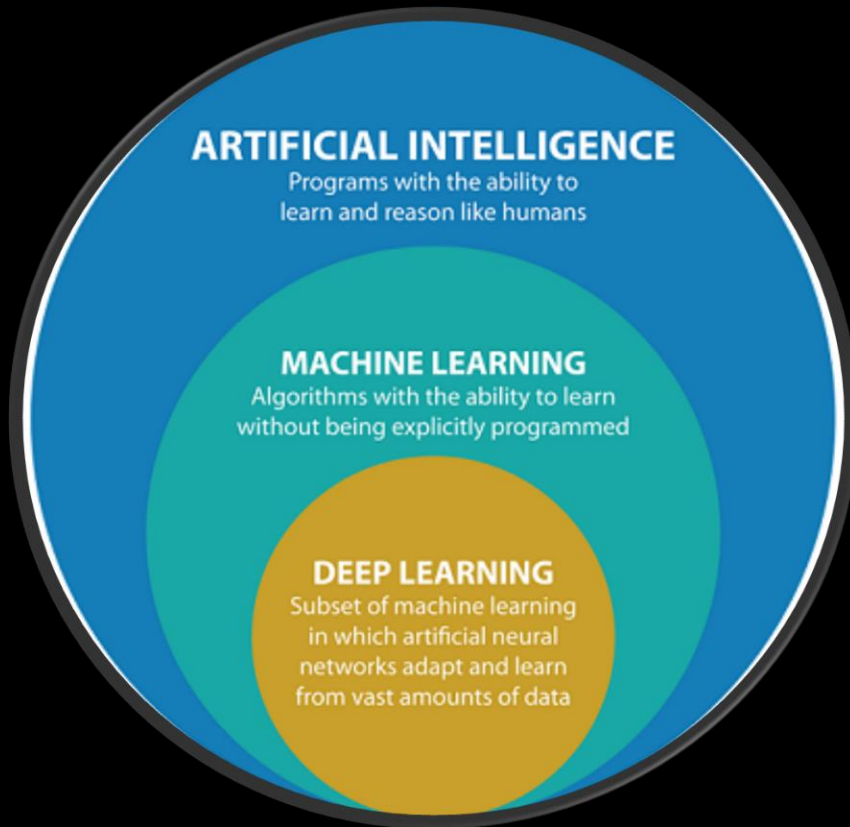


# A conferência de Dartmouth (1956)

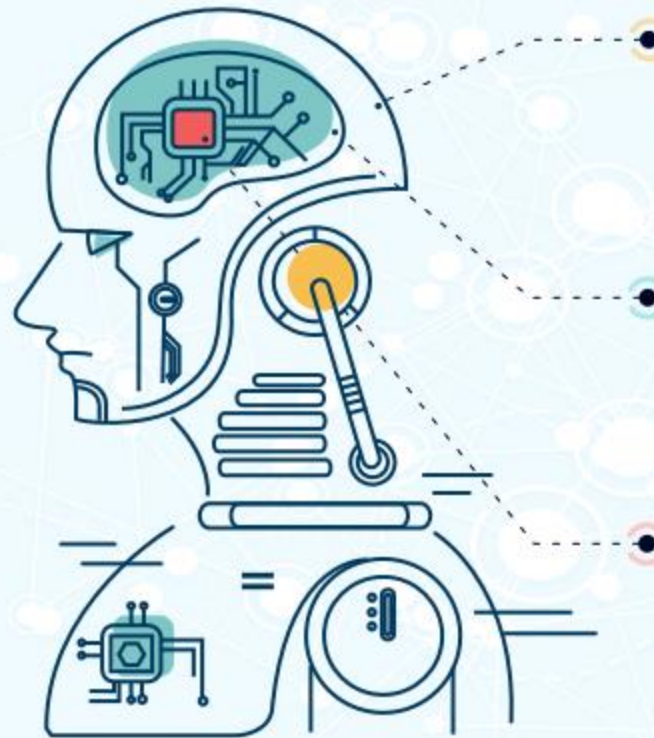
- A IA começa efetivamente numa conferência em Hanover (New Hampshire, EUA), em 1956, com um grupo de cientistas: John McCarthy, Marvin Minsky, Claude Shannon e Nathaniel Rochester. Foi patrocinado pela Fundação Rockefeller e reuniu vários teóricos da área da computação. É nessa conferência que o termo *Artificial Intelligence* é criado, por McCarthy, para nomear a nova ciência que estava a surgir, que mais parecia ficção científica.



# IA – ML - DL



## Diferenças entre Inteligência Artificial, Machine Learning e Deep Learning



### Inteligência Artificial (IA)

Campo que estuda como criar programas computacionais com a **capacidade de aprender e raciocinar como os humanos** para resolver problemas de forma criativa.

### Machine Learning (ML)

Aplicação da IA dedicada à criação de algoritmos que permitem que os sistemas **aprendam sem intervenção humana**, ou seja, sem a necessidade de programação explícita.

### Deep Learning (DL)

Subconjunto do ML focado na **criação de redes neurais artificiais**, ou seja, sistemas que imitam o cérebro humano, adaptando-se e aprendendo a partir de grandes quantidades de dados.

- O quanto um computador consegue imitar o cérebro humano?
- O cérebro humano funciona como um computador?
- Conhecer o funcionamento da máquina poderá ajudar a compreender melhor o cérebro humano e as funções mentais superiores?
- A tecnologia irá criar um novo tipo de humano?
- Quais as consequências socioeconômicas da IA?
- Poderá a máquina tornar-se mais inteligente do que o Homem?
- Irá a máquina responder às grandes questões da Humanidade?

# Diferentes abordagens ou paradigmas da IA

---

Atualmente, o estudo da IA tem várias abordagens, diferenciando-se pela forma de representação do conhecimento adquirido, que poderá ser por meio de símbolos, de conexões entre neurónios (sinapses) e de cromossomas, dando origem a diferentes abordagens.



Paradigma	Inspiração
Simbólico	Semiologia, Lógica, Lingüística
"Navegacionista"	Algorítmica, Pesquisa Operacional
"Restricionista"	Álgebra Linear, Cálculo
Probabilista	Probabilidades e Estatística
Conexionista	Neurologia
Evolucionista	Paleontologia
Multi-Agentes	Sociologia, Lingüística, Entomologia

# Inteligência Artificial: Forte e Fraca

---

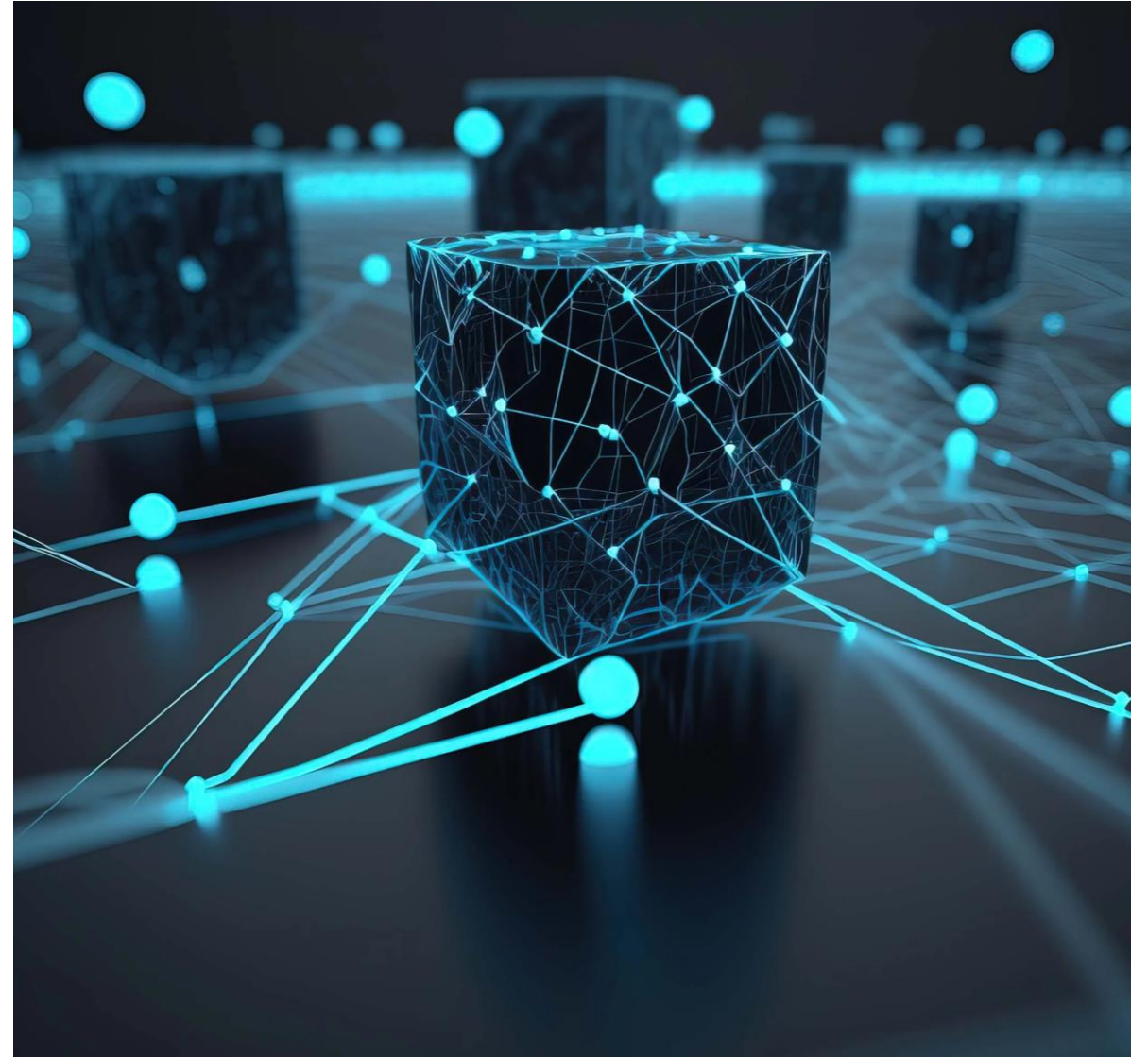


## **Forte**

As máquinas têm a capacidade de pensar e adquirir novos conhecimentos.

## **Fraca**

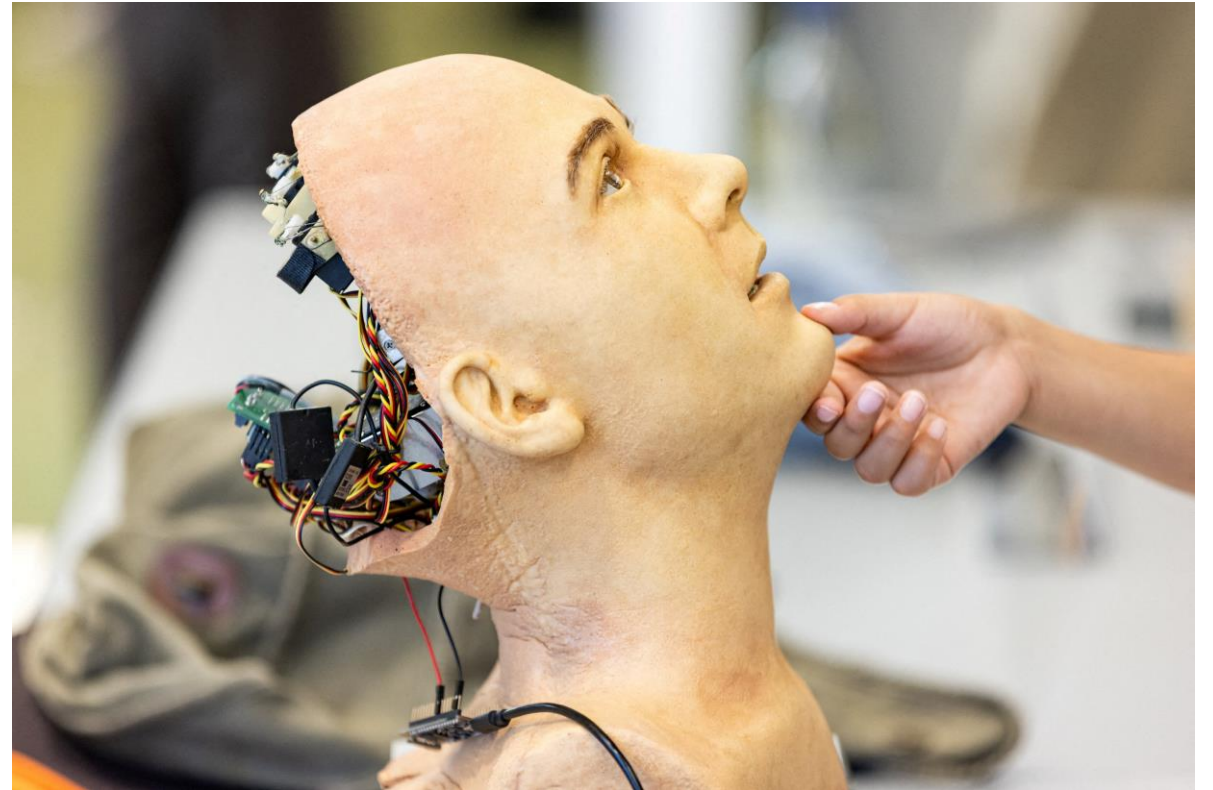
As máquinas simulam apenas comportamentos humanos programados.

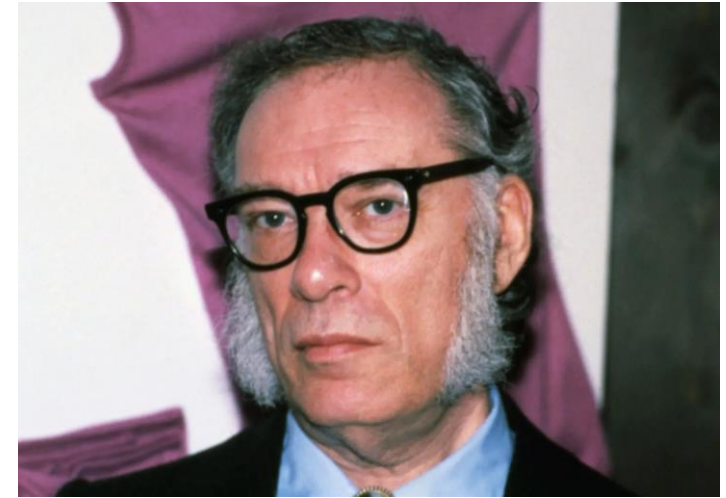
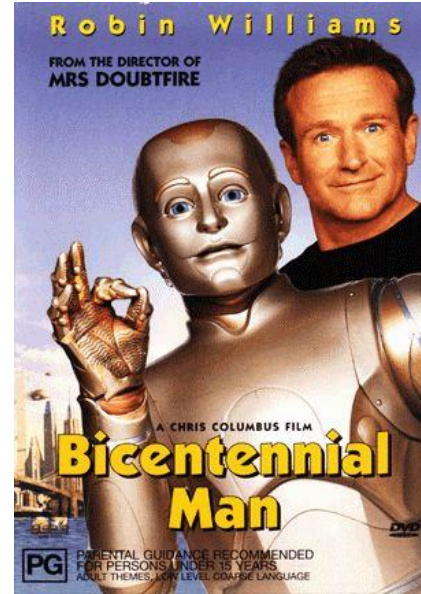
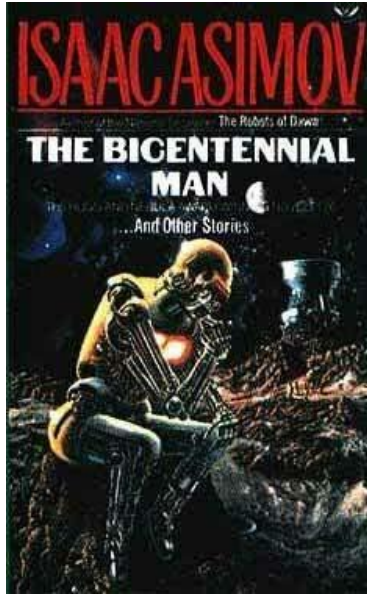


# IA Forte

---

- Enquanto a IA fraca é uma simulação do pensamento. A IA forte pretende ser uma execução do pensamento.
- Capacidade de raciocinar e resolver problemas.
- Uma forma de auto-consciência.
- Polêmica: envolve temas de consciência e ética.
- O que fazer com uma entidade cognitivamente semelhante ao ser humano?
- Tema abordado pela ficção científica: Isaac Asimov.





## IA Forte na Ficção Científica

- “O Homem Bicentenário” de Isaac Asimov – história de um *robot* consciente e inteligente que luta por possuir, na sociedade, um *status* semelhante ao de um ser humano.
- Asimov: 3 leis da robótica, reduz os *robots* a servos dos seres humanos. A etimologia da palavra *robot* vem da língua checa e significa “trabalho forçado, servidão”.



# As Três Leis da Robótica

## – Isaac Asimov

I - Um *robot* não pode ferir um humano ou permitir que um humano sofra algum mal.

II - Os *robots* devem obedecer às ordens dos humanos, exceto nos casos em que tais ordens entrem em conflito com a primeira lei.

III - Um *robot* deve proteger a sua própria existência, desde que não entre em conflito com as leis anteriores.

“*Eu, Robot*” (1950)

Lei Zero: Um *robot* não pode causar mal à humanidade ou, por omissão, permitir que a humanidade sofra algum mal.

Seguirão os humanos  
essas leis?

# Rebelião das Máquinas – R.U.R. (Karel Capek, 1920)

---

Nesta peça teatral, onde o nome *robot* é introduzido, assiste-se a uma revolta numa fábrica de produção de pessoas artificiais a partir de matéria orgânica sintética (mais próximo de clones). Inicialmente, trabalham felizes para os humanos, mas uma revolta põe fim à espécie humana.



<https://futureoflife.org/cause-area/artificial-intelligence/>

- As máquinas podem ter consciência? Podem ter objetivos e ter um comportamento orientado para objetivos.
- As máquinas podem ser más? Podem ser competentes a atingir os seus objetivos. Por isso é preciso garantir que os seus objetivos estejam alinhados com os nossos.
- As máquinas podem controlar os humanos? Não controlamos os tigres porque somos mais fortes, mas porque mais inteligentes. Se cedermos a nossa inteligência às máquinas, também poderemos estar a ceder o controlo.



RECOMMENDED READING

### Benefits & Risks of Artificial Intelligence

From SIRI to self-driving cars, artificial intelligence (AI) is progressing rapidly. While science fiction often portrays AI as robots with human-like characteristics, AI can encompass anything from Google's search algorithms to IBM's Watson to autonomous weapons.

[Read article](#) →

# Singularidade tecnológica e Superinteligência

- **Singularidade tecnológica** - um hipotético momento no futuro em que o crescimento tecnológico se tornaria incontrolável e irreversível, com consequências imprevisíveis para a civilização humana.
- **Superinteligência** - um agente hipotético detentor de uma inteligência que ultrapassa em muito as mentes humanas mais brilhantes e talentosas.

**Como seria a vida dos seres humanos num mundo  
pós-singularidade?**

# Quem tem medo do Frankenstein?

## Complexo de Frankenstein

Termo criado por Isaac Asimov, inspirado no romance de Mary Shelley – *Frankenstein* –, no qual um cientista cria um ser inteligente que acaba por se voltar contra ele, para descrever o medo diante dos avanços da robótica e da criação de inteligência artificial.

## Complexo de Frankenstein ≠ Síndrome de Frankenstein

### Síndrome de Frankenstein

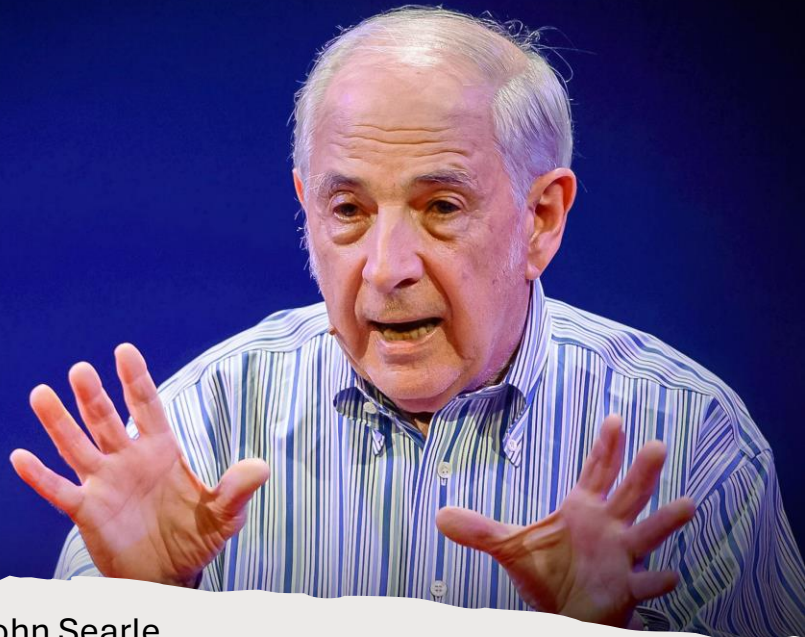
Ao sentirem-se vítimas de intolerância e preconceito, algumas pessoas assumem comportamentos de “monstros” perante a sociedade, como mecanismo de defesa contra a agressividade que sentem vir dos outros.





# Poderão as Leis da Robótica impedir o genocídio?

- Se os *robots* assumirem que, num futuro, outras vidas inteligentes e não-humanas (alienígenas) são uma ameaça para a humanidade, estarão programados para os exterminar?
- Se um grupo de indivíduos (re)definir nos *robots* o conceito de humano, excluindo alguma parte da humanidade, poderemos ter uma limpeza étnica ou algo semelhante?
- Saberão os *robots* lidar com dilemas éticos ou conflitos na aplicação das Leis da Robótica? (Ex.: *Runaround*, Asimov).



John Searle



Marvin Minsky

# Debate Searle-Minsky sobre IA Forte

O debate entre John Searle (filósofo da mente) e Marvin Minsky (cientista da computação) sobre Inteligência Artificial Forte concentra-se na questão da consciência. Searle argumenta que a simulação de inteligência por uma máquina não implica compreensão verdadeira, enquanto Minsky acredita que a simulação (sem consciência) é suficiente para afirmar que a máquina é verdadeiramente inteligente. Experimento mental do “Quarto Chinês”, de Searle.

# IA Fraca e as suas questões éticas

Perspectivas Críticas sobre Algoritmos, Dados e Máquinas Inteligentes

Nos últimos anos, tem havido uma explosão nos sistemas de IA, acompanhada por uma análise crítica crescente dessas tecnologias. O foco dessas análises tem sido principalmente os estudos feministas, que exigem que a IA seja desenvolvida e implementada de forma responsável, promovendo a busca pela justiça social. Este movimento também está atento às discussões na indústria sobre os riscos e oportunidades relacionados à adoção da IA. Esta coleção reflete a crescente interconexão entre academia, indústria e grupos de pesquisa corporativos, combinando *insights* feministas interdisciplinares com estudos pós-coloniais, teoria da deficiência e estudos críticos de raça para abordar preconceitos de idade, racismo, sexismo, capacitação e opressões de classe na IA.

OXFORD

EDITED BY

jude BROWNE

stephen CAVE

eleanor DRAGE

kerry MCINERNEY

## feminist AI

critical  
perspectives  
on data,  
algorithms and  
intelligent  
machines

A person in a dark blue pinstriped suit and tie is shown from the chest up. Instead of a head, there is a black tablet. The tablet screen displays a blue and white target icon with an arrow hitting the bullseye. A pink brushstroke is visible on the left side of the tablet screen. The background is a dark blue textured surface with a large, semi-transparent orange circle behind the person's head area.

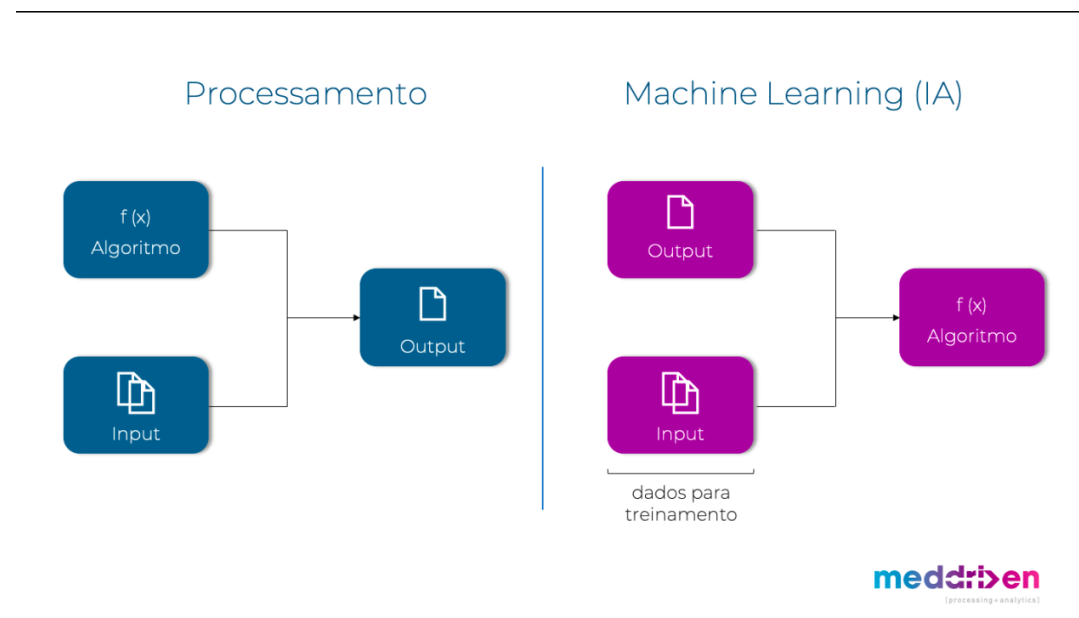
Uma máquina capaz de reproduzir os  
nossos preconceitos?

# Algoritmo – A Fórmula de Deus

Um algoritmo é um conjunto de regras ou procedimentos passo-a-passo para a solução de um problema.

Os algoritmos estão na base e são a essência da IA. São eles que definem as regras, os critérios e os processos que vão permitir que as máquinas e os dispositivos aprendam, tomem decisões e executem ações de forma autónoma e adaptativa.

Na IA, não são os humanos a criar o algoritmo, isto é, a estabelecer o procedimento pelo qual a tarefa deve ser realizada, mas ele é fruto dos dados que inserimos (*input*). Os dados aprimoram o algoritmo.





Alimentamos um Deus ou um ídolo?



# O problema do viés na IA

O viés é a tendência de distorção.

- A IA está a amplificar os vieses humanos.
- De que depende a qualidade dos dados e do algoritmo?
- Que processos se podem estabelecer para reduzir o viés na IA?
- A presença de vieses, produzindo resultados distorcidos, compromete o potencial de uso e aumenta a desconfiança na sua utilização.
- Google e IBM a trabalhar em processos de redução de vieses.

# Tipos de vieses

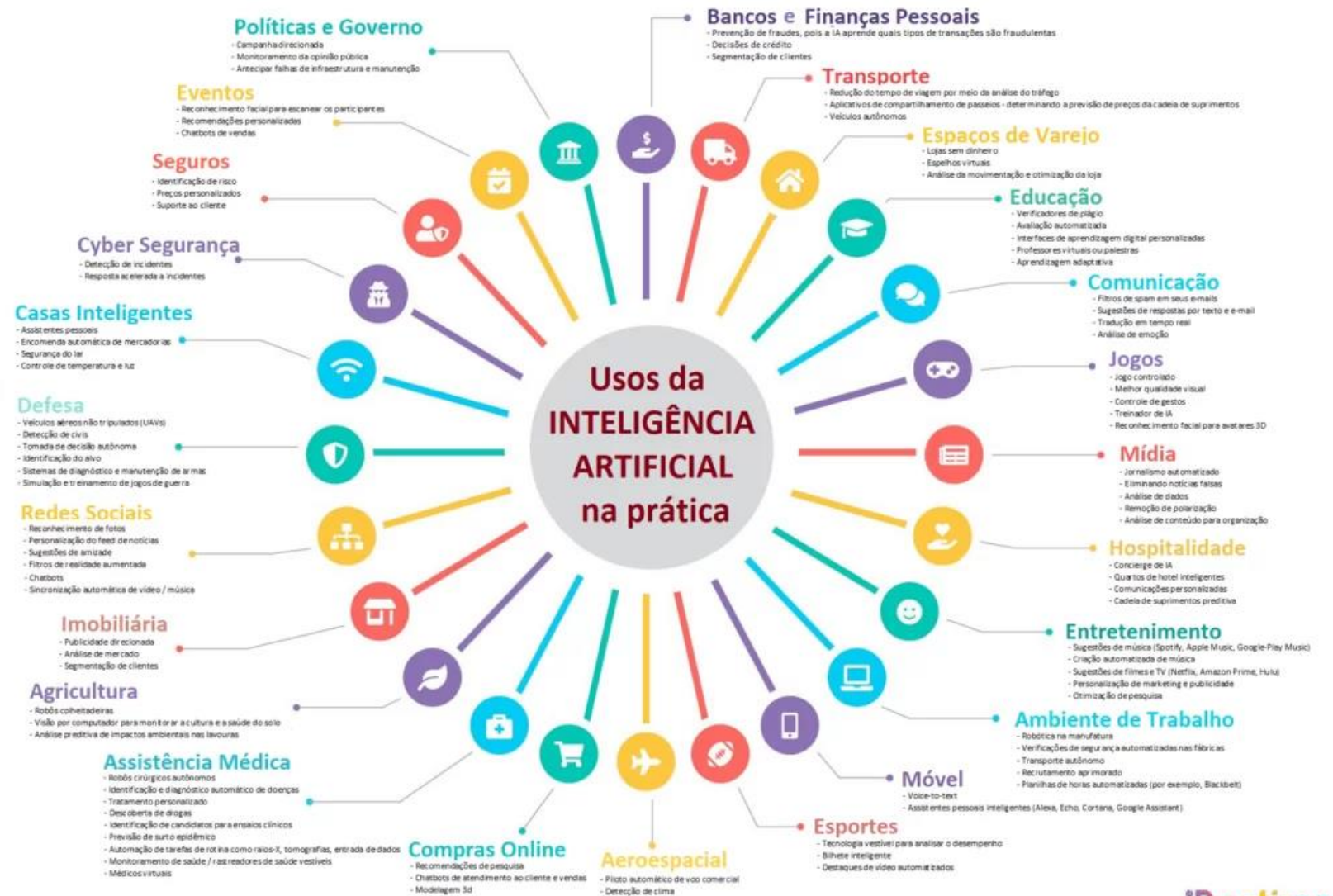
- **Viés de Confirmação** – quando se procura apenas evidências que confirmem as suas crenças, ignorando as informações que as contradigam. Visão distorcida da realidade. Difícil o diálogo.
- **Viés de Disponibilidade** – quando se dá mais importância às informações mais acessíveis à sua memória, conduzindo a conclusões precipitadas e a generalizações injustas.
- **Viés de Grupo** – quando se adota uma visão apenas por ser compartilhada pelo seu grupo, limitando a capacidade de reflexão e capacidade crítica.
- **Viés de Autoridade** – quando se aceita uma visão por ser defendida por uma autoridade reconhecida, limitando a capacidade de formar opiniões independentes.
- **Viés de Atribuição** – quando alguém atribui causas a eventos a partir das suas próprias crenças, levando a interpretações distorcidas da realidade.



# Como educar para evitar os vieses?

- Estar aberto a diferentes perspectivas e argumentos.
- Considerar diferentes pontos de vista antes de formar uma opinião.
- Questionar as suas próprias crenças e preconceitos, procurando constantemente expandir o conhecimento e a compreensão.

# Qual é o potencial da IA para o ser humano?



Traduzida por @neigrando

Marcos Bazmandegan



# Padrões de IA

- **Reconhecimento:** máquinas que podem reconhecer o mundo.
- **Conversacional:** máquinas que podem comunicar-se como seres humanos.
- **Análise preditiva:** ajudando os humanos a tomarem melhores decisões.
- **Sistemas orientados para objetivos:** resolução de problemas complexos.
- **Sistemas autônomos:** reduzindo a necessidade de trabalho manual.
- **Identificação de padrões e anomalias:** descobre o que combina e o que não combina,
- **Hiperpersonalização:** atende as características de um indivíduo.





# Comissão Europeia aprova Regulamento sobre Inteligência Artificial

26 de Março 2024

O Parlamento Europeu aprovou, este mês, o Regulamento sobre Inteligência Artificial (IA) que entra em vigor no próximo dia 21 de abril e define o período de até 21 de abril de 2025 para que as empresas se possam adaptar às novas regras.

Este é mais um passo na liderança tecnológica da UE, em total respeito pelos direitos fundamentais dos cidadãos europeus.

O regulamento original foi proposto pela Comissão Europeia, em abril de 2021, procurando definir as regras-base para garantir a segurança, a ética e a inovação da IA sem descurar.

Fonte: <https://www.compete2030.gov.pt/comunicacao/noticias-pt-informacao/comissao-europeia-aprova-regulamento-sobre-inteligencia-artificial/>

# Regulamento Inteligência Artificial

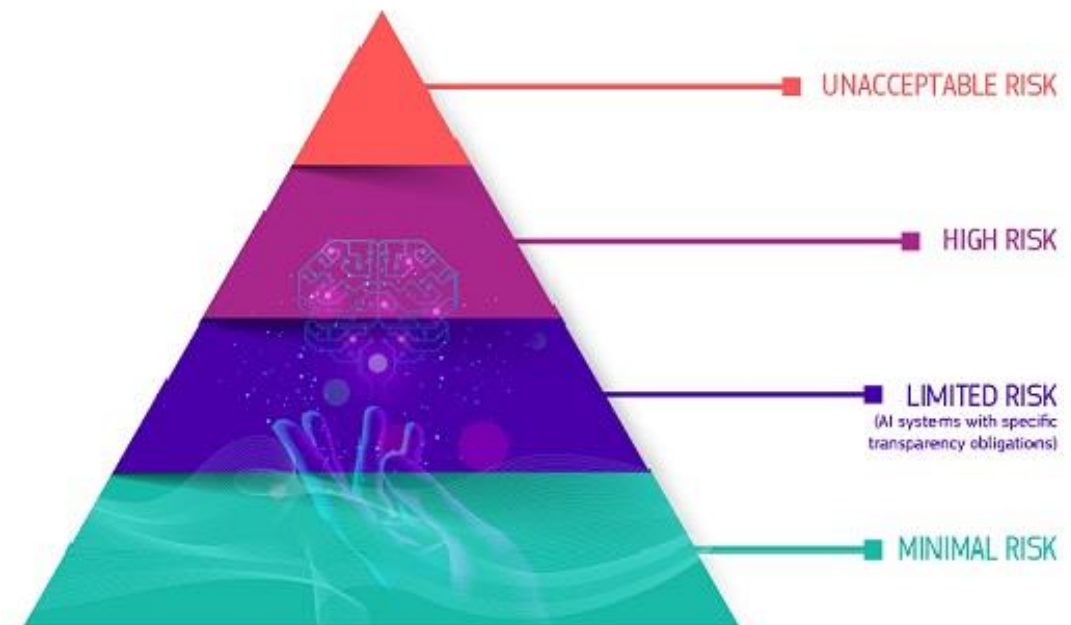
O Regulamento Inteligência Artificial é o primeiro quadro jurídico em matéria de IA, que aborda os riscos da IA e posiciona a Europa para desempenhar um papel de liderança a nível mundial.

Uma abordagem baseada no risco.

<https://digital-strategy.ec.europa.eu/pt/policies/regulatory-framework-ai>

Para ver o Regulamento:

<https://eur-lex.europa.eu/legal-content/PT/TXT/?uri=CELEX%3A52021PC0206>





# Regulate AI?

## How US, EU and China Are Going About It

Por Courtney Rozen e Jillian Deutsch

13 de março de 2024

Os governos não têm um grande histórico de acompanhamento das tecnologias emergentes. Mas o campo complexo e em rápida evolução da inteligência artificial levanta preocupações jurídicas, de segurança nacional e de direitos civis que não podem ser ignoradas. A União Europeia aprovou uma lei abrangente que colocaria barreiras de proteção à tecnologia; na China, nenhuma empresa pode produzir um serviço de IA sem as devidas aprovações. Os EUA ainda estão a trabalhar na sua abordagem regulamentar. Enquanto o Congresso analisa a legislação, algumas cidades e estados americanos já aprovaram leis que limitam o uso da IA em áreas como investigações policiais e contratações, e o Presidente Joe Biden instruiu as agências governamentais a examinarem futuros produtos de IA quanto a potenciais riscos de segurança nacional ou económica.

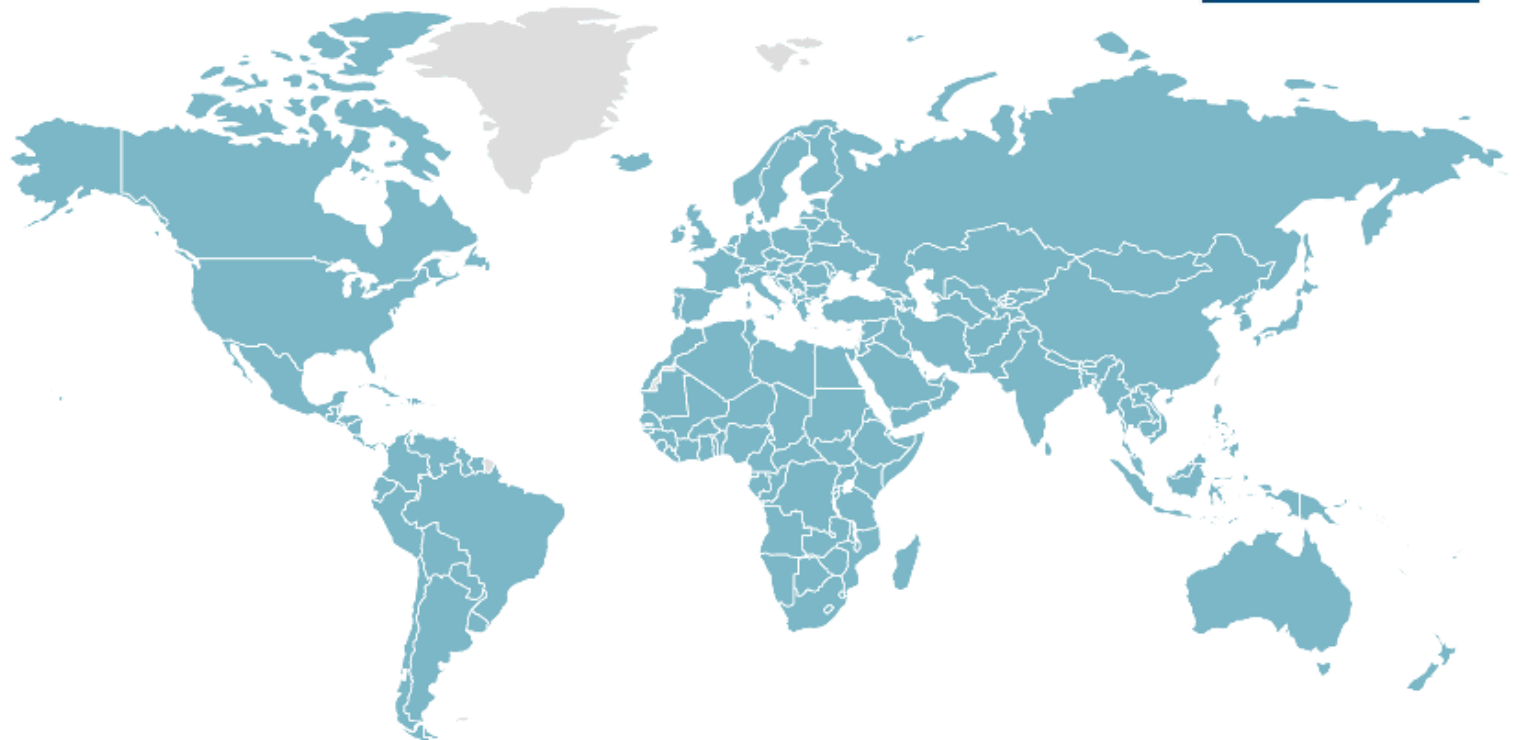
Fonte: <https://www.bloomberg.com/news/articles/2024-03-13/regulate-ai-how-us-eu-and-china-are-going-about-it>

# UNESCO – Recomendação sobre a Ética da Inteligência Artificial



A UNESCO é uma agência especializada das Nações Unidas. A abreviação significa "Organização Educacional, Científica e Cultural das Nações Unidas". Ela foi fundada em 1945 para promover a paz e a segurança através da cooperação internacional em educação, ciência, cultura e comunicação. Seu principal objetivo é contribuir para a construção de uma cultura de paz, erradicando a pobreza, o desenvolvimento sustentável e o diálogo intercultural através de seus programas e atividades.

A UNESCO conta atualmente com um total de 195 países com uma população combinada de 7,99 bilhões.



Os **objetivos** da presente Recomendação são os seguintes:

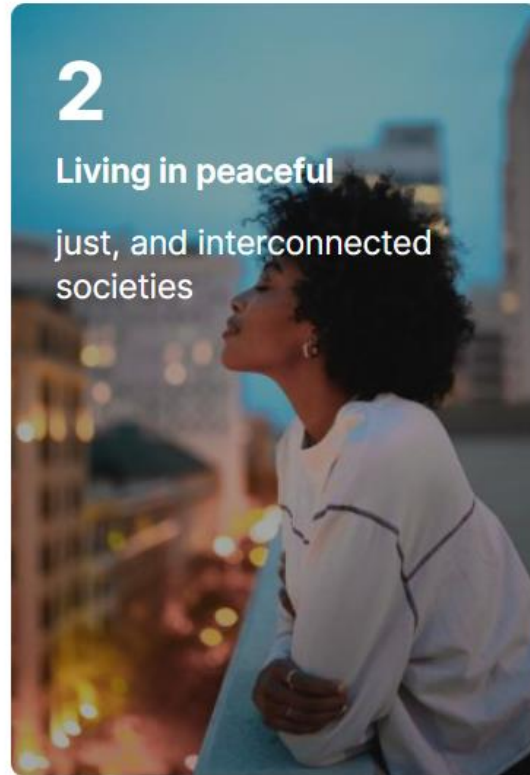


- (a) fornecer um marco universal de valores, princípios e ações para orientar os Estados na formulação de suas legislações, políticas ou outros instrumentos relativos à IA, em conformidade com o direito internacional;
- (b) orientar as ações de indivíduos, grupos, comunidades, instituições e empresas do setor privado para garantir a incorporação da ética em todas as etapas do ciclo de vida dos sistemas de IA;
- (c) proteger, promover e respeitar os direitos humanos e as liberdades fundamentais, a dignidade e a igualdade humana, incluindo a igualdade de gênero; salvaguardar os interesses das gerações presentes e futuras; preservar o meio ambiente, a biodiversidade e os ecossistemas; e respeitar a diversidade cultural em todas as fases do ciclo de vida dos sistemas de IA;
- (d) promover o diálogo multidisciplinar e pluralista com as várias partes interessadas e a construção de consenso sobre questões éticas relacionadas aos sistemas de IA;
- (e) promover o acesso equitativo a avanços e conhecimentos no campo da IA e o compartilhamento dos benefícios, com atenção especial às necessidades e às contribuições dos países de baixa renda, incluindo os países menos desenvolvidos.

# Valores fundamentais:



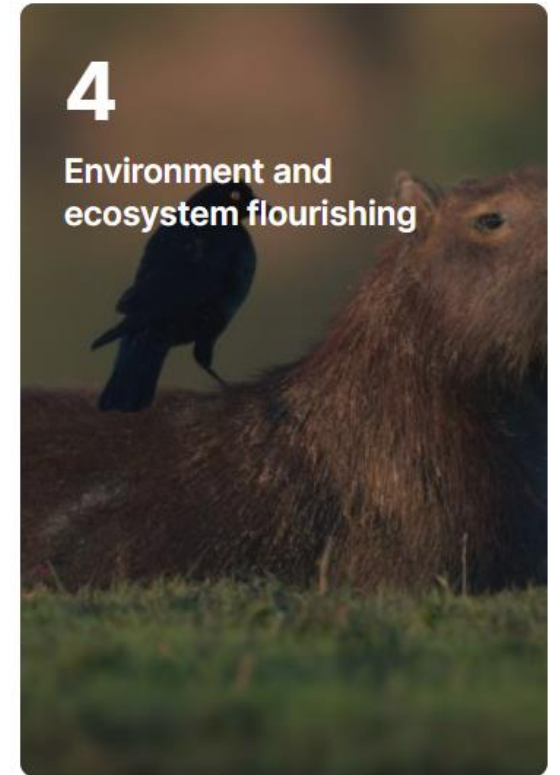
**1**  
**Human rights and human dignity**  
Respect, protection and promotion of human rights and fundamental freedoms and human dignity



**2**  
**Living in peaceful just, and interconnected societies**



**3**  
**Ensuring diversity and inclusiveness**



**4**  
**Environment and ecosystem flourishing**

## Documento da UNESCO na íntegra (em português):

[https://unesdoc.unesco.org/ark:/48223/pf0000381137\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000381137_por)

Recomenda que os Estados-membros apliquem de **maneira voluntária** as disposições da presente Recomendação tomando medidas adequadas, incluindo quaisquer medidas legislativas ou de outra natureza que sejam necessárias, em conformidade com as práticas constitucionais e os marcos governamentais de cada Estado, para efetivar, dentro de suas jurisdições, os princípios e as normas da Recomendação, em conformidade com o direito internacional, inclusive as leis internacionais de direitos humanos.





Qual é o real poder da ONU na geopolítica atual?

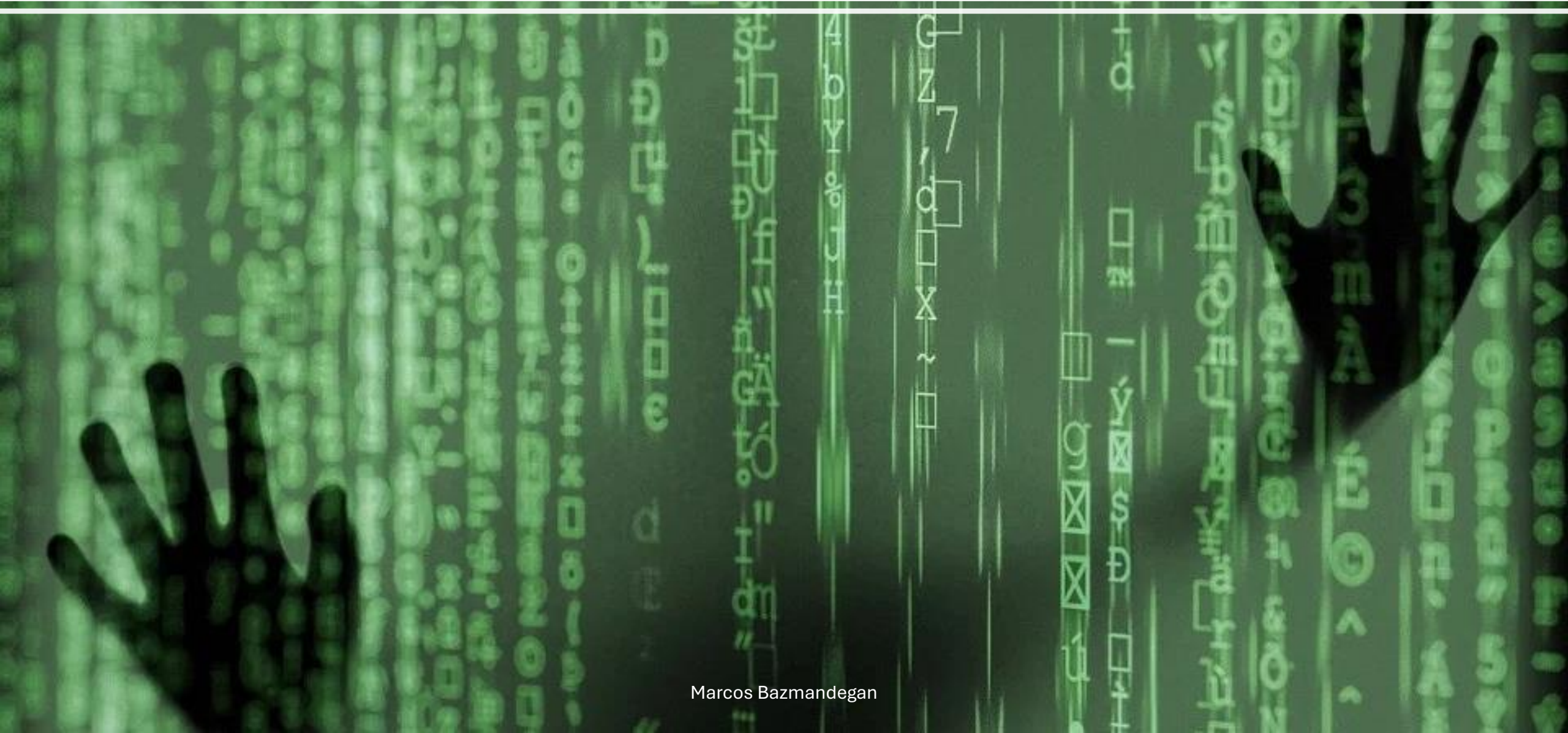
# Qual será o papel da IA na Economia da Felicidade?

A Economia da Felicidade é uma área de estudos que combina Economia, Psicologia e Sociologia para estudar como fatores económicos afetam o bem-estar e a qualidade de vida das pessoas, indo além do PIB para medir a prosperidade. Foca em indicadores como saúde, educação, equilíbrio entre trabalho e vida pessoal, e relacionamentos sociais, procurando políticas que promovam a felicidade e o desenvolvimento sustentável em vez do mero crescimento económico.



O algoritmo redefinirá o ser humano?

# Estaremos aprisionados ao virtual?





Marcos Bazmandegan


Pensar Global  
Compreender a  
Complexidade

Edgar Morin

Inteligência Coletiva  
Cibercultura

Pierre Levy





Estaremos a criar um novo Deus?  
Um Deus ex Machina?