



IA

Prof. Maurício Veloso Brant Pinheiro
Dep. Física – ICEX - UFMG

IA

Prof. Maurício Veloso Brant Pinheiro

Dep. Física – ICEX – UFMG

29/11/2024

Introdução

Definindo Inteligência e Inteligência Artificial

Uma Breve História da IA

Semântica

Perceptrons e Redes Neurais Artificiais

Arquitetura e Treinamento

Kormogorov-Arnold Networks

Exemplo: MLP-NN vs KAN – Oscilador Não-Linear

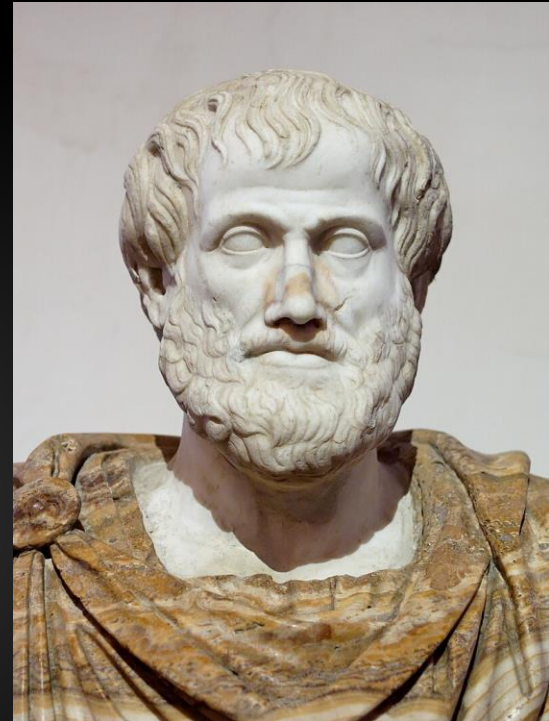
O que nos aguarda..

AGI, o lado sombrio da IA

Uma mensagem otimista (ou não?)

Literatura Recomendada

**"Se cada instrumento pudesse realizar seu próprio trabalho, obedecendo ou antecipando a vontade de outros, (...), os mestres de ofício não precisariam de servos, nem os senhores de escravos." –
*Aristóteles, Política***



Source: Wikipedia

Inteligência é a capacidade de agir de forma proposital, de pensar racionalmente e de lidar efetivamente com o ambiente - *David Wechsler* -



Inteligência Artificial (IA) é um ramo multidisciplinar da computação focado na criação de máquinas inteligentes que realizam tarefas que exigem cognição humana.

Uma Breve História da IA...

1943 McCulloch & Pitts: *A Logical Calculus of Ideas Immanent in Nervous Activity*

1950: Alan Turing: *Turing Machine & Turing Test*

1956: *Conferência de Dartmouth*

1966: John von Neumann: auto-reprodução de máquinas autônomas



1º Inverno
'70

1980: **John Hopfield** et al. *Redes Neurais Recorrentes (RNN)*



2º Inverno
'80

1986: **Geoffrey Hinton** et al. *Back-Propagation*

1997: Deep Blue da IBM vence Kasparov



3º Inverno
'00


1999: MNIST


Uma Breve História da IA...

 Meados dos anos 2000: “Deep Learning” (Geoffrey Hinton)

Internet
'05  2012: NN → ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

Business
'10  2014: Google → DeepMind → AlphaGo vence Lee Sedol

Perception
'15  2017: Adoção massiva de IA pela indústria: Data Science → Big Data → AI (Deep Learning) → 4ª Revolução Industrial

Autonomous
'20  2018: Google/Deepmind → AlphaFold: **David Baker, Demis Hassabis, John Jumper**

2020: GPT-3 da OpenAI: LLM → 175×10^9 parâmetros

2022: ChatGPT da OpenAI: Populartização da Inteligência Artificial

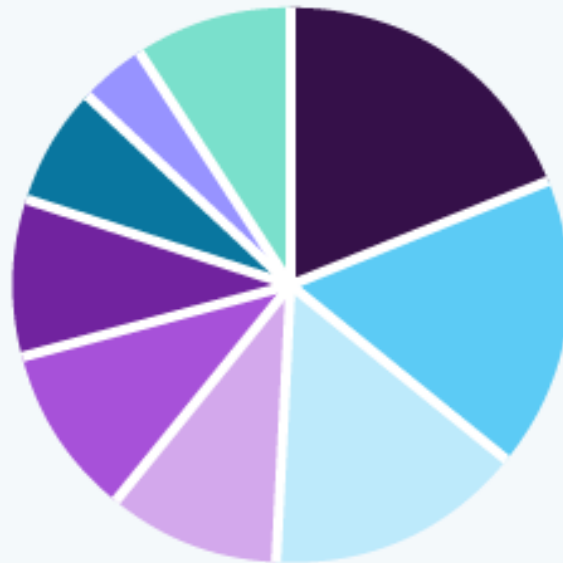
2024: GPT-4, GPT-4o, Gemini, Llama, Claude, MidJourney, Dall-E, Stable Diffusion, Milhões de Apps, Sites, Tools, Start Ups com IA



O MERCADO GLOBAL DE IA

Global Artificial Intelligence Market

Share, by End-use, 2023 (%)



● Advertising & Media ● BFSI ● Healthcare ● Automotive & Transportation
● Manufacturing ● Retail ● Agriculture ● Law ● Others



GRAND VIEW RESEARCH

\$1.5T @ 2030



\$196.6B

Global Market Size,
2023

Source:
www.grandviewresearch.com

BFSI: Banking, financial services and insurance

IA



Aprendizado de Máquina



Sistemas Especialistas
Algoritmos Genéticos
Redes Bayesianas
⋮
Redes Neurais Artificiais

“Learns from Rules/Examples/Data”



O Neurônio Artificial (*Perceptron*)

Aprendizado Profundo (*Deep Learning*)



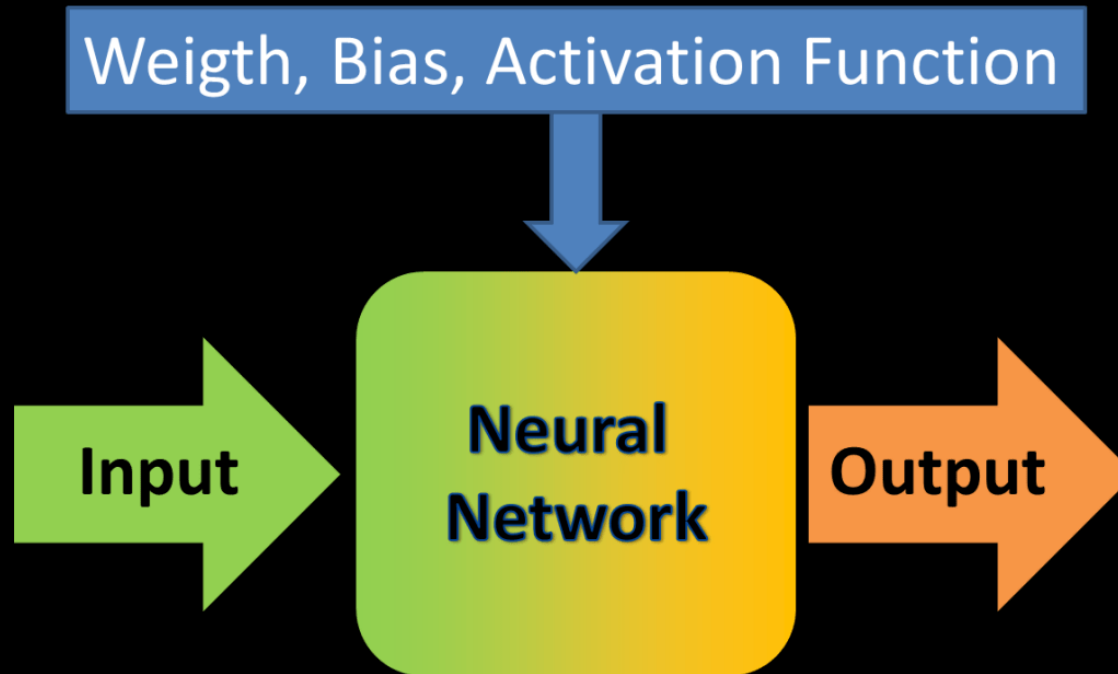
Feed Forward Multi-Layer Perceptron (MLP)
Recurrent Neural Networks (RNN)
Convolutional Neural Networks (CNNs)
Generative Adversarial Networks (GANs)
Large Language Models (LLMs)
Kormogorov-Arnold Networks (KANs)
Decision Trees
⋮

Supervisionado (Predição, Reconhecimento de padrões)
Por Reforço (*Player Agents*, Robótica, HFTs)
Não Supervisionado (*Generative IA*, *Data Analysis*)
⋮



Data is the new oil. Clive Humby, British mathematician and data science entrepreneur (2006)

A rede neural artificial

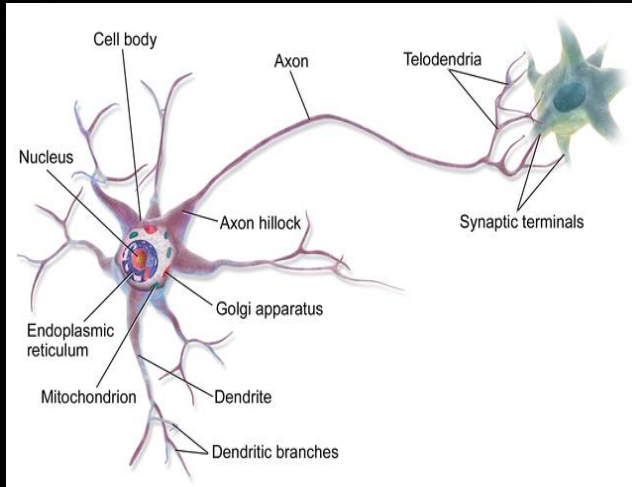


Source: AI-Talks.org

O Neurônio Artificial

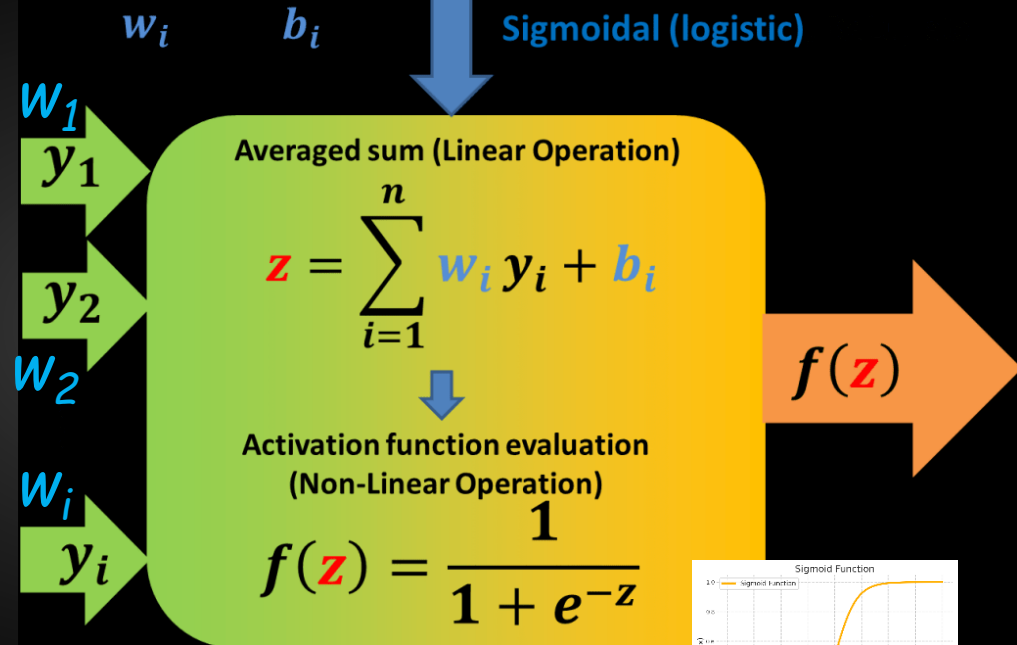
McCulloch, Warren S., and Walter Pitts. "A logical calculus of the ideas immanent in nervous activity." *The bulletin of mathematical biophysics* 5 (1943): 115-133.

The Biological Neuron



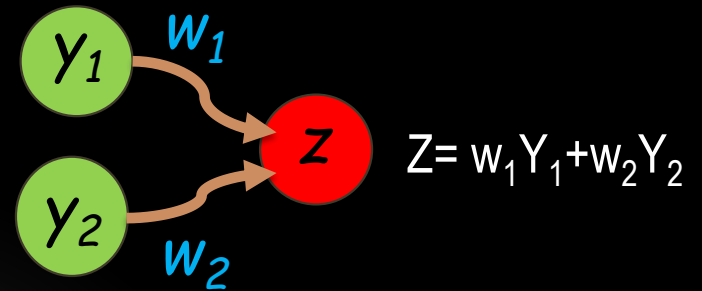
Multipolar Neuron. By BruceBlaus CC BY 3.0 September 30, 2013. Source: Wikimedia Commons.

Weight, Bias, Activation Function



Source: AI-Talks.org

Exemplo: Ensinando um *Perceptron* Linear a somar...



Hyperparameter

Learning Rate (LR) = 0,01

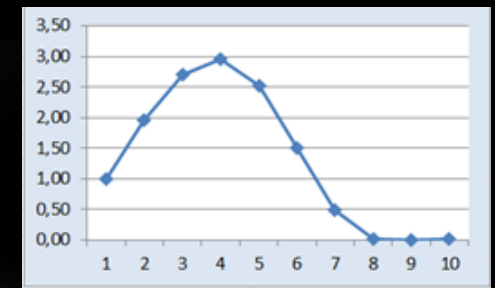
Weights

New Weights

Foward Feed + New Weights	N	y1	y2	y1 + y2	w1	w2	z= w1*y1 + w2*y2	Error = (y1 + y2) - (w1*y1 + w2*y2)	w1+LR*y1*Error	w2+LR*y2*Error
First Training	0	1	1	2	0,40	0,60	1,00	1,00	0,41	0,61
	1	2	2	4	0,41	0,61	2,04	1,96	0,45	0,65
	2	3	3	6	0,45	0,65	3,30	2,70	0,53	0,73
	3	4	4	8	0,53	0,73	5,04	2,96	0,65	0,85
	4	5	5	10	0,65	0,85	7,49	2,51	0,77	0,97
	5	6	6	12	0,77	0,97	10,49	1,51	0,86	1,06
	6	7	7	14	0,86	1,06	13,51	0,49	0,90	1,10
	7	8	8	16	0,90	1,10	15,99	0,01	0,90	1,10
	8	9	9	18	0,90	1,10	18,00	0,00	0,90	1,10
	9	10	10	20	0,90	1,10	20,01	0,01	0,90	1,10

Epoch
 Inputs Y_1, Y_2
 Label $Y_1 + Y_2$

Erro
 $E = Z - (Y_1 + Y_2)$



TEOREMA DA APROXIMAÇÃO UNIVERSAL

O teorema da aproximação universal afirma que, para qualquer função contínua g definida em um conjunto compacto, é possível encontrar uma rede neural com uma única camada oculta que aproxime g .

Para uma rede neural com n valores de entrada, N neurônios na camada oculta

$$F(x_1, x_2, \dots, x_n) = \sum_{j=1}^N \alpha_j f \left(\sum_{i=1}^n x_i w_{ji} + b_j \right)$$

Onde $F(x_1, x_2, \dots, x_n)$ é uma aproximação da função g , tal que:

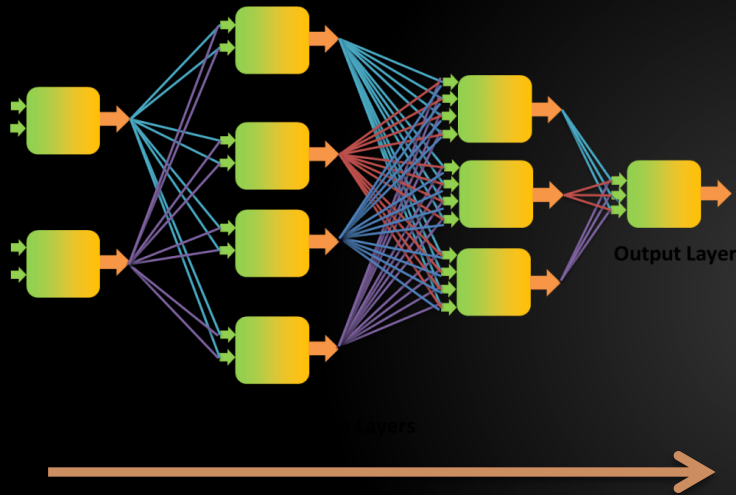
$$|F(x_1, x_2, \dots, x_n) - g(x_1, x_2, \dots, x_n)| < \epsilon$$

Em que ϵ é uma espécie de erro entre a função que se deseja aproximar e a rede.

Hornik, Kurt; Stinchcombe, Maxwell; White, Halbert (January 1989). "Multilayer feedforward networks are universal approximators". *Neural Networks*. **2** (5): 359–366.

Deep Learning Step 1: Deep Neural Network Feed-Foward

$$\tilde{y} = f^{(L)} \left(\mathbf{W}^{(L-1)} f^{(L-1)} \left(\mathbf{W}^{(L-2)} f^{(L-2)} \left(\dots f^{(1)} \left(\mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)} \right) + \mathbf{b}^{(L-2)} \right) + \mathbf{b}^{(L-1)} \right) + \mathbf{b}^{(L)} \right)$$



Deep Neural Network Feed-Foward

Step-by-Step Breakdown:

1. Compute the input to the first hidden layer:

$$\mathbf{z}^{(1)} = \mathbf{W}^{(1)} \mathbf{x} + \mathbf{b}^{(1)}$$

Apply the activation function:

$$\mathbf{a}^{(1)} = f^{(1)}(\mathbf{z}^{(1)})$$

2. Propagate through each subsequent layer $l = 2, 3, \dots, L - 1$:

$$\mathbf{z}^{(l)} = \mathbf{W}^{(l)} \mathbf{a}^{(l-1)} + \mathbf{b}^{(l)}$$

$$\mathbf{a}^{(l)} = f^{(l)}(\mathbf{z}^{(l)})$$

3. Compute the final output for the last layer L :

$$\tilde{y} = f^{(L)} \left(\mathbf{W}^{(L-1)} \mathbf{a}^{(L-1)} + \mathbf{b}^{(L)} \right)$$

Deep Learning Step 2: Backpropagation

Learning representations by back-propagating errors DE Rumelhart, GE Hinton, RJ Williams Nature 323 (6088), 533-536

Error Function

$$E = \frac{1}{2} \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

→ Backpropagation for Gradient Calculation

$$\text{Weights } W_{\text{new}} = W_{\text{old}} - \alpha \frac{\partial E}{\partial W}$$

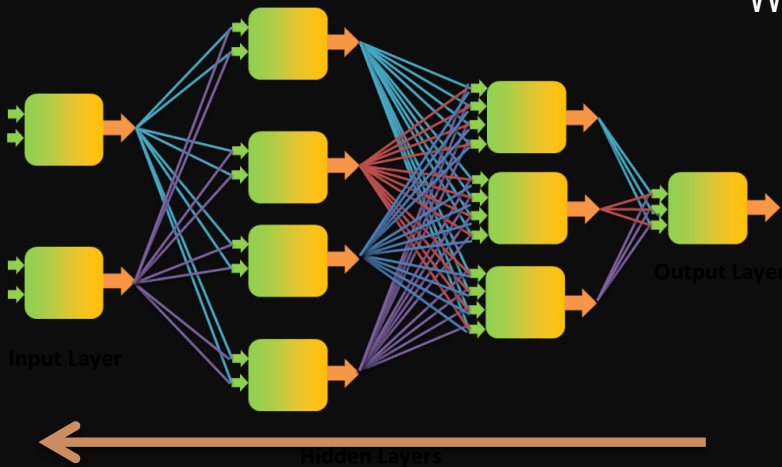
$$\text{Biases } B_{\text{new}} = B_{\text{old}} - \alpha \frac{\partial E}{\partial B}$$

Learning Rate

Partial Derivatives
(via Chain Rule)

$$\frac{\partial E}{\partial W} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial Z} \cdot \frac{\partial Z}{\partial W}$$

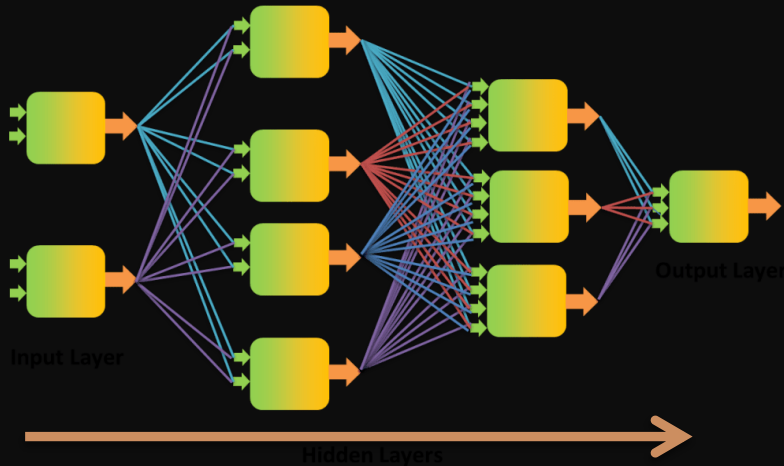
$$\frac{\partial E}{\partial B} = \frac{\partial E}{\partial \hat{y}} \cdot \frac{\partial \hat{y}}{\partial Z} \cdot \frac{\partial Z}{\partial B}$$



Deep Neural Network *Backpropagation*

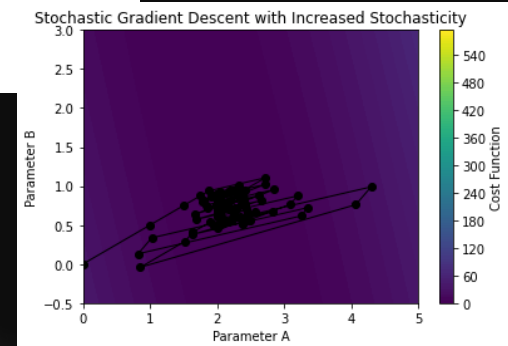
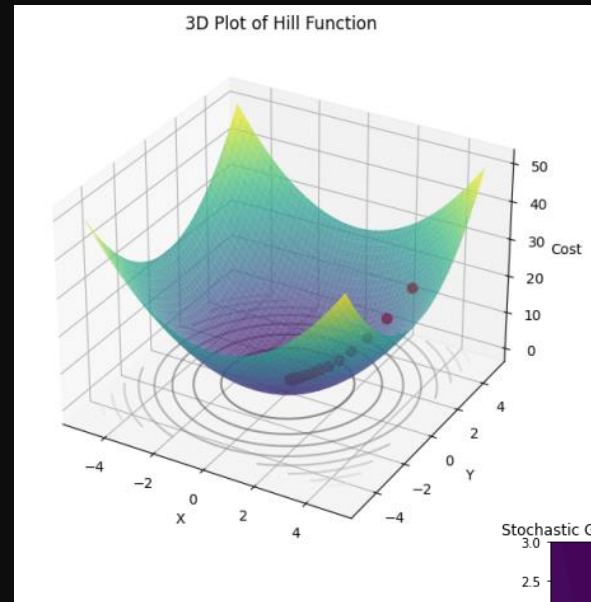
-
-
-

Deep Learning: Step N: Feed Forward + Backpropagation with Stochastic Gradient Descent until the Error reaches a minimum.



Deep Neural Network **Feed-Foward**

Deep Neural Network **Backpropagation**



Kolmogorov-Arnold Networks (KAN)

TEOREMA DE REPRESENTAÇÃO DE KOLMOGOROV-ARNOLD

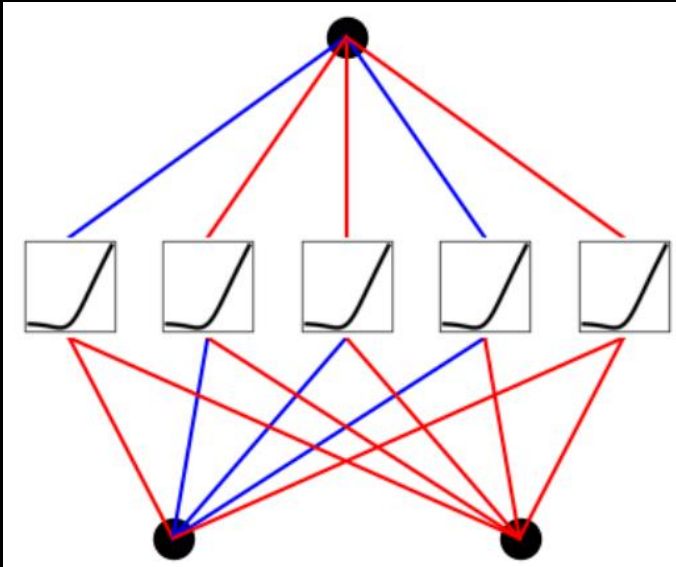
Vladimir Arnold e Andrey Kolmogorov estabeleceram que, se f é uma função contínua multivariada em um domínio limitado, então f pode ser escrita como uma **composição finita de funções contínuas de uma variável e a operação binária de adição**. Mais especificamente, para uma função suave $f: [0,1]^n \rightarrow \mathbb{R}$,

$$f(x) = f(x_1, \dots, x_N) = \sum_{q=1}^{2n+1} \Phi_q \left(\sum_{p=1}^N \varphi_{q,p}(x_p) \right)$$

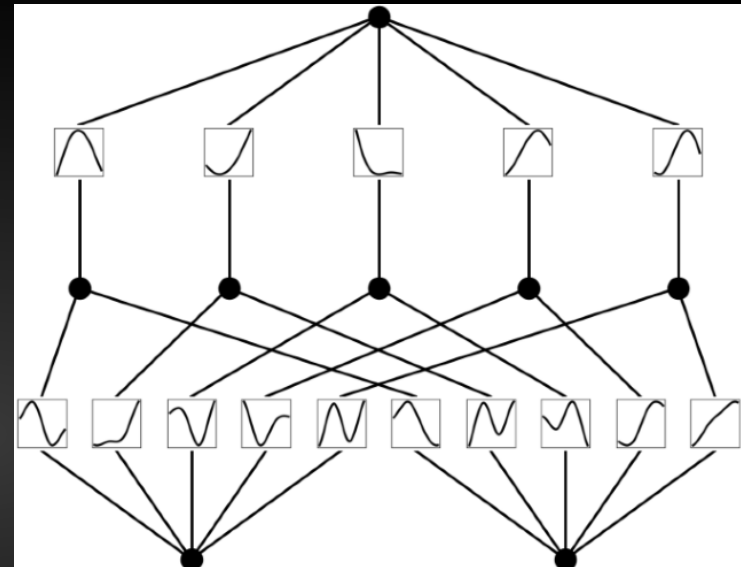
LIU, Ziming et al. KAN: Kolmogorov-Arnold Networks. *arXiv preprint*, 2024.

ARQUITETURA MLP VERSUS KAN

Multi-layer perceptrons (MLPs)



Kolmogorov-Arnold Networks (KANs)



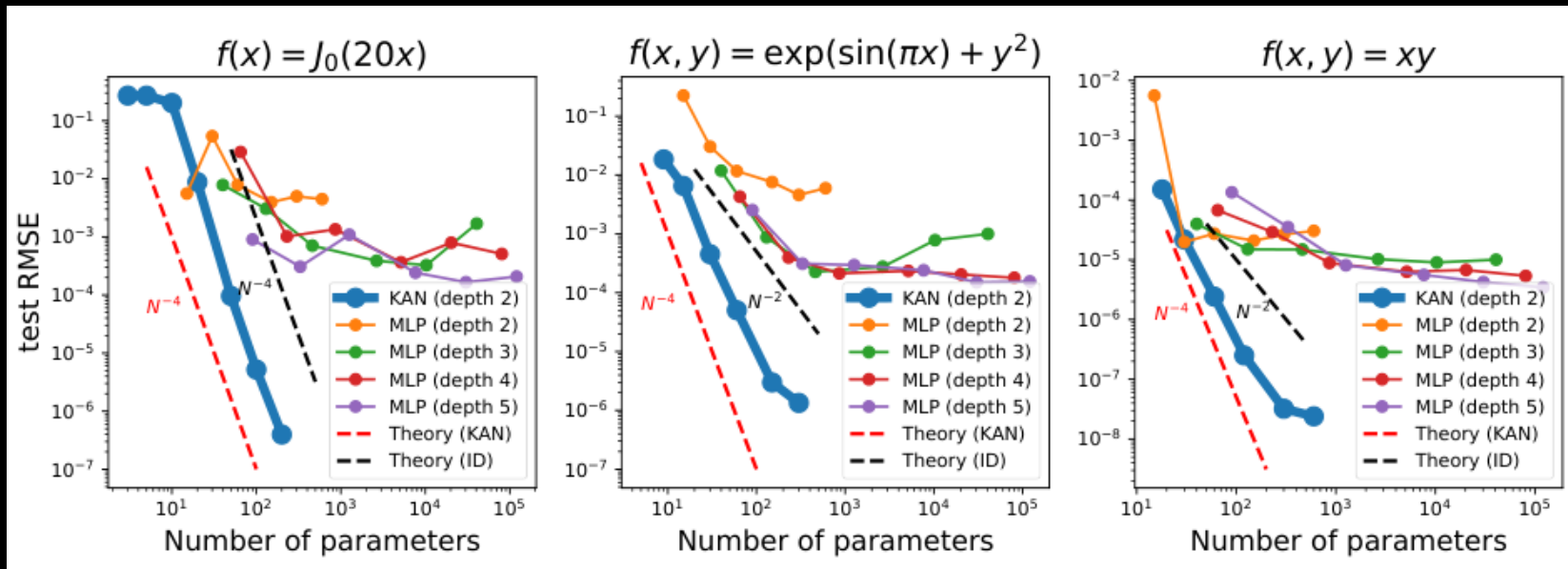
$$MLP(x) = (W_{L-1} \circ \sigma \circ W_{L-2} \circ \sigma \circ \dots \circ W_1 \circ W_0)x$$

$$KAN(x) = (\Phi_{L-1} \circ \Phi_{L-2} \circ \dots \circ \Phi_1 \circ \Phi_0)x$$

$$f(x) = \sum_{i_{L-1}=1}^{n_{L-1}} \varphi_{L-1, i_L, i_{L-1}} \left(\sum_{i_{L-2}=1}^{n_{L-2}} \varphi_{L-2, i_3, i_2} \left(\dots \sum_{i_1=1}^{n_1} \varphi_{1, i_2, i_1} \left(\sum_{i_0=1}^{n_0} \varphi_{0, i_1, i_0}(x_{i_0}) \right) \right) \right)$$

Todas as operações são diferenciáveis: permite usar backpropagation nas KANs.

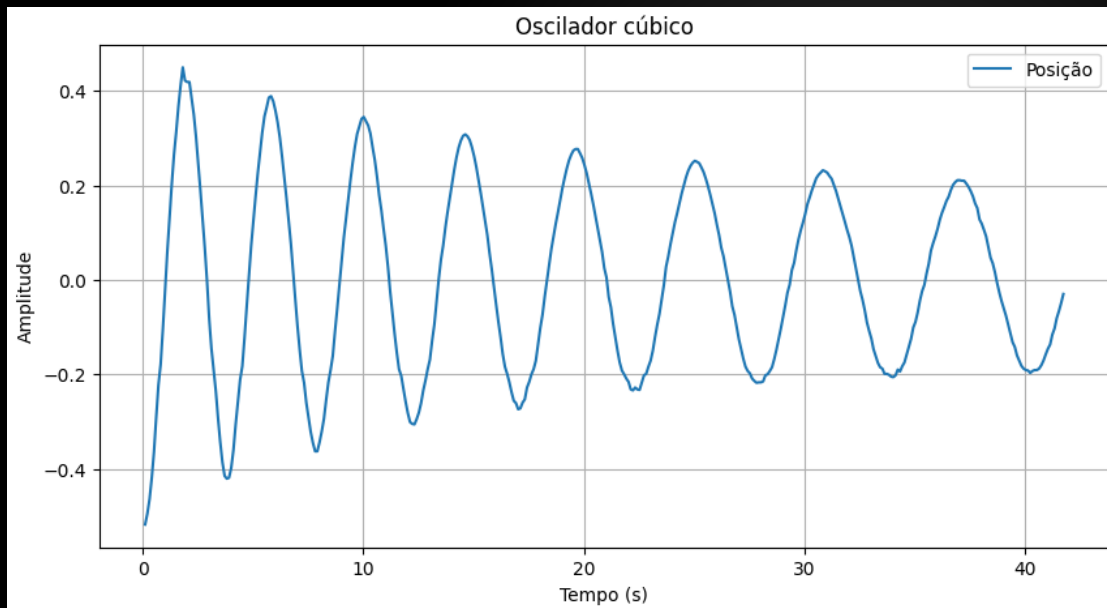
MLP versus KAN



- MLPs: adição de mais neurônios e camadas na rede para melhorar a aproximação.
- KANs tentam **entender** melhor a estrutura da função, dessa forma, conseguem fazer a aproximação de forma mais eficiente, **usando menos neurônios**.
- KANs levam cerca de **10x mais tempo** do que as MLPs para serem treinadas.
- KANs são mais eficazes na representação de funções do que as MLPs em várias tarefas (regressão e resolução de EDPs).

EXEMPLO: OSCILADOR NÃO LINEAR

Quando a força restauradora é não linear, as oscilações do sistema passam a não serem mais harmônicas. Sistemas com essa característica são conhecidos como **osciladores anarmônicos** ou **osciladores não lineares**. Nesses casos a frequência dependerá da amplitude da oscilação.



Cortesia: Rafaela e Dânthanny

O **oscilador cúbico**. Nele a força restauradora é dada por:

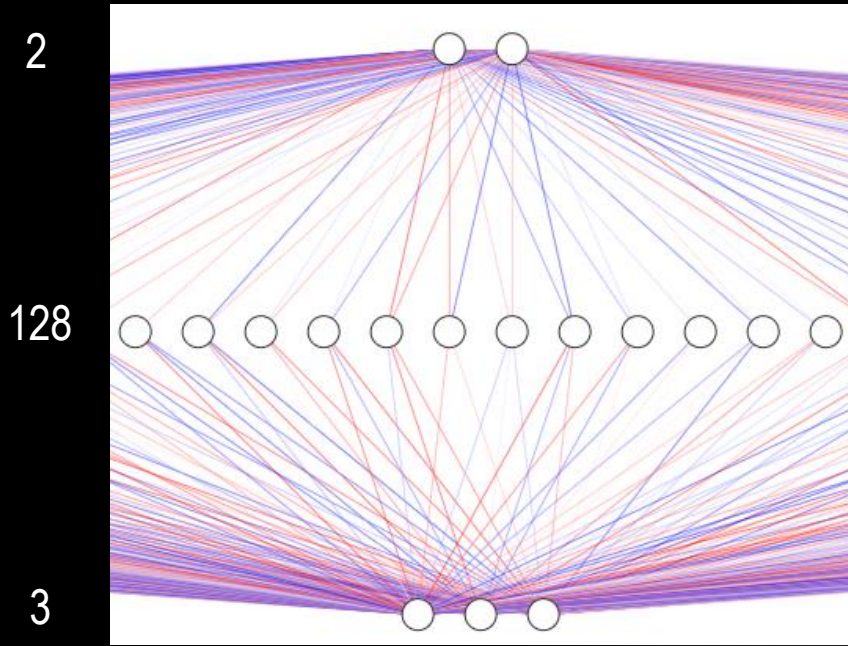
$$F = -hx^3$$

$$\Rightarrow \frac{d^2x(t)}{dt^2} = -\frac{h}{m}x^3(t)$$

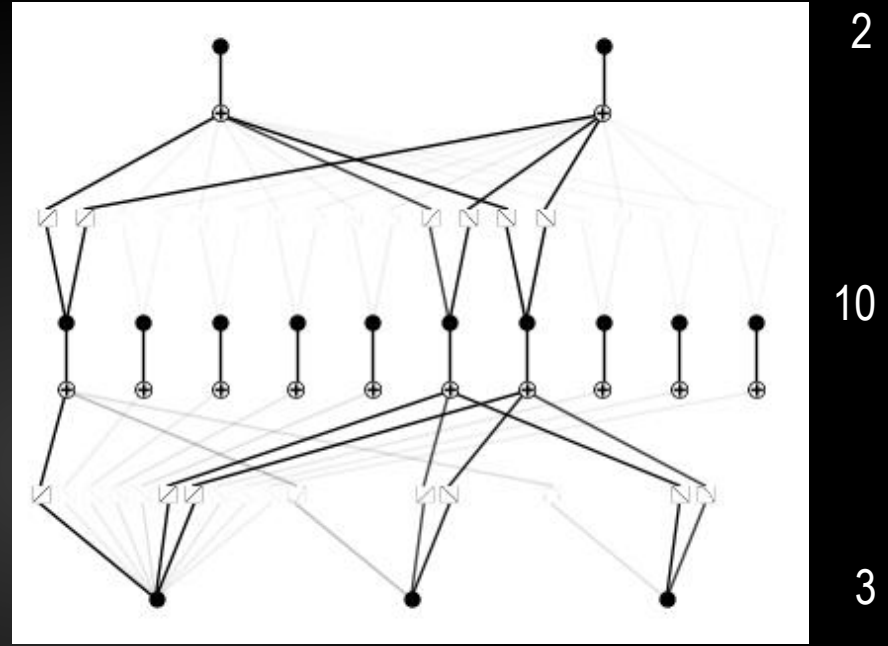
que é uma equação diferencial não linear sem solução analítica.

ARQUITETURA MLP VS KAN

Multi-layer perceptrons (MLPs)



Kolmogorov-Arnold Networks (KANs)



Training and Testing Dataset: Numerical Solution - *Runge-Kutta de 4ª ordem (RK4)*

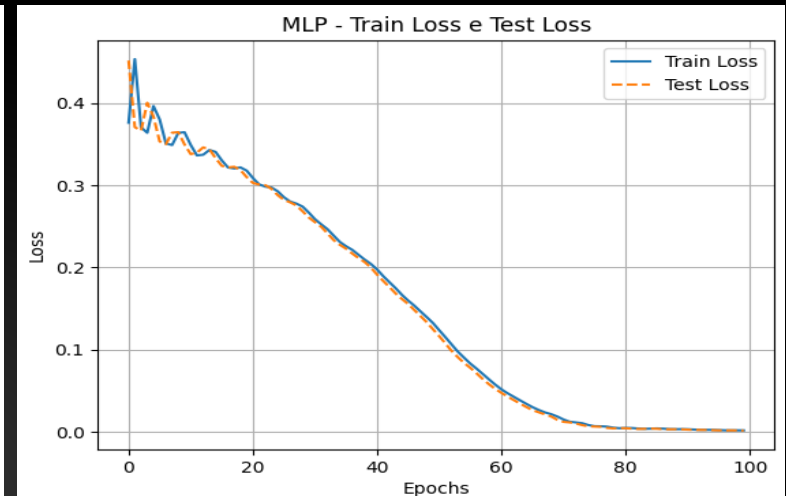
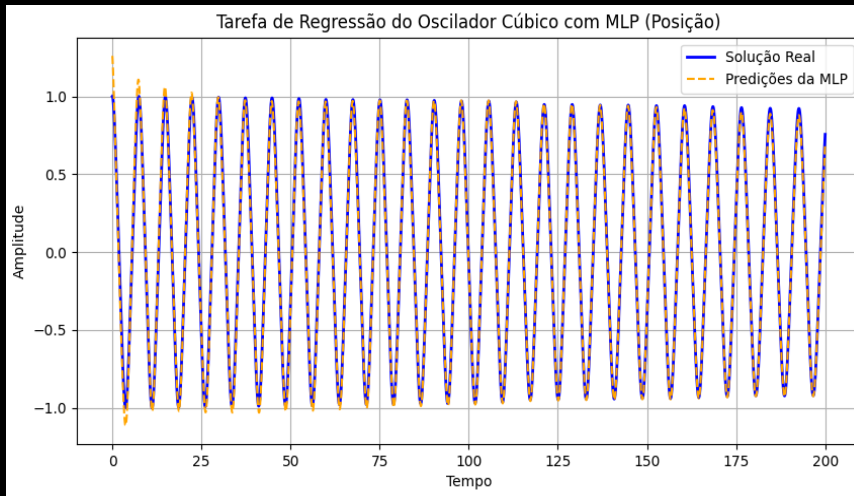
Matriz de entrada: 3x20.000

Dados de entrada: tempo (t), posição (x), velocidade (v)

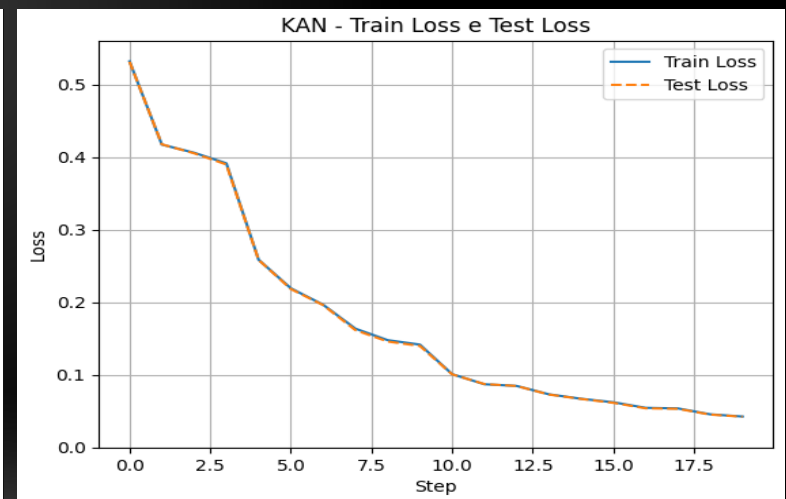
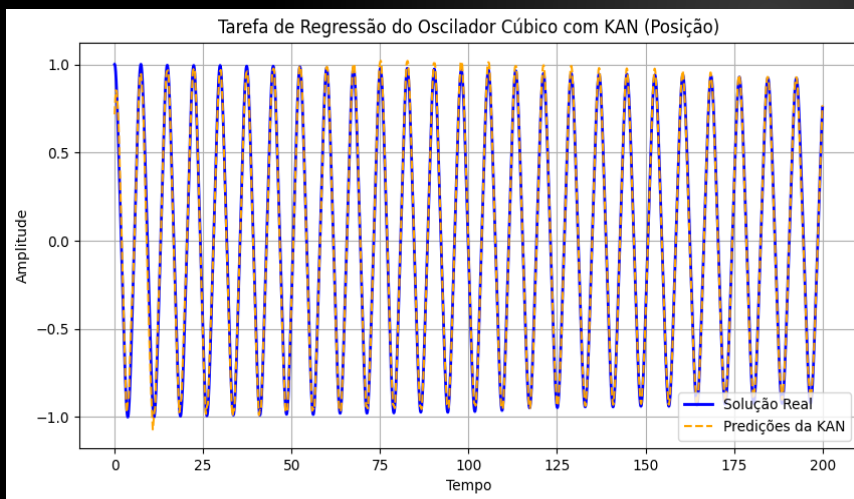
Label: posição e velocidade em um instante de tempo posterior: $x_{(t+\Delta t)}$ e $v_{(t+\Delta t)}$

SOLUÇÕES MLP VS KAN

MLP



KAN



MLP VS KAN

	KAN	MLP
Arquitetura	[3,10,2]	[3,128,2]
Tempo de execução	519,71 segundos	14,31 segundos
Erro final (<i>train loss</i>)	0,0426	0,0018
Erro final (<i>test loss</i>)	0,0423	0,0016
Generalização (diferença entre treino e teste)	0,0003	0,0002
Erro final na posição (x)	0,0254	0,0240
Erro final na velocidade (v)	0,0310	0,0164
Acurácia final (<i>training dataset</i>)	94,72%	94,62%
Acurácia final (<i>testing dataset</i>)	94,77%	94,65%
Consumo de memória (CPU)	0,01 MB	0,07 MB
Explicabilidade	Sim	Não

O Futuro

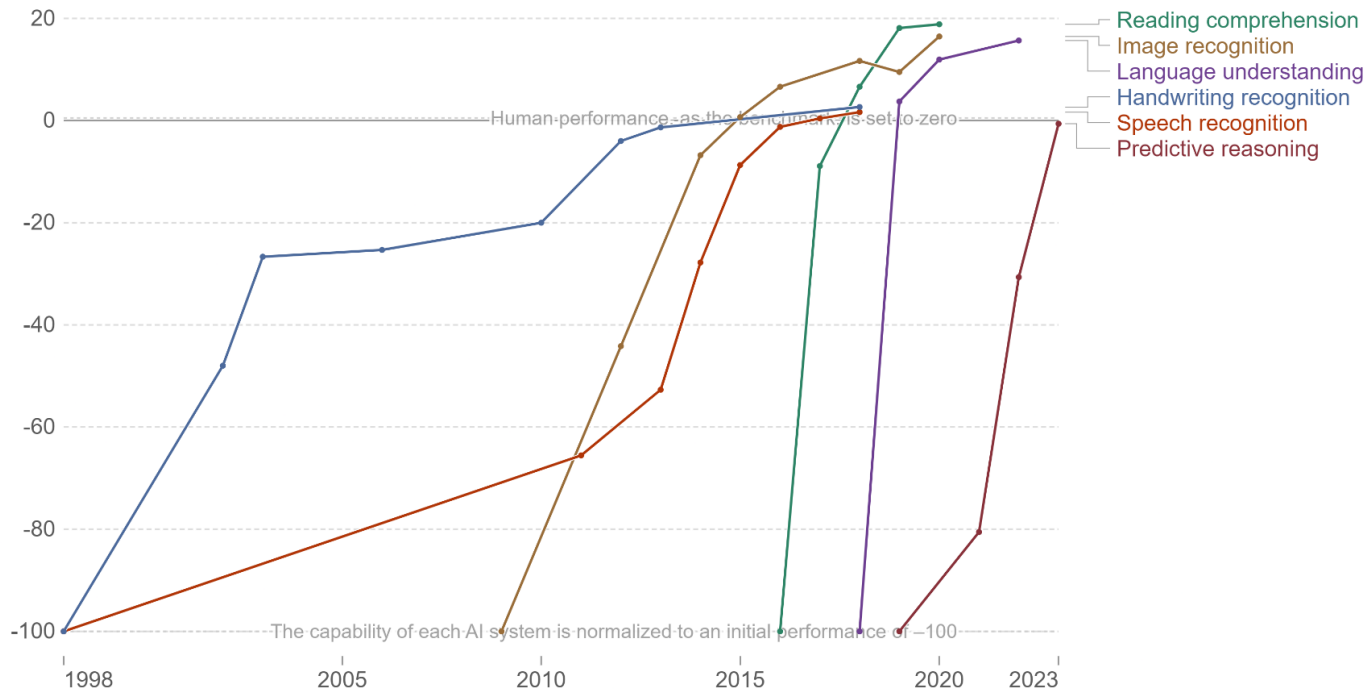


AGI (ARTIFICIAL GENERAL INTELLIGENCE)

Test scores of AI systems on various capabilities relative to human performance

Our World in Data

Within each domain, the initial performance of the AI is set to -100. Human performance is used as a baseline, set to zero. When the AI's performance crosses the zero line, it scored more points than humans.



Data source: Kiela et al. (2023)

OurWorldinData.org/artificial-intelligence | CC BY

Note: For each capability, the first year always shows a baseline of -100, even if better performance was recorded later that year.

O lado sombrio da IA



Desemprego - McKinsey Global Institute (2024)

- Até 2030, **30% das horas trabalhadas** nos EUA serão automatizadas pela IA.
- Até 2030 **14% dos trabalhadores** globalmente precisarão mudar de ocupação.
- **40% das empresas globais** utilizam IA diretamente.
- **82% das empresas globais** estão explorando IA em suas organizações.
- Espera-se que **o mercado global de IA atinja \$1,85 trilhões até 2030.**

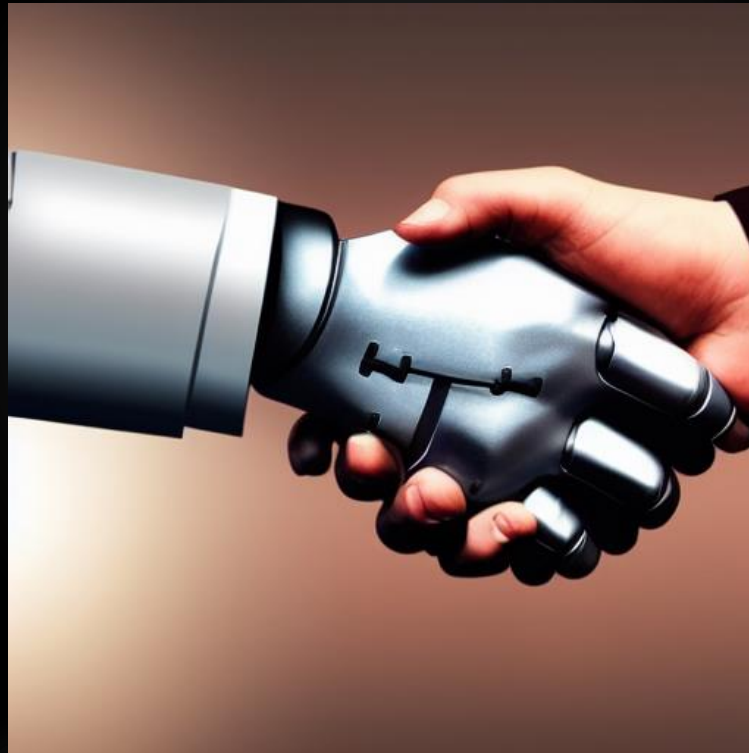
Crise Climática - Nature Briefings (Oct. 2024)

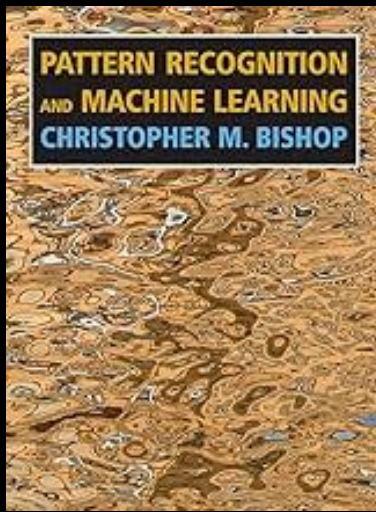
- De acordo com a AIE, em 2022, os data centers consumiram **1,65 bilhão de gigajoules** de eletricidade, cerca de **2% da demanda global.**
- **Até 2026, a AIE projeta um aumento de 35% a 128%** (consumo anual de energia da Suécia no menor estimativa e da Alemanha na pior das hipóteses).

Armas Autônomas, Colapso do Sistema Financeiro, Colapso dos Sistemas Políticos, Aumento da Desigualdade...

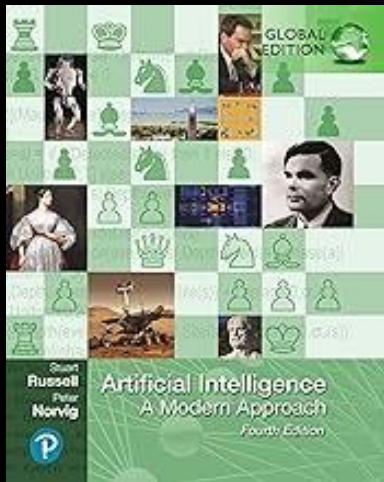
AGI, Hidden Goals, Emerging Properties, Singleton/Singularidade...

"Em tempos de mudança, o medo é natural, mas a esperança nos define. A inteligência artificial é uma ferramenta cujo propósito cabe a nós decidir. A verdadeira revolução será usar a tecnologia para expandir nossa humanidade."– ChatGPTo (Oct 2024)

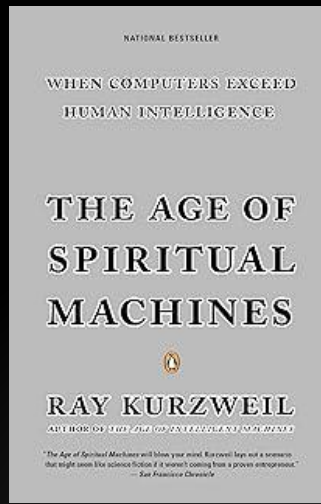




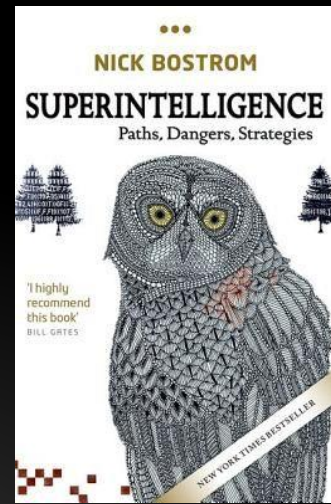
2011



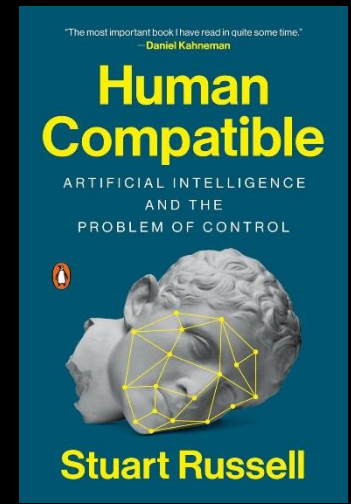
4th Ed. 2021



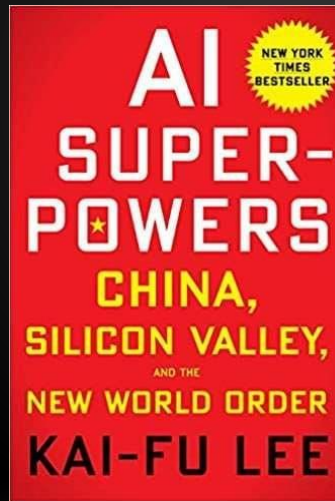
2000



2016



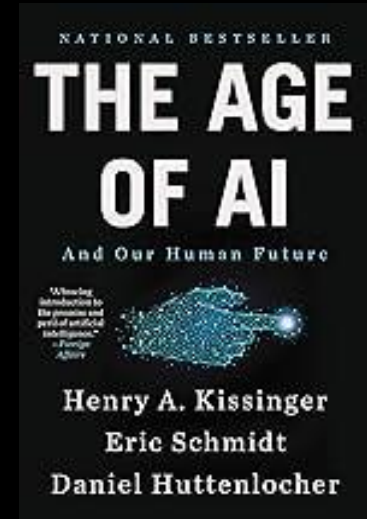
2020



2021



2021



2022



AI in the Physical Sciences,
Business, Technology, Biology,
Humanities, Art, Entertainment, and
beyond...

364 Artigos (Português/Inglês)
~50k Leitores
200+ Países



Scan me



<https://ai-talks.org/ai-talks-en/ai-talks-ufmg/>
SIEX 404764 – 04/07/2023

Obrigado!