

# Foundation Models

Prof. Luiz Eduardo S. Oliveira

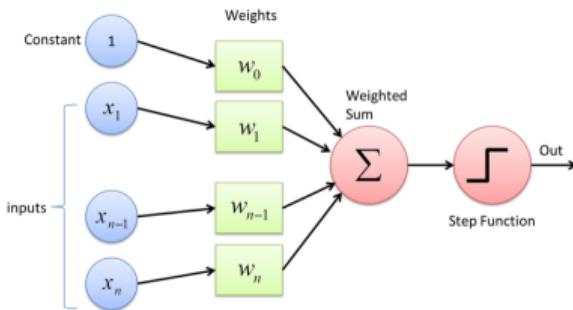
Universidade Federal do Paraná  
Departamento de Informática  
[www.inf.ufpr.br/lesoliveira](http://www.inf.ufpr.br/lesoliveira)

# Deep Learning

- Subárea da IA que utiliza redes neurais profundas para aprender a partir de grandes volumes de dados.

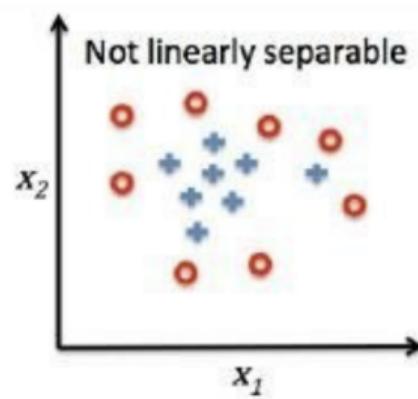
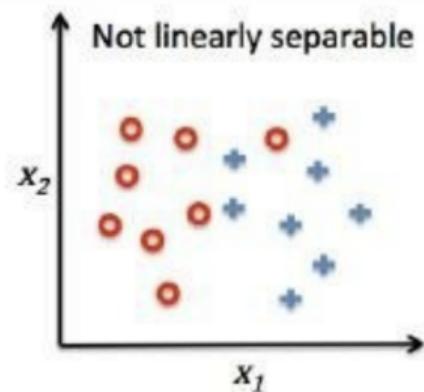
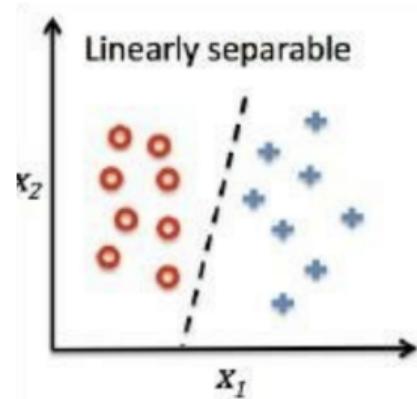
# Deep Learning

- Subárea da IA que utiliza redes neurais profundas para aprender a partir de grandes volumes de dados.
- O que é uma rede neural?
  - ▶ Modelo computacional inspirado no funcionamento do cérebro humano.
  - ▶ Composta por um conjunto interconectado de unidades de processamento chamados neurônios artificiais.

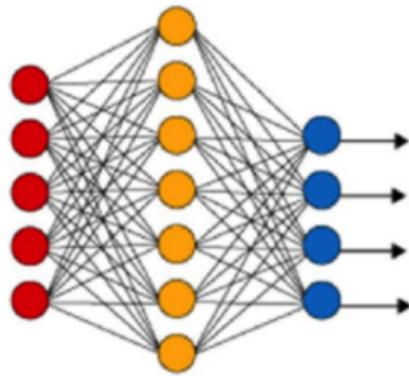


Perceptron (F. Rosenblatt, 1959): Rede neural de uma camada capaz de resolver problemas linearmente separáveis.

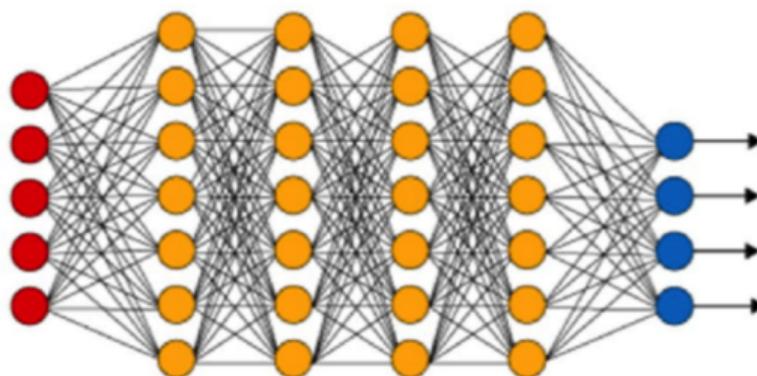
# Problema Linearmente Separável



# Multi-Layer Perceptron



Uma camada escondida



Várias camadas escondidas

Backpropagation (D. Rumelhart, G. Hinton, R. Williams, 1986 - Learning Representations by back-propagating errors): Uso do algoritmo de retro-propagação de erros para treinar redes neurais.

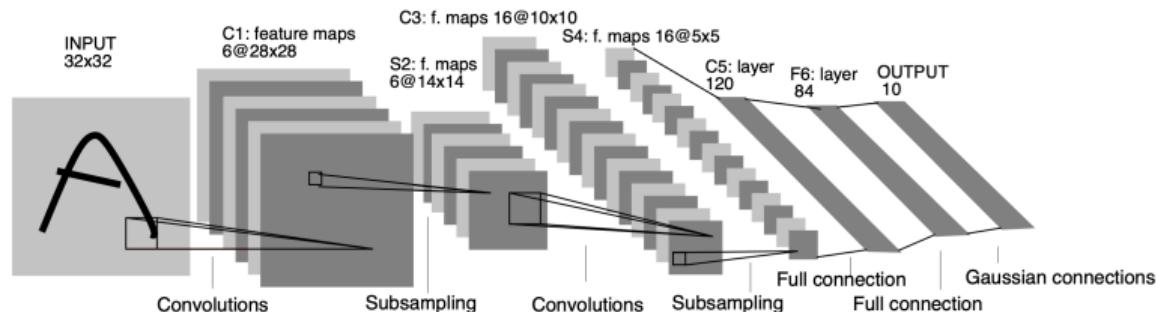
- Mais camadas (extratores de características) → Mais parâmetros → Mais dados de treinamento.

# Redes Neurais Recorrentes (RNN)

- Propostas na década de 90.
- Redes capazes de modelar dependências temporais.
  - ▶ (Hochreiter & Schmidhuber, 1997) LSTM - Long Short Term Memory.
  - ▶ (Cho et al, 2014) GRU - Gated Recurrent Unit.
- Utilizadas largamente em problemas de Processamento de Linguagem Natural.
- Dificuldades no treinamento e manuseio de sequências longas e em capturar contexto não sequenciais.

# Redes Neurais Convolucionais (CNN)

- (K. Fukushima, 1980) Neocognitron.
  - ▶ Arquitetura inspirada no processamento visual do cérebro
- (Y. LeCun & Y. Bengio, 1998) LeNet 5<sup>1</sup>
  - ▶ Aplicada com sucesso no reconhecimento de caracteres manuscritos.
  - ▶ Dificuldade no treinamento em função da grande quantidade de parâmetros (cerca de 60k)



- (A. Krizhevsky & G. Hinton, 2012) ImageNet (60 milhões de parâmetros)
  - ▶ Uso de GPU para treinar CNN

<sup>1</sup>Gradient-based learning applied to document recognition, Procs IEEE, 1998

## 'Godfathers of AI' honored with Turing Award, the Nobel Prize of computing



From left to right: Yann LeCun | Photo: Facebook; Geoffrey Hinton | Photo: Google; Yoshua Bengio | Photo: Botler AI

/ Yoshua Bengio, Geoffrey Hinton, and Yann LeCun laid the foundations for modern AI

By [James Vincent](#), a senior reporter who has covered AI, robotics, and more for eight years at The Verge.

Mar 27, 2019 at 7:02 AM GMT-3 | □ 0 Comments / 0 New

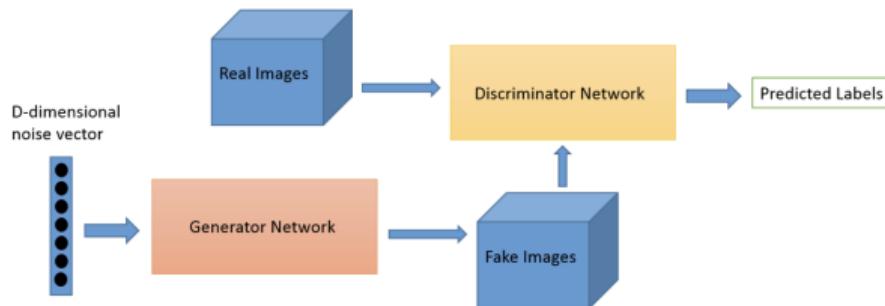


# Nobel Prize 2024



# IA Generativa

- CNNs são em geral modelos discriminativos, ou seja, usados para classificação e predição.
- Outro tipo de redes neurais profundas que ganharam bastante atenção nos últimos anos são as redes gerativas, entre elas as GANs ((I. Goodfellow, 2014) Generative Adversarial Networks).
- Compostas por um gerador e um discriminador
  - ▶ O gerador cria exemplos sintéticos e o discriminador avalia a autenticidade desses exemplos.
  - ▶ O objetivo é treinar o gerador para enganar o discriminador, gerando dados cada vez mais realistas.





Faces geradas através de uma GAN<sup>2</sup>

<sup>2</sup>T. Karas, et al, A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019. ↗ ↘ ↙



Pinturas geradas através de uma GAN<sup>3</sup>

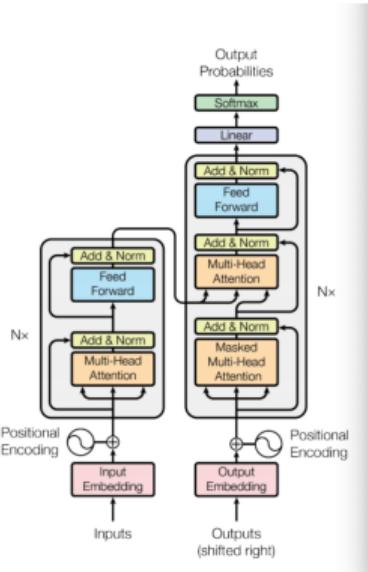
<sup>3</sup>T. Karas, et al, A Style-Based Generator Architecture for Generative Adversarial Networks, CVPR 2019. ↗ ↘ ↙

- Dado o sucesso das GANs em diversas aplicações, a aplicação desta arquitetura para geração de texto começou a ganhar mais atenção a partir de 2016.
  - ▶ Geração de sequências de palavras.
  - ▶ Diálogos e resumos de texto.
  - ▶ Geração de código.

- Dado o sucesso das GANs em diversas aplicações, a aplicação desta arquitetura para geração de texto começou a ganhar mais atenção a partir de 2016.
  - ▶ Geração de sequências de palavras.
  - ▶ Diálogos e resumos de texto.
  - ▶ Geração de código.
- Dificuldade de avaliação objetiva
  - ▶ Ao contrário de tarefas de geração de imagem, onde a qualidade visual pode ser avaliada com relativa facilidade, avaliar a qualidade do texto gerado é uma tarefa mais complexa.
  - ▶ As GANs podem enfrentar dificuldades em aprender essas estruturas complexas, resultando em textos que são gramaticalmente incorretos, incoerentes ou difíceis de interpretar.

# Transformers

- Arquitetura proposta em 2017 por pesquisadores do Google
- A ideia principal foi a introdução de um mecanismo de atenção (self-attention)
- Selecionar quais partes do texto devem ser utilizada ao invés de usar todo o texto.



## Attention Is All You Need

Ashish Vaswani\*  
Google Brain  
avaswani@google.com

Noam Shazeer\*  
Google Brain  
noam@google.com

Niki Parmar\*  
Google Research  
nikip@google.com

Jakob Uszkoreit\*  
Google Research  
usz@google.com

Llion Jones\*  
Google Research  
llion@google.com

Aidan N. Gomez\* †  
University of Toronto  
aidan@cs.toronto.edu

Lukasz Kaiser\*  
Google Brain  
lukasz.kaiser@google.com

Illia Polosukhin\* †  
illia.polosukhin@gmail.com

# Transformers

## Vantagens:

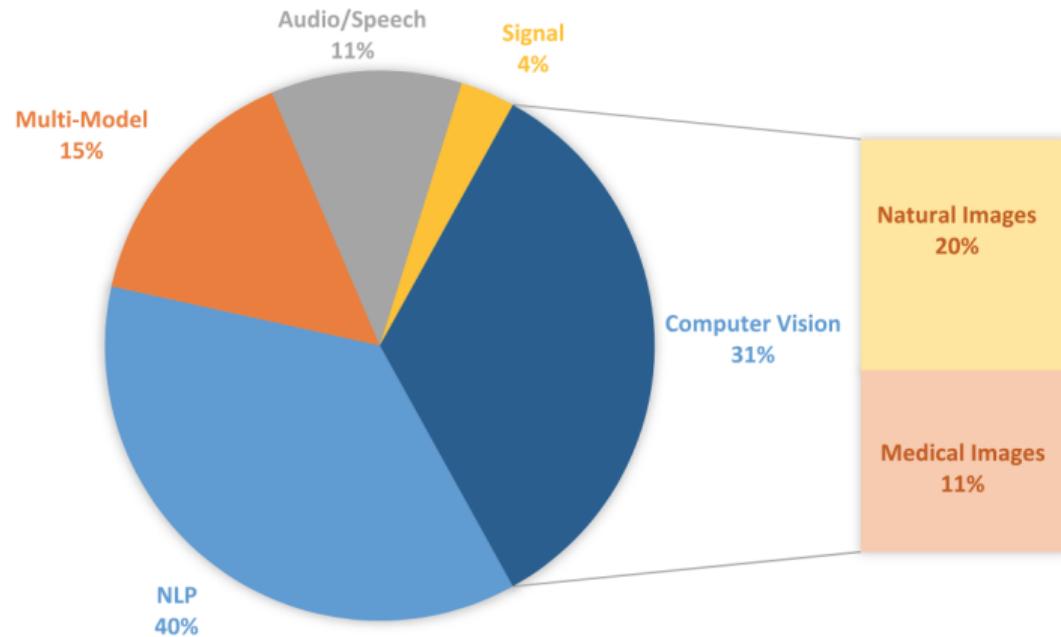
- Mecanismo de atenção permite que cada elemento de uma sequência se relacione com todos os outros elementos, independente da distância.
- Processamento paralelo pois cada elemento pode ser processado independentemente. Não precisa de uma ordem sequêncial como uma RNN, por exemplo.
- Aprende representações contextuais das palavras em uma sequência.
- Capturar contexto em textos longos.
- Adequados para transferência de aprendizado (transfer learning). Podem ser pré-treinados em grandes conjuntos de dados não rotulados, e em seguida, adaptados para tarefas específicas.



Dominou a área de PLN em pouco tempo

# Transformers

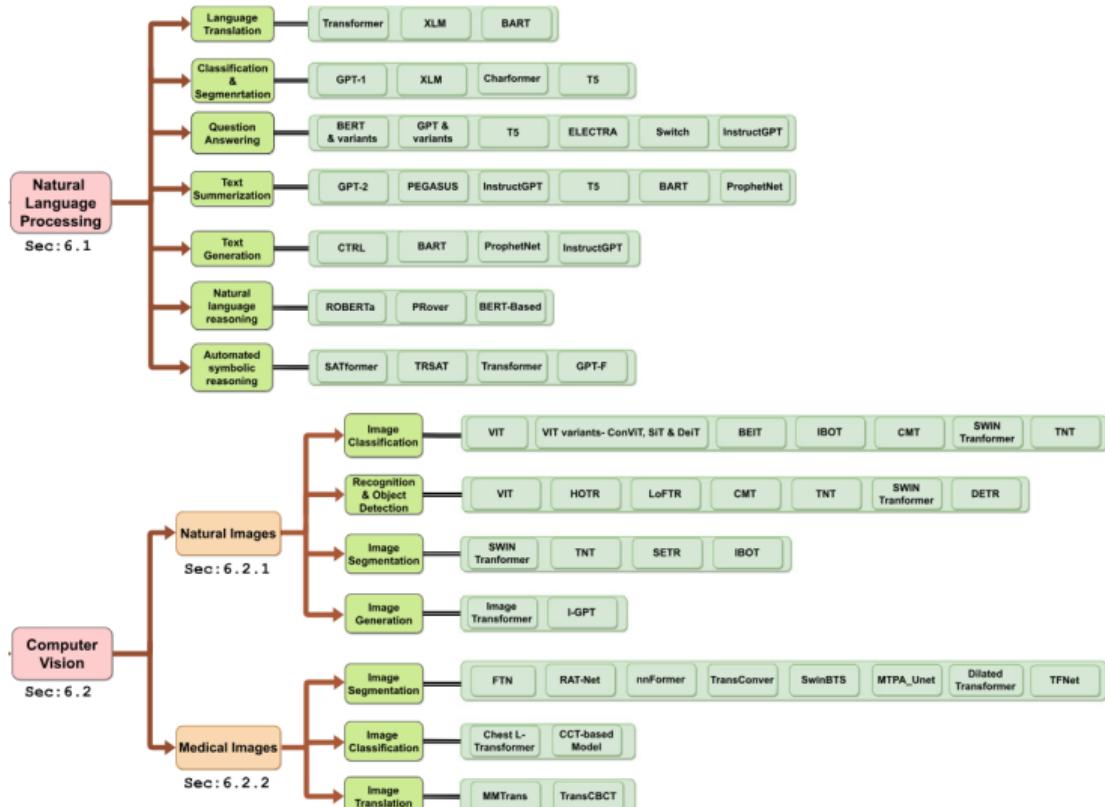
Adoção crescente em diversas aplicações<sup>4</sup>



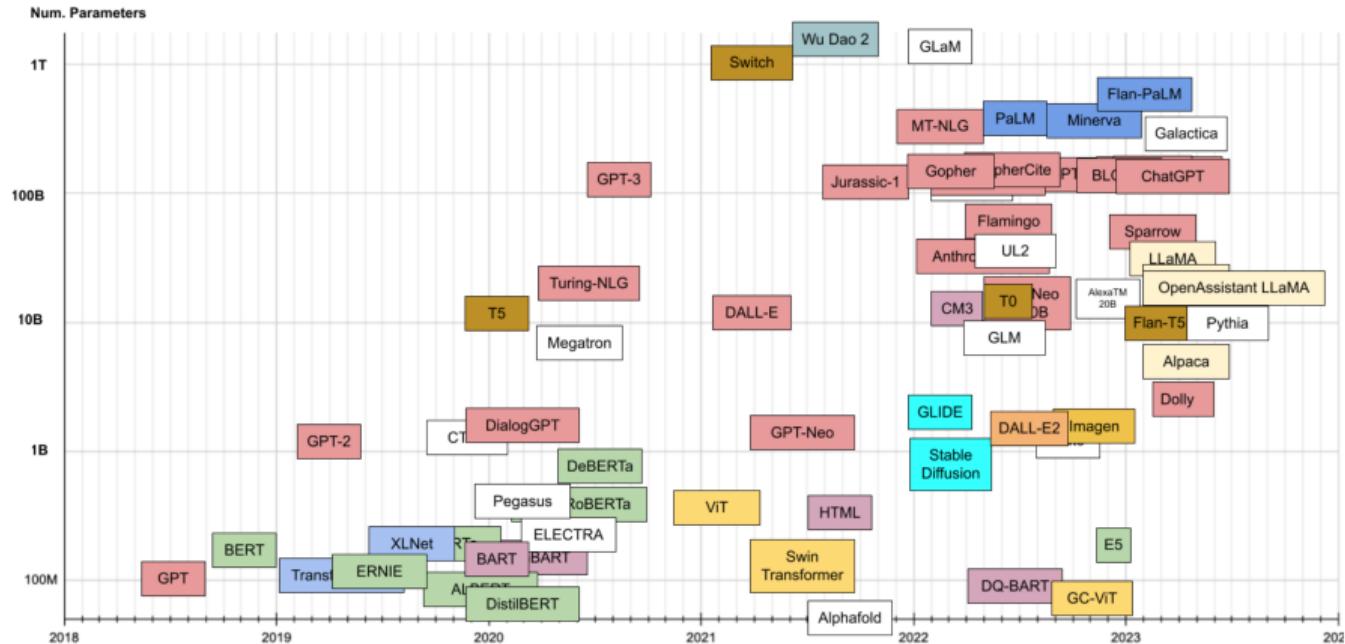
<sup>4</sup><https://doi.org/10.1016/j.eswa.2023.122666>

# Transformers

## Taxonomia



# Transformers



# Foundation Models

## On the Opportunities and Risks of Foundation Models

Rishi Bommasani\* Drew A. Hudson Ehsan Adeli Russ Altman Simran Arora  
Sydney von Arx Michael S. Bernstein Jeannette Bohg Antoine Bosselut Emma Brunskill  
Erik Brynjolfsson Shyamal Buch Dallas Card Rodrigo Castellon Niladri Chatterji  
Annie Chen Kathleen Creel Jared Quincy Davis Dorottya Demszky Chris Donahue  
Moussa Doumbouya Esin Durmus Stefano Ermon John Etchemendy Kawin Ethayarajh  
Li Fei-Fei Chelsea Finn Trevor Gale Lauren Gillespie Karan Goel Noah Goodman  
Shelby Grossman Neel Guha Tatsunori Hashimoto Peter Henderson John Hewitt  
Daniel E. Ho Jenny Hong Kyle Hsu Jing Huang Thomas Icard Saahil Jain  
Dan Jurafsky Pratyusha Kalluri Siddharth Karamcheti Geoff Keeling Fereshte Khani  
Omar Khattab Pang Wei Koh Mark Krass Ranjay Krishna Rohith Kuditipudi  
Ananya Kumar Faisal Ladhak Mina Lee Tony Lee Jure Leskovec Isabelle Levent  
Xiang Lisa Li Xuechen Li Tengyu Ma Ali Malik Christopher D. Manning  
Suvir Mirchandani Eric Mitchell Zanele Munyikwa Suraj Nair Avanika Narayan  
Deepak Narayanan Ben Newman Allen Nie Juan Carlos Niebles Hamed Nilforoshan  
Julian Nyarko Giray Ogut Laurel Orr Isabel Papadimitriou Joon Sung Park Chris Piech  
Eva Portelance Christopher Potts Aditi Raghunathan Rob Reich Hongyu Ren  
Frieda Rong Yusuf Roohani Camilo Ruiz Jack Ryan Christopher Ré Dorsa Sadigh  
Shiori Sagawa Keshav Santhanam Andy Shih Krishnan Srinivasan Alex Tamkin  
Rohan Taori Armin W. Thomas Florian Tramèr Rose E. Wang William Wang Bohan Wu  
Jiajun Wu Yuhuai Wu Sang Michael Xie Michihiro Yasunaga Jiaxuan You Matei Zaharia  
Michael Zhang Tianyi Zhang Xikun Zhang Yuhui Zhang Lucia Zheng Kaitlyn Zhou  
Percy Liang\*<sup>1</sup>

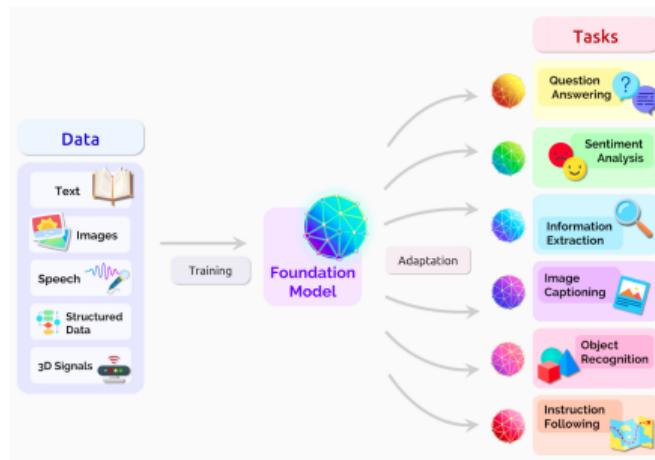
Center for Research on Foundation Models (CRFM)  
Stanford Institute for Human-Centered Artificial Intelligence (HAI)  
Stanford University

5

<sup>5</sup><https://arxiv.org/abs/2108.07258>

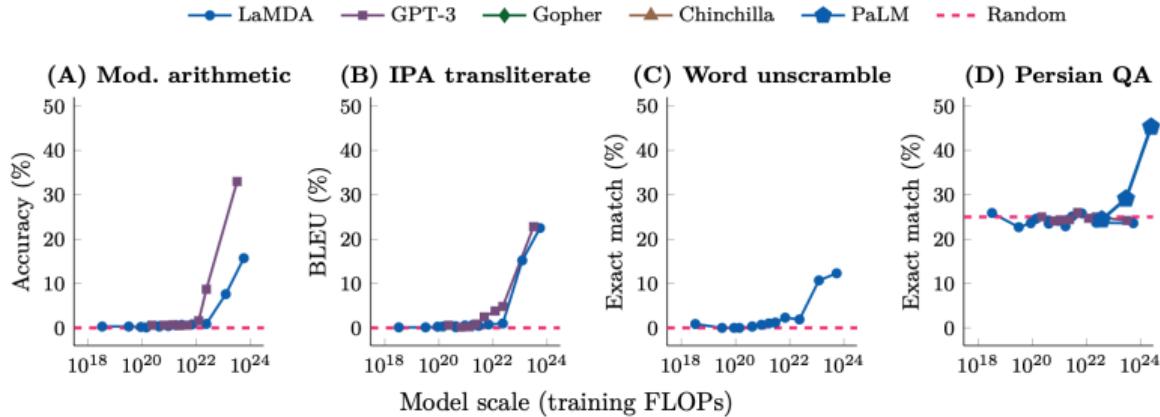
# Foundation Models

- Modelo treinado em grande volume de dados que pode ser adaptado para uma ampla gama de tarefas (transfer learning).
- Transfer learning é a base para esses modelos.
- Aprendizagem usando dados não rotulados, evitando assim o custo de rotulação.
- Reuso de modelos.



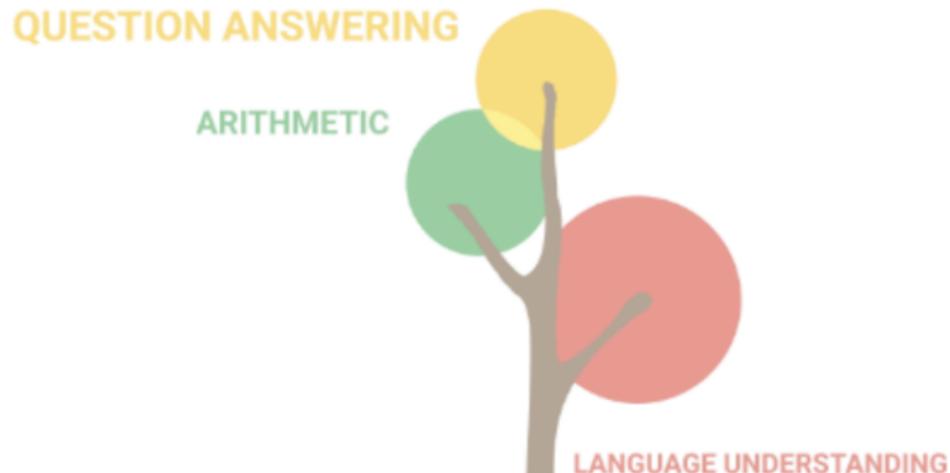
# Foundation Models

- Escala é o que os torna poderosos.
  - ▶ GPU
  - ▶ Transformers (paralelismo)



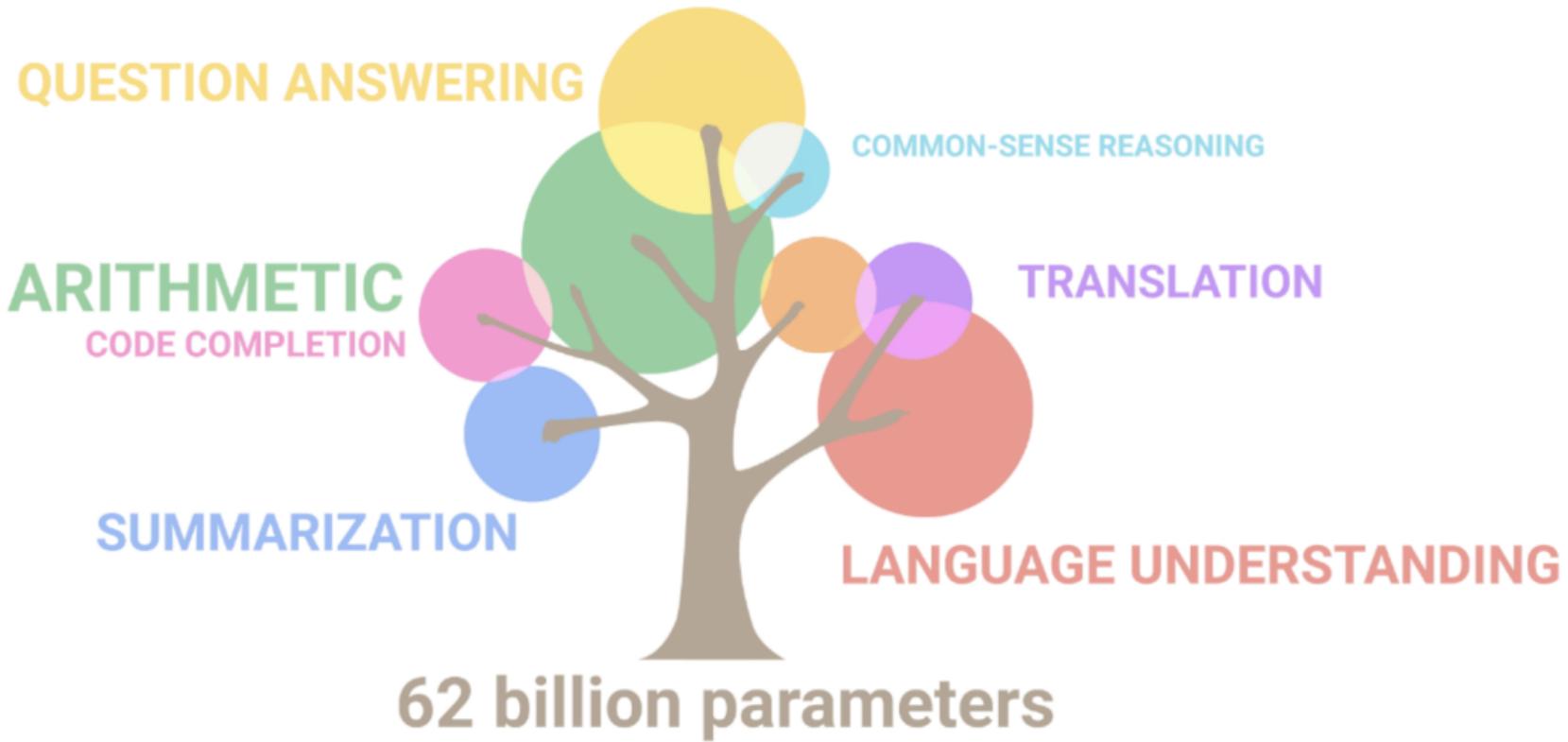
Muitas habilidades dos modelos emergem somente quando os modelos atingem um certo tamanho<sup>6</sup>

<sup>6</sup><https://arxiv.org/abs/2206.07682>

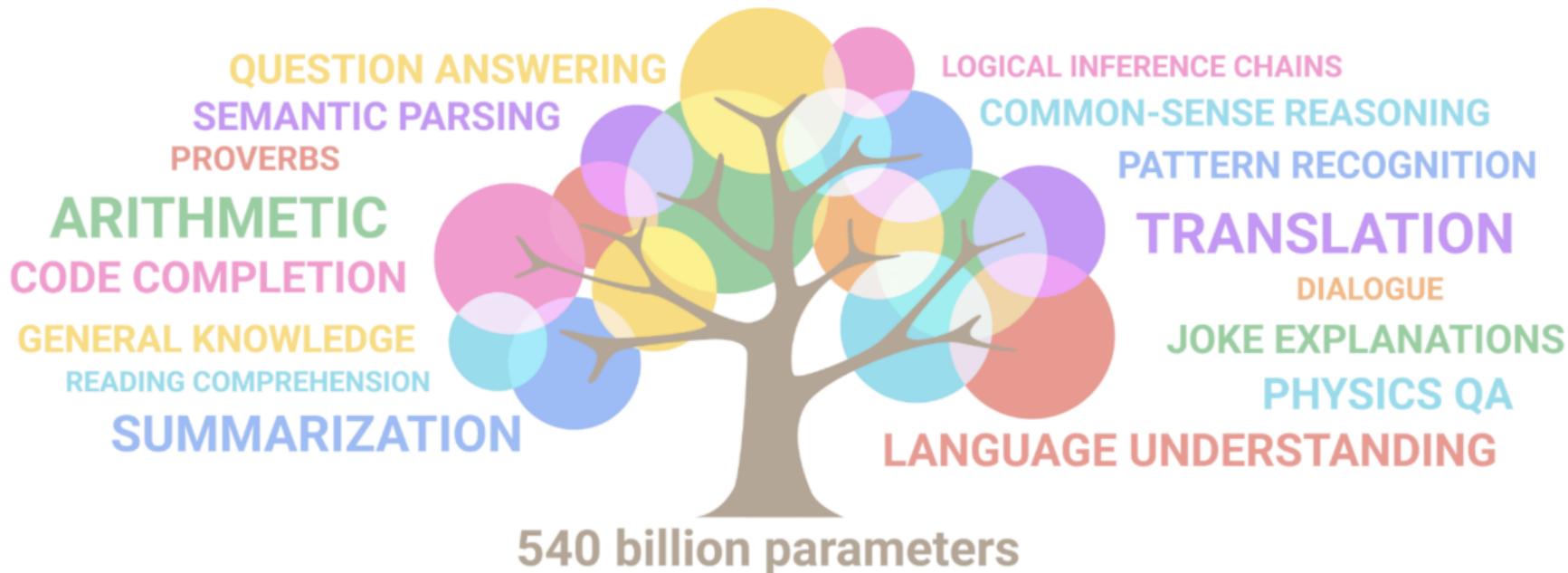


**8 billion parameters**

# Foundation Models



# Foundation Models



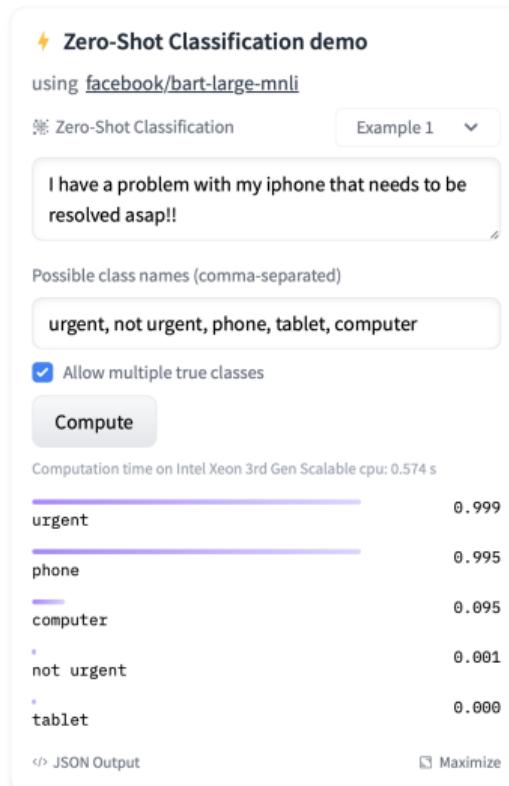
# Estratégias de Aprendizagem de Máquina

- Clássica
  - ▶ Definição das características
  - ▶ Definição do algoritmo aprendizagem
  - ▶ Treinamento do modelo
- Aprendizagem Profunda
  - ▶ Definição da arquitetura profunda (RNN, CNN, etc...)
  - ▶ Treinamento do modelo
- Modelos pré-treinados
  - ▶ Fine-tuning.
  - ▶ Não envolve definição de arquitetura
- Foundation models (LLM)
  - ▶ Zero-shot learning
  - ▶ Few-shot learning
  - ▶ In-context learning

# Foundation Models

## Zero-shot learning

- Modelo que é capaz de classificar novos exemplos de classes desconhecidas, ou seja que não foram vistas durante o treinamento.



# Foundation Models

Zero-shot learning usando Mask2Former (Swin backbone)<sup>7</sup>



<sup>7</sup><https://huggingface.co/facebook/mask2former-swin-large-cityscapes-semantic>

# Foundation Models

## Few-shot learning

- O modelo recebe alguns exemplos sobre o problema, para então inferir sobre o problema.
- Pesos do modelo não são atualizados.

Example prompt:

Tweet: "I hate it when my phone battery dies."

Sentiment: Negative

###

Tweet: "My day has been "

Sentiment: Positive

###

Tweet: "This is the link to the article"

Sentiment: Neutral

###

Tweet: "This new music video was incredibile"

Sentiment:

# Foundation Models

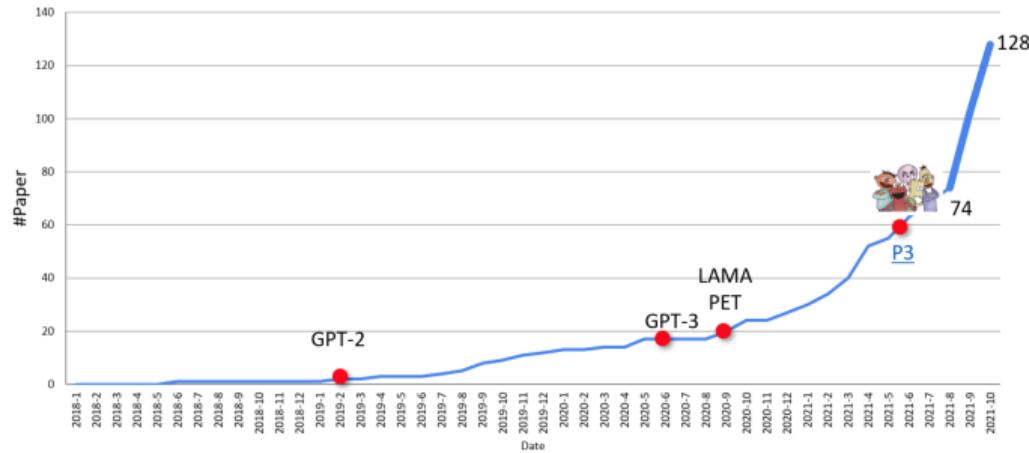
## In-Context Learning

- A ideia nesse caso é fazer o modelo aprender por analogia.
- Mostrar ao modelo as etapas (passo a passo) usadas para se chegar a uma determinada resposta.
  - ▶ Chain-of-thought

# Foundation Models

## In-Context Learning

- A ideia nesse caso é fazer o modelo aprender por analogia.
- Mostrar ao modelo as etapas (passo a passo) usadas para se chegar a uma determinada resposta.
  - ▶ Chain-of-thought



Evolução do número de artigos na literatura usando esses paradigmas

# Foundation Models

- Essas estratégias são importante pois:
  - ▶ Permitem aplicar os modelos a novas tarefas sem coletar dados adicionais,
  - ▶ Sem a necessidade de um treinamento adicional (fine-tunning).
- Reduz a quantidade de esforço necessário para construir uma aplicação.

# Foundation Models

- Essas estratégias são importantes pois:
  - ▶ Permitem aplicar os modelos a novas tarefas sem coletar dados adicionais,
  - ▶ Sem a necessidade de um treinamento adicional (fine-tuning).
- Reduz a quantidade de esforço necessário para construir uma aplicação.
- Mudança de paradigma
  - ▶ Aproxima o desenvolvedor do usuário de aprendizagem de máquina.

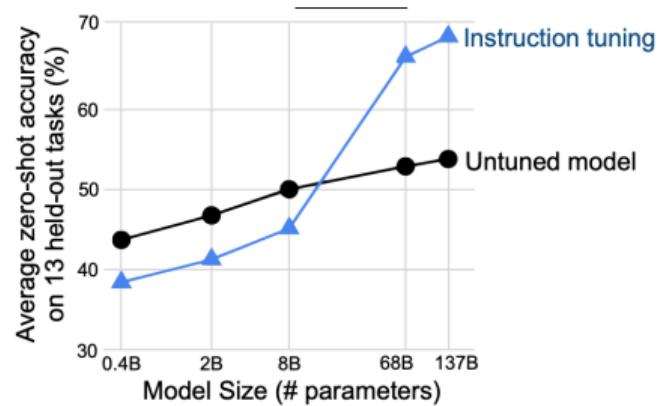


Adaptar tarefas ao modelo e não modelos às tarefas

# Foundation Models

## Evolução no desempenho dos modelos de linguagem

- Tamanho
- Treinamento dos modelos com texto e código
  - ▶ Aparentemente faz com que os modelos aprendam como identificar a estrutura do texto.
- Treinamento com dados anotados por humanos<sup>8</sup>
- Muita pesquisa em curso sobre como melhor treinar esses modelos.

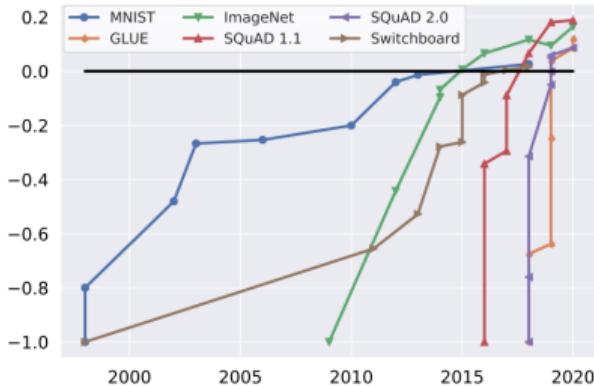


<sup>8</sup><https://arxiv.org/pdf/2109.01652.pdf>

# Foundation Models

## Benchmarks

- A velocidade da inovação está tornando benchmarks obsoletos rapidamente<sup>9</sup>
- Apensa 66% dos benchmarks receberam mais de 3 resultados.<sup>10</sup>
- Muitos deles são saturados logo após o lançamento.



<sup>9</sup><https://aclanthology.org/2021.nacl-main.324.pdf>

<sup>10</sup><https://arxiv.org/pdf/2203.04592.pdf>

# Desafios para Adoção em Empresas

- Interpretabilidade

- ▶ Redes neurais são frequentemente descritas como “caixas pretas”.
- ▶ Os humanos que os usam podem nunca entender como os modelos chegam a suas previsões ou recomendações.
- ▶ Isso pode tornar difícil para as empresas explicar ou justificar suas decisões aos clientes ou reguladores.

# Desafios para Adoção em Empresas

- Interpretabilidade

- ▶ Redes neurais são frequentemente descritas como “caixas pretas”.
- ▶ Os humanos que os usam podem nunca entender como os modelos chegam a suas previsões ou recomendações.
- ▶ Isso pode tornar difícil para as empresas explicar ou justificar suas decisões aos clientes ou reguladores.

- Segurança e Privacidade

- ▶ Exigem acesso a dados confidenciais, como informações de clientes ou dados comerciais proprietários.
- ▶ Isso pode gerar preocupações sobre privacidade e segurança, principalmente se o modelo for implantado na nuvem ou acessado por provedores terceirizados.

# Desafios para Adoção em Empresas

- Aspectos legais

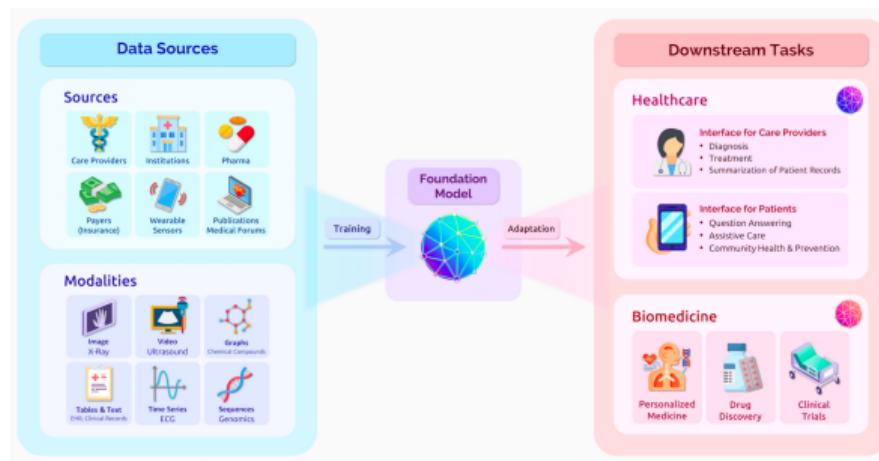
- ▶ Os modelos são treinados em uma enorme quantidade de dados da natureza, e nem todos esses dados se alinharão aos valores do seu negócio.
- ▶ O uso desses modelos pode levantar considerações legais e éticas relacionadas a preconceito, discriminação e outros danos potenciais.

# Desafios para Adoção em Empresas

- Aspectos legais

- Os modelos são treinados em uma enorme quantidade de dados da natureza, e nem todos esses dados se alinharão aos valores do seu negócio.
- O uso desses modelos pode levantar considerações legais e éticas relacionadas a preconceito, discriminação e outros danos potenciais.

- Múltiplas fontes de informação → Modelo Multimodal



# Preocupações

- Concentração de poder
  - ▶ Somente grandes empresas, ou start-ups (com grande investimento, e.g., OpenAI) com recursos disponíveis para treinar esse modelos.
  - ▶ Como o desempenho dos modelos tem alta correlação com a escala, isso levanta preocupações sobre concentração de mercado.
- Algumas alternativa para modelos aberto:
  - ▶ Huggingface<sup>11</sup>
  - ▶ Masakhane<sup>12</sup> (African Languages)
  - ▶ Eleuther.ai<sup>13</sup>
  - ▶ BLOOM (Open LLM - 176 bilhões de parâmetros - BigScience - CNRS, GENCI)
  - ▶ Ollama<sup>14</sup>

---

<sup>11</sup><https://huggingface.co>

<sup>12</sup><https://www.masakhane.io>

<sup>13</sup><https://www.eleuther.ai>

<sup>14</sup><https://ollama.com>

## Em resumo...

- Enorme potencial, mas ainda estamos no começo.
- Apesar de sua implantação no mundo real, esses modelos são protótipos de pesquisa e pouco compreendidos.
- Coloca novos desafios na área da educação.
- Passo para trás em reproduzibilidade
  - ▶ Iniciativas de disponibilizar dados e código (PyTorch, TensorFlow) com a “onda” da aprendizagem profunda não acontece mais.
  - ▶ Em muitos casos somente uma API é disponibilizada.
- Preocupação da academia nos rumos da pesquisa liderada exclusivamente pelas grandes corporações.

Obrigado pela Atenção

