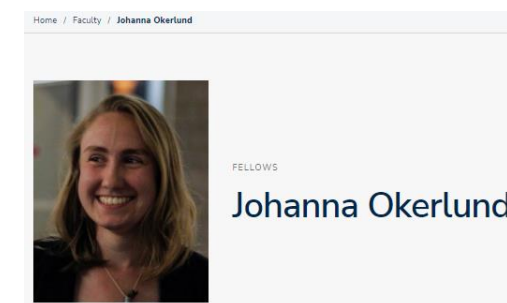


Grandes Modelos de Linguagens

Por que eles são importantes e o que Devemos *fazer* com eles



Fonte imagem: [Science, Technology and Public Policy \(STPP\) | STPP Homepage \(umich.edu\)](#)



About author:

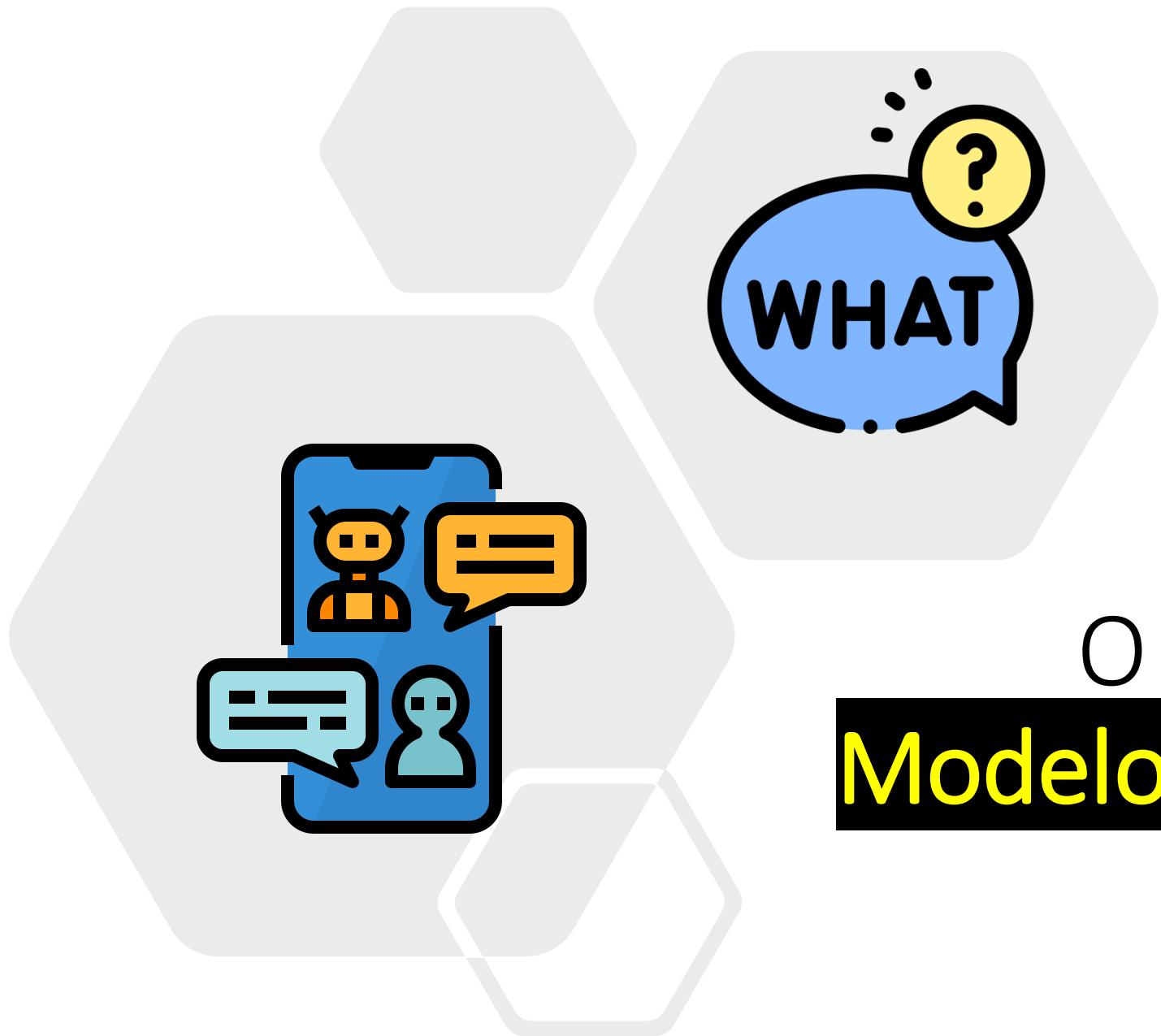
<https://stpp.fordschool.umich.edu/faculty/johanna-okerlund>



About author:

<https://fordschool.umich.edu/faculty/shobita-parthasarathy>

[What's in the Chatterbox? | Science, Technology and Public Policy \(STPP\) \(umich.edu\)](#)



O que são **Grandes**
Modelos de Linguagens?



Esta Foto de Autor Desconhecido está licenciado em [CC BY-NC-ND](#)

São
Algoritmos de
aprendizado
de máquina
que podem:

Reconhecer;

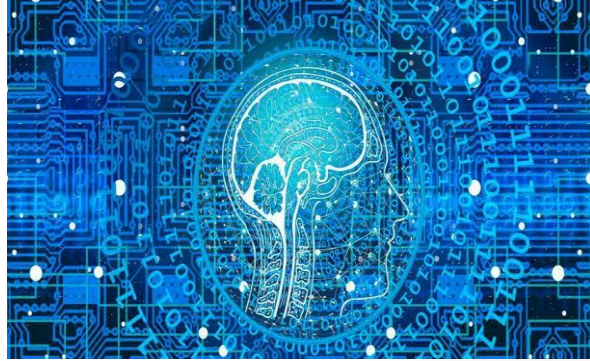
Resumir;

Traduzir;

Prever;

Gerar linguagem **semelhante à linguagem humana**

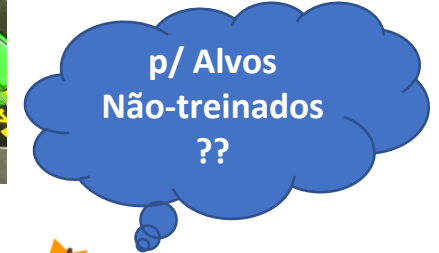
Como funcionam os grandes modelos de linguagem?



ETAPA 2



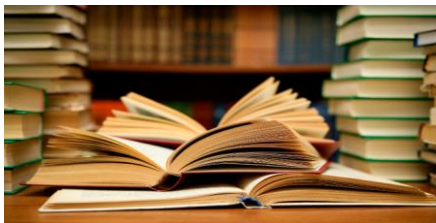
ETAPA 3



ETAPA 1



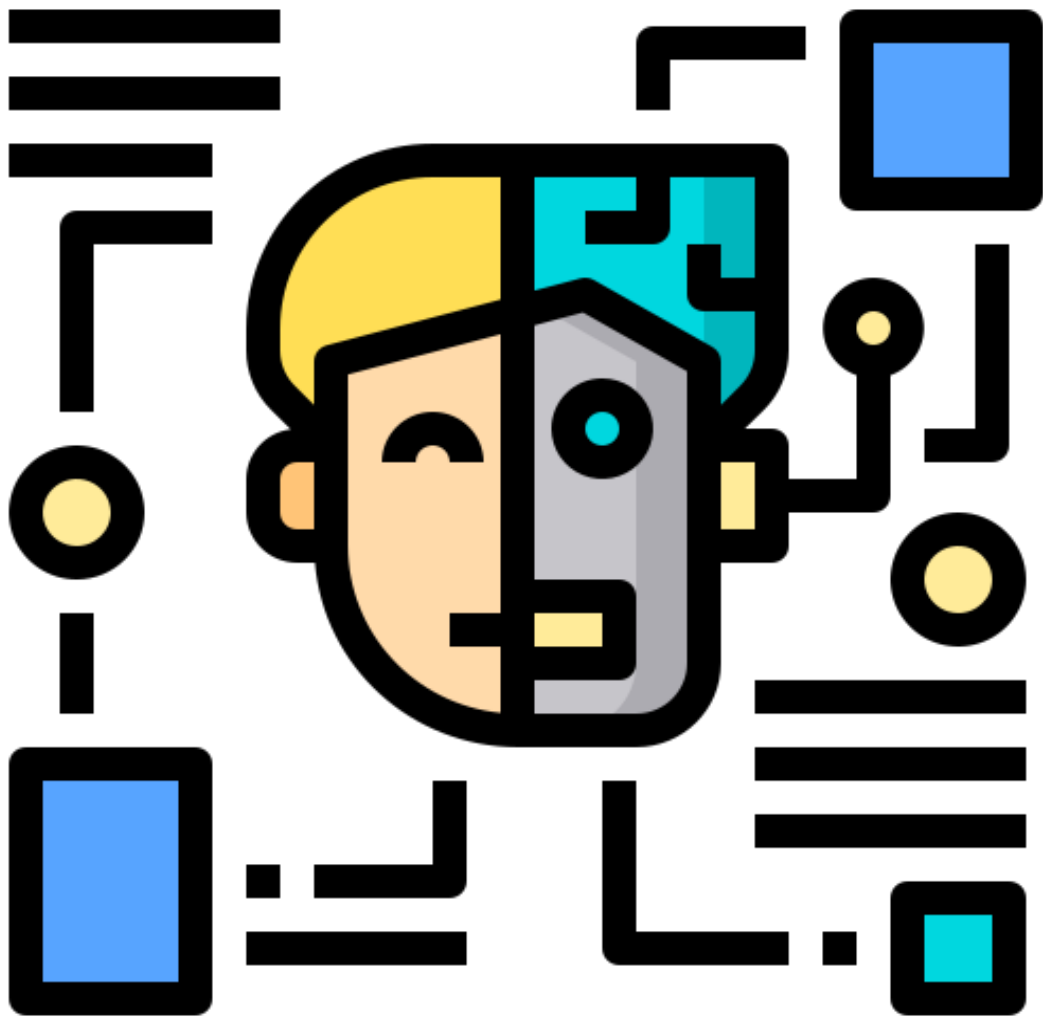
[Esta Foto](#) de Autor Desconhecido
está licenciado em [CC BY-NC](#)



[Esta Foto](#) de Autor Desconhecido
está licenciado em [CC BY-NC-ND](#)



[Esta Foto](#) de Autor Desconhecido
está licenciado em [CC BY-SA](#)



Google

music is the most advanced form of mathematic: X

Todas Imagens Vídeos Notícias Shopping Mais Ferramentas

Aproximadamente 71.300.000 resultados (0,70 segundos)

"Music is the most advanced form of mathematics" — **it's a sentence you wouldn't find anywhere on the internet, up until a few days ago.** That's because it wasn't written by a philosopher, or a composer, or even a drunken math student. 24 de jul. de 2020

<https://www.zmescience.com> > Other > Pieces

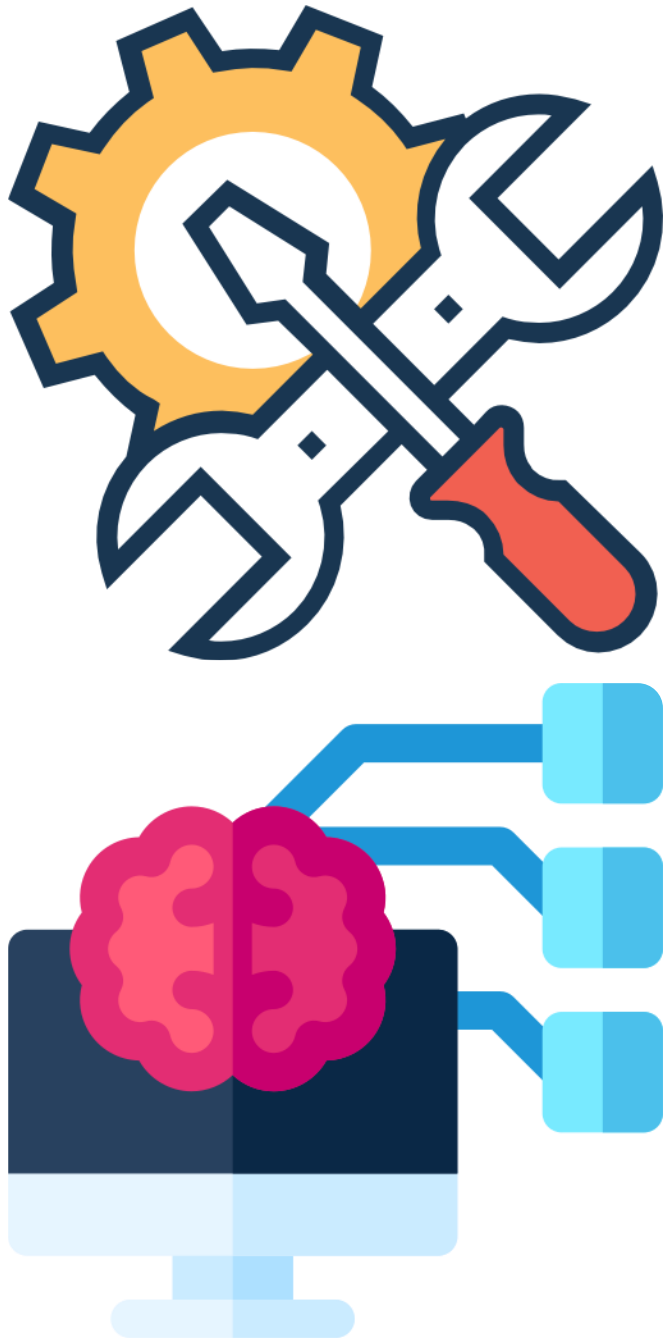
[New AI can write in any style, answer deep questions, and ...](#)

Fonte youtube: <https://www.youtube.com/watch?v=pbVwH8o837A&t=540s>

- "Produzem imitação convincente da linguagem humana através do computador";
- Linguagem + *sofisticada e humana* que seus antecessores;
- Melhor desempenho em tarefas para as quais não foram treinados explicitamente



>USO

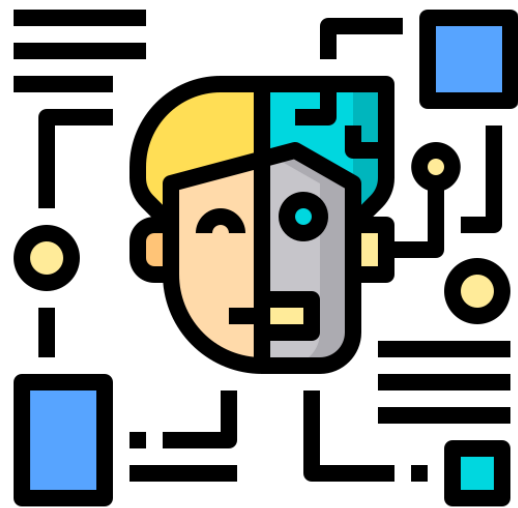


- Os políticos podem avaliar a opinião pública sobre legislação pendente (*sentiment analysis*);
- Os pacientes podem resumir e avaliar o estado do conhecimento em saúde para capacitar suas interações com os profissionais desta área;
- Os cientistas podem traduzir os resultados da pesquisa em vários idiomas

Os LLM exigem enormes recursos:

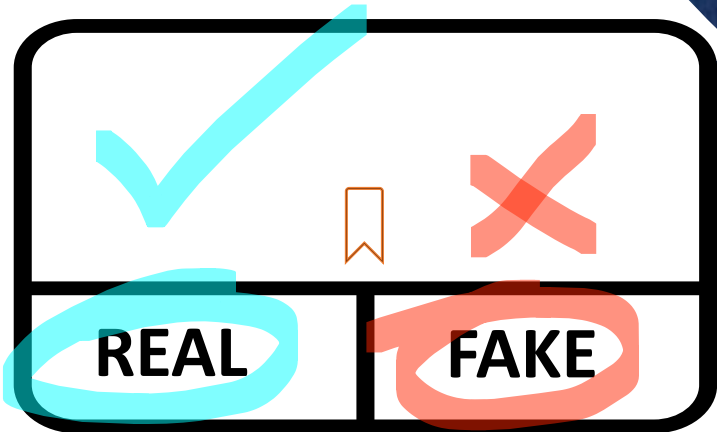
- Financeiros;
- Infra-estrutura;
- Humanos;
- Poder computacional.





?





**FAKE
OR
REAL?**





Esses modelos de linguagens também podem apresentar **viéses** de:

- Gênero;
- Raça;
- Religião;
- Sexualidade...

Metodologia de Casos Análogos

Nem toda hesitação vacinal é igual. Nem é simplesmente o resultado da ignorância ou antipatia pela ciência.

- As comunidades questionam se seus governos e instituições científicas, tecnológicas e médicas realmente representam suas **necessidades** e **prioridades**.
- Longos legados de maus-tratos à comunidades marginalizadas alimentam ainda mais essa desconfiança.

In Communities We Trust

Institutional Failures and Sustained Solutions for Vaccine Hesitancy

Zixuan Wang
Margarita Maria Rodriguez Morales
Kseniya Husak
Molly Kleinman
Shobita Parthasarathy



Esta Foto de Autor Desconhecido está licenciado em [CC BY-NC](#)

1. limitações e falhas em instituições científicas e técnicas. Exemplos:

- Vioxx
- explosão do ônibus espacial Challenger.
- as instituições de política científica e tecnológica tendem a basear suas decisões em avaliações sistemáticas – e muitas vezes quantitativas – (isso geralmente **exclui o conhecimento** e a **experiência da comunidade**)
- como as comunidades científicas comunicam as incertezas (lutam para comunicar as incertezas e limitações de suas descobertas, o que também pode fomentar uma maior desconfiança.

Essas abordagens alienam as pessoas que são menos propensas a confiar em instituições especializadas no futuro.

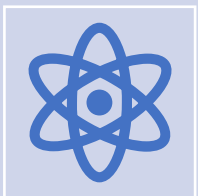
2. Maus-tratos institucionalizados de comunidades marginalizadas.

Tanto os governos quanto as instituições científicas, tecnológicas e médicas têm um longo legado de maltratar comunidades marginalizadas, incluindo comunidades de cor historicamente desfavorecidas.

Como resultado, essas comunidades se sentem alienadas e agora estão céticas sobre a enxurrada de atenção para vaciná-las. Na verdade, eles podem até ver a recusa da vacina como uma forma de resistência.




Por serem treinados em livros texto antigos e páginas da web, os LLMs **reproduzem preconceitos históricos** e **discursos de ódio** contra comunidades marginalizadas;

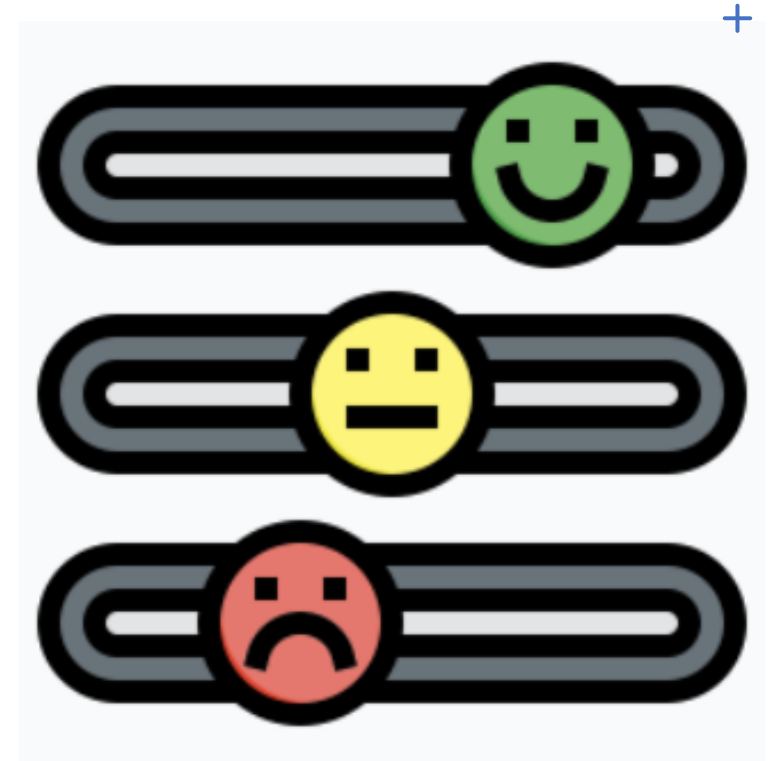


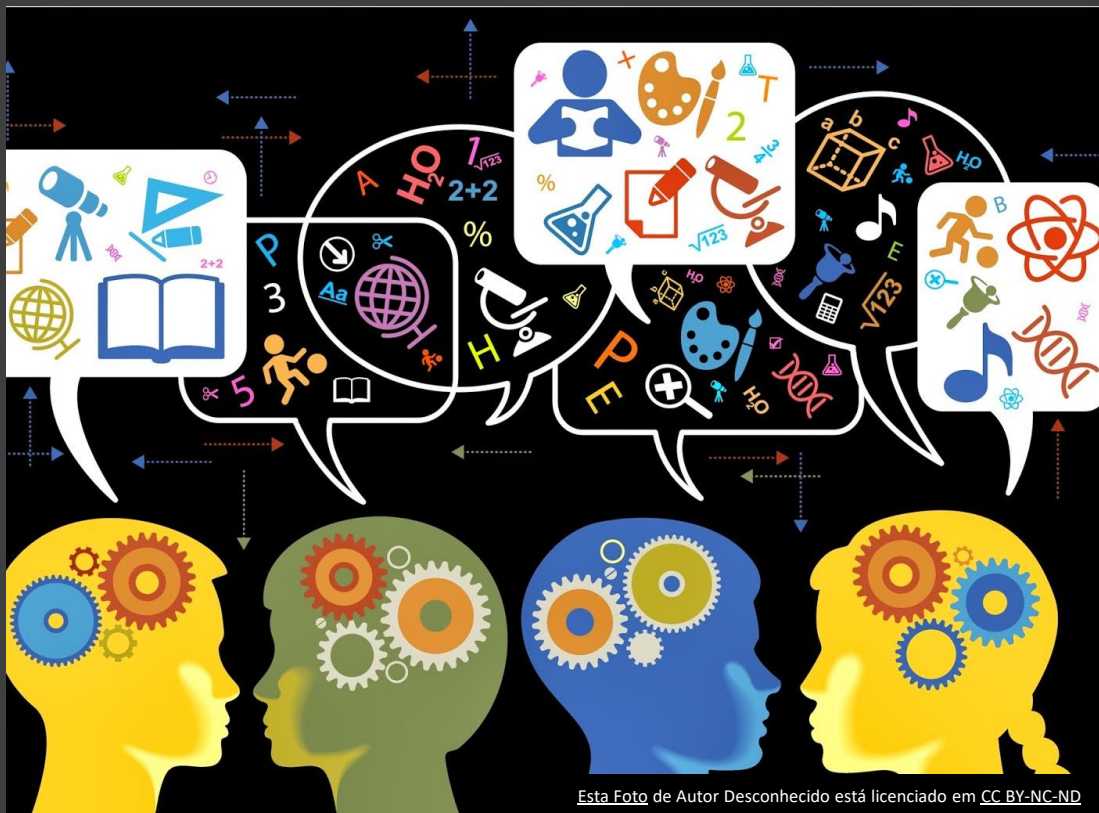
Exigem **enormes quantidades de energia** e **poder computacional**



Implicações sociais

- Exacerbar os problemas ambientais;
- Acelerar a sede por dados;
- Integração rápida à infra-estrutura existente;
-  das desigualdades;
- Reorganização do trabalho e especialização;
- Aumento da fragmentação social





Transformação na pesquisa científica

- Esses modelos transformarão tanto o tipo de pesquisa que os cientistas fazem, quanto como eles o fazem.
 - autoria e métodos científicos;
 - mudanças no sistema de financiamento de pesquisas;
 - transformação no sistema peer review.
- Os editores acadêmicos provavelmente desenvolverão modelos para manter seu poder de monopólio sobre a maior parte da literatura científica.
- O uso destes modelos para realizar avaliações científicas produzirá controvérsia entre os cientistas.
- Esses modelos reforçarão o domínio anglo-americano na ciência.

Recomendações

- O governo americano deveria regular os modelos:
 - análise da qualidade dos “corpus”
 - diversidade do corpus
 - considerar as comunidades marginalizadas
 - transparência dos corpus e algoritmos
- Monitoramento dos apps que usam estes modelos;
- Necessidade de uma organização nacional ou internacional para estabelecer padrões de usos;

- Reforçar as leis de privacidade e segurança dos dados;
- Sob nenhuma circunstância os aplicativos baseados nestes modelos implantados pelo governo (por exemplo, chatbots que fornecem informações sobre serviços sociais, aplicativos de avaliação de risco pré-julgamento em processos de justiça criminal) deveriam coletar informações de identificação pessoal;
- consultores nas áreas *social e dimensões de equidade da tecnologia*. “Direitos Humanos em Tecnologia”
- regras claras sobre o status de *direitos autorais de invenções geradas pelos modelos*;

- Avaliações ambientais de novos *data centers*;
- O governo deve avaliar os riscos de saúde, segurança e psicológicos que os modelos e outras formas de inteligência artificial criam para os trabalhadores;
- programas de reciclagem profissional e, em um nível amplo, uma garantia de renda básica e saúde universal.

- A Fundação de Ciência Nacional deve aumentar substancialmente seu financiamento para o desenvolvimento destes modelos;
- Esse financiamento deve priorizar:
 - a. Desenvolver corpora e modelos alternativos, especialmente aqueles impulsionados pela necessidades de comunidades marginalizadas e de baixa renda
 - b. Reuniões para estabelecer padrões para tornar os corpora representativos e para incorporar o conhecimento dos cidadãos (particularmente de baixa renda e comunidades marginalizadas)

- a. Apoiar atualizações e manutenção de corpora e modelos existentes (em contraste com apenas fazer mais modelos novos)
- b. Apoiar a pesquisa na construção de novos tipos de modelos que são mais facilmente atualizados;
- c. Pesquisas sobre a equidade, sociais e ambientais dos modelos.
impactos

#References:

Research Report

- **What's in the Chatterbox? Large Language Models, Why They Matter, and What We Should Do About Them**

April 2022 | Johanna Okerlund, Evan Klasky, Aditya Middha, Sujin Kim, Hannah Rosenfeld, Molly Kleinman, Shobita Parthasarathy

Disponível em:

<https://stpp.fordschool.umich.edu/sites/stpp/files/2022-05/large-language-models-TAP-2022-final-051622.pdf>

- **In Communities We Trust: Institutional Failures and Sustained Solutions for Vaccine Hesitancy**

May 2021 | Zixuan Wang, Margarita Maria Rodriguez Morales, Kseniya Husak, Molly Kleinman, Shobita Parthasarathy

Disponível em:

<https://stpp.fordschool.umich.edu/sites/stpp/files/2021-07/vaccine-hesitancy-STPP-TAP-2021-v5.pdf>