

As IA's são realmente “inteligentes”,
perigosas, e vamos poder regulá-las?

Walter Carnielli

Centro de Lógica, Unicamp
e AI2



De que vamos falar

- O que é a IA?
- O que é o ChatGPT, GPT-3, GPT-4?
- Como funcionam
- Já chegamos na verdadeira “Inteligência Geral”?
- Limitações da IA
- Riscos sociais, econômicos e humanos da IA
- O poder das BigTechs
- Dificuldades em se conceituar a IA
- Consequências para a sociedade

O que é a IA?

- **IA simbólica, ou IA baseada em conhecimento e lógica:** usa representações do conhecimento e regras simbólicas para raciocinar e tomar decisões
- **IA Conectivista, ou IA baseada em redes neurais:** usa sistemas compostos de camadas de neurônios artificiais interconectados que são treinados usando grandes conjuntos de dados para reconhecer padrões e fazer previsões locais
- **IA Evolutiva ou IA baseada em algoritmos genéticos:** constrói sistemas que podem evoluir e adaptar-se ao longo do tempo, inspirados no processo de seleção natural

Generative Pre-trained Transformer (GPT)

- Generative Pre-trained Transformer (GPT) é uma arquitetura de rede neural em camadas, projetada para gerar texto e outras sequências de dados com base em uma entrada.
- GPT completa textos baseado em bilhões de dados de treinamento
- Por exemplo:
- Numa manhã de ...X... crianças vão ao ...Y... para andar de... Z...
- Numa manhã de tempestade crianças vão ao aeroporto para andar de tanque de guerra ?

O que é o ChatGPT?

- O ChatGPT é um modelo de linguagem de grande escala (LLM) da OpenAI, baseado na arquitetura GPT (Generative Pre-trained Transformer) 3.5.
-
- O ChatGPT é capaz de compreender a linguagem natural em vários idiomas e gera respostas que *parecem* humanas.
- Foi treinado em grandes quantidades de dados textuais, permitindo que ele aprenda e capture padrões na linguagem natural.
-
- É um grande “copia- e- cola” com gramática correta

Como é a matemática do GPT?

- O modelo GPT utiliza técnicas de **álgebra linear, cálculo e estatística** para aprender a relação entre as palavras e gerar previsões com base em padrões aprendidos a partir dos dados de treinamento.
- O ChatGPT utiliza a mesma arquitetura do GPT3,5, mas é especificamente projetado para conversar e responder perguntas. E agora temos o GPT4, e logo o GPT5....
- Já se pode usar GPT4 + Google gratuitamente em [//www.sincode.ai](https://www.sincode.ai)!
- Tudo muito bem feito, mas nada de misterioso- só mostra o poder da matemática e dos algoritmos, aos quais muita gente não presta atenção!

Como pode ser usado GPT?

- O ChatGPT pode ser usado em tarefas que envolvem linguagem, como atendimento ao cliente, assistentes virtuais, fazer resumos, corrigir textos.
- Consegue entender e escrever código em diferentes linguagens de programação e pode ser usado para depurar códigos

O ChatGPT popularizou a IA e registrou um milhão de registros nos primeiros dias

- Transforma tudo em linguagem, e transforma a linguagem em probabilidades e álgebra linear (matrizes e vetores)
- O GPT-4 é 10 vezes mais avançado que seu antecessor, o GPT-3.5, muito mais do mesmo.

O que melhora no GPT4?

- Segundo a OpenAI, melhor “alinhamento com valores morais”
- Parametrização: GPT-3 usa 175 bilhões de parâmetros, enquanto GPT-4 usa 1 trilhão de parâmetros.(rumores...)
- O GPT-4 não gera imagens, mas pode receber imagens como entrada.
- Segundo a OpenAI, é 40% mais preciso que o GPT-3.5
- Pode lidar com mais de 25.000 palavras de texto, ou 50 páginas
- Muita propaganda... e há vários outros modelos das “Big Techs”

IA alucina e não sabe a taboada!

- O ChatGPT alucina- há centenas de casos
- Não sabe fazer contas com 5 dígitos:
- Qualquer calculadora sabe que
 $75257 \times 98729 = 7.430.048.353$
- O ChatGPT errou 8 vezes esta conta, e continua errando!
- *“O resultado da multiplicação entre 75257 e 98729 é 743,255,853”.*
- Quem vai deixar o ChatGPT administrar sua empresa, fazer seu Imposto de Renda, ou controlar as contas públicas??

Isso tudo é mesmo “inteligência”, ou é só “artificial”? IA's ou HCD's?

- Muita coisa rotulada como “IA” e “Machine Learning”, vendendo a ideia de algo “sobre-humano”, são simples algoritmos falíveis ou métodos estatísticos clássicos
- Poderiam ser melhor chamadas de “habilidades cognitivas digitais” -HCD's

Mas isso é inteligência? Não, é apenas *habilidade cognitiva digital* !

- Falta capacidade de generalização e abstração
- Capacidade de dedução e consequência
- Compreensão contextual
- Capacidade de abdução
- Capacidade de explicação
- Falta a noção de causa e efeito
- Aprendizado autônomo
- Interação física com o ambiente.

Limitações da IA

- Viés algorítmico: Os algoritmos de IA em geral repetem preconceitos humanos, perpetuando e ampliando as desigualdades existentes
- Dependência de dados: A IA precisa de uma grande quantidade de dados para ser treinada
- Altíssimos custos ecológicos em níveis de CO2 para treinar GPTs
- Dependência de supervisão humana- rotuladores de dados (data taggers) a US\$ 2 por hora
- Falta de compreensão do contexto: não entende linguagem figurativa, ou ironia
- Não tem a capacidade de inovação
- Limitações computacionais: Alguns problemas são muito complexos para serem resolvidos com a tecnologia atual. A IA na prática, não tem o poder computacional suficiente

Riscos da IA: econômicos, sociais, políticos

- Muita gente vai perder o emprego: jornalistas, modelos, pessoal de propaganda, de contabilidade e auditoria, professores, auxiliares de todo tipo, entregadores, trabalhadores de fábrica, caixas
- Super-gerador de “fake news”
- Violação da privacidade e vieses
- Responsabilidade e ética

DALL-E2 , Salvador Dali- o poder das imagens

- <https://openai.com/product/dall-e-2>

- **GANs (Generative Adversarial Networks), que consiste em dois modelos de rede neural trabalhando em conjunto para gerar imagens**

MidJourney



GPT4



Riscos da IA: econômicos, sociais, políticos...

- Sem regulação, a IA vai causar danos sociais, econômicos e políticos seríssimos
- Prejudicar a concorrência, a privacidade e a escolha do consumidor
- Super-automatizar o trabalho, aumentar a desigualdade, reduzir salários, “uberizar” as relações de trabalho
- Fim do discurso político real, a força da democracia
- A “IA ética” é uma utopia!
- Uso militar secreto!

“Unipólio”?

- Dados são maior ativo para a IA e para as redes sociais
- Quem fornece os dados? Nós, o “gado voluntário”
- “Unipólio”: mais que monopólio, Metaverso e outras iniciativas que virão pretenderão “fidelizar” o usuário?
- Seu carro será GoogleCar, seu banco GBank, seu plano de saúde GHealth, seu estudo Gfac... ? Ou será MetCar, MetaBank, MetaHealth, MetaFac?

Carta aberta do Instituto Future of Life - apenas retórica?

- Parada de 6 meses:
- Sistemas robustos para distinguir o real do sintético
- Financiamento público robusto para pesquisa e segurança
- Atenção para as perturbações econômicas e políticas advindas da IA

Peça de propaganda, ou estratégia para deter a concorrência?

– Pode ser, mas p.ex. Geoffrey Hinton, um dos criadores das redes neurais profundas, deixou a Google para divulgar os perigos implícitos da IA

-
- Cientistas estão sugerindo uma agência internacional reguladora, como a energia atômica!
-

As IA's das 5 Big Techs chegando...

- OpenAI GPT4 e DALL-E
- Google Bard
- Anthropic startup de IA da Google
- Microsoft Bing com GPT4
- Perplexity AI
- Amazon Bedrock
- Biointeligência
- Sistemas homem-máquina
- Tudo isso potencializado pelo 5G!

Receita Líquida: BigTechs em 2021

- Apple US\$ 492 bilhões
- Amazon US\$ 57 bilhões
- Microsoft US\$ 206 bilhões
- Facebook US\$117 bilhões
- Tesla/OpenAI US\$ 16 bilhões
- Alphabet/ Google US\$65 bilhões

Total US\$ 1 trilhão

- **PIB do Brasil em 2022: US\$ 1,92 trilhão**

Grandes dificuldades

- A IA é realmente difícil de se definir
-
- Projeto de Lei 21/20 no Art 2º, I considera:
 - I – sistema de inteligência artificial ...utilizando abordagens baseadas em aprendizagem de
- máquina e/ou lógica e representação do conhecimento só?, por meio de dados de entrada provenientes de máquinas ou humanos, com o objetivo de produzir previsões, recomendações ou decisões *imitar?*
Influenciar? Ler sua mente? que possam influenciar o ambiente virtual ou real;

Sexto Empírico, século II d.C.

- **Ceticismo:** A única maneira segura é aprender com um filósofo de 1800 anos atrás:
- "*Pode ser e pode não ser*" - Sexto Empírico, Hipotiposes Pirrônicas



Referências

[1]- GPT-4 Technical Report
OpenAI
ArXiv on-line 27/03/2023
<https://arxiv.org/abs/2303.08774>

[2]-Sophie Bushwick. What the New GPT-4 AI Can Do
Scientific American Online 16/03/2023
<https://www.scientificamerican.com/article/what-the-new-gpt-4-ai-can-do/>

[3]- Walter Carnielli. How AI can be surprisingly dangerous for the philosophy of mathematics— and of science
Circumscribere Vol. 27 (2021)
<https://revistas.pucsp.br/index.php/circumhc/article/view/55033>

[4]- Hanmeng Liu, Ruoxi Ning, Zhiyang Teng, Jian Liu, Qiji Zhou, Yue Zhang.
Evaluating the Logical Reasoning Ability of ChatGPT and GPT-4
ArXiv Online 05/05 2023
<https://arxiv.org/abs/2304.03439>

[5]- Matt O'Shaughnessy. One of the Biggest Problems in Regulating AI Is Agreeing on a Definition
Carnegie Endowment for International Peace
<https://carnegieendowment.org/2022/10/06/one-of-biggest-problems-in-regulating-ai-is-agreeing-on-definition-pub-88100>

- [6]- Kevin Roose. A.I. Poses 'Risk of Extinction,' Industry Leaders Warn
<https://www.nytimes.com/2023/05/30/technology/ai-threat-warning.html>
- New York Times, 30/05/2023