

# Cátedra Oscar Sala

## Grupo de Pesquisa

## Inteligência Artificial Responsável

### **Grupo de Trabalho: Educação**

Márcia Azevedo Coelho

Claudia Helena dos Santos Araújo

Danielle Soares e Silva Bicudo Ferraro

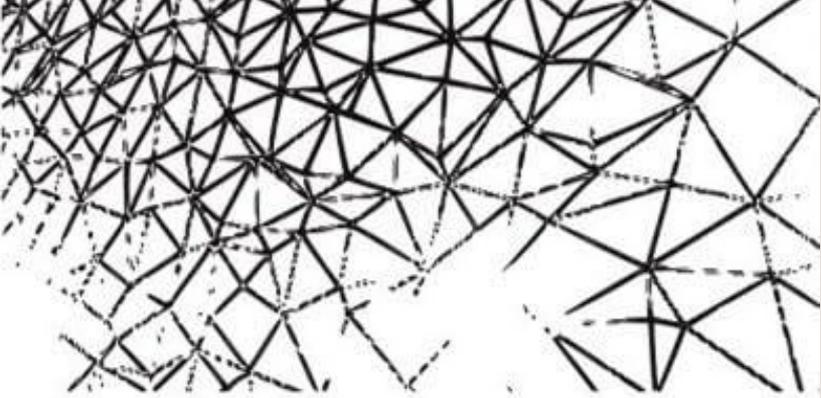
Evelyn Christina Peres Barrelin

Lívia Carolina Vieira

Maikel Pons Giralt

Paulo da Silva Quadros





# Há como deter a invasão do ChatGPT?

Lucia Santaella

Coleção  
**Interro**ações  
Coordenadora  
Lucia Santaella

 Estação  
das Letras  
e Cores

Fichamento para o GT Educação  
Danielle Ferraro

12 de dezembro de 2023

**ie]**<sup>A</sup>  
Instituto de  
Estudos  
Avançados da  
Universidade de  
São Paulo

 Catedra Oscar Sala

# A autora



**Lucia Santaella** é pesquisadora 1 A do CNPq, professora titular no programa de Pós-Graduação em Comunicação e Semiótica e no programa de Pós-graduação em Tecnologias da Inteligência e Design Digital, ambos da PUCSP. É Coordenadora da Pós-graduação em Tecnologias da Inteligência e Design Digital, Diretora do CIMID, Centro de Investigação em Mídias Digitais e Coordenadora do Centro de Estudos Peirceanos, na PUCSP. É presidente honorária da Federação Latino-Americana de Semiótica e Membro Executivo da Asociación Mundial de Semiótica Massmediática y Comunicación Global, México, desde 2004.

ie] <sup>A</sup>  
Instituto de  
Estudos  
Avançados da  
Universidade de  
São Paulo

 C S Cátedra Oscar Sala

Fonte: Biblioteca Virtual da FAPESP

# Introdução

"Há como deter a invasão do ChatGPT", de Lucia Santaella, é o resultado de vasta pesquisa bibliográfica da autora, com referências muito recentes (muitas de 2023). O foco do livro é compreender, a partir do ChatGPT, os inúmeros efeitos que a Inteligência Artificial Generativa está produzindo em diversas camadas da sociedade: arte, educação, comunicação etc.

Capítulos do livro

Capítulos 1: Conhecer e compreender o funcionamento do ChatGPT

Capítulo 2: Habilidades e limites do ChatGPT

Capítulo 3: Abalos na autoria, criatividade e autonomia

Capítulo 4: Produção científica

Capítulo 5: Educação

Capítulo 6: Os dados estão lançados/Conclusão

# Alguns questionamentos iniciais

O avanço do ChatGPT pode ser detido? É possível regular a IA? É possível propor sistemas de IA transparentes e passíveis de explicação, confiáveis e seguros a riscos, eticamente responsáveis, democráticos e inclusivos? É preciso informar que é a máquina, e não os humanos, que produzem tais conteúdos?

Brasil: 2 visões sobre a IA (pró-mercado X riscos)

UE: regulamentação, obrigações de acordo com os riscos aos direitos fundamentais da humanidade, incluindo a educação. ChatGPT e outras ferramentas de IAG: capazes de produzir textos, imagens, códigos, áudios.

Ambivalência das IAGs: não temos respostas negativas ou afirmativas. A autora busca pensar a partir de uma moldura de referências: com o que estamos lidando? Quais suas implicações sociais e consequências para o humano?

IAG, para a autora, está no campo de estudo da linguagem - linguística e da semiótica verbal.

# Capítulo 1: O cérebro do ChatGPT

A autora traz muitas referências técnicas. Conhecer, cercar-se de informações, para compreender de que, afinal, estamos falando.

Os estudos de IA iniciaram-se há 70 anos, mas o campo desenvolveu-se, muito, nos últimos anos, de tal forma que as aplicações de IA são onipresentes.

O avanço da IA, dos últimos anos, deve-se à aprendizagem de máquina e, em especial, ao deep learning, possível a partir das redes neurais artificiais, que imitam o funcionamento das redes neurais de nosso cérebro.

## Redes neurais artificiais (RNA):

- Estruturas complexas interligadas por elementos de processamento simples que realizam operações como cálculos, em paralelo, para processamento de dados e representação do conhecimento.
- Compostas por várias unidades de processamento, conectadas por canais de comunicação associados a determinado peso.
- O comportamento inteligente vem das interações entre as unidades de processamento de rede.
- Série de equações matemáticas cujos parâmetros são ajustados para realizar regressões logísticas de forma eficaz.
- Técnica que levou a uma variedade de aplicações: carros autônomos, reconhecimento facial e processamento de linguagem natural, potente interação humano-computador presente em chatbots, tradutores, apps para resumos de textos.

# Panorama do ChatGPT

O ChatGPT, em si, é uma IA generativa, que funciona a partir de redes adversariais generativas (GAN).

GANs (2014): algoritmo de aprendizagem de máquina que usa 2 redes neurais que competem entre si para realizar previsões mais precisas.

Rede neural X rede neural: geram novas instâncias de dados sintéticos que melhoraram algumas aplicações, como a dublagem de filmes, além de conteúdos educacionais. Por outro lado, graças a elas surgiram as deep fakes, modalidade mais complexa de fake news.

IAG + Transformers = surgem modelos de linguagem revolucionários

- **Transformers:** tipo de machine learning que permitiu treinar modelos de dados cada vez maiores, sem que fosse necessário rotular os dados. Assim, novos modelos puderam ser treinados em bilhões de páginas de texto, resultando em respostas mais profundas.
- Sua arquitetura de dados usa mecanismos de auto-atenção ou regressão, treinado em um conjunto de dados de textos não rotulado, o que potencializa o aperfeiçoamento do algoritmo com treinos sem intervenção direta humana. Por ex.: retirar algumas palavras do texto e treinar o programa para acertar o que falta, assim o modelo de linguagem aprende a preencher as frases.
- Modelo poderoso e generalizável que permite estabelecer conexões entre palavras em páginas, capítulos e livros. Rastreia conexões para analisar códigos, proteínas, produtos químicos e DNA.

# Panorama do ChatGPT

O ChatGPT é um Large Language Models, modelo algorítmico que aprende a realizar associações estatísticas entre bilhões de palavras e frases para executar tarefas como gerar resumos, responder perguntas e classificar textos, o que representa um grande avanço na NLP (processamento de linguagem natural).

## **G: Generative - P: Pre-trained - T: Transformer**

O pré-treinamento generativo melhora as taxas de acerto, mas depende da base de dados que gerou a rede neural. O Transformer permitiu a aceleração e economia de tempo, pois o treinamento ocorre de forma paralela, em muitos processadores.

Desenvolvida pelo Google, essa técnica permitiu o avanço de muitos outros modelos.

- ChatGPT: aprende a dar respostas relacionando palavras, contexto (é treinado a relacionar as palavras).
- GPT3: absorve todos os padrões que puder: gramática, estrutura, gênero...
- GPT 3.5: avanço foi adquirir a maleabilidade da linguagem
- Até então, os modelos funcionavam com base no que aprendiam (Prompt de entrada → saída), de forma supervisionada. O GPT3 aprende tarefas a partir de poucos exemplos, mesmo sem treinamento supervisionado.
- GPT4: tem mais parâmetros (mais valores de uma rede neural). Responde à linguagem e às imagens.
- A OpenAI (ao contrário do Google) não compartilhou a técnica.

# Os demônios ocultos da linguagem

A caixa-preta do ChatGPT: 175 bilhões de pesos.

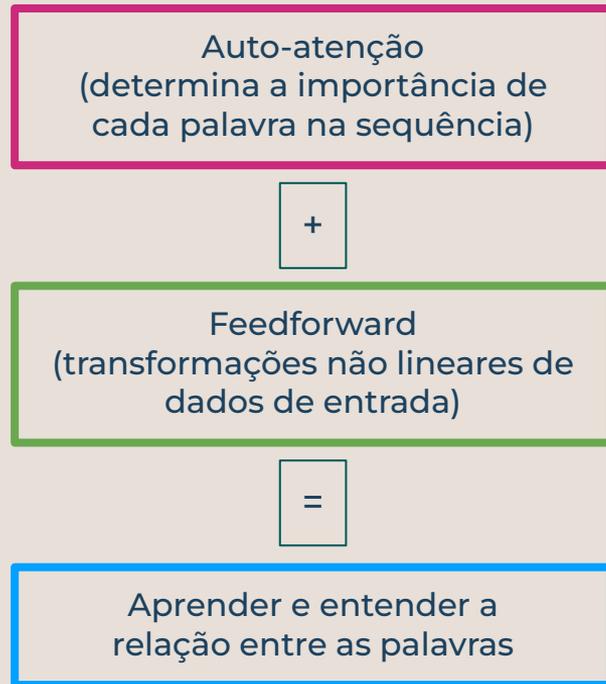
Como foram determinados?

Treinamento em grande escala, de bilhões de páginas escritas por seres humanos.

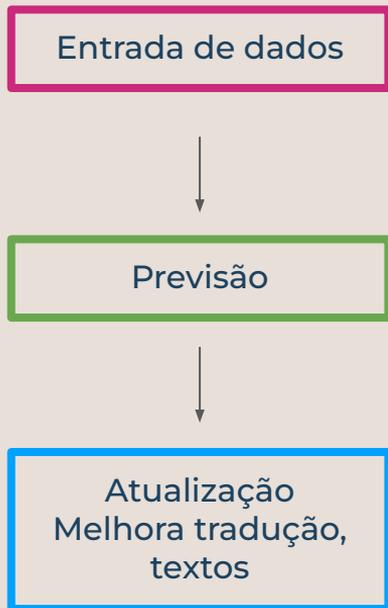
Transformers: não precisam de dados rotulados/supervisionados. Grandes volumes de dados que desenvolveram correlações estatísticas entre textos, sem serem rotulados.

ChatGPT dá continuidade ao que alguém está escrevendo com base no que existe na Web. Dado o texto X, a próxima palavra (ou probabilidade de palavras) é determinada pela sua relevância na web.

Camadas/subcamadas do transformer:



# Os demônios ocultos da linguagem



Treinamento: à medida em que respostas corretas são inseridas no sistema, o ChatGPT aprende as melhores opções. Essas respostas são ordenadas pela relevância. Desta forma, o texto vai se aprimorando e se tornando mais significativo.

O ChatGPT, inicialmente, foi treinado por humanos (trabalho terceirizado pela OpenIA) - palavras/modelos classificados e corrigidos. Porém, não há controle sobre as respostas, pois o sistema não tem auto-advertência.

Imprevisibilidade para situações com as quais não foi treinado: Black Box do ChatGPT.

# Os demônios ocultos da linguagem

Black Box do Chat GPT, as redes neurais, funcionam com um volume gigantesco de dados, que interagem de maneiras complexas, sem que haja conhecimento sobre suas estatísticas. Como as máquinas chegaram a determinados resultados? Podem acontecer:

- Tarefas novas e imprevisíveis
- Comportamentos emergentes auto-organizativos
- Origem a novas combinações

Redes neurais recorrentes: analisam palavra por palavra, por isso são previsíveis, aprendem com a própria experiência.

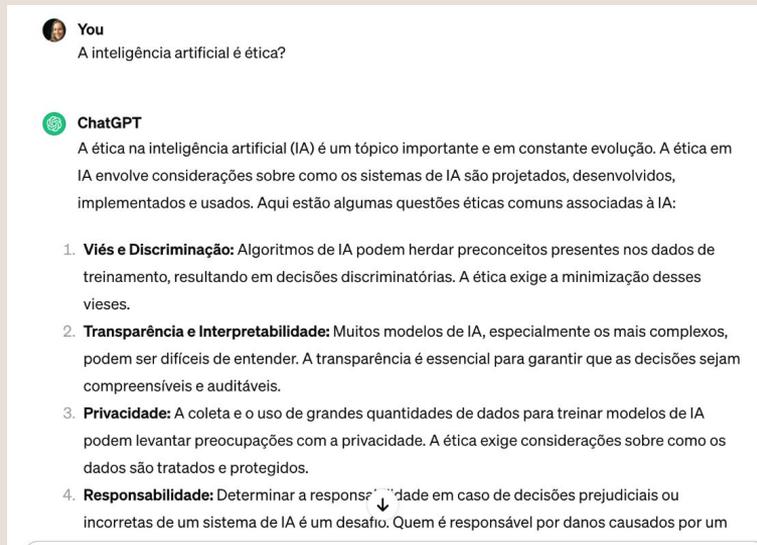
Transformer: analisa grandes volumes de textos, de forma paralela, graças às redes neurais. Mais memória, mais conexões e mais capacidade de imitar a linguagem humana.

Capacidade de solucionar novos problemas sem quaisquer dados de treinamento, o que pode gerar vieses ideológicos, preconceitos e imprevisibilidade, ou seja, aumento da margem de risco do sistema.

# Capítulo 2: Proezas e Armadilhas do ChatGPT

ChatGPT: sutiliza internas, inacessíveis aos leigos

Capacidade de responder por escrito a perguntas com grande maleabilidade para desenvolver conversações aparentemente humanas - com a diferença de não demonstrar qualquer sinal de cansaço... um companheiro de conversas capaz, inclusive, de fornecer conselhos sobre questões pessoais, ainda que não tenha sido treinado para isso...



The image shows a chat interface with a white background. On the left, there is a small circular profile icon for 'You'. The user's question is 'A inteligência artificial é ética?'. On the right, there is a green circular icon for 'ChatGPT'. The response discusses the ethics of AI, mentioning its evolution and the need for ethical considerations. It lists four key points: 1. **Viés e Discriminação:** Algorithms can inherit biases from training data. 2. **Transparência e Interpretabilidade:** Complex models are hard to understand, making transparency essential. 3. **Privacidade:** Large amounts of data are used for training, raising privacy concerns. 4. **Responsabilidade:** Determining responsibility for AI decisions is a challenge.

**You**  
A inteligência artificial é ética?

**ChatGPT**  
A ética na inteligência artificial (IA) é um tópico importante e em constante evolução. A ética em IA envolve considerações sobre como os sistemas de IA são projetados, desenvolvidos, implementados e usados. Aqui estão algumas questões éticas comuns associadas à IA:

- Viés e Discriminação:** Algoritmos de IA podem herdar preconceitos presentes nos dados de treinamento, resultando em decisões discriminatórias. A ética exige a minimização desses vieses.
- Transparência e Interpretabilidade:** Muitos modelos de IA, especialmente os mais complexos, podem ser difíceis de entender. A transparência é essencial para garantir que as decisões sejam compreensíveis e auditáveis.
- Privacidade:** A coleta e o uso de grandes quantidades de dados para treinar modelos de IA podem levantar preocupações com a privacidade. A ética exige considerações sobre como os dados são tratados e protegidos.
- Responsabilidade:** Determinar a responsabilidade em caso de decisões prejudiciais ou incorretas de um sistema de IA é um desafio. Quem é responsável por danos causados por um

Imagem: diálogo com o ChatGPT

# Habilidades do ChatGPT

Imensas possibilidades de geração de conteúdo, desde receitas culinárias até dicas para encontrar trabalho, escrever textos acadêmicos, poemas e histórias, redigir modelos jurídicos, uma carta. A autora denomina o ChatGPT de "tarefeiro multitalentos".

Tais habilidades vem das associações estatísticas entre bilhões de palavras e frases, aprendidas por seus algoritmos. Ao contrário dos chatbots anteriores, o ChatGPT pode lembrar o que um usuário disse anteriormente, abrindo a possibilidade de criar bots de terapia (antigo ELIZA), com velocidade e coerência.

O ChatGPT 3 já tinha grande capacidade conversacional, o que dificultava distinguir sua prosa da humana, pois tem um toque lúdico na interação.

→ performance que agrega as mais variadas atividades comunicacionais das pessoas:

- Capacidade de ajudar programadores a detectar e a corrigir erros em códigos.
- Elaborar e responder perguntas analíticas das tarefas escolares
- Diagnósticos médicos
- Criar jogos baseados em textos
- Explicar conceitos científicos
- Resumir documentos legais
- Dar respostas de atendimento ao cliente
- Identificar indivíduos em risco em uma comunidade

# O ChatGPT não trabalha sozinho

ChatGPT só se inicia com um prompt (comando): texto em que o usuário exprime o que deseja obter. Assim, a conversação é conectada, instaurando "uma performance a dois".

O chat foi treinado para ser dócil e polido em suas respostas, mesmo quando não consegue dar conta da solicitação do usuário, o que a autora classifica como fator humano, "de relevância fundamental na performance do chat".

Suas frases apresentam coerência sintática (foi treinado para isso).

O nível de complexidade de suas respostas, numa primeira experiência, pode deixar os usuários impressionados. O ChatGPT tem um grande poder imitativo: ele não cria nada por si, mas imita modelos pré-existentes.

O Chat não tem compreensão sobre o que escreve (aqui a autora recorre aos pós estruturalistas para falar sobre significado e a Peirce).

ChatGPT só se inicia com um prompt (comando): texto em que o usuário exprime o que deseja obter. Assim, a conversação é conectada, instaurando "uma performance a dois".

# O ChatGPT não trabalha sozinho

O chat foi treinado para ser dócil e polido em suas respostas, mesmo quando não consegue dar conta da solicitação, o fator humano, "de relevância fundamental na performance do chat".

Suas frases apresentam coerência sintática (foi treinado para isso).

O nível de complexidade de suas respostas pode deixar os usuários impressionados. O ChatGPT tem um grande poder imitativo: ele não cria nada por si, mas imita modelos pré-existentes. O Chat não tem compreensão sobre o que escreve.

Essa questão é uma das armadilhas do ChatGPT.

Desenvolvedores e usuários devem injetar ética num sistema que fala como gente, sem ter ideia do que fala, alimentado por palavras sem conexão com a realidade que representam ou indicam. Porém, dá respostas que parecem ser dotadas de entendimento, em discursos que parecem humanos. Ao agir como se fosse humano, o ChatGPT induz o seu interlocutor ao esquecimento de que não passa de um simulador de humano. Com respostas rápidas, pode-se perder a capacidade de uma mediação lenta, de uma ponderação reflexiva. Emprega-se um "pragmatismo para a prática automatizada e robotizada em que o humano passa a imitar a própria automatização do robô".

# Internalidades negativas do ChatGPT

O ChatGPT comete erros ao gerar respostas baseadas em suposições probabilísticas, de não ter informações atualizadas para todos os usuários do sistema. Não tem opiniões sobre debates políticos, oferece um resumo imparcial do que cada lado acredita. Foi programado para recusar pedidos inapropriados (o que os usuários conseguem burlar, reformulando solicitações).

Sua maior armadilha, entretanto, provém de sua coerência sintática.

Sabemos dos riscos da coleta de dados constante a que somos submetidos, mas nos esquecemos de que o Chat age como se fosse gente, sua "função encantatória", o que esconde sua maior fragilidade: a falta de senso comum.

O ChatGPT é um modelo matemático, aprisionado no reino da linguagem, que se comporta como se aprendesse e interagisse como humano.

Desconectado do contexto externo e por não ter filtros éticos próprios, devido à sua incapacidade de compreender o que diz, o Chat pode passar para frente discursos de ódio, gerar estereótipos racistas e sexistas, refletindo as associações que seus dados de treinamento lhe concedem (assim como todas as IAs).

Para alguns autores, o ChatGPT pode levar a um apagão da inteligência se as pessoas confiarem cegamente nele.

# Internalidades negativas do ChatGPT

Sem as devidas precauções, os usuários podem transformar esse sistema numa máquina de gerar Fake News.

As dificuldades éticas ocorrem quando confiamos aos algoritmos autônomos tarefas com consequências para os seres humanos.

Algoritmos não têm a liberdade de ser moral ou imoral, pois são apenas uma sucessão de estados inequívocos. A questão está em quem projeta e em quem usa os seus sistemas: são as pessoas que obedecem a padrões morais. É preciso que esses padrões sejam injetados no sistema e que os usuários sejam capazes de antecipar as consequências éticas de quando passam à frente resultados falsos de suas consultas ao ChatGPT.

A arte da suspeita e a arte do cuidado.

Ética e regulamentação seriam as soluções para IA e para o Chat. Qualquer sistema de IA deveria ter supervisão humana durante todo o processo de iniciação de operação, coleta e aplicação de dados.

Seria o Chat tão perigoso, como já se pensou que seriam os carros, para a humanidade? Para a autora, o Chat estende nossa capacidade de pensar e falar, porém, despido de autoentendimento e de ética - os dois constituintes mais complexos da humanidade. Assim, impacto social, questões éticas e filosóficas devem ser levadas em consideração.

# Capítulo 3: Abalos na autoria, criatividade e autonomia

IA Generativa: as 4 aplicações mais utilizadas para esse novo tipo de tradução de texto em imagem: Dall-E (OpenIA); Midjourney (via extensão do Discord); Stable-Difusion (open source) e Imagen (Google).

ChatGPT causou mais impacto pois afeta todos os seres humanos e letrados, levando nisso todas as profissões e formações educacionais. Afinal, onde a linguagem não está envolvida?

As mas IAGs de imagens estão abalando a questão de autoria, criatividade e autonomia.

# IAGs de Imagens

Modelos com bases técnicas no Transformer, rede neural baseada na tecnologia de Modelos de Difusão, que utiliza o aprendizado multimodal para conectar a semântica entre textos e imagens (produzem imagens a partir de comandos verbais). Os modelos tem dois componentes principais: uma rede neural para emparelhar uma imagem com o que o texto descreve e outra para gerar imagens do zero.

Os modelos de difusão são treinados para limpar as imagens removendo o ruído pixelado adicionado no treinamento, até que as imagens originais sejam apagadas e restem somente os pixels aleatórios. A rede neural é treinada para prever como seria a versão menos pixelizada de uma determinada imagem - após várias vezes, o modelo resultará numa imagem de alta resolução.

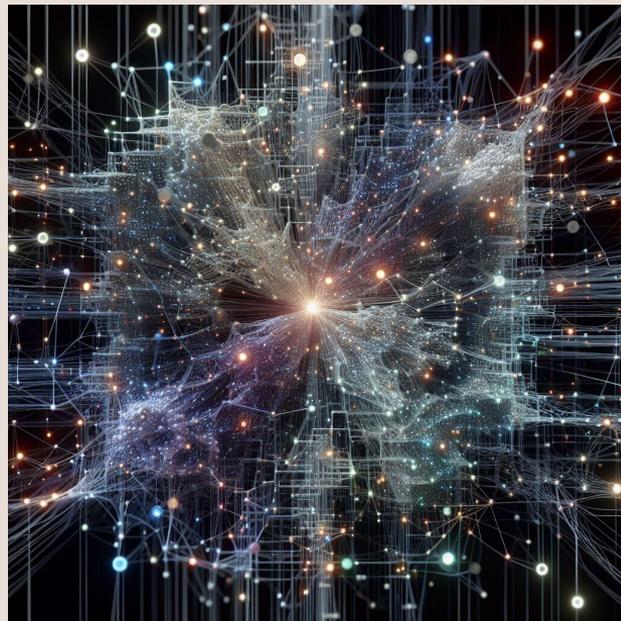


Imagem gerada no Dall-e

# Um cenário em ebulição

Como fica o agenciamento e a autonomia das produções artísticas?

Enquanto alguns artistas digitais levaram um susto, outros estão vendo os modelos de conversão de texto em imagem como uma inovação única, pois podem levar da ideia a um primeiro esboço em segundos. Velocidade para criar e explorar.

Trabalho de vários departamentos de animação digital em poucos minutos.

A longo prazo, essas tecnologias podem gerar projetos para quase tudo (cinema, moda, arquitetura, publicidade...).

E a questão de autoria e propriedade intelectual de imagens muito semelhantes às que já existiam?

# Capítulo 4: A produção do conhecimento

O ChatGPT provocará mudanças no modo como se produz pesquisa científica.

Metodologias e coletas seletivas de informação.

Wikipedia - hoje é fonte de consulta, ainda que não seja confiável, é um direcionamento para pesquisas.

Google: fonte de informações, filtrar conteúdos.

Como ficam as produções acadêmicas com o ChatGPT?

# Reações Instantâneas

ChatGPT: deve ser incorporado como autor?

O ChatGPT pode ser responsabilizado por seu conteúdo, pela validade e integridade dos artigos científicos?

Co-autor: o ChatGPT pode fazer contribuições relevantes?

Cuidados:

- O verniz da objetividade e a fluência da linguagem podem gerar uma falsa credibilidade ou rigores da ciência, que o ChatGPT não possui.
- Falsas referências, textos que parecem bem escritos, porém são superficiais, estereotipados.
- Mediador, considerando suas limitações: ausência de fontes, opacidade do modelo, de seus dados de treinamento e de seu código.
- Já existem regras para publicações.

# Para que serve o ChatGPT na ciência?

Já existem assistentes automatizados de pesquisas com modelos de IA que podem:

- Digitalizar bancos de dados de artigos
- Produzir resumos sucintos e precisos de literatura existentes
- Busca e seleção de arquivos
- Mapas de referências de artigos e citações
- Mostrar impacto de determinadas pesquisas
- Leitura de artigos
- Escrita (correção gramatical e ortográfica, linguagem, estrutura, dados, tabelas e referências, além de sugestões para melhorar o texto)
- Gerar tabelas e infográficos
- Criar apresentações, unindo o ChatGPT e Dall-e para imagens

Pontos negativos: autoria e direitos autorais, custos, colonialismo de dados, internacionalização da pesquisa, mantendo a relação centro e periferia na produção do conhecimento.

# Ponderações da autora

A ciência é longe de ser feita apenas com textos escritos.

A IA pode revolucionar a forma como os componentes das ciências se relacionam. Sistemas de IA poderão ajudar a formular hipóteses mais fortes, com modelos de candidatos mais promissores a novos medicamentos; simulações de experimentos (antes de levar para o mundo "real"); condução de experimentos de forma mais rápida, barata e em maior escala.

É preciso cuidado e promover regulamentação adequada, que permita acesso igualitário a recursos, que controle questões éticas, vieses nos algoritmos.

# Capítulo 5: Dilemas e desafios da educação

A ciência é longe de ser feita apenas com textos escritos.

A IA pode revolucionar a forma como os componentes das ciências se relacionam. Sistemas de IA poderão ajudar a formular hipóteses mais fortes, com modelos de candidatos mais promissores a novos medicamentos; simulações de experimentos (antes de levar para o mundo "real"); condução de experimentos de forma mais rápida, barata e em maior escala.

É preciso cuidado e promover regulamentação adequada, que permita acesso igualitário a recursos, que controle questões éticas, vieses nos algoritmos.

# Primeira fase

- Instituições e professores precisam compreender como essa nova tecnologia funciona e o que ela pode tornar possível.
- Não há educação sem a incorporação de mediações técnicas ou tecnológicas que contemplem os sistemas e processos educacionais como um todo.
- Os sistemas educacionais, ainda presos a uma formatação nascida na cultura do livro e da transmissão sob as luzes do Iluminismo, tiveram que buscar mudanças por meio de metodologias ativas, sala de aula invertida, metodologia de projetos, mudanças no papel do professor, inserções de recursos digitais, ainda que sujeitos às lógicas das plataformas.
- Educação midiática: educação sob o ponto de vista das novas tecnologias. Termo inadequado, pois estamos na era das redes e da performatividade algorítmica.
- Inércia da educação diante da emergência de tecnologias de IA



Instituto de  
Estudos  
Avançados da  
Universidade de  
São Paulo

# Primeira fase

- A introdução de recursos tecnológicos nas escolas foi a partir de laboratórios de informática, envelhecidos pela inutilidade atual do computador, abrindo caminho para cursos online e a evolução de ambientes virtuais de aprendizagem.
- Resumindo: utilização de recursos oferecidos pelo computador e seus complementos na educação.
- Termo genérico: e-learning: aprendizagem que se desenvolve por meio eletrônico. Surgiram então m-learning (aprendizagem movel) e a u-learning (aprendizagem ubíqua).
- Aprendizagem adaptativa: ambiente de aprendizagem é construído para atender às necessidades e estilos de aprendizagem de cada estudante.
- Aplicativos de IA estão sendo usados como facilitadores para a aprendizagem e sua gestão.
- Digitalização, convergência tecnológica, ubiquidade e, por fim, a entrada do ChatGPT: reviravolta na educação, em que modelos tradicionais de educação foram postos à prova.

# Segunda fase

Após o pânico inicial, ficou claro que:

- os estudantes vão, inevitavelmente, trabalhar em um ambiente de produção ubíqua de conteúdos;
- veículos de produção acadêmica já estabeleceram políticas de co-autoria com o Chat;
- é inevitável repensar as práticas pedagógicas;
- os chatbots vieram para ficar e bani-los não é a solução.

Desta forma,

- muitos professores passaram a acreditar que o Chat pode ajudar a tornar o ensino melhor, com potencial para aumentar a interatividade nas aulas;
- usar o Chat para educar os alunos para o uso das redes;
- preparar planos de aula mais personalizados;
- substituir o professor em tarefas burocráticas.
- Pensar em reviravolta pedagógica!

# Segunda fase

E no Brasil (e outros países com o mesmo cenário)?

- professores sobrecarregados;
- escolas precarizadas, sem recursos elementares, como conectividade.

... A autora não responde, mas destaca que:

- o uso da IAG por alunos é gratuito e bastante facilitado;
- são mudanças que não dependem de instalações tecnológicas sofisticadas;
- o celular abriu as portas para a aprendizagem ubíqua;
- a IAG só intensifica os dilemas educacionais;
- a sequência de emergência tecnológica hoje alcança seu cume com a IAG, tendendo a criar uma brecha entre alunos de prontidão para o uso e professores despreparados para mudanças no planejamento;
- a única maneira de normalizar e regular o Chat é mais discussão, orientação e compreensão.

# Sugestões

- Existe uma brecha entre alunos e professores que instituições estão tentando estreitar. Diversas publicações propõem solução para o dilema:
- o ChatGPT pode levar os professores a se desviarem do foco excessivo nos resultados finais, pois o engajamento cooperativo com a IA poderia levar a um pensamento crítico capaz de tornar o ensino mais humano - e não formar alunos robôs.
- o ChatGPT pode alimentar a desinformação - o que também pode ser tomado como estratégia pedagógica, despertando o pensamento crítico/questionador dos alunos.
- as "alucinações" do Chat podem ser tomadas como uma oportunidade para discussões produtivas sobre informações equivocadas e vieses.

# Plágio, Informações Falsas

- É possível detectar se um texto foi escrito por humano ou pelo Chat?
- Ainda não temos sistemas capazes de detectar, com precisão, a procedência dos textos. A efetividade piora em casos de textos produzidos por IA e ajustados por humanos;
- Outra questão é que frases produzidas pelo Chat parecem corretas. Seus modelos funcionam prevendo a próxima palavra, mas não tem a menor ideia se algo é correto ou falso. Também apresenta confiança discursiva nas afirmações, mesmo quando não são verdadeiras;
- Ferramentas de IA podem distorcer ainda mais as informações que consumimos;
- Precisamos aprender a diferenciar a escrita humana da escrita de IA para combater usos indevidos de tecnologia.

# Papel dos professores

- Nenhuma das questões citadas anteriormente justifica a "inércia pedagógica". É preciso uma reviravolta mais ampla no papel do professor já que a informação, antes exclusiva das salas de aula, hoje está em todos os lugares: primeiro online, depois em chatbots.
- Educadores são facilitadores da informação: devem mostrar aos alunos não apenas como encontrar conteúdos relevantes, mas em quais informações confiar ou não, sabendo diferenciá-las.
- Trazer mais interatividade para a sala de aula, sendo que o ChatGPT pode funcionar como um contra-argumentador de um aluno.
- Não é o fim da educação, mas um recomeço.

# Das sugestões à sistematização

Enfrentamento dos problemas educacionais que emergem com os grandes modelos de linguagem a partir do documento da Harvard Business Publishing, "Por que todas as aulas de repente se tornaram aulas de IA?" e de iniciativa da Fundação Dom Cabral:

- focar nos problemas apenas distrai das oportunidades que a IA pode oferecer;
- começar com os educadores se familiarizando com as ferramentas da IA;
- esses dispositivos farão parte da vida estudantil, o que implica treinar os alunos para o uso adequado dessa tecnologia, trabalhando com modelos de IA;
- experimentar a IA para aumentar as expectativas dos alunos e aprimorar o aprendizado;
- criar uma política de sala de aula que preserve o uso adequado da IA;
- aprender onde e como é apropriado usar um sistema de IA constitui-se em uma etapa fundamental que os educadores podem tomar para mitigar o potencial de danos.

# Ponderações pontuais

- Enquanto não existirem detectores de IA, não existe alívio para as instituições de educação. Por outro lado, a IA está presente e se impregnando cada vez mais na educação.
- O mundo digital já havia provocado a emergência de um novo ecossistema de transferência de conhecimento, cuja complexidade só tem crescido.
- Em meio à convergência tecnológica e à ubiquidade, os modelos tradicionais de educação foram postos à prova.
- Os jovens estão utilizando ferramentas de IA e o ChatGPT - para tarefas educativas ou não. Ao invés de proibir, é urgente estabelecer regras para sua utilização, adequadas à cada tipo de comunidade. Policiamento, proibição de uso, é o menos efetivo nesse processo.
- Por outro lado, os princípios da Educação, que visam o desenvolvimento humano, e estão na base da pedagogia, devem ser mantidos.
- Não há uma estratégia única: é preciso levar o nível de aprendizagem em conta, sem grandes generalizações.
- Promover a saída do amadorismo pelas escolas: desenvolver mecanismos de regulação que visem às fontes, à concepção e à certificação desse sistema. Assim como o Chat deve estar de acordo com os códigos éticos das instituições de ensino.
- É preciso redesenhar a relação ensino-aprendizagem, do processo educativo até as formas de avaliação. As mudanças colocam o educador em estado de alerta permanente e infatigável.

# Os dados estão lançados

- O choque com as IAGs se refere à capacidade que esses sistemas têm de exibir algo muito similar ao modo como operam o pensamento, a compreensão e a criatividade humana.
- A maioria das consequências das relações do ChatGPT com o humano se refere a um tronco comum: o tronco da autoria.
- Ao competir com a habilidade humana de conversar, ler, escrever e responder, o ChatGPT fere aquilo que pensamos quando dizemos: Eu escrevo, Eu crio, Eu falo.
- As considerações finais deste livro se voltam, então, à questão: esses sistemas podem, de fato, pensar, compreender e criar? O Chat, afinal, tem inteligência?

# Ladainhas minimizadoras

- É preciso fugir das velhas dicotomias ocidentais para entender o conceito de autoria e inteligência que está sendo discutido em textos e imagens produzidos por IAG. Alguns autores defendem que as LLMs não são inteligentes: são sistemas treinados para dar a aparência externa da inteligência humana (Goff, 2023);
- A ideia de que programas como o ChatGPT podem representar uma forma reconhecível de inteligência é ainda mais prejudicada por sua arquitetura... um sistema que obedece a regras para cada ato mecanicista de produção gramatical... nem sequer faz sentido falar sobre o Chat como uma entidade singular (Newport, 2023);
- A IA está sendo usada para o marketing de si mesma e de terceiros, segundo Nicoletti (2023). Para o autor, inteligência tem que ser orgânica, portanto, a IA não é tão inteligente, não se compara à inteligência humana. Voltamos aqui, segundo Santaella, para a ideia de Descartes, de que as máquinas são burras;
- Carnielle (2023) diz que o ChatGPT não tem criatividade, capacidade de inovação ou imaginação humana, nem a capacidade intuitiva de paralelismo de causa e efeito. Também falta ao Chat a capacidade de dedução e de abdução, ao contrário dos humanos.

# Ladainhas minimizadoras

- Ossmy (2023) diz que o Chat tem algoritmos muito sofisticados, mas basicamente coleta muita informação da internet para conseguir atender diferentes perguntas e solicitações. O que falta ao Chat é a capacidade de se adaptar, de aprender certas coisas dentro de um cenário de um ambiente diferente. Os algoritmos não são flexíveis como o cérebro dos bebês.
- Para a autora, nega-se a inteligência às máquinas, como se todos soubessem o que é a inteligência. Noção a partir de um modelo de inteligência; ou a inteligência é tomada como sinônimo de consciência.
- Discutir a inteligência da IA é uma maneira retroativa de discutir os impactos resultantes das interações humanas com esses sistemas. Embora distintas, essas inteligências levam os humanos a fazer coisas que não fariam sem elas, coisas das mais diversas ordens: desde auxiliar um diagnóstico até atolar as redes de fake news. Depreciar as formas de inteligências não humanas é minimizar os efeitos que os sistemas algorítmicos podem provocar.

# Inteligência vs agenciamento e sua refutação

- Teoria de causalidade de Peirce para refutar a ideia do filósofo Luciano Floridi sobre inteligência x agenciamento.
- Relação triádica entre acaso, causação eficiente, causação final
- Resumo: a autora sugere que precisamos parar de antropofomizar a IA
- Autores que tiram do foco a questão da inteligência da IA:
- Betanzos (2023) diz que a IA excede nossa inteligência em campos específicos, como em diagnósticos de câncer (mas são péssimas generalizadoras);
- Para Komlosy (2023), a IA é capaz de processar dados de forma mais rápida e muito mais abrangentes que os humanos. Não devemos confundir a inteligência da IA com a capacidade criativa e não linear dos humanos;
- Ramos (2023) destaca a capacidade de aprendizagem autônoma das IAs, identificando padrões e relacionando-os, sem necessidade de intervenção humana - fenômeno de aprendizagem contextual.

# Sem ética, a inteligência degenera

Bandeiras vermelhas para o futuro:

- controle cada vez maior, a partir dos dados, de nossas mentes, corpos e experiências, pelas empresas detentoras de IA;
- a IA está pronta para remodelar nossas vidas, e nós não temos regulação definida;
- estruturas de governança, valores éticos e morais que moldam o desenvolvimento e uso das IAs;
- a IA não reforçará apenas a inteligência humana, mas também nossa estupidez, no individual e no coletivo;
- consumo massivo de notícias falsas como fonte de informação, usuários deixam de pensar por eles mesmos.

Conclusão:

Sem pessimismo, nem otimismo, evitando as dualidades, é preciso conhecer para compreender o que nos espera.