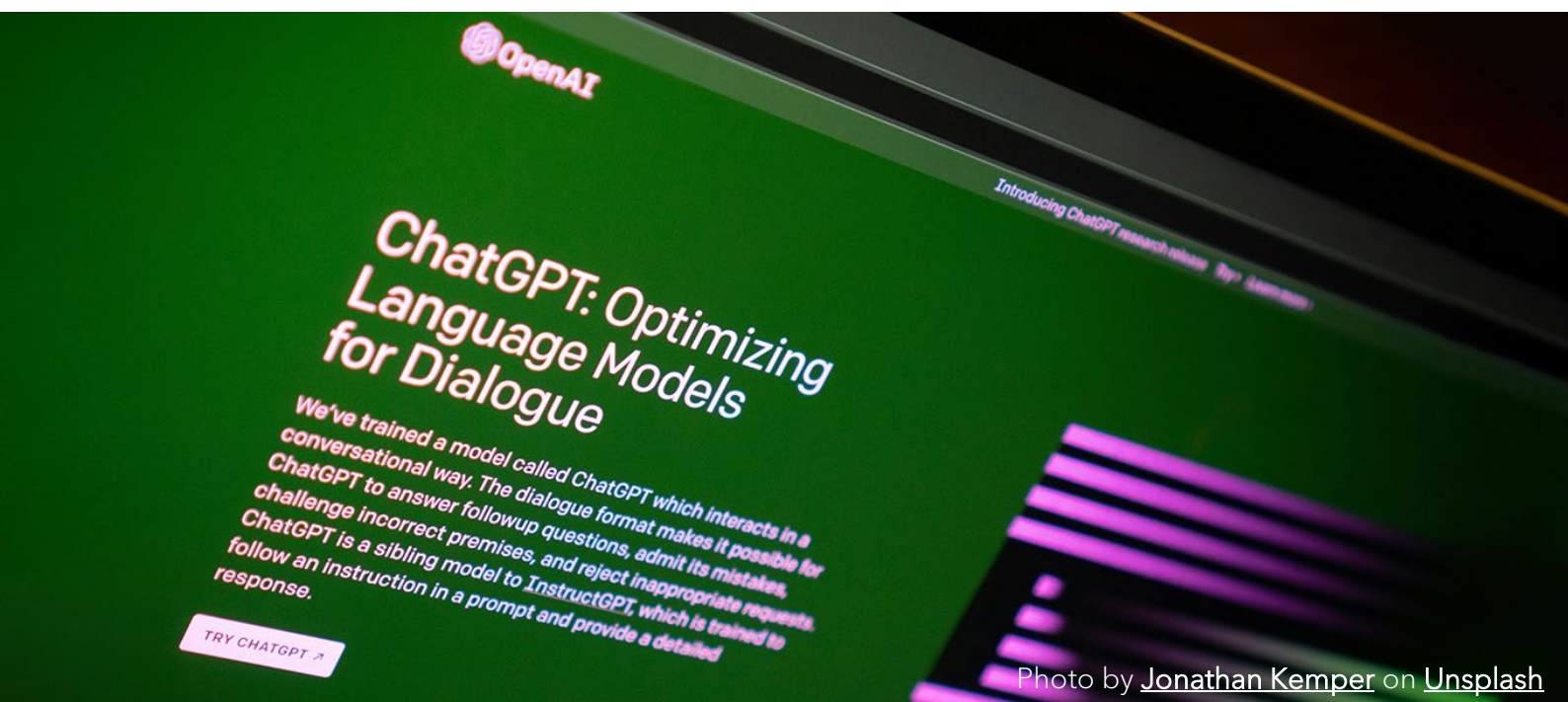


ChatGPT em Ciência de Dados

Análise Crítica do Artigo



UC Inteligência Artificial Simbólica para Ciência de Dados

Licenciatura Ciência de Dados

Grupo 17, CDC1 e CDC2

Docente

Isabel Machado Alexandre

André Filipe Gomes Silvestre N°104532

Maria Margarida Ramos Pereira N°105877

Sumário Executivo

A Inteligência Artificial (IA) tem redefinido o panorama tecnológico, e o *ChatGPT*, surge como uma peça-chave neste cenário. Este modelo avançado de linguagem, baseado na arquitetura *GPT*, destaca-se pela sua capacidade única de compreender e gerar texto de forma natural.

Este trabalho explora a relevância do *ChatGPT* na Ciência de Dados (CD), destacando o papel crucial da utilização destas tecnologias em tarefas de análise, interpretação e manipulação de dados. Ao mergulhar nas áreas emergentes de *Explainable AI* e *Responsible AI*, analisamos como o *ChatGPT* impulsiona a evolução desses campos.

Destacam-se ainda as aplicabilidades desta tecnologia, bem como as implicações éticas e considerações inerentes ao seu uso, com principal foco na transparência do processo de autorização de direitos de autor.

Por fim, abordamos os rumores do futuro do *ChatGPT*, as inovações das empresas concorrentes e perspetivamos as potenciais utilizações futuras dessas tecnologias.

Palavras-chave: Inteligência Artificial; *ChatGPT*; Ciência de Dados; Processamento de Linguagem Natural; Direitos Autorias; Programação; Ética

Índice

Introdução.....	4
Inteligência Artificial.....	4
OpenAI	4
ChatGPT	4
Objetivo do Trabalho	7
Síntese do Artigo	8
Tese do Artigo.....	8
Problema Resolvido.....	9
Persuasão dos Autores.....	10
Técnica/Abordagem do Artigo.....	11
Técnica/Abordagem que Artigo não faz.....	11
Relevância para Ciência de Dados	12
Crítica ao Artigo.....	13
Comparação com Outros Modelos	14
O Problema do ChatGPT	15
Processo de Autorização de Direitos de Autor	15
Regulamento Europeu de Inteligência Artificial.....	17
Oportunidade de Negócio.....	17
ChatGPT-4 e Futuro do ChatGPT	19
GPT-5.....	19
Aplicações Futuras Possíveis	20
Novidades da Concorrência	21
Google	21
Microsoft	21
Conclusão.....	22
Bibliografia.....	23

Índice de Figuras

Figura 1 Fontes do dataset de treino do ChatGPT-3.	5
Figura 2 Ilustração dos Modelos Transformer.....	6
Figura 3 Exemplo da Interação com o ChatGPT numa questão ligada a programação.	10
Figura 4 Exemplo de um erro comum do ChatGPT.....	12
Figura 5 Gráfico comparativo dos principais modelos linguísticos de IA 2018-2023.....	14
Figura 6 Conferência da Amy Webb no Nordic Business Forum.....	15
Figura 7 Infografia das 50 plataformas de IA	18
Figura 7 Campanha da CocaCola na Índia celebrativa do Diwali.	20

Introdução

Inteligência Artificial

A Inteligência Artificial (IA) é um campo da ciência da computação que se dedica a desenvolver sistemas capazes de realizar tarefas que normalmente requerem inteligência humana. Estes sistemas são concebidos para aprender, raciocinar, resolver problemas e até mesmo compreender linguagem natural. Utilizando algoritmos avançados, a IA procura simular a capacidade cognitiva humana, permitindo a máquina realizar ações autónomas e adaptar-se a novas situações. [1]

A Inteligência Artificial é uma área em incessante evolução que tem revolucionado a forma como interagimos com a tecnologia. No epicentro desta revolução destaca-se o lançamento do *ChatGPT* pela *OpenAI* em novembro de 2022, uma notável revelação, cuja utilização é amplamente utilizado em diversas áreas. [2][3]

OpenAI

Historicamente, em primeiro lugar é fundada a *OpenAI* em dezembro de 2015, originalmente criada como uma organização sem fins lucrativos. Mais tarde, em 2019, a Microsoft tornou-se um dos principais investidores, injetando 1bilião de dólares nesse mesmo ano e mais recentemente, este ano, injetou mais 10biliões. O seu principal foco é investigar sobre IA de modo a promover e desenvolver ferramentas que tenham consideração os princípios de *Responsible AI*. [4]

ChatGPT

O *ChatGPT* assume um papel proeminente na revolução da comunicação assistida por IA, fornecendo interfaces de conversação que transcendem as expectativas convencionais.

O *ChatGPT* é uma ferramenta avançada desenvolvida pela *OpenAI*, sendo uma versão da família de modelos de linguagem *Generative Pre-trained Transformer (GPT)*. Baseado na arquitetura *GPT*, utiliza algoritmos de *Deep Learning* para compreender e gerar texto de forma natural. Este modelo é treinado em grandes conjuntos de dados, permitindo-lhe aprender padrões linguísticos complexos.

Segundo um artigo de pesquisa da *OpenAI* [5], o *ChatGPT - 3* foi treinado com um conjunto de dados que incluía dados textuais de 5 fontes, cada uma com a ponderação proporcional representada na **Figura 1**.

Para compreender o gráfico é necessário esclarecer que

Common Crawl consiste em dados de páginas da web, extrações de metadados e extratos de texto de mais de 8 anos de rastreamento da web;

WebText2 consiste em *posts* do *Reddit* com 3 ou mais votos positivos;

Livros 1 e **Livros 2** consiste em duas coleções de livros baseadas na Internet.

Fontes de Dados do Treino do ChatGPT-3

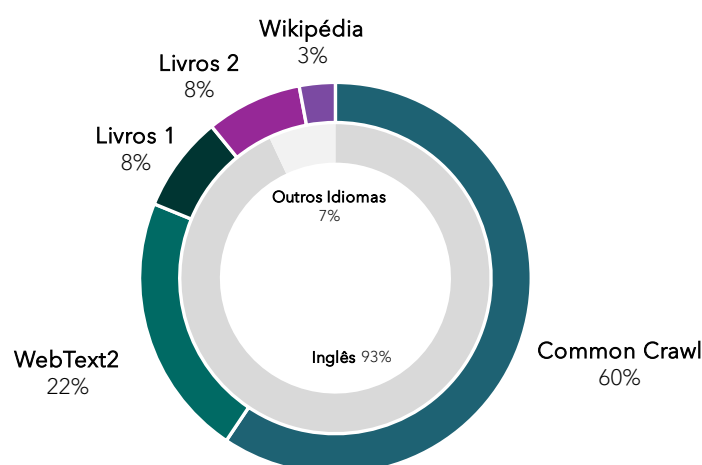


Figura 1 | Fontes do *dataset* de treino do *ChatGPT-3*.

Fonte: Representação Gráfica da Tabela 2.2 de [5]

Adicionalmente é de referir que todo este treino foi feito com um *dataset* em que 93% do conteúdo é de língua inglesa.

Quanto ao domínio específico dos algoritmos, esta tecnologia destaca-se pelo uso de uma arquitetura de rede neuronal baseada no modelo *Transformer*. Esta abordagem inovadora utiliza conhecimentos previamente adquiridos num contexto específico, para aprimorar o desempenho do modelo em novas tarefas. Durante o treino, o modelo é ensinado a prever a palavra seguinte numa sequência de texto com base nas palavras anteriores. O uso estratégico desta arquitetura é uma das características inerentes que preconiza o *ChatGPT* como uma ferramenta avançada na geração de texto e interações assistidas por IA, com diversas aplicações nas mais variadas tarefas do quotidiano. [6]

Deste modo, este modelo (Fig. 2) destaca-se na compreensão de perguntas, fornecendo respostas contextualmente relevantes e interagindo de forma convincente com os utilizadores, tornando-se uma notável aplicação de IA na geração de texto em vários idiomas e comunicação assistida por computador. [6]

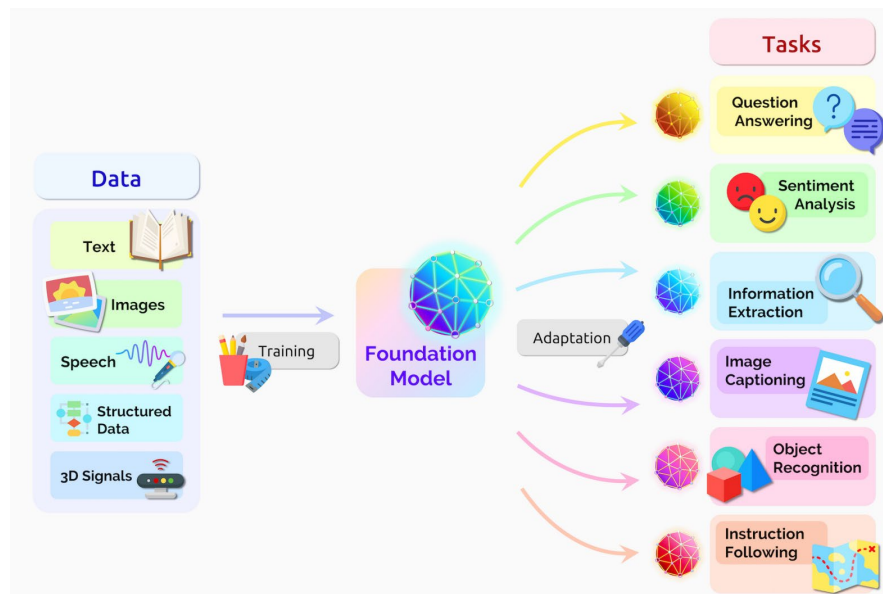


Figura 2 | Ilustração dos Modelos *Transformer*.
Fonte: [7]

No que diz respeito às linguagens utilizadas e suportadas, atualmente o *ChatGPT* suporta 95 idiomas, apesar de ter sido projetado principalmente para compreender e gerar texto em inglês. Independentemente de o foco estar na compreensão de linguagens naturais, este produz respostas nas mais conhecidas linguagens de programação, podendo codificá-las e compreendê-las. [8]

Objetivo do Trabalho

O domínio da Ciência de Dados (CD) está a evoluir exponencialmente, impulsionada pelo avanço tecnológico e pela crescente necessidade de conhecimento obtido através da análise de dados. Assim, podemos defini-lo como uma abordagem multidisciplinar que analisa e interpreta conjuntos de dados, para extrair informações valiosas e *insights* significativos. [9]

Num mundo cada vez mais digital, a CD desempenha um papel crucial em diversas áreas, tornando possível fazer tomadas de decisão mais informadas, identificar oportunidades de otimização e resolver problemas complexos. Assim, esta desempenha um papel vital na condução de inovações, impulsionando o desenvolvimento de tecnologias emergentes como a inteligência artificial e o *Machine Learning*. [9]

O aumento da quantidade de dados disponíveis e a evolução das ferramentas tecnológicas têm contribuído para o rápido crescimento e importância desta área, tornando-a uma disciplina indispensável em diversos setores.

Neste sentido, foi proposto no âmbito da UC de *Inteligência Artificial Simbólica para Ciência de Dados* inserida na Licenciatura de Ciência de Dados, uma análise de um artigo científico que visa a aprendizagem sobre duas novas áreas da inteligência artificial: *Explainable AI* e *Responsible AI*.

A *Explainable AI* refere-se a ferramentas e *frameworks* para entender e interpretar os seus modelos de *Machine Learning*; enquanto que a *Responsible AI* define-se como a prática de conceber, desenvolver e implementar a IA com boas intenções para capacitar o Homem, e ter um impacto justo na sociedade. [10][11]

Atendendo ao artigo de perspectiva escolhido, intitulado " *The Role of ChatGPT in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field*", explora-se as oportunidades e os desafios associados à utilização do *ChatGPT* na ciência dos dados e examina-se como pode ser utilizado para melhorar várias tarefas neste domínio. [12]

Síntese do Artigo [12]

Tese do Artigo

O artigo explora o impacto do *ChatGPT* na Ciência de Dados, destacando como esta ferramenta influencia a programação, a solução de problemas e a ética na área.

Esta pesquisa evidencia de maneira inequívoca o impacto significativo do *ChatGPT* em setores específicos, revelando uma compreensão mais profunda sobre como a inteligência artificial influenciará diretamente a educação e as oportunidades de emprego, especialmente no campo da CD.

Ademais, salienta a necessidade de iniciativas inovadoras por parte das universidades para desenvolver programas de formação que para além de complementarem os cursos existentes, também habilitem as futuras gerações a utilizar estas tecnologias generativas de forma eficiente, integrando princípios éticos.

A utilização do *ChatGPT*, embora ofereça avanços notáveis, enfrenta desafios consideráveis que não podem ser ignorados. Entre estes, destacam-se questões relacionadas com imprecisões, privacidade, parcialidade e plágio.

Este, como qualquer modelo linguístico, não é sempre preciso, especialmente em situações que envolvem complexidade, ambiguidade ou domínios específicos, como a matemática e a medicina. Vários fatores influenciam a precisão do *ChatGPT*, incluindo a qualidade dos dados de treino, a natureza da pergunta e a presença de viés nos dados. É fundamental reconhecer que ao lidar com questões matemáticas complexas, a supervisão humana e fontes adicionais de informação podem ser necessárias para garantir a precisão e confiabilidade.

No contexto mais amplo das inovações tecnológicas, o artigo destaca que a responsabilidade recai sobre a humanidade, dotada de criatividade, pensamento crítico e discernimento, para direcionar o uso do *ChatGPT* em prol da sociedade. A meta é criar um ambiente em que a IA seja um instrumento para potenciar a inteligência humana, em vez de substituí-la.

Em suma, o artigo conclui que o *ChatGPT* tem o potencial de revolucionar a Ciência de Dados, mas é importante usá-lo de forma responsável e informada, nunca esquecendo que se trata de um modelo que não consegue explicar como é que o seu resultado foi gerado.

Problema Resolvido

O problema central que o artigo aborda consiste em compreender como o *ChatGPT* influencia as práticas de programação, a resolução de problemas e as necessidades éticas a ter em consideração nesta área.

Segundo os mesmos, esta ferramenta pode ser utilizada para criar código de automatização de processos de recolha, formatação ou limpeza de dados; definir estruturas de dados; orientar-nos sobre as infografias que devem ser produzidas e as informações que devem conter; criar material de formação; identificar as fontes de dados necessárias para tarefas específicas; criar dados sintéticos; dar conselhos sobre conformidade, regulamentação e medidas práticas que podem ser tomadas para garantir que as operações de dados são legais, imparciais, e éticas; e ajudar a identificar processos analíticos que conduzam a melhores práticas.

Tendo em conta todo o potencial referido, à medida que a tecnologia continua a evoluir, na perspetiva do artigo, é provável que se torne uma ferramenta cada vez mais eminente no domínio da CD.

Todavia, é igualmente realizada uma análise detalhada do impacto da sua utilização, considerando setores específicos, como a educação e as oportunidades de emprego, especialmente no âmbito da Ciência de Dados, onde destacam a necessidade de adaptação e inovação.

Os autores argumentam que o *ChatGPT* tem o potencial de automatizar tarefas, analisar dados não estruturados e gerar novas ideias. No entanto, eles também destacam os potenciais riscos do mesmo, como a perda de empregos e a geração de conteúdos enganosos ou prejudiciais.

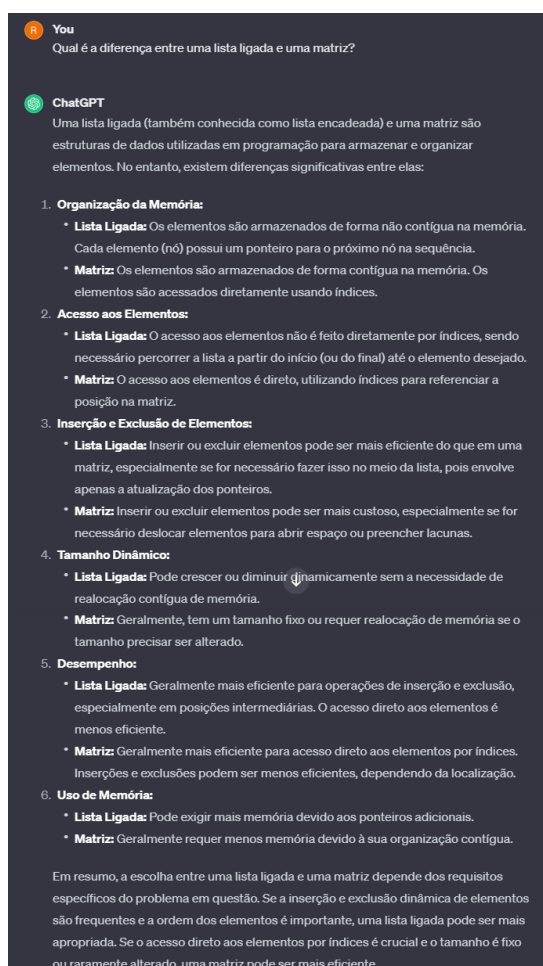
Ao salientar a necessidade de iniciativas inovadoras por parte das universidades, como o desenvolvimento de programas de formação que integrem princípios éticos, os autores indicam uma abordagem proativa para capacitar as futuras gerações a o utilizar de maneira eficiente.

De um modo geral, os autores abordam a falta de especialistas em Ciência de Dados e como o *ChatGPT* pode ser uma solução, *"dada a sua capacidade de formar leigos para codificar e programar soluções para problemas analíticos"* (Hassani & Silva, 2023). Estes, alegam que pode ter um impacto positivo, mas que também é importante considerar os potenciais riscos.

Persuasão dos Autores

Como visto anteriormente, a convicção transmitida pelos autores reside na visão de que o *ChatGPT* é uma ferramenta valiosa para aprimorar habilidades de programação, antecipar desafios e fornecer soluções através da análise de dados existentes.

Dada a fácil e intuitiva utilização, o aparecimento deste tipo de ferramentas suprime a necessidade de extensa formação em matemática, estatística, programação ou IA, tornando-o mais acessível, eficiente e eficaz na análise de dados, modelação preditiva e tradução de línguas.



Como exemplo, os autores apresentaram casos práticos nos quais a ferramenta pode ser particularmente benéfica, como é o caso da **Figura 2**, de modo a testar a veracidade dos resultados apresentados aquando da assistência do seu uso no esclarecimento de uma questão ligada à programação: *Qual é a diferença entre uma lista ligada e uma matriz?*

Figura 3 | Exemplo da Interação com o *ChatGPT* numa questão ligada à programação.

Fonte: *ChatGPT*

A influência do *ChatGPT* é vasta e a preparação adequada é vital. Ao incorporá-lo nos diferentes processos, é possível otimizar tempo e recursos, bem como obter *insights* mais profundos e precisos. Os autores também reconhecem os riscos potenciais do seu uso, mas argumentam que eles podem ser mitigados através de uma utilização responsável. Assim, concluem que os benefícios de utilização na Ciência de Dados são superiores aos custos.

Técnica/Abordagem do Artigo

Analisando a estrutura e abordagem do artigo, este explora exemplos práticos de como esta tecnologia pode ser utilizada para automatizar tarefas, destacando a facilidade de interação com a linguagem natural.

Inicialmente, é contextualizado o que é o *ChatGPT*, a sua estrutura e as suas aplicações, direcionando o foco da perspectiva para CD. Após exemplificar as suas visões, os autores comparam este modelo com outros semelhantes.

A abordagem do artigo é eficaz na medida em que fornece aos leitores uma compreensão tangível do potencial desta tecnologia. No entanto, a falta de uma discussão aprofundada das possíveis limitações éticas e viés da mesma é uma limitação.

Técnica/Abordagem que o Artigo não faz

O artigo apresenta uma contribuição valiosa para o campo da Ciência de Dados, proporcionando uma visão abrangente do potencial do *ChatGPT*.

No entanto, como qualquer modelo, é suscetível a limitações éticas e viés. Os autores, não abordam completamente estas limitações, deixando espaço para discussões mais aprofundadas sobre estas.

A discussão ética é crucial quando se trata de tecnologias de inteligência artificial, e a falta de abordagem nesse sentido no artigo pode deixar o leitor ávido por uma consideração mais abrangente das questões associadas.

Em projetos de Ciência de Dados, a integridade dos dados e a ética na manipulação e interpretação dos mesmos são cruciais. A utilização deste tipo de ferramentas, se não considerar a fundo essas implicações éticas, pode resultar em desafios associados, como possíveis distorções nos resultados, parcialidade nos dados e até mesmo plágio.

Relevância para Ciência de Dados

Corroborando tudo o que foi anteriormente exposto, é evidente que a relevância é destacada, com ênfase na assistência em desafios de programação, otimização de código e a criação de aplicativos interativos em ambientes como *Shiny*.

Na prática, a relevância para a Ciência de Dados pode ser observada em exemplos concretos de como o *ChatGPT* facilita a interação na linguagem natural, proporcionando um ambiente mais acessível para programadores e analistas de dados. Neste contexto específico, a sua integração pode simplificar tarefas relacionadas com a programação em linguagens como *Python* ou *Java*.

Contudo é impreterível ter atenção à precisão das suas respostas. E este é um dos aspetos críticos desta tecnologia em CD, pois os exemplos dados de possíveis erros que o *ChatGPT* falha são em cálculos, estatísticas e outros conceitos matemáticos avançados, que são áreas fulcrais em CD.

Como exemplificado abaixo (**Fig. 4**), quando pedimos a esta ferramenta para contar o número de vezes que aparece o número 100 na lista de números fornecida, o resultado não corresponde à realidade.

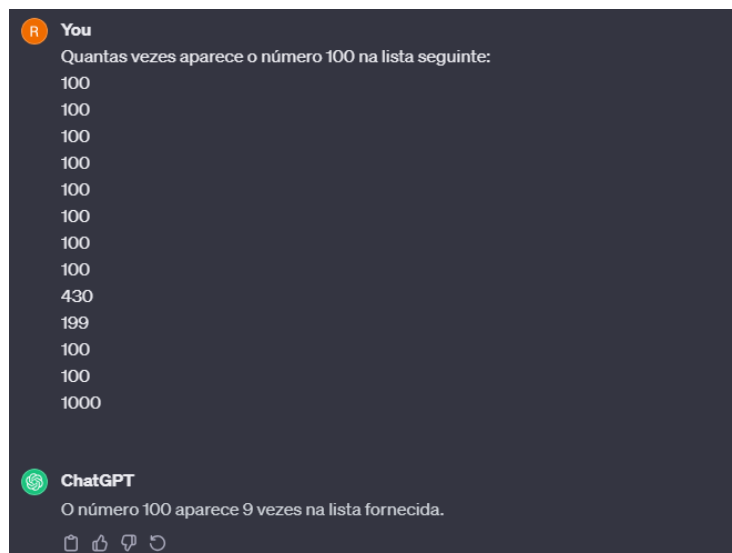


Figura 4 | Exemplo de um erro comum do *ChatGPT*.

Fonte: *ChatGPT*

Crítica ao Artigo

O artigo apresenta-se bem escrito e informativo, fornecendo uma visão abrangente do potencial impacto do *ChatGPT* na área da Ciência de Dados. Os autores discutem os benefícios potenciais, como a automatização de tarefas, a análise de dados não estruturados e a geração de novas ideias; e discutem os desafios e limitações do mesmo, como a confiabilidade, a imparcialidade e o potencial de viés.

Todavia face à incessante metamorfose desta temática e sendo o artigo datado de março de 2023, já se evidenciaram progressos notáveis nesta área e discussões das problemáticas a si intrínsecas.

Dos pontos em falta, destacamos de seguida alguns que consideramos relevantes referir.

Em primeiro lugar, os autores poderiam ter fornecido mais informações sobre as metodologias utilizadas para avaliar o *ChatGPT*. Por exemplo, poderiam ter discutido mais detalhadamente como foi treinado e como os resultados foram comparados a outros modelos de linguagem. Sendo esta uma IA generativa, subsiste a possibilidade de gerar conteúdo factualmente incorreto ou enganoso, contribuindo para o incremento da desinformação.

Em segundo lugar, os autores poderiam ter explorado mais aprofundadamente os potenciais impactos sociais e éticos do *ChatGPT*. A título de exemplo, relativamente ao treino com conjuntos de dados provenientes de conteúdos online, emergem questões sobre direitos autorais e responsabilidades decorrentes de eventuais danos originados pelos resultados produzidos por estes sistemas de IA.

Em terceiro lugar, os autores poderiam ter fornecido mais recomendações específicas sobre como integrar a formação *ChatGPT* em programas de CD ou, de modo mais abrangente, nos sistemas educativos. Esta é uma das significativas problemáticas associadas, digna de profunda reflexão.

Comparação com Outros Modelos

No que diz respeito à comparação com outros modelos, apresenta-se de seguida a **Figura 5**, que ilustra a comparação do número de parâmetros entre os modelos de linguagem até março de 2023, destacando o *GPT-3* à esquerda e o *GPT-4* à direita, representado a **vermelho**.

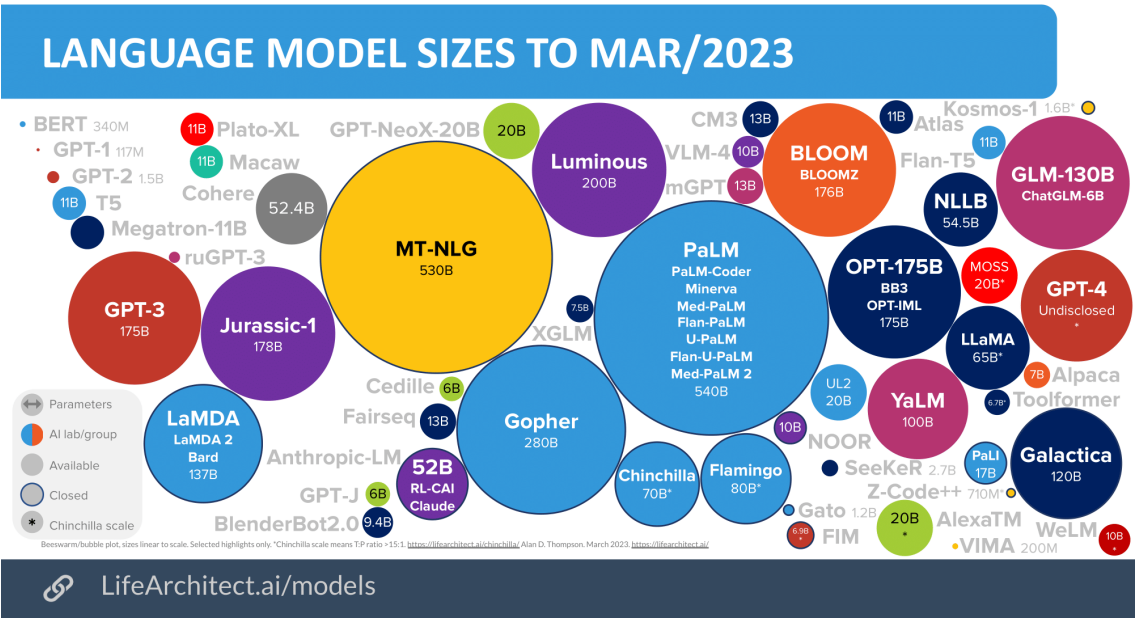


Figura 5 | Gráfico comparativo dos principais modelos linguísticos de IA 2018-2023
Fonte: [13]

Este gráfico foi produzido pelo Dr. *Alan D. Thompson*, perito mundial em IA, especializado no aumento da inteligência humana. Nele pode-se observar uma análise comparativa à quantidade de parâmetros, tornando a comparação mais intuitiva e informativa do que a **Tabela 1** do artigo em análise [12].

No entanto, o artigo mantém a sua relevância, uma vez que contribui para a compilação de informações numa temática que continua em desenvolvimento.

De seguida faremos uma análise a algumas problemáticas que considerámos pertinentes de aprofundar acerca do *ChatGPT* de modo a complementar o que foi referido no artigo.

O Problema do *ChatGPT*

Processo de Autorização de Direitos de Autor

O *ChatGPT* tem sido usado para uma variedade de propósitos, incluindo a criação de conteúdo criativo, a tradução de idiomas e a resposta a perguntas de forma informativa. No entanto, o seu uso levanta preocupações sobre os direitos autorais.

Recentemente, o evento *Nordic Business Forum*, que ocorreu em outubro de 2023, contou com a presença da futurista e autora *Amy Webb* (Fig. 6).

Na sua conferência [14], a mesma discutiu as complexidades das fontes de dados, a transparência e as considerações éticas que moldam o futuro da IA.

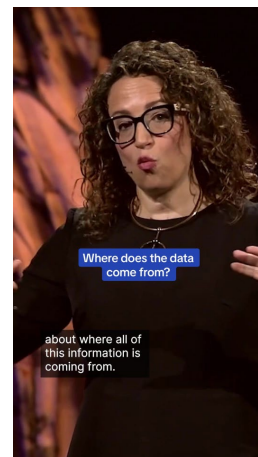


Figura 6 | Conferência da *Amy Webb* no *Nordic Business Forum*.
(Nota: Clicar para visualizar o vídeo)
Fonte: [14]

Webb revelou que o *ChatGPT*, e versões subsequentes, foram treinados em conjuntos de dados expandidos, incluindo fontes como o *Book Corpus*, milhões de páginas da web, *Wikipedia*, *Reddit* e, de forma menos clara, dados de redes sociais e outros espaços.

O exemplo destacado por *Webb* foi a existência de um *corpus* denominado "*Book Three*", composto por 187.000 livros que foram ilegalmente extraídos, sem o consentimento dos editores. Surpreendentemente, *Webb* descobriu que todas as obras que já escreveu fazem agora parte do *ChatGPT*, levantando questões sobre a origem e a ética do uso desses dados.

Esta revelação destaca a necessidade urgente de debates e reflexões sobre o panorama em rápida evolução das tecnologias baseadas em IA, especialmente no que diz respeito à crescente utilização de conjuntos de dados extensos e à transparência das fontes. A conferencista destacou a importância de desenvolver tecnologias de IA que sejam justas, inclusivas e imparciais, evitando vieses e discriminação.

As preocupações expressas por *Amy Webb* ressaltam a importância de uma abordagem ética e responsável na integração dessas tecnologias, sublinhando que estas questões devem ser discutidas de forma urgente e imediata.

Assim, a utilização destes *chatbots* e ferramentas generativas na criação de conteúdo levanta preocupações, especialmente quanto à originalidade e distinção são requisitos essenciais.

No âmbito da propriedade intelectual, a singularidade das respostas produzidas por estas tecnologias torna-se crucial. À medida que esses sistemas geram conteúdos que se assemelham aos criados por humanos, a necessidade de estabelecer diretrizes claras para a criação e envio de obras torna-se evidente, especialmente em domínios sensíveis como o jornalismo e a escrita criativa.

Existem inúmeras formas de abordar a problemática do processo de autorização dos direitos de autores. Uma abordagem é exigir que os utilizadores obtenham autorização do autor original do texto antes de o usar. Outra abordagem seria desenvolver tecnologias que possam identificar texto protegido por direitos autorais e impedir que a tecnologia IA o gere.

No entanto, ambas as abordagens apresentam desafios. Exigir que os utilizadores obtenham autorização é difícil e logisticamente complexo, bem como desenvolver tecnologias que possam identificar texto protegido por direitos autorais é um árduo desafio.

Uma terceira abordagem é desenvolver novas leis que definam como os direitos autorais se aplicam ao conteúdo gerado por estas ferramentas. Essas leis poderiam exigir que os utilizadores atribuam crédito ao autor original do texto.

De acordo com o artigo da *European Journal of Risk Regulation* [15] publicado pela Universidade de Cambridge, o *ChatGPT* é treinado com enormes *datasets*, que incluem uma grande quantidade de obras protegidas por direitos autorais.

No entanto, o *ChatGPT* não é capaz de entender o significado do texto que é gerado. Ele simplesmente usa algoritmos para identificar padrões e correlações entre palavras. A lei dos direitos autorais, intrinsecamente ligada à criatividade e autoria humanas, levanta barreiras à elegibilidade de proteção quando falta o elemento humano no processo criativo. Isto significa que, *"de acordo com o critério de originalidade da UE, as obras geradas por IA podem não ser elegíveis, uma vez que carecem das escolhas criativas e da expressão pessoal de um autor humano."* (Lucchi, 2023). [15]

Em suma, apesar das capacidades atuais do *ChatGPT* em atender a padrões mínimos de criatividade, a proteção de direitos autorais permanece inacessível sem a contribuição criativa humana. No entanto, considerando o progresso tecnológico, vislumbra-se um futuro onde a redução gradual do envolvimento humano no processo criativo pode resultar em criações artísticas sem atribuição clara a um autor específico.

Regulamento Europeu de Inteligência Artificial

No seu mais recente comunicado de imprensa, datado de 9 de dezembro de 2023, a *Comissão Europeia*, o *Parlamento Europeu* e o *Conselho da União Europeia* chegaram a um acordo político sobre o *Regulamento de Inteligência Artificial*, que estabelece um quadro jurídico abrangente para a IA na União Europeia. [16][17]

O acordo político prevê uma abordagem baseada nos riscos para a regulamentação da IA. Os sistemas de IA serão classificados em três categorias, de acordo com o seu nível de risco mínimo, elevado e inaceitável. [16]

Assim, a União Europeia torna-se pioneira a estabelecer restrições ao uso da IA. A legislação deverá entrar em vigor no início do próximo ano, após a sua ratificação formal por ambas as partes, e deverá ser aplicada dois anos depois. [17]

Todavia, fica por mencionar, qual será a entidade incumbida de supervisionar, fiscalizar e avaliar esses riscos.

Oportunidade de Negócio

Relativamente às potencialidades do *ChatGPT*, este pode ser uma oportunidade de negócio em uma variedade de setores.

A título de exemplo, segundo um artigo da *Forbes* [18], pode ser utilizado por:

- **Criação de conteúdo:** usado para escrever artigos, histórias, roteiros e artes de conteúdo criativo, contribuindo assim para que as empresas gerem conteúdos novos de forma rápida e menos custosa.
- **Marketing:** utilizado na criação de conteúdo personalizado para clientes, como e-mails, mensagens de texto e anúncios, promovendo a fortificação da relação entre a marca e os clientes.
- **Serviço ao cliente:** potencializado para criar *chatbots*, é capaz de responder a perguntas e resolver problemas de clientes, permitindo às empresas não só reduzir custos operacionais, mas também elevar a satisfação do cliente mediante uma resposta eficiente e personalizada.

Além do *ChatGPT*, atualmente, há uma proliferação de tecnologias e plataformas que utilizam a IA.

No infograma seguinte (Fig.7), destacam-se as 50 tecnologias mais visitadas entre setembro de 2022 a agosto de 2023, permitindo que a criatividade e a destreza sejam o limite do humano no uso da IA em tudo o que faz, de forma a potenciar a produtividade e facilidade de certas tarefas.

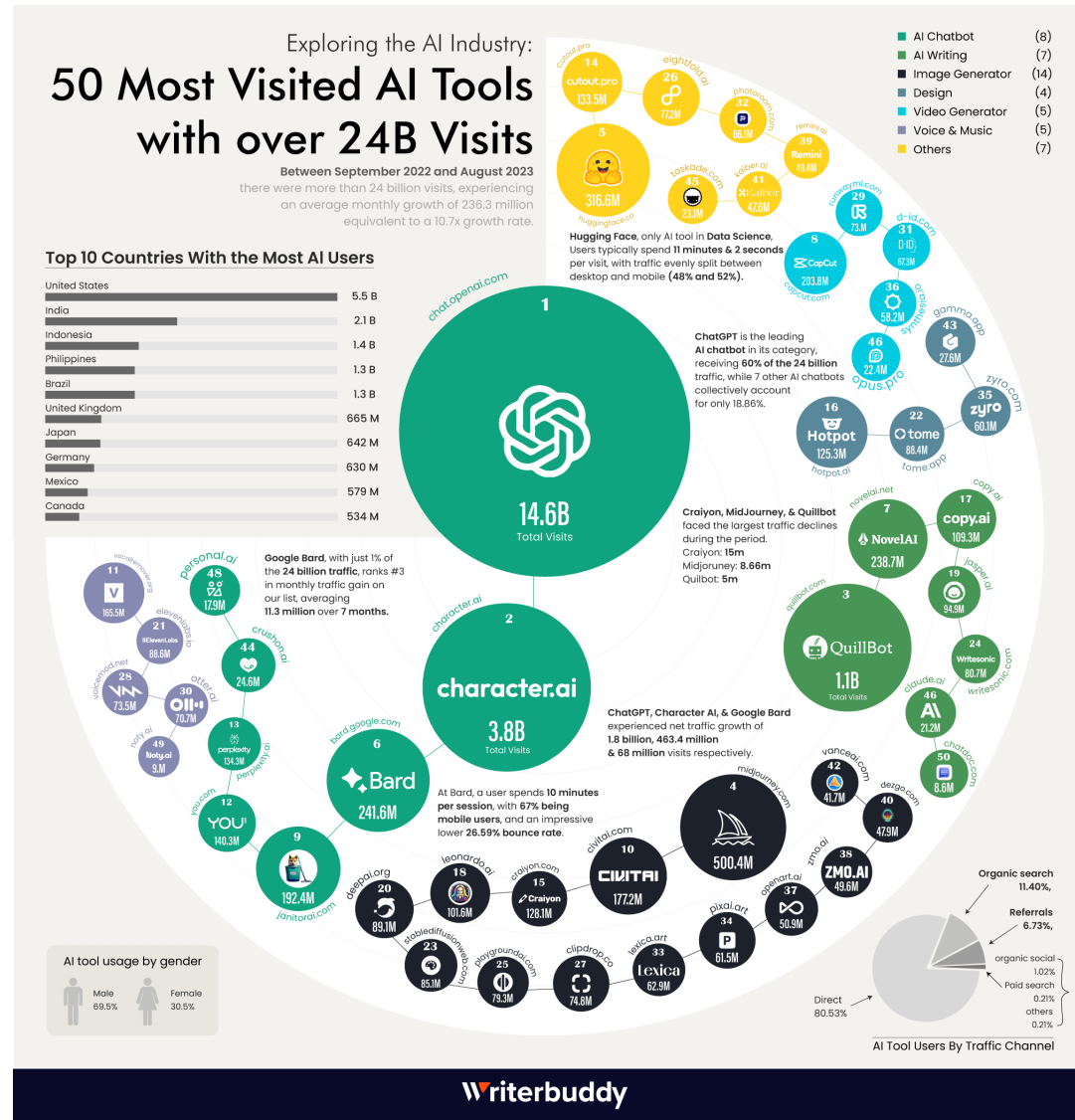


Figura 7 | Infografia das 50 plataformas de IA mais visitadas entre Set de 2022 a Aug de 2023. Fonte: [19]

Da análise do infograma (Fig.7), salienta-se o facto do ChatGPT ser a ferramenta de IA mais popular, tendo liderado com mais de 14 bilhões de visitas, representando mais de 60% do tráfego analisado. Até ao momento, os EUA são o principal mercado para ferramentas de IA, com 5,5 bilhões de visitas. Por fim, os dados de gênero revelam uma disparidade: 69,5% são utilizadores do sexo masculino, em comparação com 30,5% que são utilizadores do sexo feminino. [19]

ChatGPT-4 e Futuro do ChatGPT

Aquando da publicação do artigo analisado, ainda pouco se sabia sobre o GPT-4. Lançado a 13 de março de 2023, esta é uma versão mais recente do ChatGPT, segundo a OpenAI, é *"mais criativo e colaborativo do que nunca"*, podendo aceitar imagens como *inputs* e gerar legendas, classificações e análises. [20]

Dados partilhados pela OpenAI [20], revelam que *"passaram 6 meses a tornar o GPT-4 mais seguro e mais alinhado, [...] com 82% menos probabilidade de responder a pedidos de conteúdo não permitido e 40% mais probabilidade de produzir respostas factuais do que o GPT-3.5"* (Open AI, 2023).

Ser **10x** mais rápido, ter **8x** mais parâmetros, treinado com mais dados, assegurar mais segurança e mais versatilidade, são algumas das vantagens desta nova versão. Contudo, infelizmente para muitos utilizadores, com esta nova atualização, está associada também com a necessidade de uma subscrição mensal, ao contrário da versão anterior que é de uso gratuito.

GPT-5

Apesar de ainda não lançado, esta será, provavelmente, a próxima versão do ChatGPT. Segundo publicado pelo Dr. Alan D. Thompson, já referenciado anteriormente, atentou no seu blogue em [21] um comentário de Manolo Arroyo (*Global Chief Marketing Officer for The Coca-Cola Company*) proferido a 24 de novembro de 2023 no podcast *The Morning Brief* do jornal *The Economic Times* [22] no minuto 17m31s:

"Posso dar-vos talvez um insight, algumas notícias novas que ninguém partilhou até agora. Temos uma parceria com a Bain e a OpenAI...

Na verdade, fomos a primeira empresa a combinar o GPT, que é o motor que permite o ChatGPT, e o DALL-E. Na altura, ninguém sabia que, devido à parceria com a OpenAI, éramos a primeira empresa a utilizar o GPT-4 e o DALL-E 2 numa experiência digital integrada para o consumidor. [...]

Ninguém sabe, porque ainda não foi lançado, que o Coca Cola Diwali foi feito com o GPT-5, que ainda não está disponível comercialmente, e o DALL-E 3, que também não foi lançado..."

Assim, foi confirmado que a Coca-Cola está já a utilizar a nova versão da *OpenAI* na sua campanha na Índia que produziu imagens como a ilustrada na **Figura 7**.



Figura 8 | Campanha da CocaCola na Índia celebrativa do Diwali.

(Nota: A imagem contém uma hiperligação que permite visualizar o vídeo da campanha no *Twitter*, demonstrando a utilização da IA na geração de imagens únicas e personalizadas.)

Fonte: [23]

De forma iminentemente atual, foi lançada uma campanha de Natal com a mesma tecnologia, que permite a geração de composições artísticas alusivas a esta época, com a possibilidade de inserção de elementos personalizados conforme a criatividade de cada utilizador. [24]

Aplicações Futuras Possíveis

Com este caso, é indubitável que as aplicações destas tecnologias e ferramentas podem ser inúmeras e vir a alterar o paradigma mundial em certas áreas.

Como exemplo futurístico, podemos pensar num mundo em que cada indivíduo tem o seu próprio *Assistente Pessoal*, treinado com o seu próprio *GPT*; ou cada empresa ter um *Gestor de Projetos* utilizando estas ferramentas que avalia a performance de cada trabalhador. Estes e inúmeros outros exemplos podem ser imaginados, num mundo que será cada vez mais orientado por estas tecnologias.

Novidades da Concorrência

Próximo ao término do prazo para a entrega deste trabalho, as duas proeminentes empresas tecnológicas, *Google* e *Microsoft*, lançam os seus produtos baseados em IA.

Google

Concorrente direto do *ChatGPT*, o ***Gemini***, lançado a 8 de dezembro de 2023, foi apresentado pela Google como sendo “o seu maior e mais capaz modelo de IA”, cuja missão é *construir a IA de forma responsável para beneficiar a humanidade* (Google, 2023) [25][26]

Com uma pontuação de **90%**, o *Gemini Ultra* é o primeiro modelo a superar os peritos humanos no MMLU (*Massive Multitask Language Understanding*), que utiliza uma combinação de 57 disciplinas como matemática, física, história, direito, medicina e ética para testar tanto o conhecimento do mundo como as capacidades de resolução de problemas. [26]

Microsoft

Simultaneamente ao investimento multibilionário na *OpenAI*, a *Microsoft* desenvolve internamente as suas próprias soluções em IA.

Contrariamente ao *ChatGPT* e ao *Gemini*, a *Microsoft* opta por investir num SLM (*Small Language Models*), o ***Phi-2***, focalizando-se na resolução de problemas em domínios específicos com maior confiabilidade. Desenvolvido pela *Microsoft Research*, o modelo demonstrou excelentes capacidades de raciocínio e de compreensão da linguagem. [27]

Para a *Microsoft*, a qualidade dos dados de treino desempenha um papel fundamental na performance do modelo. Ao contrário do *GPT-4*, a *Microsoft* efetua uma curadoria de dados da *Web* que são filtrados de acordo com o seu valor educativo. [27]

Ainda que numa escala muito menor, o ***Phi-2*** pode representar uma alternativa para diversas aplicações em IA generativa e uma solução para a multinacional caso a parceria com a *OpenAI* chegue a um impasse no futuro.

Ambos os modelos deverão estar disponíveis em breve para o público, e então poderemos observar o que o futuro reserva em termos de resultados e impacto que possam ter.

Conclusão

Em síntese, o presente trabalho abordou de forma abrangente a temática da Inteligência Artificial, com especial destaque no impacto da tecnologia *ChatGPT* no campo da Ciência de Dados, destacando-se como um modelo de linguagem avançado, propiciando uma revolução na comunicação assistida por IA.

No contexto da Ciência de Dados, vital na análise de dados para extrair *insights* valiosos, o *ChatGPT* apresenta-se como uma ferramenta promissora. O artigo analisado, focado nas áreas emergentes da *Explainable AI* e *Responsible AI*, ressalta a necessidade de formação e adaptação às inovações tecnológicas.

No cenário atual, a evolução para o *ChatGPT-4* e as perspectivas do *ChatGPT-5* evidenciam uma trajetória ascendente, marcada por avanços tecnológicos, maior segurança e versatilidade. Contudo, a utilização responsável e a consideração de implicações éticas permanecem imperativas.

Em suma, o *ChatGPT* representa um marco notável no campo da Inteligência Artificial, com repercussões profundas na Ciência de Dados. O seu potencial inovador é inegável, todavia, o caminho para sua integração eficaz requer uma abordagem equilibrada entre inovação, ética e responsabilidade. O futuro, cada vez mais orientado por essas tecnologias, apresenta desafios e oportunidades que exigirão um enquadramento ético e regulatório, para assegurar o seu uso benéfico na sociedade.

Bibliografia

- [1] IBM. (n.d.). O que é inteligência artificial (IA)? <https://www.ibm.com/br-pt/topics/artificial-intelligence>
- [2] Soares, M. (2023, abril 30). *Impacto do Chat GPT na sociedade*. LinkedIn. <https://www.linkedin.com/pulse/impacto-do-chat-gpt-na-sociedade-margarida-soares/>
- [3] OpenAI. (2023). *ChatGPT*. OpenAI. <https://openai.com/ChatGPT>
- [4] Brockman, G., Sutskever, I., & OpenAI. (2015, December 11). *Introducing OpenAI*. Openai.com. <https://openai.com/blog/introducing-openai>
- [5] Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., & Hesse, C. (2020). *Language Models are Few-Shot Learners*. <https://arxiv.org/pdf/2005.14165.pdf>
- [6] AWS. (n.d.). *O que é GPT AI? - Explicação sobre os transformadores generativos pré-treinados - AWS*. Amazon Web Services, Inc. <https://aws.amazon.com/pt/what-is/gpt/>
- [7] Merritt, R. (2022, março 25). *What Is a Transformer Model?* NVIDIA Blog. <https://blogs.nvidia.com/blog/what-is-a-transformer-model/>
- [8] Flensted, T. (2023, agosto 26). *How Many Languages Does ChatGPT Support? The Complete ChatGPT Language List*. <https://seo.ai/blog/how-many-languages-does-ChatGPT-support>
- [9] AWS. (n.d.). *O que é ciência de dados?* <https://aws.amazon.com/pt/what-is/data-science/>
- [10] Google. (n.d.). *Google Responsible AI Practices*. Google AI. <https://ai.google/responsibility/responsible-ai-practices/>
- [11] Google. (n.d.). *Explainable AI*. Google Cloud. <https://cloud.google.com/explainable-ai?hl=pt-br>
- [12] A Hassani, H., & Silva, E. S. (2023). The Role of *ChatGPT* in Data Science: How AI-Assisted Conversational Interfaces Are Revolutionizing the Field. *Big Data and Cognitive Computing*, 7(2), 62. MDPI AG. Acedido em <http://dx.doi.org/10.3390/bdcc7020062>
- [13] Thompson, A. (2021, agosto 26). *GPT-4*. Dr Alan D. Thompson - Life Architect. <https://life architect.ai/gpt-4/>
- [14] Nordic Business Forum. (2023, novembro 23). *Nordic Business Forum on TikTok*. TikTok. <https://www.tiktok.com/@nbforumhq/video/7304646436099083552>

- [15] Lucchi, N. (2023, agosto 29). *ChatGPT: A Case Study on Copyright Challenges for Generative Artificial Intelligence Systems*. <https://www.cambridge.org/core/journals/european-journal-of-risk-regulation/article/ChatGPT-a-case-study-on-copyright-challenges-for-generative-artificial-intelligence-systems/CEDCE34DED599CC4EB201289BB161965>
- [16] Comissão Europeia. (2023, dezembro 9). *Press corner*. European Commission - European Commission. https://ec.europa.eu/commission/presscorner/detail/pt/ip_23_6473
- [17] Mukherjee, F. Y. C., Martin Coulter, Supantha. (2023, dezembro 9). *União Europeia torna-se pioneira a estabelecer restrições ao uso da IA*. PÚBLICO. <https://www.publico.pt/2023/12/09/mundo/noticia/uniao-europeia-tornase-pioneira-estabelecer-restricoes-uso-ia-2073107>
- [18] Cook, J. (2023, novembro 10). *5 ChatGPT Prompts To Identify Your Next Big Business Opportunity*. Forbes. <https://www.forbes.com/sites/jodiecook/2023/11/10/5-chatgpt-prompts-to-identify-your-next-big-business-opportunity/>
- [19] Sarkar, S. (2023). *Uncovering the AI Industry: 50 Most Visited AI Tools and Their 24B+ Traffic Behavior - WriterBuddy*. Writerbuddy.ai. https://writerbuddy.ai/blog/ai-industry-analysis?utm_source=www.theaivalley.com&utm_medium=newsletter&utm_campaign=the-most-advanced-video-ai-to-date
- [20] OpenAI. (n.d.). GPT-4 is OpenAI's most advanced system, producing safer and more useful responses. <https://openai.com/gpt-4>
- [21] Thompson, A. (2023). *GPT-5*. Dr Alan D. Thompson - Life Architect. <https://lifearchitect.ai/gpt-5/>
- [22] The Economic Times. (2023). *The Morning Brief | Coca-Cola's Mega Marketing Transformation*. Amazon.in. <https://music.amazon.in/podcasts/f378141f-9afb-495d-83da-9b08ec5047ac/episodes/106de3df-f030-482c-a9de-5738f9356b39/the-morning-brief-coca-cola%E2%80%99s-mega-marketing-transformation>
- [23] Coca-Cola India. (2023, novembro 10). X. https://twitter.com/CocaCola_Ind/status/1722969986921013355
- [24] Coca-Cola. (2023). *Create Real Magic*. <https://www.createrealmagic.com/>
- [25] Google. (2023, December). *Gemini - Google DeepMind*. Deepmind.google. <https://deepmind.google/technologies/gemini/#introduction>
- [26] Google. (2023, dezembro 6). *Introducing Gemini: our largest and most capable AI model*. Google. <https://blog.google/technology/ai/google-gemini-ai/#sundar-note>
- [27] Hughes, A. (2023, December 12). *Phi-2: The surprising power of small language models*. Microsoft Research. <https://www.microsoft.com/en-us/research/blog/phi-2-the-surprising-power-of-small-language-models/>