

## **Inteligência artificial generativa: questões jurídicas e éticas em torno do ChatGPT**

### **Generative artificial intelligence: legal and ethical issues around ChatGPT**

Regina Rossetti\*

Kethly Garcia\*\*

#### **RESUMO**

O objetivo deste artigo é discutir a Inteligência Artificial (IA) generativa a partir de quatro pontos principais: o uso da Inteligência Artificial generativa na redução de esforços humanos na execução de determinadas tarefas; o uso de conteúdos inéditos pela Inteligência Artificial generativa e direito autorais sobre conteúdo; a verdade como um desafio frente as novas tecnologias; questões éticas relacionadas ao uso dessas tecnologias de Inteligência Artificial generativas. A metodologia envolve pesquisa bibliográfica e documental. Os resultados apontam que, embora as IAs auxiliem nas atividades humanas, são inúmeros os impasses e riscos da utilização da Inteligência Artificial generativas sem regulação jurídica e princípios éticos norteadores.

**Palavras-chave:** Inteligência Artificial generativa; ética, direito e tecnologia; ChatGPT.

#### **ABSTRACT**

The purpose of this article is to discuss generative Artificial Intelligence (AI) from four main points: the use of generative Artificial Intelligence in reducing human efforts in the execution of certain tasks; the use of unpublished content by generative Artificial Intelligence and copyright on content; the truth as a challenge in the face of new technologies; ethical issues related to the use of these generative Artificial Intelligence technologies. The methodology involves bibliographical and documentary research. The results point to the impasses and risks of using generative Artificial Intelligence without legal regulation and guiding ethical principles.

**Keywords:** Generative Artificial Intelligence; ethics, law and technology; ChatGPT.

---

Artigo submetido em 10 de junho de 2023 e aprovado em 13 de novembro de 2023.

\* Pós-Doutorado e Doutorado em Filosofia pela Universidade de São Paulo, Mestrado e Graduação em Filosofia pela Pontifícia Universidade Católica de São Paulo, Especialista em Direito Digital pela Escola Brasileira de Direito. Atualmente é docente e gestora do Programa de Pós-Graduação em Comunicação da Universidade Municipal de São Caetano do Sul - USCS. Tem experiência na área de Comunicação com ênfase em Ética, Epistemologia e Filosofia da Comunicação. Atua na investigação interdisciplinar entre direito, comunicação e filosofia contemporânea, pesquisando os seguintes temas: Comunicação e Inovação, Esfera Pública, Comunicação de Interesse Público, Ética e Tecnologia, Direito e Inteligência Artificial. Email: [regina.rossetti@online.uscs.edu.br](mailto:regina.rossetti@online.uscs.edu.br).

\*\* Mestranda pelo Programa de Pós-graduação em Comunicação da USCS. Email: [kethly.garcia@online.uscs.edu.br](mailto:kethly.garcia@online.uscs.edu.br).

## 1 INTRODUÇÃO

Ao permitir seu acesso ao público em geral, o ChatGPT causou um grande alvoroço entre professores, acadêmicos e profissionais de diversas áreas. Em poucas semanas, esse *chatbot* ganhou destaque na mídia mundial pela infinita gama de possibilidades que oferece. A produção de conteúdo é realizada de forma rápida e estruturada e a cada solicitação do usuário, novos resultados são gerados. Por meio do ChatGPT é possível, por exemplo, organizar apresentações, fazer *brainstormings*, elaborar textos de pesquisa e traduções.

O *chatbot* pode ser classificado como uma Inteligência Artificial generativa, que por definição, é uma tecnologia que utiliza categorias de algoritmos de aprendizagem (*machine learning*) autônoma para criar conteúdo originais (WE Fórum, 2023). O ChatGPT utiliza a LLMs (*large language models*), baseada em redes neurais e, mais especificamente, artificiais (ANNs) (International Business Machines-IBM, 2023) treinadas com uma grande quantidade de dados e que se baseiam no aprendizado não supervisionado (sem categorias pré-definidas direcionadas ao modelo para aprendizagem) (WEI, et al., 2022). O aprendizado profundo ou *deep learning* é uma subcategoria das redes neurais. As redes neurais possuem até três camadas e o *deep learning* mais de três camadas. As lógicas de camadas e nós de entradas e saídas e camadas ocultas e são semelhantes aos neurônios humanos (IBM, 2023).

O ChatGPT é uma criação da empresa OpenAI, empresa fundada em São Francisco, Califórnia, que tem como missão garantir que a humanidade se beneficie com a Inteligência Artificial. Atualmente a empresa disponibiliza uma versão premium paga do *chatbot*, O ChatGPT *plus*. Em 15 de março de 2023 foi lançado o GPT-4, uma espécie de atualização de modelo usado para o ChatGPT, que agora ao invés de processar apenas textos, processa imagens e textos como *inputs* e conta com uma maior precisão de atuação. Além dele, a empresa também conta com o Dall-e2, um sistema gerador de imagens e artes exclusivas e Whisper, um sistema automático de reconhecimento que faz a conversão de áudio em texto e a tradução para diversos idiomas (Open AI, s/d).

Logo após a explosão de notícias sobre o ChatGPT, outras IAs surgiram ou tiveram visibilidade, como por exemplo, a Taskade, um assistente de produtividade; a Durable, que permite a criação de sites; a Rationale, que auxilia a tomar decisões; a MidJourney, que transforma textos em produções artísticas dentre muitas outras de empresas diversas que vem ganhando destaque.

Neste artigo, são analisadas as discussões emergentes sobre o tema, a partir de quatro pontos principais: o uso da Inteligência Artificial generativa na redução de esforços humanos na execução de determinadas tarefas; a verdade como um desafio para o futuro e impasses; o uso de conteúdos inéditos gerado pela Inteligência Artificial generativa e direito autorais sobre conteúdo; questões éticas relacionadas ao uso das tecnologias de Inteligência Artificial generativas.

Qualquer previsão sobre o tema no presente momento tem caráter fortemente especulativo e provisório, uma vez que as indagações se referem a um determinado tempo na história e conforme as evoluções tecnológicas e sociais acontecem, novas possibilidades de julgamento podem surgir. Previsões mais assertivas são feitas com base no que é conhecido,

porém, quando se analisa algum fenômeno sem precedentes muito claros, a complexidade de uma análise mais assertiva aumenta (Zuboff, 2019).

## 2 USOS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA NA REDUÇÃO DOS ESFORÇOS HUMANOS

A Inteligência Artificial não é uma novidade, surgida na década de 1950, no intuito de emular a inteligência humana, ela teve que aguardar o momento certo para se desenvolver. Isso porque, a IA necessita de uma quantidade enorme de dados digitais disponíveis para poder processar esses dados e isso somente começa a acontecer no final da década de 1990 e início dos anos 2000 com a digitalização da vida e o uso massivo da Internet comercial que dá origem à big data.

A IA é idealizada para otimizar processos que podem ou poderiam ser feitos pela inteligência humana. Enquanto os humanos demorariam horas, dias ou meses para executar determinadas tarefas e fazer determinados raciocínios, a Inteligência Artificial é capaz de concretizar essas ações em poucos minutos. Atualmente, a IA recorre à aprendizagem autônoma, a chamada *machine learning*, percorrendo caminhos que muitas vezes fogem ao controle de seus próprios programadores e que podem superar processos cognitivos humanos em relação ao tempo em que são executados ou em relação a complexidade. As IAs são múltiplas, e fala-se no plural, pois aqui se trata não do conceito, mas de cada IA como única e com identidade própria, uma vez que pertencem às fontes privadas e se desenvolvem de forma exclusiva a depender de sua base de dados e suas correlações com novos dados. Com o advento das diversas IAs, preditivas, analíticas e generativas, especula-se sobre se existe uma tendência de que os seres humanos utilizem menos o raciocínio lógico e que ocorram mudanças na construção cognitiva dos seres. Sendo assim, o cérebro humano deixaria de executar determinadas tarefas, que seriam executadas pelas IAs e dessa forma, teria foco em outras. Essa mudança pode ser analisada de diversas perspectivas.

Pode-se inferir, por exemplo, que essa facilidade trará mudanças nas relações de trabalho, uma vez que as técnicas facilitarão processos que podem diminuir a carga horária despendida. Um dos questionamentos já conhecido sobre o tema é: as máquinas substituirão os humanos? Não se tem certeza do status que as máquinas alcançarão a longo prazo, mas a curto prazo, nos dias atuais, se poderia mudar essa indagação para: alguns trabalhadores serão substituídos por humanos que projetam, programam e alimentam as máquinas? Muito provavelmente, haverá grande destaque para esses profissionais e suas respectivas profissões no mercado de trabalho. Talvez as mãos de obras especializadas não sejam mais tão necessárias ou talvez possam vir a ser para alimentar as IAs especializadas.

Todas essas são implicações a serem consideradas. Além da reflexão da Inteligência Artificial como possível substituta, deve-se refletir sobre ela também a partir do ponto em que atua de maneira simbiótica com humanos nas relações de trabalho. Essa automatização e trabalho colaborativo exige que, inclusive, gestores, sindicatos, empregados e demais envolvidos repensem as relações de trabalho e direitos dos trabalhadores (Silva, 2022).

Esse esforço reduzido ou modificado também ocorrerá em tarefas diárias. Outro destaque na discussão em relação ao ChatGPT foi a possível redução de esforços nas produções relacionadas ao ensino e aprendizagem. Um documento recente da Unesco analisa

os usos do ChatGPT no ensino e mostra as possibilidades de uso do recurso em alguns rols. A ferramenta pode ser usada para consultas, debates, colaboração em grupo, tutoria e exploração de dados. Essas possibilidades aplicam-se não só aos alunos, mas também aos docentes. No entanto, o documento orienta para o uso da ferramenta apenas se o usuário tiver disponibilidade de cruzar e validar informações, além de se responsabilizar pelo uso delas. O documento ainda menciona o fato de que o ChatGPT já foi usado para gerar artigos completos para revistas e que esse fato abre um debate ético importante. Poderia um artefato ser considerado um criador de conhecimento? A Unesco afirma não existir um consenso, pois alguns editores acreditam que o ChatGPT não pode ser considerado um autor, uma vez que não pode ser responsabilizar pelos conteúdos produzidos, já outros aceitam que o ChatGPT pode desempenhar o papel de autor (Unesco, 2023).

### **3 FAKE NEWS PROPULSIONDAS POR INTELIGENCIA ARTIFICIAL GENERATIVAS**

Para Arendt (2005), no mundo contemporâneo antes da digitalização da vida, as distinções entre verdade e mentira já tendiam a ser mais difíceis de serem realizadas, considerando que as novas técnicas de comunicação de massa levam a uma tendência de manipulação, uma vez que atuavam massivamente. Essa manipulação da verdade pode ser feita de diversas formas. Ao contar uma história, fatos podem ser manipulados para atender à determinados interesses. Atualmente, nas eleições políticas de diversos países, não é incomum o uso de *fake news*, cuidadosamente projetadas e arquitetadas, às vezes contando com grandes equipes para a produção de conteúdo falsos para atingir determinados fins. Santaella (2019), aborda o tema e discute uma era da pós-verdade, em que ela já não é mais considerada como um elemento importante politicamente, mas que foi subjugado a um papel secundário na sociedade.

Para além das *fake news* criadas manualmente, com as inteligências artificiais generativas se pode elevar essa discussão a outro patamar. Agora, notícias, imagens e áudios falsos também são gerados por *bots* em grande escala e exemplos não faltam.

Em março de 2023, o artista Boris Eldagsen venceu o Sony World Photography Awards e ao ser anunciado como vencedor, o artista anunciou em seu site pessoal que recusaria o prêmio pois sua arte havia sido criada através de Inteligência Artificial. Além disso, afirmou que sua participação tinha sido única e exclusivamente com o intuito de testar se os avaliadores estavam preparados para diferenciar imagens geradas por IA e conforme o resultado final, não houve diferenciação (BBC, 2023).

A revista alemã Die Aktuelle, há poucas semanas, divulgou uma entrevista com Michael Schumacher, que supostamente seria a primeira desde o acidente que ocorreu em 2013 e que causou uma lesão cerebral no piloto. A entrevista continha declarações de Michael Schumacher discutindo temas pessoais e supostamente inéditos. A editora chefe da revista foi a responsável pela criação de conteúdo e ao final do texto, informava que a entrevista havia sido criada com Inteligência Artificial. Embora as fontes da criação tenham sido informadas, esse tipo referência pode facilmente passar despercebida por muitos dos leitores. A família

tenta manter a privacidade do caso da imagem de Schumacher desde o acidente e notificou que vai mover um processo contra os responsáveis (CNN,2023).

Fatos como esses suscitam questionamentos em relação a como a verdade deve ser tratada. Nos dois casos citados, os devidos créditos foram atribuídos aos geradores de conteúdo, no entanto, essa não é nem será uma regra clara para todos os usuários de IA. Conforme Goldstein *et al.* (2023) para atores com foco em intenções e interesses individuais atuando em detrimento a princípios coletivos, os modelos de linguagem generativa podem também atuar como recursos afim de persuadir e influenciar a opinião pública. As tecnologias têm potencial para atuar nas operações de influência, que degradam a confiança da sociedade inclusive em formas autênticas de manifestação. Para mitigar esses problemas, o ideal é uma abordagem que trabalhe em cooperação com mídias sociais, agências governamentais e desenvolvedores de IA, além da participação da sociedade civil, para que modelos sejam compreendidos e identificáveis.

No Brasil, o PL da Fake News ou PL 2630/2020 (em tramitação), é um exemplo de lei que surge para aplicar os princípios de liberdade, transparência e responsabilidade na Internet. O projeto em seu texto original, prevê o combate a desinformação, desencorajando o uso de contas inautênticas nos provedores de aplicação sediadas inclusive no exterior, em que considera desinformação os conteúdos sem veracidade, manipulados, com potencial a causar danos. Conforme o PL 2630/2020, ficam vedadas as contas inautênticas, disseminadores artificiais e imagens manipuladas que imitam a realidade. O projeto prevê como responsabilidade dos provedores publicizarem postagem, contas e disseminadores artificiais removidos e suspensos; transparência em relação às interações com conteúdo verificados como desinformação (visualizações, comportamento, alcance), além de remuneração por conteúdo jornalístico artístico ofertado nas plataformas. No Capítulo V do texto original do projeto são previstas as sanções para o descumprimento das cláusulas em forma de advertência e multa, cuja graduação é analisada com base na gravidade, reincidência e capacidade econômica dos infratores. Foram apresentadas 152 emendas e 86 requerimentos (Brasil, 2023). A inspiração do o PL da Fake News vem de uma Lei da União Europeia, a leis e serviços digitais (DAS).

Em contrapartida, algumas *Bigh Techs*, como Meta e Google, afirmam que as remunerações obrigatórias prejudicam a oferta de gratuidades para os usuários finais e que com as ameaças de multas e pressões, discursos legítimos poderiam ser excluídos das plataformas gerando censura (BBC, 2023). O Google, em publicação em seu blog no Brasil, afirma que o PL acaba protegendo o usuário final, que produz a desinformação, além de dar poder para que om órgão governamental decida o que será visto ou não, fato que prejudica empresas e anunciantes e quem nem todas as questões previstas são possíveis de aplicação, sendo o diálogo necessário para criação de leis exequíveis de fato (Google, 2023).

O tema de regulação levanta pontos divergentes no debate feito pela sociedade civil. De um lado, a regulação é alegada como sendo uma espécie de censura, principalmente, pelos provedores e, por outro lado, como o único meio possível, pelo menos no momento, para garantir um funcionamento coeso e harmônico na relação de empresas e interesses privados autônomos e seres humanos.

O ciberespaço é um espaço difícil de controlar. Por falta de conhecimento técnico, muitas vezes acredita-se que as coisas impostas são como devem ser e não se consegue enxergar mudanças possíveis ou exequíveis. Quais protocolos, regras e códigos fazem com que esse espaço não possa ser regulado? E eles não poderiam ser diferentes? É necessário, apenas, compreender e prever mudanças na arquitetura para que essas regulações sejam possíveis e também aplicáveis (Lessig, 2006).

É verdade sim, que brechas legislativas podem ser usadas para atingir fins particulares, não é incomum que instituições dos mais diversos tipos monopolizem possibilidades que beneficiem sua existência e imagem (Paganotti, 2016). No entanto, esses fatos estão ligados à ideologia por trás das redes, tecnologias e sociedade de forma geral. O real ponto de discussão não é simplesmente ou tão somente a tecnologia em si, mas sim a ideologia e a lógica por trás dos usos das tecnologias, que as direcionam para determinados fins. Essas mesmas ideologias podem ocorrer na elaboração e aplicação de documentos regulatórios. Isso não significa, de forma alguma, que se deve deixá-los de lado, mas sim trabalhar em diálogos para construir regulações plausíveis, aplicáveis e éticas, que visem o bem social e bem comum com a pretensão de construir possibilidades de melhores vivências em sociedade.

#### **4 REGULAÇÃO JURÍDICA DOS CONTEÚDOS GERADOS PELA INTELIGENCIA ARTIFICIAL GENERATIVA**

A maioria das IAs que geram conteúdos no formato de texto, imagem e som, possui a mesma característica: a cada solicitação feita, uma resposta diferente é gerada, ainda que em todas às vezes tenha sido feita a mesma pergunta. Os conteúdos gerados são individualizados e a organização dos conteúdos disponibilizados pela ferramenta não podem, muitas vezes, ser encontrados exatamente iguais na Internet, ainda que as fontes sejam citadas. Sendo assim, é possível imaginar que muitos usuários podem aproveitar os materiais sem os devidos créditos, sendo isso feito de maneira intencional ou não. A OpenAI, por exemplo, ao falar sobre regras para uso da marca no marketing e comunicações, oferece em seu site diretrizes para a atribuição de conteúdo. A empresa não obriga, até o momento, a atribuição de textos ou imagens geradas por seus modelos, mas caso o usuário opte, é possível atribuir conforme os exemplos: “escrito com ChatGPT, legenda escrita com ChatGPT, criado com DALL·E, imagem criada com DALL·E” (Open AI, 2023).

Depois de seu lançamento, o ChatGPT já aparece na Amazon como coautor de centenas de livros, segundo um levantamento feito pela *Reuters*. Entretanto, do ponto de vista jurídico, considera-se como autoral apenas o conteúdo produzido pelo autor original da obra e o ChatGPT não cria conteúdo, ele apenas compila e reorganiza conforme sua lógica programada. Para a legislação brasileira que regula os Direitos Autorais (Lei 9.610/98) o autor é a pessoa física criadora da obra, seja ela obra literária, artística ou científica e essas obras intelectuais são protegidas exatamente porque são criações do espírito, mesmo que expressas por qualquer meio. Nesse sentido, não se pode atribuir a uma IA, mesmo generativa, que ela seja autor de uma obra porque ela não possui um espírito criador, sequer uma consciência. Portanto, uma IA não pode ser autor nos termos da Legislação brasileira vigente. Algumas revistas, como *A Civil Procedure Review* (CPR), uma revista eletrônica e de acesso gratuito,

vetaram a atribuição de autoria a IA em textos acadêmicos. O ChatGPT como ferramenta pode até auxiliar os autores humanos, mas os textos gerados devem ser supervisionados pelos autores porque esses sistemas “alucinam”, ou seja, cometem erros e trazem informações equivocadas que devem ser revisadas pelos autores.

No entanto, a própria empresa OpenAI abre mão dos direitos gerados, quando não obriga os usuários a citar conteúdos extraídos de seus recursos. Os códigos da OpenAI podem ser privados, no entanto, os conteúdos gerados podem ser utilizados livremente por todos. O código fonte do recurso já foi aberto, mas pouco tempo depois voltou a ser fechado considerando os grandes potenciais econômicos do recurso e a concorrência que certamente enfrentaria. Observando essa questão, já há estudos que trabalham modelos para identificar textos produzidos por humanos e pelos modelos GPT, como por exemplo o GPT-Sentinel, que funciona com base em métodos estatísticos, classificação zero-shot (um LLM treinado para identificar semelhantes) e um modelo de ajuste fino com base em um modelo de linguagem existente ajustado para identificação e detecção do tipo de conteúdo (Chent, et al., 2023). Modelos como esse podem se desenvolver e auxiliar nessa nova demanda de identificação de autoria, quando for necessária.

A polêmica surge porque não há consenso sobre se um texto gerado por essa tecnologia pode ser considerado uma produção nova ou meramente derivada de um trabalho já existente, que fora anteriormente produzido por outros autores. Nesse sentido, quem seria o verdadeiro proprietário do conteúdo gerado pela IA? O usuário da ferramenta, o desenvolvedor que treinou o modelo ou os criadores que produziram os exemplos?

Após a comercialização do ChatGPT, o Parlamento Europeu lançou as bases para uma regulamentação da Inteligência Artificial, com objetivos de regular o uso de fontes e dados que alimentam os modelos de IA generativas e impedir a violação de direitos de proteção de dados e privacidade dos cidadãos. As *Big Techs* devem ser transparentes com a indicação das fontes dos textos, imagens e sons utilizados na geração das criações das IAs generativas, garantindo a autenticidade dos créditos da autoria original dos dados. Outro aspecto da Lei de Inteligência Artificial da UE é a indicação das fontes de dados utilizadas para o treinamento dos sistemas de IA generativas com vistas a proteger os direitos autorais e a propriedade intelectual.

No Brasil, tramita no Congresso Nacional a Lei de Inteligência Artificial, que segue uma linha semelhante a Lei de IA da UE. Mas, assim como a UE, depara-se com as dificuldades que países que não produzem sistemas de IAs, mas apenas consomem essa tecnologia, enfrentam para regular os sistemas de IA que são internacionais. A abrangência legislativa fica circunscrita ao território nacional, dificultando o alcance as *Big Techs* americanas e chinesas, que operam com regras e princípios próprios de seus Estados.

## 5 ÉTICA E INTELIGENCIA ARTIFICIAL GENERATIVA

Problemas de ordem ética estão presentes nas tecnologias digitais, IA e algoritmos e podem vir a crescer conforme esses recursos se desenvolvem. No site do Google, ao tratar da IA generativa Bard, a empresa cita princípios com os quais se preocupa em relação à aplicativos de Inteligência Artificial e seu futuro. Espera-se que sejam benéficos à sociedade,

evitem criar ou reforçar preconceitos e injustiças, sejam seguros e responsáveis com design de privacidade, mantenham altos padrões de excelência científica e que seus usos sempre atendam a esses princípios (Google IA, s/d).

No entanto, sabe-se que as *Big Techs* oferecem contratos e condições *on-line* que não permitem diálogos com o usuário final, os seus termos de uso são conhecidos como *click-wrap* (clicar-embrulhar), pois os usuários se enrolam nesses contratos de adesão. Os documentos e condições para usos de tecnologia costumam ser complexos extensos e nem sempre facilmente compreensíveis (Zuboff, 2019).

Ao acessar as políticas de privacidade da OpenAI, a empresa afirma que pode coletar informações associadas a conta como nome, informações do cartão de pagamento, uploads, feedbacks, conteúdos de mensagem feitos com a empresa, logs, endereço, IP, conteúdos que visualiza, informações dos dispositivos e outros. Essas informações podem ser usadas para melhorar, administrar e fornecer serviços. A empresa ainda cita que, ocasionalmente, pode analisar comportamentos e características dos usuários e compartilhar informações agregadas com terceiros. As informações fornecidas a terceiros, sem aviso prévio, podem ocorrer em algumas circunstâncias: para atender determinadas necessidades de provedores de serviços e fornecedores, caso seja necessário a transferência de negócios para outro provedor, para cumprir exigências legais e no compartilhamento com afiliadas. No tópico “seus direitos”, informa que dependendo da localização, os usuários podem ter determinados direitos em relação aos dados pessoais, como acessar as informações, excluir, corrigir, retirar o consentimento para processamento das informações (openai, 2023). Essas informações referem-se às políticas de privacidade analisadas em 18 de maio de 2023.

Ferreira, Pinheiro e Marques (2021), em um estudo sistemático de 11 artigos do Portal Capes apresentam que os termos e políticas das empresas geralmente apresentam uma linguagem complexa. Por esse motivo, nem sempre são lidos por seus usuários finais. Yamauchi, Maciel e Pereira (2018), aplicaram um questionário em 180 pessoas, e obtiveram resultados de que apenas 4% leem as cláusulas, 44% as vezes leem e 52% não leem. No entanto, ao analisar brevemente as políticas de privacidade da OpenAI, por exemplo, se vê a importância de que seus usuários estejam cientes de como seus dados podem ser usados e seus direitos em relação a eles.

Garantir que a comunicação dos usuários e a IA se pautem em princípios éticos, deve ser uma preocupação frente aos novos desafios impostos pela tecnologia. Os interesses e direitos coletivos devem ser considerados nesse processo, indo muito além dos interesses individuais e comerciais. Nesse sentido, a Unesco publicou um guia explicativo sobre o ChatGPT, cita algumas implicações éticas e entre elas, está a falta de regulação e possíveis problemas relacionados a proteção de dados. O guia explicativo cita também que o ChatGPT não se rege por princípios éticos, sem capacidade de fazer as distinções de verdadeiro e falso (UNESCO, 2023). A Itália, por exemplo, bloqueou o recurso devido a preocupações relacionadas à privacidade de seus usuários o que também é problemático, porque privou os italianos de utilizarem o ChatGPT.

Uma ética deontológica baseada em princípios é a mais adequada para a Inteligência Artificial. Os princípios éticos podem ser definidos como orientações abstratas que direcionam ações que considerem interesses compartilhados em relação à comunidade

(Comparato, 2006). Uma IA construída respeitando alguns princípios éticos, como não discriminação, responsabilidade e transparência, é mais segura e coloca o ser humano como fim e não como meio.

Para além da necessária regulação jurídica das IAs, a ética é uma forte aliada na prevenção de injustiças causadas pela tecnologia. Conforme Gomes (2006), com base na teoria moral de Kohlberg, uma sociedade bem desenvolvida passa a depender menos de aparatos legais e pode ser guiada por princípios de justiça. Por meio desse nível de desenvolvimento é possível julgar a sociedade e seus aparatos. Embora a sociedade conte com a legislação, muitas vezes esta possui aspectos e sanções generalistas e ocasionalmente pode ser interpretada afim de privilegiar interesses individuais. A orientação por princípios éticos indica que a sociedade se encontra no nível pós-convencional, que seria o mais elevado da teoria.

Muitas empresas já se preocupam em garantir que a IA seja guiada e pautada em princípios éticos. Corrêa et al (2022), em um estudo sistemático, analisou 200 documentos que se relacionam a governança na Inteligência Artificial e destacou dezessete princípios elementares que constam nesses documentos. Entre eles, estão a prestação de contas/responsabilidade; beneficência/não maleficência, direitos da criança e do adolescente; dignidade/direitos humanos; diversidade/inclusão/pluralismo/ acessibilidade; liberdade/autonomia/valores democráticos, soberania tecnológica; formação/educação humana; centramento/alinhamento humano; propriedade intelectual; justiça/Equidade/Igualdade/não discriminação; direitos trabalhista; código aberto/concorrência leal/cooperação; privacidade; confiabilidade/segurança/confiabilidade; sustentabilidade; transparência/ explicabilidade/ auditoria e veracidade.

Nas implicações éticas do uso da IA, parte da responsabilidade está com o usuário e mais do nunca, o desenvolvimento de uma consciência ética, que trabalhe com aspectos e preocupações coletivas e princípios como a responsabilidade são necessários.

Na ética clássica, as preocupações humanas estão direcionadas as relações de humanos com humanos. A ética tradicional é considerada antropocêntrica, entretanto, nos tempos atuais, diferentemente do que estava presente nos conceitos da ética clássica, a preocupação ética alcança actantes não humanos, como as IAs. A técnica deixou de ser apenas um recurso da humanidade para satisfazer e auxiliar em necessidades diversas e passa a estar atrelada a uma cobrança por superações e evoluções cada vez mais significativa com ações se perpetuam no tempo e tornam os humanos atuais responsáveis, não somente por seus atos e de seus artefatos, mas também pelas gerações futuras (Jonas, 2006). Sendo assim, se pode esperar progressos cada mais significativos, que demandarão análises e resoluções éticas complexas.

Sistemas de IA não possuem *status* moral, segundo Bostrom e Yudowski (2014). Seres humanos os manipulam e controlam e suas restrições morais estão ligadas à a sua interação com outros seres, como por exemplo, outros humanos. Para os autores, dois critérios estão ligados ao *status* moral: a sensiência, que é a capacidade de *qualia* ou sentimento, e a sapiência, conceito descrito e atrelado à uma consciência e inteligência mais elevada, que levariam a IA a ser autoconsciente ou um ser racional responsável e até o presente momento essas características não podem ser atribuídas as IAs generativas. As IAs generativas se alimentam e são treinadas por dados produzidos por seres humanos. Esses dados carregam os

valores e julgamentos humanos sobre o mundo. Portanto, as IAs generativas reproduzem os valores humanos, elas não possuem uma moral própria. Nesse sentido, a ética continua sendo humana, mas a sua abrangência ampliou-se de forma exponencial junto com o desenvolvimento das IAs. O ser humano deve ser responsável por si, mas também pelos resultados de suas criações no mundo tecnológico.

## 6 CONSIDERAÇÕES FINAIS

As tecnologias de Inteligência Artificial generativas estão se desenvolvendo velozmente e de forma promissora e ainda pouco se sabe sobre seus efeitos e *status* futuro. As previsões são complexas, uma vez que sua exponencialidade depende de questões sociais e políticas. Refletir sobre as questões éticas e jurídicas que envolvem o tema é fundamental para tentar compreender essas tecnologias emergentes e mitigar seus riscos. Uma vez que as tecnologias influenciam o agir humano e atuam em sua relação simbiótica com ele, considerar a ética na análise desse panorama é mais do que nunca um fator necessário. Assim como pensar a regulação é importante, a aplicação de princípios éticos é fundamental, porquanto eleva a sociedade a um alto nível de uma avaliação pautada em justiça em relação aos fenômenos que a rodeiam.

Momentos de escolhas éticas requerem não somente a reflexão, mas a construção de sistemas que garantam a restauração da integridade da sociedade. Trata-se da instauração de um conjunto ordenado de normas e práticas que visam garantir a integridade dos comportamentos dos diferentes agentes sociais. No setor tecnológico se refere a um conjunto de mecanismos que visam garantir que as ações dos membros integrantes dos sistemas midiáticos (humanos e não-humanos) estejam em conformidade com metas, missão, valores e papel social dentro de uma perspectiva de constante revisão e melhoramento dos resultados com vistas ao aperfeiçoamento das práticas éticas. Esses mecanismos criados podem incluir: comitês de ética nos setores de tecnologia, códigos de conduta ética, regulação ética, legislação, institucionalização de controle interno, certificações de *compliance*, além da efetiva contribuição para a existência de um mercado autorregulado e responsável. Além disso tudo, as IAs generativas exigem um ambiente operacional adequado para apoiar seu desenvolvimento o que inclui um setor de tecnologia robusto, infraestrutura de dados adequada e segura, visão estratégica e atenção à governança e ética.

## REFERÊNCIAS

ARENDDT, Hannah. **Entre o passado e o futuro**. São Paulo: Perspectiva, 2022.

BARD EXPERIMENT. **Bard Faq**. Disponível em: <https://bard.google.com/faq?hl=en>. Acesso em 17 de mai. 2023

BBC News Brasil. **A imagem feita por Inteligência Artificial que enganou jurados de um grande prêmio de fotografia**. Disponível em: <https://www.bbc.com/portuguese/articles/cqe5edn6997o>. Acesso em 17 de mai. 2023

BOSTROM, Nick.; YUDKOWSKY Eliezer. A ética da Inteligência Artificial. **Fundamento – Rev. de Pesquisa em Filosofia**, 1(3): 200-226, 2011.

BRASIL. Senado Federal. Projeto de Lei N° 2630, DE 2020. Institui a Lei Brasileira de Liberdade, Responsabilidade e Transparência na Internet. Brasília: Senado Federal, 2013. Disponível em: [https://legis.senado.leg.br/sdleg-getter/documento?dm=8110634&ts=1648639813988&disposition=inline&\\_gl=1\\*mq9hv9\\*\\_ga\\*MTA0ODE1ODQ1NS4xNjg0MzIzNjM3\\*\\_ga\\_CW3ZH25XMK\\*MTY4NDMyMzYzNy4xLjEuMTY4NDMyMzc1MS4wLjAuMA](https://legis.senado.leg.br/sdleg-getter/documento?dm=8110634&ts=1648639813988&disposition=inline&_gl=1*mq9hv9*_ga*MTA0ODE1ODQ1NS4xNjg0MzIzNjM3*_ga_CW3ZH25XMK*MTY4NDMyMzYzNy4xLjEuMTY4NDMyMzc1MS4wLjAuMA). Acesso em 17 de mai. 2023.

CHEN, Yutian; KANG, Hao et. al. **GPT-Sentinel: Distinguishing Human and ChatGPT Generated Content**. Disponível em: <https://arxiv.org/abs/2305.07969>. Acesso em 17 de mai. 2023.

COMPARATO, Fabio Konder. **Ética: direito, moral e religião no mundo moderno**. São Paulo: Companhia das Letras, 2006.

GOLDSTEIN, et al. **Generative Language Models and Automated Influence Operations: Emerging Threats and Potential Mitigations**. Disponível em: <https://arxiv.org/abs/2301.04246>. Acesso em 17 de mai. 2023

GOMES, Pedro Gilberto. **A Filosofia e a ética da comunicação na mediação da sociedade**. São Leopoldo: UNISINOS, 2006.

GOOGLE AI. **Nossos Princípios**. Disponível em: <https://ai.google/responsibility/principles/>. Acesso em 17 de mai. 2023

JONAS, Hans. **O Princípio Responsabilidade: ensaio de uma ética para a civilização tecnológica**. Rio de Janeiro: Editora PUC- RIO, 2006.

LACERDA, Marcelo. Blog do Google Brasil. **PL das Fake News pode aumentar a confusão sobre o que é verdade ou mentira**. Disponível em: <https://blog.google/intl/pt-br/novidades/iniciativas/pl2630-2/>. Acesso em 17 de mai. 2023

LESSIG, Lawrence. **Code: And Other Laws of Cyberspace, Version 2.0**. Basic Books, 2006.

MNCWABE, Sammy; SCHMIDT, Nadine. CNN BRASIL. **Família de Michael Schumacher planeja ação legal por entrevista falsa de IA**. Disponível em: <https://www.cnnbrasil.com.br/esportes/familia-de-michael-schumacher-planeja-acao-legal-por-entrevista-falsa-de-ia/>. Acesso em 17 de mai. 2023.

OPENAI. **About**. Disponível em: <https://openai.com/about>. Acesso em 17 de mai. 2023

OPENAI. **Brand**. Disponível em: <https://openai.com/brand>. Acesso em 17 de mai. 2023

OPENAI. **Privacy Policy**. Disponível em: <https://openai.com/policies/privacy-policy>. Acesso em 17 de mai. 2023

PAGANOTTI, Ivan. **A discussão da visibilidade e a revisão da censura na esfera pública**. USP, 2006. Disponível <https://sdpscp.fflch.usp.br/sites/sdpscp.fflch.usp.br/files/inline-files/76-237-1-PB.pdf>. Acesso em 17 de mai. 2023

SABZALIEVA, Emma; VALENTINI, Arianna. **ChatGPT and Artificial Intelligence in higher education**. Unesco, 2023. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000385146>. Acesso em 17 de mai. 2023

SCHEREIBER, Mariana. BBC News Brasil. **5 pontos polêmicos do PL das Fake News**. Disponível em: <https://www.bbc.com/portuguese/articles/cyeyxje7r9go>. Acesso em 17 de mai. 2023.

ROUTLEY, Nick. **What is generative AI? An AI explains**. World Economic Forum. 2023. Disponível em: <https://www.weforum.org/agenda/2023/02/generative-ai-explain-algorithms-work/>. Acesso em 17 de mai. 2023.

STOKEL-WALKER, Chris. ChatGPT listed as author on research papers: many scientists disapprove. **Nature**. 613(7945): 620-621. doi: 10.1038/d41586-023-00107-z, 2023.

SILVA, Ulisses Juliano da. Realidades adaptadas: perspectiva dialética entre o homem e a Inteligência Artificial nas relações de trabalho. **Revista da Faculdade de Direito de Uberlândia**, 1(50): 207-228, 2022.

SANTAELLA, Lucia. **A Pós-Verdade é Verdadeira ou Falsa?** Barueri: Estação das Letras e Cores, 2019.

WEI, Jason et al. **Emergent abilities of large language models**, 2022. Disponível em: <https://arxiv.org/abs/2206.07682>. Acesso em 17 de mai. 2023

YAMAUCHI, Eduardo Akimitsu; MACIEL, Cristiano; PEREIRA, Vinícius Carvalho. An Analysis of Users' Preferences on Pre-Management of Digital Legacy. In: **brazilian symposium on human factor in computing systems**. Belém. 45: 1-5, 2018. Disponível em: <https://doi.org/10.1145/3274192.3274237>. Acesso em 27 jan. 2023.