

# INTELIGÊNCIA ARTIFICIAL: um risco iminente ou uma dádiva do progresso tecnológico?

## ARTIFICIAL INTELLIGENCE: an imminent risk or a gift of technological progress?

JÚLIO CÉSAR DA SILVA <sup>1</sup>

Graduando em história pela faculdade Unyleya

FERNANDA CORREA GUIMARAES <sup>2</sup>

Coordenadora do Curso de História da Faculdade Unyleya.

Email: fernanda.correia@unyleya.edu.br

**Resumo:** Este estudo visa investigar se os avanços recentes em inteligência artificial, como o ChatGPT, representam uma ameaça à vida humana, conforme sugerido pela carta do *Future Of Life Institute*. Para isso, foram analisados os aspectos históricos da IA, os possíveis riscos dessa tecnologia emergente e os protocolos de segurança existentes. Como base teórica, contamos com a contribuição de especialistas renomados, incluindo o filósofo e historiador israelense Yuval Noah Harari e os brasileiros Muriel-Torrado (professor doutor da Universidade Federal de Santa Catarina) e Ronaldo Lemos (diretor do Instituto de Tecnologia e Sociedade do Rio). Utilizamos o método de revisão bibliográfica-documental e a história do tempo presente. Através do estudo, observamos as implicações éticas e econômicas que essa tecnologia pode gerar quando monopolizada e não regulada. Também identificamos que já existem protocolos estabelecidos, necessitando apenas de atualização, processo que já está em andamento.

**Palavras-chave:** ChatGPT. Ética. História. Inteligência Artificial. Protocolos de Segurança.

**Abstract:** This work's primary objective is to investigate whether the latest advances in artificial intelligence, such as ChatGPT, genuinely pose a threat to human life, as suggested by the recent letter from the Future of Life Institute. To this end, we examined the historical aspects of AI, the potential risks associated with this emerging technology, and the existing security protocols. As a theoretical foundation, we drew on the insights of notable experts in the field, including the Israeli philosopher and historian Yuval Noah Harari, and the Brazilians Muriel-Torrado (Professor at the Federal University of Santa Catarina) and Ronaldo Lemos (Director of the Institute of Technology and Society of Rio). Our approach was guided by the bibliographic-documentary review method and contemporary history. Through this study, we observed the ethical and economic ramifications that this technology could entail if monopolized and unregulated. Additionally, we noted that existing protocols merely require updates, a process that is already underway.

**Keywords:** ChatGPT. Ethic. History. Artificial intelligence. Security Protocols.

Da SILVA, Julio Cesar; GUIMARAES, Fernanda Correa. INTELIGÊNCIA ARTIFICIAL: um risco iminente ou uma dádiva do progresso tecnológico? *Educação Sem Distância*, Rio de Janeiro, v.5, n.1, jun. 2025.

## 1 Introdução

A inteligência artificial tem se tornado uma realidade cada vez mais perceptível a muitas pessoas na sociedade contemporânea (Sichman, 2021). Seja pelo uso de aplicativos, programas, softwares cada vez mais sofisticados e acessíveis ou pelas últimas pesquisas divulgadas nos meios de comunicação, alertando sobre riscos iminentes à humanidade. O que é certo é que essa tecnologia vem se consolidando nas últimas décadas, alimentando o imaginário das pessoas, inclusive na ficção científica (Barbosa, 2020). Sobre esse imaginário, vale adiantar que ele é mais antigo do que se imagina.

Partindo disso, estudaram-se os possíveis riscos (elencados pela recente carta do Future of Life Institute) que as novas tecnologias de inteligência artificial, como o ChatGPT, podem trazer à humanidade. Tomamos como base esse documento e a posição de figuras emblemáticas da atualidade, que divergem sobre o assunto, como o historiador Yuval Noah Harari. Fizemos também um mergulho na história para ver o que os pesquisadores da Dartmouth College, na conferência de 1957, imaginavam quando criaram o termo inteligência artificial, e os pontos de maior relevância dessa tecnologia que foram se desenrolando nas décadas seguintes. Nesse aspecto da história, fomos mais além: estudamos alguns elementos do imaginário antigo sobre inteligência não humana.

### 1.1 Justificativa

Vivemos em um mundo imerso em tecnologia. Nunca houve tantos avanços em tão curto espaço de tempo como nas últimas décadas. A internet e os aplicativos têm revolucionado a vida moderna. Com isso, uma inovação que tem chamado bastante atenção é a inteligência artificial, seja pelos potenciais inimagináveis de aprimoramento da vida ou pelos riscos éticos, ainda incalculáveis, para a humanidade. Certamente há algo de muito inovador, mas também de muito assustador nisso tudo.

Nessa perspectiva, a presente investigação tem como objetivo estudar melhor esse cenário, que ora se mostra promissor, ora destrutivo. Examinar as implicações éticas dessa tecnologia, alertadas pelo documento já mencionado, é, sem sombra de dúvidas, fundamental, e motivo pelo qual se emprega, neste artigo, os esforços de análise.

### 1.2 Objetivos

O principal objetivo deste estudo é investigar se os avanços recentes em inteligência artificial, como o ChatGPT, representam uma ameaça à vida humana, conforme indicado pela carta do Future of Life Institute. Para tanto, este trabalho se dividirá em três partes menores. Primeiramente, será estudado o imaginário antigo sobre inteligência não humana, com foco em um levantamento breve dos pontos relevantes na pesquisa sobre inteligência artificial, desde a criação do termo na década de 1950 pelos pesquisadores da Dartmouth

---

College até os dias atuais, destacando a conferência de Nova York em 2017. Em segundo lugar, será analisada a carta mencionada, na qual especialistas pedem a paralisação das pesquisas em IA, e serão discutidas as controvérsias geradas, tanto pelos estudiosos favoráveis quanto pelos contrários. Por fim, o terceiro objetivo, igualmente relevante, é examinar os protocolos de segurança recentes e a responsabilidade ética sobre o tema.

### 1.3 Organização do Trabalho

Na segunda seção, será apresentada a fundamentação teórica do nosso trabalho, com contribuições de pesquisadores que estudam o tema da inteligência artificial há anos e oferecem reflexões relevantes sobre os avanços recentes e as polêmicas envolvidas. Na terceira parte, detalharemos a metodologia e a historiografia utilizadas no estudo. Na seção seguinte, serão enunciadas algumas considerações, que esperamos não sejam finais, mas que apontem para outros possíveis caminhos de pesquisa.

## 2. Fundamentação Teórica

Antes de tudo, é essencial destacar que as concepções sobre inteligência artificial, ao contrário do que se imagina, não são tão recentes. Desde a antiguidade, há um imaginário alimentado sobre esse tema. A ideia de um ser não humano, mas com características humanas, inclusive dotado de inteligência, esteve presente em muitas mitologias (Nunes, 2021, p. 10). Nesta perspectiva, “o mito seria, de algum modo, o modelo matricial de toda narrativa, estruturado pelos esquemas e arquétipos fundamentais da psique dos *sapiens sapiens*, a nossa” (Durand, 1996 apud Nunes, 2021, p. 11). Cientes desses elementos psicológicos que circundam a mente humana primitiva, não podemos negar “a hipótese de uma origem mitológica para a inteligência artificial, e que a partir da compreensão dos mitos possamos traçar possíveis origens simbólicas de tais ideias” (Ibidem, 2021, p.11). Na Grécia antiga, filósofos já se interessavam por essa temática. É certo que não existiam os sistemas complexos da IA como hoje, mas a ideia de uma mente não humana autônoma, que pudesse se autogerir, já estava presente nas imaginações criativas desses pensadores (Marar, 2018 apud Prado, 2018).

De certo modo, esse aspecto embrionário da inteligência artificial foi se redimensionando ao longo dos séculos, tomando a ciência renascentista de Leonardo da Vinci, no século XVI, com um protótipo *sophia* (Unicap, 2020), e à literatura, com obras como *Pinóquio* de Carlo Collodi e *Frankenstein* de Mary Shelley (ambas do século XVIII) como disseminadoras dessas ideias (Baldessin, 2006, p. 12). Fosse na oficina do grande gênio italiano ou na oficina do personagem Gepeto ou, até mesmo, no castelo enigmático do cientista Victor Frankenstein, imaginação e realidade se confundiam, inteligência artificial já se projetava de um potencial criativo e imaginário para uma realidade futura possível.

Como se pode ver, já havia um germinar dessa tecnologia na história humana, mas sua consolidação só ocorreria no século XX, mais precisamente na década de 50 nos Estados

---

Unidos. Vale destacar que é nesse momento que o termo inteligência artificial é criado propriamente (Sichman, 2021). E podemos entendê-lo como um “conjunto de técnicas para a construção de máquinas inteligentes, capazes de resolver problemas que requerem inteligência humana” (Nilson, 2009 apud Barbosa, 2020, p. 92). A partir daí, teremos eventos marcantes que materializam, digamos assim, o desenvolvimento desta, que vai se sofisticando cada vez mais pelas décadas seguintes até os dias atuais, assumindo uma dimensão colossal, se assim podemos dizer. É sobre esse recorte temporal que iremos nos debruçar agora.

Para início, começemos pela conferência na universidade de Dartmouth, no verão de 1956, com os pesquisadores John MacCarthy, Marvin Minsky, Alan Newell e Herbert Simon que desbravaram a área da computação com a IA chamando atenção, na época, da mídia televisiva, da comunidade científica e até mesmo do público leigo em geral, suscitando altas expectativas com essa tecnologia. Criou-se um verdadeiro alvoroço com as possibilidades de mudanças que a IA poderia trazer para a vida humana. Mas isso não durou muito, pois os poucos recursos na área, nos anos seguintes, frustraram os possíveis avanços. Os anos de 1975 a 1980 e 1987 a 1993 são conhecidos como o inverno da IA, dada a frustração nas pesquisas (Sichman, 2021). Esses anos são percebidos como os de muitas oscilações, mesmo uma tecnologia que fascinava o interesse de muitos, inclusive explorada gigantescamente pelo cinema, não encontrava terreno fértil; o paradigma era outro, o tempo (com suas demandas técnicas, científicas, políticas, morais e culturais) era bem diferente do nosso atual. Em outras palavras, o que hoje se torna viável economicamente e culturalmente, nesse passado recente não era.

As primeiras décadas não foram fáceis para pesquisa com IA. Herbert Simon que, em 1958, profetizava a vitória certa de uma máquina no xadrez com qualquer humano, viu em 1965 uma criança de 10 anos vencê-la facilmente. No mesmo período, o senado americano emite um relatório com sérias críticas a essa tecnologia (Unesco, 2018). Vemos sinais de um progresso lento e nenhum pouco linear. Acumulando oscilações e muito descrédito, só quando as pesquisas ganham novas reformulações é que o jogo vira. Isso ocorre a partir da década de 80, mas com uma solidez maior nos anos 90.

Nessa época, com as pesquisas se voltando para os mecanismos de entendimento, ou seja, para a psicologia da memória, a inteligência artificial vai se dimensionando, ampliando cada vez mais suas aplicações, inclusive em diagnósticos médicos. Perto dos anos 2000, a interface homem-máquina começa a assimilar sentimentos e emoções. Observamos um grau de sofisticação sem precedentes da computação afetiva, segundo o professor da universidade de Sorbonne, em Paris, Jean Gabriel Ganascia (Idem, 2018). Os chatbots, agentes conversacionais, são um exemplo disso.

De 2010 à data presente, aprendendo com os erros do passado e potencializando os acertos dos novos rumos que tomou, a IA tem se apropriado de técnicas de aprendizagem profundas (big data), tomando as redes neurais como base. Sem sombra de dúvidas, um avanço gigantesco e incomum.

Frente ao novo panorama de evolução nas pesquisas em IA, as implicações éticas também se dimensionam. Afinal, a linha que separa a inteligência artificial da humana

---

parece estar se tornando cada vez mais tênue. Preocupados com essa nova realidade, filósofos da universidade de Nova York, David Chalmers e Ned Block, realizaram uma conferência no ano de 2017 para tratar dos princípios éticos e dos valores humanos a serem introduzidos nos sistemas inteligentes. Neste evento foram abordados quatro conceitos essenciais ao tema: “Machine Morality”, “Machine Ethics”, “Artificial Morality”, “Friendly IA” (Barbosa, 2018, p. 96).

Esses autores alertaram a sociedade sobre a aprendizagem das máquinas estar negligenciando os valores éticos, podendo “apresentar respostas insatisfatórias e inadequadas quanto a esse aspecto, ou seja, capazes de ofender e agredir moralmente, sem falar no risco imposto em relação ao trabalho (ameaça de desemprego em massa)”, ressalta Barbosa (2018). Problemáticas de como garantir o controle humano sobre esses sistemas e de como adicionar aos sistemas artificiais, valores e princípios da nossa espécie, foram a tônica da conferência. São questões complexas e desafiadoras da sociedade da informação que especialistas da IA ainda em 2023 se debruçam, principalmente agora com a carta da Future Of Life Institute (2023) que pede a paralisação das pesquisas em IA, dado os últimos progressos dessa tecnologia. É sobre isso que nosso estudo pretende avançar nas próximas linhas.

Há alguns meses, uma notícia ganhou as manchetes de todos os veículos de comunicação. Era a carta assinada por vários especialistas em IA e algumas celebridades como Bill Gates e Elon Musk pedindo a paralisação urgente de pesquisas pelo mundo sobre IA, mais precisamente sobre uma nova cadeia evolutiva. O documento mencionava a falta de políticas regulatórias sobre essas novas tecnologias, que estavam alcançando um nível de inteligência altíssimo e, por isso, perigoso. Essa manifestação causou um alarde mundial.

Por semanas não se via outra coisa nos noticiários da TV, da rádio e das mídias digitais se não o grande acordo de impedimento das pesquisas em IA. No DW (2023) do dia 30 de março sai o artigo “Musk e especialistas pedem pausa em inteligência artificial”; no mesmo dia, na BBC News (2023) sai “Inteligência artificial: o alerta de mil especialistas sobre ‘risco para humanidade’”; no dia 31 de março, saiu na Folha de São Paulo (2023) “Elon Musk erra ao pedir pausa em IA, dizem especialistas”; no G1 da Globo (2023), 1 de abril, o artigo destaca “Inteligência artificial: entenda porque cientistas estão preocupados com avanço rápido da tecnologia”. Não demorou para autoridades nas áreas políticas, econômicas e acadêmicas se posicionarem. Sobre esta última, trataremos da perspectiva antagônica de pesquisadores logo mais. Antes, nos atentemos na análise de alguns pontos desta carta.

Os autores, no início da carta, chamam atenção para a mudança profunda que a IA pode causar na história da vida na Terra. Dizem que essa tecnologia deve ser planejada com todo cuidado; algo que, segundo eles, não está acontecendo nos laboratórios. Citam ainda que não há entendimento, prevenção ou controle pelos seus criadores. Alertam sobre a circulação de informações falsas, problematizam se automatizar todos os trabalhos não seria um erro e se a humanidade não está perdendo os rumos da própria civilização para essa inteligência. Ressaltam que não há confiança sobre os efeitos serem positivos e os riscos gerenciáveis. No meio do documento, pedem uma pausa de seis meses no desenvolvimento dessa IA avançada e o empenho, nesse período, de produzirem protocolos de segurança com

---

auditoria rigorosa e supervisão de especialistas internacionais independentes. Não para por aí. A carta ainda pede “financiamento público robusto para pesquisa técnica de segurança de IA” (Future of life institute, 2023). Ela finaliza lembrando as benesses da IA para a humanidade, citando o seu grande auxílio à ciência para enfrentar a pandemia. Vejamos o trecho que menciona isso: “a humanidade pode desfrutar de um futuro florescente com a gripe aviária, tendo conseguido criar sistemas poderosos de IA, podemos agora desfrutar de um ‘verão de IA’ no qual colhemos as recompensas” (Idem, 2023).

O professor Enrique Muriel-Torrado (2023) da universidade federal de Santa Catarina, especialista na área, fala que há tempos que estudiosos alertam o mundo sobre os perigos com essa tecnologia, e que muito lhe admira de só agora lançarem uma carta do tipo. Na visão dele, a preocupação maior deveria ser com o monopólio da inteligência artificial, “que pode concentrar em poucas mãos bilhões de benefícios e controle sobre inúmeros produtos e serviços” (Muriel-Torrado, 2023). O estudioso ainda ressalta a insensatez dos autores desse documento, bilionários da tecnologia, em pedir “financiamento público robusto para pesquisa técnica em segurança de IA” (Idem, 2023). Como se sabe, não é de hoje que grandes corporações se valham de dinheiro público, com narrativas de que é para interesse coletivo, quando, na verdade, não o é. Os aspectos mercadológicos desses grandes empresários vêm sempre em primeiro lugar. Parece, inclusive, que esse seja um grande ponto de controvérsia desse documento. Muriel-Torrado ainda ressalta o conflito de interesses entre assinantes da carta, como a META e o Google, “atuais perdedores da corrida pela IA”, e a OpenAI, umas das desenvolvedoras dessa nova tecnologia (Idem, 2023).

Um dos signatários da carta é o filósofo e historiador israelense, também especialista no assunto, Yuval Noah Harari. Em uma entrevista concedida a um jornal inglês, ele menciona, de forma catastrófica, os riscos que essa nova tecnologia pode trazer à humanidade. O estudioso trás com apreensão as ações do software ChatGPT, “um modelo de inteligência artificial que pode interagir com humanos redigindo textos criativos, o que despertou fascínio e preocupação global” (Unisinos, 2023). Citou também os programas de inteligência artificial que produzem vídeos que simulam traços, características, imagens e sons, com uma perfeição inigualável, de pessoas reais. Para o pensador, nessas novas configurações, fica difícil distinguir o que é falso do que é verdadeiro. Segundo Harari, “precisamos entender que a inteligência artificial é a primeira tecnologia na história que pode tomar decisões por si só” (apud Unisinos, 2023). Em resumo, parece que o que nos cabe agora é enfrentar uma crise civilizatória sem precedentes.

Harari também relata, na entrevista, que o problema não é apenas cultural ou econômico, mas também político. Ele alerta que, em mãos erradas, essa tecnologia pode causar sérios danos à humanidade. Comparando regimes totalitários do passado, o autor afirma que “um novo regime no século XXI terá ferramentas muito mais poderosas. Sendo assim, as consequências podem ser muito mais desastrosas. Isso é algo que não sei se a humanidade pode sobreviver” (Idem, 2023). Harari destaca outros perigos, como uma massa gigantesca de pessoas desempregadas. O pesquisador vê as implicações psicológicas, econômicas, culturais, políticas e principalmente éticas desse novo paradigma. Para frear esse impulso que, na visão dele, é destrutivo, é necessário criar regulação e protocolos confiáveis de

---

segurança. Sem isso, a humanidade não caminharia para outro lugar que não para o seu próprio fim.

Pesquisadores brasileiros questionam os propósitos do documento. Um deles é o renomado cientista Ronaldo Lemos, diretor do Instituto de Tecnologia e Sociedade do Rio, que em uma coluna de jornal disse que “a carta de Musk mais celebra o suposto poder da tecnologia do que pede sua restrição” (Lemos, 2023). Segundo ele, tudo não passaria de uma peça publicitária, até mesmo porque o tempo sugerido seria ínfimo comparado às possíveis demandas pela regulação. O especialista menciona que houve momentos na história da tecnologia em que funcionalidades não eram bem aceitas inicialmente, mas logo depois foram incorporadas e hoje fazem parte do comportamento social. Lemos é muito cético quanto à aplicabilidade desse dispositivo. O cientista diz que “o foco no debate ético acaba ofuscando o problema real da IA, que na verdade é econômico” (Idem, 2023). Para tanto, ele cita as contribuições do professor Richard Freeman na elaboração de três leis da economia dos robôs:

Lei nº1: a inteligência artificial cria substitutos robóticos ao trabalho humano (maior elasticidade de substituição); lei nº 2: o custo dessa tecnologia é decrescente (como mostra o ChatGPT); lei nº 3: a questão passa a ser quem serão os donos dos robôs que vão nos substituir. Se você é dono dos robôs, isso vai melhorar sua renda e sua vida. Mas se você não for dono, prepare-se para o pior (Freeman apud Lemos, 2023).

Como o pesquisador sugere, esse documento precisaria fomentar uma outra discussão, ou seja, trazer à tona as implicações econômicas no mercado de trabalho e na manutenção das desigualdades que já são latentes, mas podem ficar ainda piores. A história é testemunha de que os benefícios dos avanços tecnológicos, ao longo dos séculos, terminam sendo monopolizados por grupos bem afortunados economicamente. E não tocar nesse assunto nos parece improfícuo, para não dizer maldoso.

Outros brasileiros veem com descrédito essa carta. O ex-vice-presidente da Google e também engenheiro da companhia, Luiz André Barroso (apud Amancio, 2023), vê boas intenções no documento, mas acha um tanto problemático. Ele e outros colegas que discursaram sobre o futuro da IA na Brazil Conference, evento organizado pelas universidades de Harvard e MIT (Massachusetts Institute of Technology), se mostraram incrédulos com os efeitos da carta. Um desses é Marcus Fontoura, diretor executivo de tecnologia da Stone, que afirmou que “existe governança sobre os dados, regulamentação, a gente está muito longe de um cenário de ficção científica em que robôs vão dominar o mundo” (Idem, 2023). Ele ainda acha “a carta muito infeliz e triste”. Mencionando o ChatGPT e reforçando o coro, digamos assim, o outro colega, Flávio Calmon, que também é professor de engenharia elétrica de Harvard, disse que “certamente a carta não terá nenhum efeito no desenvolvimento desses algoritmos” (Idem, 2023).

Tendo visto as contribuições de especialistas sobre a carta, inclusive as controvérsias geradas, partamos agora para o estudo dos protocolos de segurança das pesquisas em IA. É sabido por todos que esse evento tem mobilizado governos pelo mundo todo para formulação de políticas de regulação na área. É nítido que a preocupação não é exclusividade da ciência. Em maio deste ano, a União Europeia e os Estados Unidos

---

publicaram em conjunto um código de conduta sobre a inteligência artificial. O código, na realidade, ainda não se encontrava finalizado, mas seu rascunho já dava indícios de uma não obrigatoriedade dos grupos que lidam com essa tecnologia em segui-lo, uma vez que o secretário de Estado americano Antony Blinken disse que o “objetivo é estabelecer códigos de condutas voluntários” (Carta Capital, 2023).

No Brasil, há projetos de lei tramitando no congresso sobre a regulação. Estima-se que até o fim de 2023 todos eles serão discutidos. O senado federal, um ano antes, já havia instalado uma comissão de juristas para elaborar o marco regulatório da inteligência artificial. Associado à UE, aos Estados Unidos e ao Canadá, nosso país se mostra engajado na resolução desse gargalo mundial. Contudo, sabe-se que o trabalho é árduo: envolve várias dimensões da vida humana. “A tarefa de estabelecer regras para controlar o uso de programas desse tipo é complexa. A IA incorporou-se à ciência, ao sistema financeiro, à segurança, à saúde, à propaganda e ao entretenimento”. Tudo isso, “na maioria das vezes sem que o usuário perceba” (Schmidt, 2023). Depois de meses de trabalho, a comissão de juristas trouxe um relatório com uma série de recomendações, entre elas: questões como segurança, transparência e discriminação se sobressaíram, dada a realidade nacional. Em entrevista ao Jornal da Globo, em março deste ano, a advogada e relatora Laura Schertel disse: “a gente tem uma série de riscos que a comissão identificou (...). Nós temos sistemas que são pouco transparentes (...). A gente tem uma sociedade muito desigual, e o sistema acaba reproduzindo as desigualdades” (Jornal O Globo, 2023).

Em 2021, as Nações Unidas, com sua 41ª conferência-geral, já traziam essas preocupações sobre transparência, privacidade de dados e desigualdade social frente aos dilemas contemporâneos da IA. Este evento resultou “no primeiro instrumento global de definições de padrões sobre a ética da inteligência artificial” (Unesco, 2022).

Recentemente, a ONU, por questões que afligem a humanidade já há alguns anos, como a proliferação de fake news, e também pela polêmica atual da carta que pede paralisação das pesquisas em IA, propôs um código de conduta. O secretário-geral da instituição, António Guterres, falou que as implicações da IA precisam ser “seguras, responsáveis, éticas e cumprir com obrigações dos direitos humanos” (Onu, 2023). Na visão de Guterres, essa tecnologia, se não regulada, ameaça a existência humana, seus riscos são equiparáveis a uma guerra nuclear.

Uma autora que traz ótimas contribuições para o desenvolvimento e a utilização de uma IA de forma responsável é Virginia Dignum. Para ela, há três dimensões que precisam ser consideradas na adoção de uma postura ética quanto à tecnologia. A primeira seria no processo em que o projeto dos sistemas se dá, garantindo que os pesquisadores “tenham em mente e antevejam as possíveis consequências”; a segunda consistiria no comportamento dos sistemas, verificando se no projeto as “capacidades de raciocínio ético nos agentes inteligentes” estão colocadas adequadamente; e a terceira diz respeito ao código de conduta dos projetistas e desenvolvedores, orientados, é claro, por uma regulação adequada, algo que já existe em outras profissões (Dignum, 2019 apud Sichman, 2021).

Nessa primeira instância (ética no projeto), a autora ressalta a transparência e a sistematização dos valores humanos e princípios éticos sendo incluídos por meio: primeiro,

---

de uma espécie de prestação de contas que o sistema precisaria dar à sociedade, justificando suas decisões e ações; segundo, de uma responsabilidade assumida, ou seja, o sistema precisa responder pelas suas decisões e identificar os possíveis erros; e terceiro, o sistema precisa ser transparente, ter seus mecanismos inspecionados e sujeitos à governança das instâncias reguladoras, afinal os usuários e a sociedade, de uma forma geral, querem explicações e clareza sobre os dados usados (Idem, 2021).

Quanto à ética no comportamento, sabe-se que as decisões precisam ser embasadas por valores morais. As sociedades, mesmo enfrentando questões conflituosas que põem, por vezes, os valores civilizatórios em xeque, prezam por uma base moral que seja a favor da dignidade humana. “Trata-se de embasar tais agentes autônomos com mecanismos de decisão que possam ser também baseados em sentimentos e valores morais” (Bazzan et al., 2002 apud Sichman, 2021). É o caso de uma racionalidade que não seja meramente instrumental, mas movida por princípios humanos que se julgam importantes.

Dois estudiosos no assunto, Thomas Dietterich e Eric Horvitz (2015), trazem uma observação envolvendo os riscos no uso de sistemas de IA. Eles listam cinco classes de riscos. São elas:

falhas (Bugs), qualquer software pode apresentar falhas, mas isso não o impede de ser testado e validado (com a IA pode ser feito o mesmo); segurança (Cybersecurity), os ataques cibernéticos que ocorre aos softwares são também suscetíveis aos sistemas de IA; aprendiz de feiticeiro (sorcerer’s apprentice), em sistemas que interage com humanos, a IA precisa saber raciocinar, julgar se um pedido humano é aceitável, dentro de uma normalidade ou não; autonomia compartilhada (Shared autonomy), os sistemas colaborativos criados trazem desafios sobre a fluidez de engajamento, pois se valendo de sua produção em tempo real, as implicações no seu controle precisam mudar rapidamente; impactos socioeconômicos, é preciso mensurar o quanto a IA influencia na distribuição de empregos e, amplamente, na economia com programas que potencialize os benefícios para todos (Dietterich; Horvitz, 2015 apud Sichman, 2021).

Na União Europeia, um projeto de lei, criado após o evento da carta, traz uma classificação de risco que se divide em quatro partes. Dos mais graves ao menos, temos o inaceitável (sujeito a proibição); alto; limitado, pode ter uma flexibilidade; e mínimo, sem nenhum tipo de obrigação legal adicional, a não ser a legislação que já vigora. “Os programas de carro autônomo, por exemplo, estão na categoria de alto risco” (Schmidt, 2023).

Seguindo as últimas orientações internacionais sobre o assunto, influenciado principalmente pela comunidade europeia, no Brasil têm tramitado alguns projetos de lei no congresso que buscam regulamentação e trazem uma classificação de risco. Dentre eles, o PL nº 2338/23 evoca mais atenção pela sua completude, de acordo com o jurista Antonio Carlos Morato, da faculdade de direito da USP (Idem, 2023). Esse PL menciona como:

Proibido - programas de classificação social ou capazes de manipular o comportamento de populações vulneráveis; Alto risco - Devem ser avaliados e monitorados antes e durante o seu uso - Programas de classificação automática de estudantes, candidatos a emprego, pedidos de crédito ou benefícios da previdência social, diagnósticos médicos, risco de crimes e comportamento criminal, e veículos autônomos (Schmidt, 2023).

Como se pôde observar, há uma série de medidas sendo tomadas por governos no mundo todo. Algumas discussões parecem estar sendo trilhadas já há algum tempo, outras ainda se mostram tímidas. Certo é que não dá para nos precipitarmos frente a uma situação tão crucial da vida moderna como as pesquisas em inteligência artificial. Protocolos que garantam a segurança nesses sistemas parecem ser sempre bem-vindos, o que não pode é se fazer uso de mecanismos, com aparência de legalidade, para barrar avanços ou mitigar benefícios para a grande maioria da população. Enfim, ao se tratar de responsabilidades éticas, há ainda um longo caminho a ser percorrido.

Na seção dois, será apresentada a fundamentação teórica do nosso trabalho, com pesquisadores que estudam o tema da inteligência artificial há anos e trazem contribuições significativas sobre os últimos avanços e reflexões sobre as polêmicas envolvendo o assunto. Na seção três, mostramos a metodologia e a historiografia empregadas no estudo. Já na seção posterior, serão enunciadas algumas considerações, que esperamos não serem finais, mas que apontem para outros possíveis caminhos de estudo.

### 3. Metodologia

A presente pesquisa se pautou pelo método bibliográfico-documental, valendo-se de livros, artigos científicos, jornais e da carta em questão. Estes foram tratados da seguinte forma: leituras, fichamento em editor eletrônico de textos e separados por tipo e data. No que concerne à compreensão desse método, vale ressaltar que Gil (2002, p. 45) afirma que as pesquisas bibliográfica e documental são muito semelhantes. Enquanto a primeira tem a contribuição de vários autores, a segunda se refere a materiais que ainda não foram analisados, como a carta em discussão (detectamos uma escassez de artigos científicos a respeito; o que conseguimos encontrar foram artigos de jornais e blogs, criteriosamente escolhidos, de professores doutores que lecionam em universidades, especialistas em IA).

Quanto à abordagem historiográfica, adotou-se a da história do tempo presente, pois o evento em estudo tem um marco temporal preciso: a atualidade. Esta abordagem é bem explicitada pelo estudioso François Dosse (2011, p. 6), que diz que “a história do tempo presente está na intersecção do presente e da longa duração. Esta coloca o problema de se saber como o presente é construído no tempo”. O autor, nessa mesma obra, ainda coloca a problemática dos historiadores, que se lançam a estudar esse recorte temporal, de serem frequentemente desacreditados. Ele lembra Heródoto e Tucídides como os primeiros historiadores, testemunhas oculares da história, que inseridos nos acontecimentos contemporâneos a eles, não se deixaram de valer do rigor científico, mesmo que ainda rudimentar ao que se tem hoje. Quanto a isso, pode-se dizer que “a própria definição de história do tempo presente é ser a história de um passado que não está morto, de um passado que ainda se serve da palavra e da experiência de indivíduos vivos” (Rousso, 1998, p. 63 apud Dosse, 2011, p. 16). É possível dizer também que, “por seus motivos, seus métodos, suas fontes, a história do tempo presente não diferem em nada da história do século XIX”

---

(Sirinelli, 1999). Nesse sentido, acreditamos ter guiado nossa pesquisa com o máximo de objetividade científica necessária.

No que se refere à análise dos dados, construída à medida que íamos costurando o referencial teórico, ela foi sistêmica e qualitativa, impulsionada pela seguinte problemática: das pesquisas realizadas ao longo das últimas décadas, seria esta última com a tecnologia do ChatGPT a que representaria um risco real à humanidade, como ressalta a carta assinada por vários pesquisadores e personalidades como Elon Musk e Bill Gates, pedindo a paralisação das pesquisas? Guiados por essa problemática, lançamo-nos ao trabalho de levantar as bibliografias e documentos pertinentes ao assunto e analisá-los.

#### 4. Considerações Finais

Partindo para a finalização dessas reflexões, a primeira coisa que podemos dizer é que a inteligência artificial já está inserida dentro de uma espécie de matriz histórica humana. O imaginário, estudado por nós na antiguidade e desenvolvido por técnicas ao longo dos séculos, chegando à sua materialização e complexidade nas últimas décadas, nos prova isso. Essa tecnologia se torna indispensável à vida humana. É inimaginável viver sem os benefícios trazidos por ela à vida moderna. Seja na medicina, com procedimentos cirúrgicos guiados milimetricamente pela IA; na ciência, com a sofisticação de pesquisas e desenvolvimento de vacinas cada vez mais precisas e eficazes; na indústria, com o aprimoramento da automação de suas máquinas e autonomia de seus veículos; ou até mesmo na esfera governamental, com o desenvolvimento de bancos de dados capazes de ampliar o alcance de políticas públicas a um número gigantesco de pessoas; enfim, seria ingênuo, se não maldoso, negar as benesses dessa tecnologia e a dependência que o mundo contemporâneo tem por ela. Por outro lado, como a pesquisa pôde mostrar, não se pode negar que com tantos avanços, muitas implicações éticas não tenham surgido e desafiado governos e corporações por todo o mundo. Porque isso tem ocorrido, não à toa é objeto de nosso estudo.

#### Referências

AMÂNCIO, Thiago. **Elon Musk erra em pedir pausa, dizem especialistas**. Folha de São Paulo, São Paulo. 31 mar. 2023. Disponível em: <https://www1.folha.uol.com.br/tec/2023/03/elon-musk-erra-ao-pedir-pausa-em-desenvolvimento-de-ia-dizem-especialistas.shtml>. Acesso em: 28 set. 2023.

BALDESSIN, Marcell Giglioli Stoppa. **A ficção científica como derivação da utopia: a inteligência artificial**. 2006. 152p. Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Estudos da Linguagem, Campinas, SP. Disponível em: <https://repositorio.unicamp.br/Busca/Download?codigoArquivo=491897>. Acesso em: 02 out. 2023.

---

BARBOSA, Xênia de Castro. Breve introdução à história da inteligência artificial. **Jamaxi, Revista de história** 4(1). Junho, 2020. Disponível em: <https://periodicos.ufac.br/index.php/jamaxi/article/view/4730>. Acesso em: 18 ago. 2023.

DOSSE, François. História do tempo presente e historiografia. **Revista Tempo e Argumento**. Revista do Programa de Pós-graduação em História, Florianópolis, v. 4, n. 1, p. 5-22, jan./jun. 2011. Disponível em: <http://www.revistas.udesc.br/index.php/tempo/article/view/2175180304012012005>. Acesso em: 10 abr. 2025.

GIL, Antônio Carlos. **Como Elaborar Projetos de Pesquisa**. São Paulo: Atlas, 2002.

JORNAL O GLOBO. **Juristas pedem regulamentação da inteligência artificial em relatório entregue ao Senado**. Jornal da Globo. 02 mar. 2023. Disponível em: <https://g1.globo.com/jornal-da-globo/noticia/2023/03/02/juristas-pedem-regulamentacao-da-inteligencia-artificial-em-relatorio-entregue-ao-senado.ghtml>. Acesso em: 26 set. 2023.

LEMOS, Ronaldo. **Problema da inteligência artificial será econômico, não ético**. Disponível em: <https://itsrio.org/pt/artigos/problema-da-inteligencia-artificial-sera-economico-nao-etico/>. Acesso em: 06 out. 2023.

MURIEL-TORRADO, Enrique. Carta aberta para pausar as inteligências artificiais. **Future of Life Institute**. 2023. Disponível em: <https://www.techtudo.com.br/noticias/2023/03/inteligencia-artificial-tudo-sobre-a-carta-aberta-assinada-por-elon-musk-edsoftwares.ghtml>. Acesso em: 16 ago. 2023.

NUNES, L. J. D. L. **Um estudo introdutório dos aspectos psicossociais da artificialização das inteligências**. Disponível em: [https://www.teses.usp.br/teses/disponiveis/47/47134/tde-21122021-120511/publico/nunes\\_corrigeida.pdf](https://www.teses.usp.br/teses/disponiveis/47/47134/tde-21122021-120511/publico/nunes_corrigeida.pdf). Acesso em: 17 ago. 2023.

ONU. **Código de conduta da ONU quer combater violência e desinformação na internet**. Nações Unidas. ONU News. Disponível em: <https://news.un.org/pt/story/2025/04/1847116>. Acesso em: 27 set. 2023.

PRADO, Jean. **A inteligência artificial é mais antiga do que você imagina**. Disponível em: <https://tecnoblog.net/meiobit/2023/03/a-inteligencia-artificial-e-mais-antiga-do-que-voce-imagina/>. Acesso em: 28 set. 2023.

SIRINELLI, Jean-François. **Abrir a História**. Belo Horizonte: Autêntica, 2014.

---

SCHMIDT, S. **Os desafios para regular o uso da inteligência artificial**. Disponível em: <https://revistapesquisa.fapesp.br/os-desafios-para-regulamentar-o-uso-da-inteligencia-artificial/>. Acesso em: 04 out. 2023.

SICHMAN, Jaime Simão. Inteligência artificial e sociedade: avanços e riscos. **Revista SciELO - Brasil**. Inteligência Artificial. Estud. av. 35 (101). Jan-Apr 2021. Disponível em: <https://doi.org/10.1590/s0103-4014.2021.35101.004>. Acesso em: 11 jul. 2023.

UNESCO. **Inteligência artificial**: entre o mito e a realidade. Disponível em: <https://pt.unesco.org/courier/2018-3/inteligencia-artificial-o-mito-e-realidade>. Acesso em: 27 set. 2023.

UNICAP. **Inteligência artificial**: origem, dilemas e contemporaneidades. Recife: UNICAP, 2020. Web-Jornal. Disponível em: <https://webjornalismo.unicap.br/inteligenciaartificial/historico-da-inteligencia-artificial/>. Acesso em: 22 set. 2023.

UNISINOS. Harari alerta sobre a inteligência artificial: “Não sei se os humanos podem sobreviver” [entrevista cedida ao] Instituto Humanitas Unisinos. **Revista Adital**. Abril, 2023. Disponível em: <https://www.ihu.unisinos.br/categorias/628113-harari-alerta-sobre-a-inteligencia-artificial-nao-sei-se-os-humanos-podem-sobreviver>. Acesso em: 18 ago. 2023.

AFP. **UE E EUA publicarão ‘código de conduta’ sobre Inteligência Artificial**. Carta Capital. 31 mai. 2023. Disponível em: <https://www.cartacapital.com.br/tecnologia/ue-e-eua-publicarao-codigo-de-conduta-sobre-inteligencia-artificial/>. Acesso em: 02 out. 2023.