

# Por que é imprescindível um manual ético para a Inteligência Artificial Generativa?

Por Lucia Santaella<sup>1</sup>

**Resumo:** A tendência pervasiva da inteligência artificial, que já se anunciava há alguns anos, foi se intensificando cada vez mais, especialmente depois da emergência da Inteligência Artificial Generativa (IAG), em especial o Chat GPT, que foi colocado, pela start up Open AI, financiada pela Microsoft, para o uso público a partir do final de 2022. Protocolos de acesso muito facilitados, um quase nada, a partir de simples comandos verbais, tornam disponível um sistema capaz de atender às requisições informativas de um usuário, por meio de conversações sobre os mais variados assuntos, em uma linguagem que simula a capacidade humana de falar e de escrever. Distinta da Inteligência Artificial Preditiva (IAP), cujo desenvolvimento estava retido à competência dos especialistas, a IAG vem adquirindo as características de uma companheira disponível e solícita para a realização das mais variadas tarefas. Todas as áreas de produção humana relacionadas com linguagem estão sendo abaladas. Não haveria razões para que a educação ficasse de fora. Ao contrário. Entretanto, o Chat traz consigo um detalhe de suma importância: ele pode ser usado para fraudar ou para um uso honesto. Para atender a esse dilema ético, este manual, acompanhado de um guia foi elaborado.

**Palavras-chave:** manual ético; guia; Chat GPT; Inteligência Artificial Generativa

---

<sup>1</sup> É pesquisadora IA do CNPq, professora titular na pós-graduação em Comunicação e Semiótica e em Tecnologias da Inteligência e Design Digital (PUC-SP). Doutora em Teoria Literária pela PUC-SP e Livre-docente em Ciências da Comunicação pela USP. Publicou 56 livros e organizou 33, além da publicação de quase 500 artigos no Brasil e no exterior. Recebeu os prêmios Jabuti (2002, 2009, 2011, 2014), o prêmio Sergio Motta (2005) e o prêmio Luiz Beltrão (2010). Orcid: <https://orcid.org/0000-0002-0681-6073>.

## **Why an ethical manual is essential for Generative Artificial Intelligence?**

**Abstract:** The pervasive trend of artificial intelligence, which had already been announced a few years ago, intensified more and more, especially after the emergence of Generative Artificial Intelligence (IAG), in particular Chat GPT, which was made available by the start-up Open AI, financed by Microsoft, for public use from the end of 2022 on. Very easy protocols, almost nothing, from simple verbal commands, make accessible a system capable of meeting the user's informational requests, through conversations on the most varied subjects, in a language that simulates the human ability to speak and write. Distinct from Predictive Artificial Intelligence (IAP), whose development is left to the competence of specialists, IAG has been acquiring the characteristics of an available and helpful companion for carrying out the most varied tasks. All areas of human production related to language are being shaken. There would be no reason for education to be left out. On the contrary. However, Chat brings with it an extremely important detail: it can be used for fraud or for honest endeavors. To address this ethical dilemma, this Manual, accompanied by a guide, was created.

**Keywords:** manual of ethics; guide; Chat GPT; Generative Artificial Intelligenc

A partir de dezembro de 2022, o ChatGPT caiu no mundo feito um meteoro. A metáfora não é sem razão. No decorrer desse mês e naqueles que se seguiram, foi desmedido o volume de notícias, de colunas jornalísticas, de blogs e dos primeiros textos de natureza interpretativa. Até hoje o impacto não cessou, ao contrário, as tendências avaliativas são diversificadas, tornando o consenso cada vez mais perto do impossível, particularmente porque o Chat passou de 3.5 para a sofisticação do 4, prometendo o 5 e tornando-se inclusive apenas a ponta do iceberg de uma grande quantidade de sistemas em competição, além de outros subsidiários. Em suma: não é mais o mesmo aquilo que conhecíamos como revolução digital que, de resto, sempre foi uma revolução da ordem da progressão mesclada à dirupção. Nem a inteligência artificial (IA) é a mesma que ficou mais popularmente conhecida há uns quinze anos. Ela mudou e continua mudando. Quais as razões para tudo isso? Vamos por partes.

Embora o grande tema do momento, popularmente chamado de hype, seja de fato a IA, trata-se de uma área de investigação que teve seu início explícito em meados do século passado. Todos os textos que tratam do tema da IA com alguma seriedade, recorrem, mesmo que brevemente, ao seu histórico, uma recorrência importante para se evitar a displicência com o fato de que se trata de uma área de investigação científica que se desenvolveu com todos os rigores que são próprios da ciência. A IA nasceu com os mesmos propósitos que tem até hoje: desenvolver sistemas artificiais capazes de simular propriedades e habilidades que são próprias da inteligência humana. Seu ambiente de nascimento foi muito propício.

### **O nascimento da IA simbólica**

Conforme já desenvolvido em Santaella (2004, p. 73-92) e aqui retomado, quando as ciências da computação estavam emergindo, em 1956, foi realizada em Dortmund, USA, uma conferência com duração de seis semanas, que contou com a presença dos maiores especialistas em ciência da computação, tendo como tarefa estabelecer as bases de uma ciência da mente sob o modelo do computador digital. Foi dessa ideia de que o computador poderia ser um bom modelo para entender o funcionamento do cérebro humano que brotou a inteligência artificial cuja expansão interdisciplinar deu origem àquilo que passou a ser chamado de ciências cognitivas ou ciência cognitiva.

Segundo Teixeira (1998, p. 11-12), essa conferência rendeu frutos nos laboratórios de IA que foram fundados por John McCarthy e Marvin Minsky no MIT, Massachusetts, depois em Stanford, na Califórnia e tam-

bém em Pittsburgh, na Universidade Carnegie-Mellon, este sob iniciativa de Allan Newell e Herbert Simon (1972). Esses laboratórios criaram máquinas de jogar xadrez e outras máquinas inteligentes. O mais importante, contudo, era o que estava por trás da criação dessas máquinas, a saber, a busca das condições formais da atividade cognitiva, capazes de indicar o que é comum a todos os sistemas que exibem essa atividade, quer ela apareça em animais, máquinas ou humanos. Na época, o modelo computacional foi o escolhido devido à sua habilidade para simular processos cognitivos, com a possibilidade de se modelar a mente. Durante certo tempo, o “modelo computacional da mente”, isto é, o computador como metáfora da mente, constituiu-se como o paradigma clássico unificador das ciências cognitivas.

O modelo assentava-se pelo menos sobre dois pressupostos: (1) a relativa autonomia entre o *hardware* e o *software* das máquinas utilizadas para simular a inteligência, o que permite explicar o comportamento inteligente de qualquer sistema complexo sem pressupor o tipo físico ou biológico da inteligência de seus componentes; (2) a compreensão da mente como um conjunto de representações de tipo simbólico e regidas por um conjunto de regras sintáticas. Desse modo, o pensamento seria o resultado da ordenação mecânica de uma série de representações ou símbolos, ordenação esta que não pressupõe necessariamente a existência de um cérebro. Por isso, o aparato mental concebido como dispositivo lógico pode ser descrito por meio de um conjunto de computações abstratas. Simular a inteligência não implica a construção de máquinas com *hardware* específicos, mas sim o desenvolvimento de programas computacionais operando sobre dados ou representações. Chamado de IA simbólica, esse paradigma sedimentou-se a partir de final dos anos 1960, tanto nos trabalhos do grupo liderado por Newell e Simon quanto nos de Marvin Minsky e Seymour Papert (Teixeira, 1998, p. 36, 43).

O que está aí implicada é também uma mudança no conceito de inteligência que passa a ser definida como capacidade para produzir e manipular símbolos, tendo em vista a resolução de problemas. Em 1972, Newell e Simon haviam desenvolvido o conceito de “sistema físico de símbolos” para compreender como as pessoas resolvem problemas, uma vez que elas próprias são sistemas que manipulam símbolos. Mais tarde, em 1980, Newell reafirmou os fundamentos dos sistemas simbólicos físicos de modo mais sistemático. O conceito de sistema simbólico físico foi definido como “uma classe muito grande de sistemas capazes de produzir e manipular símbolos, sendo realizáveis dentro de nosso universo físico”. A hipótese é a de que esses símbolos, que são internos ao conceito

de sistema, são, “de fato, os mesmos símbolos que nós, seres humanos, produzimos e usamos todos os dias em nossas vidas”, o que significa que “os humanos são exemplos de sistemas simbólicos físicos, e, em virtude disso, a mente se insere no universo físico” (Newell, 1980, p. 136).

Conclusão, o pensamento era, então, visto como um sistema físico de símbolos, um tipo especial de máquina Turing que pode manipular símbolos. O símbolos são padrões físicos que podem ocorrer como elementos de um outro tipo de entidade que Newell e Simon chamaram de expressão (ou uma estrutura de símbolos), composta de um número de exemplares de símbolos que estão relacionados um depois do outro, de algum modo físico. Assim, os símbolos são como as letras de um alfabeto e as expressões como palavras e sentenças (Fetzer, 1991, p. 38). Depois de descrever o funcionamento de um sistema simbólico físico exemplar e depois de definir sua natureza essencial, Newell (1980, p. 172-173), considera o computador digital como um exemplo-chave para a realização de um sistema simbólico no nosso universo físico.

A teoria da mente que está por trás do paradigma computacional é o funcionalismo que exerceu domínio quase exclusivo nas ciências cognitivas até os anos 1980. Para o funcionalismo, “os estados mentais, tais como crenças e processos mentais, por exemplo, considerar e decidir, não são senão estados físicos descritos funcionalmente. O mesmo estado físico em sistemas diferentemente organizados pode levar a estados mentais diferentes; o mesmo estado mental pode ser realizado diferentemente em sistemas físicos diferentes”. O funcionalismo está baseado, portanto, na ideia de que a essência da natureza psicológica do estado ou processo mental não está na sua realização física particular, mas sim no seu papel computacional no sistema processador de informação (Garfield, 1987, p. 313-323). Assim, um mesmo papel funcional que caracteriza um determinado estado mental pode estar presente em sistemas nervosos completamente distintos. Esse é o chamado funcionalismo turinguiano para o qual a mente atualiza uma máquina de Turing no substrato biológico do cérebro.

Esse retrospecto dos primeiros tempos da IA é importante não só como documento histórico, mas também para lembrar que a IA simbólica não morreu, mas continua em novas formas convivendo com o sucesso retumbante alcançado pela atual IA baseada em redes neurais da aprendizagem profunda. Na verdade, a IA simbólica e a IA conexionista são hoje apenas duas entre outras três, a IA evolucionista, a IA bayesiana e a IA analógica, conhecidas como as cinco tribos da IA (ver Santaella, 2023, p. 37-43). Mas o sucesso atual alcançado pela IA conexionista não surgiu de um flash momentâneo, pois suas pesquisas tiveram início nos anos 1980.

## A emergência do conexionismo

Continuando com Santaella (2004, p. 73-82), a partir dos anos 1980, uma abordagem competitiva na concepção da mente começou a adquirir força nas ciências cognitivas. Trata-se do conexionismo que, sob o influxo de uma grande renovação das pesquisas sobre redes de neurônios formais, propôs a replicação da inteligência por meio da construção de redes neurais artificiais. Usando técnicas dotadas de propriedades que podem ser interpretadas em termos cognitivos, essas redes são capazes de aprender, reconhecer formas, memorizar por associações etc.

Já no movimento cibernético dos anos 1940, os cientistas estavam divididos entre duas concepções alternativas: estudar a mente ou o cérebro. Do lado do cérebro, estavam McCulloch e Pitts. Para eles, não se tratava de estudar o substrato físico do cérebro, mas as relações entre a lógica e o cérebro. Por isso mesmo, é nas pesquisas desenvolvidas por esses cientistas que o conexionismo encontra a sua paternidade, apesar de que essa origem não seja hoje devidamente lembrada. A falta de lembrança se deve, muito provavelmente, à grande repercussão que o modelo computacional da mente obteve nas ciências cognitivas durante décadas, o que deve ter provocado o esquecimento do bifurcamento das raízes que presidiu às suas origens.

Foram os trabalhos de G. E. Hinton, J. R. Anderson, D. E. Rumelhart e J. L. McClelland que, nos anos 1980, voltaram a chamar atenção para o estudo das redes neurais artificiais. Enquanto o modelo computacional da mente, também chamado de cognitivismo, estuda os processos mentais como computações abstratas, independentemente de suas formas específicas de concretização, o conexionismo pretende simular o cérebro como meio para emular a atividade mental. Os conexionistas tomam o cérebro humano como um dispositivo computacional em paralelo que opera com milhões de unidades similares aos neurônios.

Se computadores e cérebros têm como função principal processar informações, então redes neurais artificiais podem ser construídas para simular esse processo. As redes neurais constituem uma intrincada rede de conexões entre suas unidades que são dispostas em camadas hierarquicamente organizadas. Conectadas umas com as outras, unidades estimuladas via *inputs* externos excitam ou inibem outras unidades gerando padrões de conectividade. Diferentemente dos sistemas computacionais simbólicos, os conexionistas são sistemas dinâmicos compostos por um “conjunto de processos causais através dos quais as unidades se excitam

ou se inibem, sem empregar símbolos e tampouco regras para manipulá-los”. Pretende-se que esse conjunto de neurônios artificiais modele a cognição (Teixeira, 1998, p. 84).

Embora infinitamente menos complexo do que o cérebro, esse modelo da mente inspira-se na estrutura e modo de funcionamento do cérebro, chegando assim mais próximo da realidade biológica da mente. O que importa em sistemas desse tipo são os complexos padrões de atividade entre as múltiplas unidades que constituem uma rede. Para Teixeira (*ibid.*, p. 103), o conexionismo alinha-se com concepções filosóficas materialistas da mente, antecipando questões que viriam se tornar candentes com o desenvolvimento da neurociência cognitiva, da vida artificial e da robótica na década de 1990.

As pesquisas do conexionismo apontavam para futuros promissores, mas nos anos 1980 e 90 sua hora não havia ainda soado. Foi preciso esperar pela era do crescimento desmedido dos dados e pelo aumento da potência computacional para que elas pudessem passar a funcionar em simulações relativamente mais similares às operações do cérebro. Tendo como paradigma mestre o aprendizado, considerado como função magna do cérebro, o conexionismo encontrou o seu triunfo no aprendizado de máquina e no aprendizado profundo.

### **O triunfo do aprendizado de máquina e do aprendizado profundo**

Embora a aprendizagem de máquina e a sua sub-área de aprendizagem profunda sejam aquelas que mais recebem atenção, a IA constitui-se hoje em um campo hipercomplexo de investigações e de aplicações que apresenta como suas principais áreas, segundo Verma (2018, p. 5), os sistemas especialistas, instruções inteligentes auxiliadas por computador, processamento de linguagem natural, compreensão de fala, robótica e sistemas sensoriais, visão computacional e reconhecimento de cena, computação neural. Tudo isso, em rápido crescimento, está provocando um imenso impacto em vários campos da vida. As diversas técnicas aplicadas em IA são: “Rede Neural, Lógica Fuzzy, Computação Evolutiva, Instruções Auxiliadas por Computador e Inteligência Artificial Híbrida” (*ibid.*), entre outras.

Em uma definição abusivamente reduzida, inteligência significa “a capacidade de raciocinar, de desencadear novos pensamentos, de perceber e aprender” (*ibid.*, p. 6). Quando adicionado à palavra inteligência, o adjetivo “artificial” significa desenvolver computacionalmente máquinas capazes de funcionar à maneira da inteligência humana, tais como apren-

der (aquisição de informação e regras para usar as informações), raciocinar (usar as regras para chegar a conclusões aproximadas ou definitivas), autocorrigir-se e resolver problemas. Quando qualquer sistema se adapta às condições impostas por um ambiente, esse sistema é considerado inteligente. O que se extrai dessas considerações é o fato de que não é a condição psicologia humana e suas complexidades que a inteligência artificial busca simular, mas muito mais a capacidade de raciocinar e planejar para atingir determinados objetivos.

São muitas as tecnologias de IA. Conforme (Rouse, 2022), entre elas as mais conhecidas são:

- Automação como, por exemplo, a automação robótica de processos.
- Aprendizado de máquina, a ciência que faz um computador funcionar sem programação.
- Aprendizado profundo, subconjunto do aprendizado de máquina que, em termos muito simples, pode ser pensado como a automação da análise preditiva. Existem três tipos principais de algoritmos de aprendizado de máquina: 1. aprendizado supervisionado, no qual conjuntos de dados são rotulados para que padrões possam ser detectados e usados para rotular novos conjuntos de dados; 2. aprendizagem não supervisionada, na qual os conjuntos de dados não são rotulados e são classificados de acordo com semelhanças ou diferenças; e 3. aprendizagem por reforço, em que os conjuntos de dados não são rotulados, mas, após realizar uma ação ou várias ações, o sistema de IA recebe feedback.
- Visão computacional, concentrada no processamento de imagens baseado em máquina.
- Processamento de linguagem natural (PNL) é o processamento da linguagem humana por um programa de computador. Suas abordagens são baseadas no aprendizado de máquina e incluem tradução de texto, análise de sentimento e reconhecimento de fala.
- Robótica, um campo da engenharia focado no projeto e fabricação de robôs usados para realizar tarefas que são difíceis para os humanos realizarem ou executarem de forma consistente.

## Aplicações da IA

São também muitos os campos de incidência da IA. Os mais citados são:

- IA na saúde: o aprendizado de máquina está sendo usado para fazer diagnósticos melhores e mais rápidos do que os humanos e fornecer feedback médico básico.
- IA nos negócios: a automação robótica de processos está sendo aplicada a tarefas altamente repetitivas normalmente executadas por humanos. Algoritmos de aprendizado de máquina estão integrados em plataformas analíticas para descobrir informações para a melhoria dos atendimentos. Os chatbots foram incorporados aos sites para fornecer atendimento imediato aos clientes.

- IA na educação: usada para automatizar a avaliação, adaptando-se às suas necessidades e ajudando o aprendiz a trabalhar no seu próprio ritmo.
- IA nas finanças: aplicada a financiamentos por meio do recolhimentos de dados pessoais, inclusive fornecendo aconselhamento financeiro. O funcionamento geral dos sistemas bancários está impregnado de sistemas de IA.
- IA na lei: usada para automatizar análise de documentos e outras tarefas passíveis de serem programadas.
- IA na manufatura: incorporação de robôs no fluxo de trabalho humano (Rouse, 2022).

Entretanto, os setores de aplicações da IA são vastíssimos e tendem a se expandir. Um panorama das aplicações em vários setores e domínios da IA tão completo e atual quanto possível nos é fornecido por Duggal (2024, n.p.) ao qual recorreremos para complementar os quadros acima, agora com ênfase nas aplicações.

- Processamento de Linguagem Natural (PNL), usada para analisar e compreender a linguagem humana por meio de aplicativos como reconhecimento de fala, tradução automática, análise de sentimento e assistentes virtuais como Siri e Alexa.
- Análise de imagem e vídeo inclui a visão computacional para a análise e interpretação de imagens e vídeos aplicadas ao reconhecimento facial, detecção e rastreamento de objetos, moderação de conteúdo, imagens médicas e veículos autônomos.
- Sistemas de recomendação, baseados em IA são usados em comércio eletrônico, plataformas de streaming e mídias sociais para personalizar as experiências do usuário ao analisar as preferências, o comportamento e os dados históricos do usuário para sugerir produtos, filmes, músicas ou conteúdos relevantes.
- Assistentes Virtuais e Chatbots, com tecnologia de IA interagem com os usuários para lhes dar suporte e assistência personalizada ou executar tarefas.
- Jogos Algoritmos de IA são capazes de criar personagens virtuais realistas, tomando decisões inteligentes. A IA também pode otimizar gráficos de jogos, simulações físicas e testes de jogos.
- Casas Inteligentes e IoT são sistemas domésticos inteligentes para automatizar tarefas, controlar dispositivos e aprender com as preferências do usuário. A IA auxilia na funcionalidade e eficiência dos dispositivos e redes da Internet das Coisas (IoT).
- Ciber segurança diz respeito à detecção e prevenção de ameaças cibernéticas, ao analisar o tráfego de rede, identificando anomalias e prevenindo possíveis ataques.

As aplicações não param aí. Elas se expandem até se tornarem partes tão integrantes de nossa vida a ponto de se naturalizarem. Podem ser citadas como exemplo, o google maps, o waze, os filtros de realidade aumentada, Snapchat, ou “Lentes” que reconhecem características faciais, rastreiam movimentos e sobrepõem efeitos interativos nos rostos

dos usuários em tempo real. Bastante citados e discutidos são os carros autônomos e os dispositivos vestíveis rastreadores que monitoram e analisam dados de saúde. Por fim, o MuZero, desenvolvido pela DeepMind, alcançou um sucesso notável em jogos de tabuleiro complexos como xadrez, Go e shogi em um nível sobre-humano porque se auto-aprimora por meio do autojogo e do planejamento (Dugall, *ibid.*).

Todas essas aplicações são aqui citadas com o propósito de fornecer um panorama dos avanços da IA. Entretanto, existe um outro lado da IA que não pode deixar de ser lembrado. Trata-se das externalidades negativas da IA que também são muitas e preocupantes.

### As externalidades negativas da IA

Externalidade negativa é uma expressão proveniente da economia e ela significa a imposição de um custo a uma parte como efeito indireto de outra parte (Eldridge, 2024). A expressão passou a ser empregada com frequência no campo da IA para o qual o significado foi transposto com a finalidade de indicar os conflitos entre os limites da técnica e os direitos humanos (Kaufman *et al.*, 2023). Frequentemente mencionadas como externalidades negativas são, por exemplo, riscos aos direitos fundamentais, danos à democracia e ao meio-ambiente, reforço a discriminações não justificadas nas mais variadas esferas, coadjuvante em campanhas de desinformação e a intensificação do extrativismo, ou seja, da extração de recursos naturais. As externalidades negativas têm provocado uma necessária agitação em busca da regulamentação da IA que possa impedir ou minimizar riscos e efeitos colaterais. Segundo Bioni, Garrote e Guedes,

o contínuo surgimento de novas regulações direcionadas à AI, seja por meio de projetos de lei/regulamento ou de documentos internacionais de atores globais de relevância, revela a tendência global em que não se discute mais se, mas como se regular o uso desta tecnologia. Pela continuidade de produção de externalidades negativas de forma transversal, regulações setoriais não são suficientes. Isso, contudo, não afasta a necessidade de um arranjo de governança que navegue entre o geral e o específico justamente para traduzir normas de governança gerais às particularidades de um determinado contexto. Dito de outra forma, uma lei geral não exclui, mas, muito pelo contrário, abre espaço para que a regulação setorial floresça a partir de fundações comuns a diferentes setores da economia. (Bioni, Garrote e Guedes, 2023, p. 6)

O estudo realizado por Bioni *et al.* (2023, p. 611) levanta alguns tópicos capazes de auxiliar na organização do debate regulatório. São eles: (1) A regulação por meio da navegação entre o setorial e o geral. (2) A inovação responsável e resiliente socioeconomicamente. (3) Alvo regulatório plástico e uma regulação dinâmica e equilibrada (regulação assimétrica com base no risco). (4) Os vários modelos de regulação de risco. (5) Os

variados degraus da escada do risco. (6) O risco enquanto elemento dinâmico. (7) A difícil conciliação de uma regulação baseada em risco e em direitos – taxonomia de risco como um dos possíveis indicadores (proxy). (8) Avaliações de Impacto Algorítmico (AIA) públicas, inclusivas e sobre direitos sociais e não apenas individuais. (9) Uma regulação atenta aos aspectos sócio-técnicos-econômicos locais. (10) IAs Generativas (IAGs) e teste de stress das propostas de regulações de IA.

Como não poderia deixar de ser, dado o estado da arte da IA, os tópicos acima cessam na IAG, seguidos pela declaração de que, “o desafio das IAs Generativas é que, por se prestarem a diferentes finalidades (nem sempre previsíveis), tensionam o modelo regulatório baseado no risco, atualmente predominante no campo da IA, já que é inerentemente contextual” (Bioni *et al.*, p. 11). Entretanto, a hipótese que orientou o manual e o guia aqui presentes vão muito além de uma possível previsibilidade de riscos, pois a IAG não deve ser confundida com o que veio antes dela. Embora a IA precedente continue seu curso de investigações, aplicações e regulamentações, sem menosprezar a necessidade de regulamentação que também cabe à IAG, esta apresenta uma outra realidade, regida não só pelos riscos que apresenta, mas muito mais pelo uso ético que exige, cujas razões serão discutidas mais adiante, justificando a necessidade de um manual que aqui apresentamos, seguido de um guia para o uso da IAG.

## A IA preditiva e a IA generativa

Tanto a IAG apresenta uma realidade originalmente distinta que existem hoje dois títulos distintos de IA: a IA preditiva, de um lado, e IA generativa, de outro. Como funciona a IA preditiva e por que se chama preditiva? Imediatamente é possível inferir que ela é preditiva porque é capaz de prever resultados futuros. Como ela funciona para que isso seja possível?

Segundo Medha (2024), ela ingere grandes volumes de dados históricos de diferentes fontes, relevantes para o problema que lhe é colocado. Então os algoritmos de aprendizado de máquina analisam esses dados buscando tendências, padrões e relacionamentos entre variáveis.

Para isso, ela depende da modelagem estatística, ou seja, várias técnicas estatísticas e de aprendizado de máquina são usadas para, a partir dos dados, treinar modelos que sejam preditivos, ou seja, modelos que sejam treinados com o propósito de alcançar determinado resultado. Os métodos de modelagem mais utilizados são análise de regressão, árvores de decisão, redes neurais, previsão de séries temporais e modelagem de conjuntos.

Então, vem a fase da validação do modelo. Para isso, a exatidão e precisão dos modelos são não apenas rigorosamente testadas, quanto também os modelos são refinados até que o nível desejado de desempenho preditivo seja alcançado. A seguir, com os modelos razoavelmente precisos, passa-se para a simulação de cenário, quando diferentes cenários são simulados para o ajustamento dos parâmetros de entrada de modo a estimar previsões sob diversas condições. A etapa seguinte é a da implantação do modelo em ambientes de produção o que não impede que novos dados sejam continuamente inseridos nos modelos para gerar *insights* preditivos atualizados..

Por fim, vem a integração de processos dos *insights* preditivos “aos processos de negócios e fluxos de trabalho por meio de painéis, alertas APIs, etc., para permitir a tomada de decisões orientada por dados com base nas previsões do modelo”. Todo esse percurso torna a IA preditiva poderosa e valiosa para as corporações e organizações atuais, com o *surplus* de que os modelos tornam-se mais inteligentes com o tempo, à medida que processam mais informações (Medha, 2024, n.p.).

A IAG, por sua vez, diferencia-se da preditiva porque segue preceitos relativamente distintos, embora ainda se utilize de aprendizagem de máquina e redes neurais. Segundo Bharath (*apud* Lowton, 2023), a IA generativa e a IA preditiva representam paradigmas distintos no domínio da IA e do aprendizado de máquina, pois a generativa está voltada para a criação de conteúdo novo e original, como imagens, texto e outras mídias, aprendendo com os padrões de dados existentes. Por isso, pode-se dizer que auxilia em campos criativos e na resolução de novos problemas. A IA preditiva, por outro lado, usa padrões em dados históricos para prever resultados futuros ou classificar eventos futuros. Por isso, ajuda na tomada de decisões e na formulação de estratégias.

Isso não significa que sejam abordagens isoladas, pois podem se hibridizar em algumas situações. Assim, de acordo com Lawton (*ibid.*) a IAG pode ajudar a projetar recursos do produto, enquanto a IA preditiva pode prever a demanda do consumidor ou a resposta do mercado para esses recursos. Ainda também a IAG pode sintetizar dados realistas para aprimorar o conjunto de treinamento de um modelo preditivo, melhorando as capacidades preditivas. Enquanto a IA preditiva prevê eventos futuros analisando tendências históricas de dados para atribuir pesos de probabilidade aos modelos, a IA generativa cria novos dados, que podem estar na forma de texto e imagens.

Marcada a distinção entre esses dois paradigmas, o que importa ao manual que se segue é, antes de tudo, compreender qual é o modo de funcionamento da IAG que a capacita a produzir a variedade de resultados que produz, todos eles relacionados com a simulação de capacidades

semióticas humanas. A partir disso, pode-se compreender a explosão sociocultural e política provocada pela IAG e, diante disso, as razões que tornam imprescindível a existência de um manual ético para o seu uso na educação.

### **O que está por baixo do capô da IAG**

A IAG é um subconjunto do aprendizado profundo, mas de um tipo diferente, chamado de Modelo Gerativo que aprende com um conjunto subjacente de dados para gerar novos dados que imitam de perto os dados originais. Por meio do emprego de aprendizagem não supervisionada, esses modelos são usados principalmente para criar novos conteúdos, como imagens, texto ou até mesmo música, semelhantes àquilo que pode ser criado por humanos.

Os modelos generativos mais comuns são: Autoencodificadores Variacionais (VAEs), Redes Adversariais Gerativas (GANs), Máquinas Boltzmann Limitadas (RBMs) e Modelos de linguagem baseados em transformadores (*Transformers*). Os modelos geradores de textos estão baseados em grandes modelos de linguagem (*Large Language Models* – LLM) que é um tipo de aprendizagem de máquina treinado em um grande conjunto de dados de texto e que usa arquiteturas de redes neurais avançadas para gerar ou prever textos semelhantes aos humanos.

Os LLMs estão ligados ao processamento de linguagem natural (*Natural Language Processing* – NLP), que é um subconjunto da IA focado na interação entre computadores e humanos por meio da linguagem natural. Ele está habilitado a entender a língua humana e a se comunicar conosco na mesma língua (preferencialmente o inglês, vale lembrar), graças a algoritmos que ajudam os computadores a entender o contexto e o sentimento por trás das palavras e sentenças.

Por isso, os LLMs podem ser considerados como uma evolução dos modelos de processamento de linguagem natural. Embora estes últimos incluam uma ampla gama de modelos e técnicas para processar a linguagem humana, os grandes modelos se concentram na compreensão e geração de texto semelhante ao humano. Eles são especialmente projetados para prever a probabilidade de uma palavra ou frase com base nas palavras que a precedem, permitindo-lhes gerar textos coerentes e contextualmente relevantes. O processamento de linguagem natural utiliza uma ampla gama de técnicas, que vão desde métodos baseados em regras até aprendizado de máquina e abordagens de aprendizado profundo. O

LLM, por seu lado, usa principalmente técnicas de aprendizagem profunda para compreender padrões e contexto em dados de texto para prever a probabilidade da próxima palavra na sequência. Eles são um subconjunto da IAG que pode gerar muitos tipos de conteúdo como texto, imagem, vídeo, código, música etc., concentrando-se, portanto, na geração de texto.

A IAG, entretanto, é uma família mais ampla do que os sistemas voltados para a geração de texto. São conhecidos e bastante usados os sistemas de geração de imagem, como o Midjourney que usa um modelo de difusão, um tipo de modelo generativo que gradualmente adiciona ruído em uma imagem até que ela alcance o nível desejado de realismo. DALL-E usa uma rede neural treinada para produzir imagens a partir de comandos de textos. Foi anunciado com alarido um novo sistema, Sora, para geração de vídeos, muito mais potente do que eram capazes os sistemas anteriores.

Embora o ChatGPT (Transformador Gerativo Pré-treinado) tenha estado e continue estando na crista da onda desde dezembro de 2022, é bom saber que as aplicações dos LLMs não se reduzem a ele. São variados os setores e vale a pena conhecê-los. Gupta (2024) nos forcece uma lista:

- Assistentes Virtuais são os modelos LLM que analisam o comando humano e interpretam o seu significado, permitindo que os assistentes realizem ações requeridas pelos usuários.
- Tradução de idiomas são treinados em uma grande quantidade de dados de texto multilíngue, “o que lhes permite capturar distinções sutis, variações, contexto e complexidade de diferentes idiomas” (ibid.).
- Resumo: os modelos LLM podem resumir documentos ou artigos extensos, preservando as principais informações e pontos principais.
- Análise de sentimento com base em grandes quantidades de dados de texto, os LLMs conseguem compreender o contexto, as nuances e o tom da linguagem, identificando polaridades de sentimentos.
- Recomendações de conteúdo são realizadas pelas grandes plataformas a fim de fornecer aos usuários sugestões personalizadas e relevantes a eles, tendo como fonte a análise das interações dos usuários com conteúdos prévios.
- Bots de diálogo são aplicações para gerar texto coerente e contextualmente relevante semelhante ao humano, como reação a comandos dados pelos usuários. Dentre eles, o mais conhecido e usado é o ChatGPT cuja origem se reporta à introdução da arquitetura do transformador em 2017 que levou a OpenAI a lançar o GPT-1 em 2018. Vieram depois as versões GPT-2, GPT-3, GPT-3.5 e, então o GPT-4, um modelo pago, multimodal, o que significa que pode receber imagens e também texto como entrada. Já se fala em GPT 5.

### **Por que o ChatGPT 3.5 explodiu?**

É também educativa a razão pela qual o panorama histórico e situacional da IA, que foi elaborado acima, introduz este manual. É um modo de evitar uma tendência que vem tomando conta de muitos daqueles que se pronunciam sobre a IA caindo na mania do presentismo. Há poucos anos introduzi a ideia do presentismo como a reclusão no presente em si, um presente sem passado e sem futuro (Santaella, 2021, p. 121). Algum tempo depois encontrei essa mesma crítica tanto em Beck quanto em Crary. Para Beck (2018, p. 31), o mundo está sofrendo uma surpreendente metamorfose que exige a transformação do horizonte de referências e das coordenadas de ação, que são tomadas como constantes e imutáveis por posições que se mantêm aprisionadas no presentismo, ou seja, “posições que abolem o tempo para funcionar em tempo real, privilegiando o agora e nutrindo a ilusão da instantaneidade e da disponibilidade imediatas” (Crary, 2023, p. 85).

Infelizmente, na maior parte das apreciações e julgamentos socio-culturais que recebe, a IAG está sendo vítima de um presentismo agudo, como se tivesse surgido de um vácuo no tempo e no espaço. É certo que a IAG e mais particularmente o ChatGPT e congêneres produzem uma boa dose de inquietação e temores, pois com eles a IA chegou muito perto do humano, está roçando o cerne do humano devido à sua capacidade de dialogar como se fosse gente, penetrando, inclusive, em atividades concebidas como criativas em competição com aquilo que o humano considera seu tesouro e exclusividade magna. Diante disso, são até naturais as fantasias atemorizantes de que o humano está sendo usurpado de sua potencialidade mais preciosa. Contudo, só o conhecimento da história, do contexto mais amplo e, sobretudo, dos limites atuais da IAG pode evitar demonizações ou euforias despropositadas.

De outro lado, felizmente, o falatório acerca do fato de que a IA não é inteligente, baseado em noções muito antropocêntricas de inteligência, cessou até certo ponto. Ao mesmo tempo, intensificou-se o temor, por parte de alguns ex-desenvolvedores, hoje dissidentes, quanto ao desenvolvimento da IA geral ou forte, a saber, a IA que ultrapassaria o potencial humano até o ponto de levar nossa espécie a uma condição inócua ou mesmo ao extermínio. Embora não possamos descartar as incertezas e inseguranças quanto ao futuro, em nosso papel de pensadores, temos que olhar de frente para o modo como a IAG está penetrando em todas as nervuras de nossas vidas e para a maneira como crescem suas externalidades

negativas em relação aos direitos humanos, em especial, nesse caso, aos direitos autorais, perdas de empregos e outras consequências previsíveis e imprevisíveis.

Em nosso papel mais específico de educadores, papel em que especificamente este manual está contextualizado, olhar de frente significa reconhecer que a IAG está sendo explícita ou implicitamente usada pelos estudantes, sem que haja, pelo menos em nosso meio, guias claros que orientem os educadores em suas tarefas e tratamento minimamente seguro da questão.

Por que os estudantes estão inevitavelmente usando não é difícil de compreender, pois esse uso está ligado à razão primordial capaz de explicar por que a IAG explodiu com a força que demonstra, uma força que se localiza na grande e verdadeira diferença a se estabelecer entre a IA preditiva e a IAG. Estranhamente poucos têm reconhecido essa diferença, a saber: sem deixar de produzir seus efeitos em nossas vidas, cuja visibilidade mais imediata encontra-se nas semioses algorítmicamente guiadas dos nossos smartphones, a IA preditiva está reservada aos recintos dos desenvolvedores para a entrega às organizações e corporações. A IAG, por outro lado, caiu direto no colo de qualquer ser humano que disponha de um computador minimamente equipado, que saiba ler e escrever e que tenha disponibilidade para o diálogo, quando pode ver tarefas, que lhe custavam tempo, sendo realizadas rapidamente por um sistema solícito e prestativo.

A estratégia da Open AI foi magistral ao entregar, de modo gratuito e com facilidade ímpar de uso, o ChatGPT nas mãos dos usuários. Com isso, não apenas testou a receptividade dessa nova forma de IA quanto também se aproveitou dos comportamentos de uso para o aprimoramento do próprio sistema.

Depois de pouco mais de um ano o sistema ganhou aperfeiçoamento, sofisticou-se no GPT 4 e são poucos aqueles que não o estão experimentando. No campo da educação, proibir deve estar entre os piores caminhos, pois acaba por incentivar o uso oculto, cuja detecção, apesar de promessas, não é ainda segura. Ignorar significa alienar-se de um problema que exige atenção e encaminhamentos. A questão mandatória que se coloca é ética, o que envolve, antes de tudo, informar-se, conhecer, experimentar e avaliar para melhor agir. Este manual, seguido de um guia nasceu com esse propósito.

## Referências

BIONI, Bruno; GARROTE, Marina; GUEDES, Paula. *Temas centrais na regulação de IA: o local, o regional e o global na busca da interoperabilidade regulatória*. São Paulo: Associação Data Privacy Brasil de Pesquisa, 2023.

DUGGAL, Nikita. *What is artificial intelligence?* Simplilearn, 2 abr., 2024. Disponível em: <https://www.simplilearn.com/tutorials/artificial-intelligence-tutorial/what-is-artificial-intelligence>. Acesso em: 20 abr., 2024.

ELDRIDGE, Stephen. Negative externality. *Encyclopedia Britannica*, 15 mar., 2024. Disponível em: <https://www.britannica.com/topic/negative-externalitynegative-externality>. Acesso em: 17 mai., 2024.

FETZER, James. *Aspects of artificial intelligence*. Dordrecht: Kluwer, 1991.

GARFIELD, Jay L. (ed.). *Modularity in knowledge representation and natural language understanding*. Cambridge, MA: MIT Press, 1987.

GUPTA, Raja. Generative AI for beginners. Disponível em: <https://medium.com/@raja.gupta20/generative-ai-for-beginners-part-1-introduction-to-ai-eadb5a71fo7d>. Acesso em: 10 abr., 2024.

KAUFMAN, Dora; REIS, Patricia; JUNQUILHO, Tainá. Externalidades negativas da inteligência artificial: conflitos entre limites da técnica e direitos humanos. *Revista de Direitos e Garantias Fundamentais*, Sabta Lúcia. Vitória, v. 24, n. 3, p. 43-71, setembro/dezembro, 2023.

LAWTON, George. Generative AI vs. predictive AI: understanding the differences. TechTarget, 18 set. 2023. Disponível em: <https://www.techtarget.com/search/query?q=Generative-AI-vs-predictive-AI-Understandingthe-differences>. Acesso em: 2 abr., 2024.

MEDHA. Demystifying predictive AI: Definition and use. Fireflies.ai, 15 fev., 2024. Disponível em: <https://fireflies.ai/blog/predictive-ai>. Acesso em: 15 abr., 2024.

MERRITT, Rick. *What is a transformer model*. NVIDIA, 25 mar., 2022. Disponível em: <https://blogs.nvidia.com/blog/what-is-a-transformer-model/>. Acesso em 10 jun., 2023.

NEWELL, Allen. Physical symbol system. *Cognitive Science*, v. 4, p. 135-183, 1980.

NEWELL, Allen; SIMON, Herbert A. *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall, 1976.

ROUSE, Margaret. AI (Artificial Intelligence), 2022. Disponível em: <https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence>. Acesso em: 10 dez., 2022.

SANTAELLA, Lucia. *Navegar no ciberespaço*. O perfil cognitivo do leitor imersivo. São Paulo: Paulus, 2004.

SANTAELLA, Lucia. Desafios e dilemas da ética na inteligência artificial. In: GUERRA FILHO, Willis S. *et al.* (org.). *Direito e Inteligência Artificial: fundamentos*, v. 1 – Inteligência Artificial, ética e direito. Rio de Janeiro: Lumens Juris, 2021. p. 109-136.

SANTAELLA, Lucia. *A inteligência artificial é inteligente?* São Paulo: Almedina, 2023.

TEIXEIRA, João de Fernandes. *Mentes e máquinas*. Uma introdução à ciência cognitiva. Porto Alegre: Artes Médicas, 1998.

VERMA, Mudit. *International Journal of Advanced Educational Research*, Delhi, v. 3, n. 1, p. 5-10, jan., 2018.