




C A P Í T U L O 1

DA INTELIGÊNCIA ARTIFICIAL CLÁSSICA À INTELIGÊNCIA ARTIFICIAL GENERATIVA: UMA ANÁLISE EVOLUTIVA

 <https://doi.org/10.22533/at.ed.080112613011>

Márcio Mendonça

Universidade Tecnológica Federal do Paraná
PPGEM-CP - Programa de Pós-Graduação em Engenharia Mecânica CP/PG
Cornélio Procópio - PR
<http://lattes.cnpq.br/5415046018018708>

Vitor Blanc Milani

Universidade Tecnológica Federal do Paraná
Mestrando - PPGEM-CP - Programa de Pós-Graduação em Engenharia Mecânica CP/PG
Cornélio Procópio - PR
<http://lattes.cnpq.br/4504374098250296>

Fabio Rodrigo Milanez

UniSENAIPR-Campus Londrina
Londrina-PR
<http://lattes.cnpq.br/3808981195212391>

Emerson Ravazzi Pires da Silva

Universidade Tecnológica Federal do Paraná
Departamento Acadêmico de Engenharia Elétrica (DAELE)
Cornélio Procópio - PR
<http://lattes.cnpq.br/3845751794448092>

Andressa Haiduk

Dimension Engenharia
Rio Negro - PR
<http://lattes.cnpq.br/2786786167224165>

Francisco de Assis Scannavino Junior

Universidade Tecnológica Federal do Paraná Departamento Acadêmico de
Engenharia Elétrica (DAELE)
Cornélio Procópio - PR
<http://lattes.cnpq.br/4513330681918118>

Cintya Wedderhoff Machado

Universidade Tecnológica Federal do Paraná (UTFPR) – Campus Cornélio
Procópio/Londrina, Paraná – Brasil
Mestranda no Programa de Pós-Graduação em Ensino de Ciências Humanas,
Sociais e da Natureza – PPGEN
Londrina - PR
<http://lattes.cnpq.br/4604594140489347>

Vera Adriana Huang Azevedo Hypólito

Centro Estadual de Educação Tecnológica Paula Souza
Etec Jacinto Ferreira de Sá
Ourinhos – SP
<http://lattes.cnpq.br/6169590836932698>

Armando Paulo da Silva

Universidade Tecnológica Federal do Paraná
Programa de Pós-Graduação em Ensino de Ciências
Humanas, Sociais e da Natureza (PPGEN-CP/LD)
Cornélio Procópio-PR
<http://lattes.cnpq.br/6724994186659242>

Paulo Alexandre Lourenço Jesus

Afiliação acadêmica: Aluno externo no Programa de Pós-Graduação em Ensino
de Ciências Humanas, Sociais e da Natureza – PPGEN – Curso de Mestrado
Instituição: Universidade Tecnológica Federal do Paraná
(UTFPR) – Campus Cornélio Procópio/Londrina – PR
<https://lattes.cnpq.br/1346404833044235>

Adriano da Silva Moreira

Afiliação acadêmica: Mestrando no Programa de Pós-Graduação em Educação – PPEDU
Instituição: Universidade Estadual de Londrina (UEL) –
Campus Londrina – PR
<https://lattes.cnpq.br/0686904669527189>

Marcos Dantas de Oliveira

Consultoria, Projetos e Serviços em Energia Solar
Londrina – PR
<http://lattes.cnpq.br/5329306535174160>
Angelo Feracin Neto
Universidade Tecnológica Federal do Paraná
Departamento Acadêmico de Engenharia Elétrica (DAELE)
Cornélio Procópio-PR
<http://lattes.cnpq.br/0580089660443472>

Fabio Nogueira de Queiroz

Centro Paula Souza
Professor Fatec
Ourinhos – SP
<http://lattes.cnpq.br/4466493001956276>

Kazuyochi Ota Junior

Mestre - PPGEM-CP - Programa de Pós-Graduação
Multicampi em Engenharia Mecânica CP/PG
Cornélio Procópio – PR
<http://lattes.cnpq.br/3845751794448092>

Keila Felipe do Carmo Lima

Universidade Tecnológica Federal do Paraná (UTFPR) – Campus Cornélio
Procópio/Londrina, Paraná – Brasil
Mestranda no Programa de Pós-Graduação em Ensino de Ciências Humanas,
Sociais e da Natureza – PPGEN
Cornélio Procópio - PR
<http://lattes.cnpq.br/8692187746106065>

Guilherme Cyrino Geromel

Instituto Federal de Educação, Ciência e Tecnologia de São Paulo – IFSP
São Paulo - SP
<http://lattes.cnpq.br/7535398878830738>

Ana Clara Augusto Jesus

Acadêmica do curso de Pedagogia da Universidade Estadual do Norte
do Paraná (UENP) Campus de Cornélio Procópio - Paraná – Brasil
Cornélio Procópio – PR
<https://lattes.cnpq.br/5330914992553099>

Vicente de Lima Gongora

UniSENAIPR-Campus Londrina
Londrina-PR
<http://lattes.cnpq.br/6784595388183195>

RESUMO: Este artigo apresenta uma análise evolutiva da inteligência artificial (IA), examinando sua trajetória desde os paradigmas clássicos, fundamentados em sistemas simbólicos, regras lógicas e modelos determinísticos, até o advento contemporâneo das redes neurais profundas e dos modelos generativos. Inicialmente, discute-se a IA clássica sob a perspectiva histórica e conceitual, destacando suas contribuições, limitações computacionais e dependência de conhecimento explicitamente formalizado. Em seguida, analisa-se a transição para abordagens estatísticas e conexionistas, impulsionadas pelo aumento do poder computacional, pela disponibilidade massiva de dados e pelo desenvolvimento de algoritmos de aprendizado profundo. O estudo enfatiza a emergência da inteligência artificial generativa como um novo paradigma, caracterizado pela capacidade de produzir conteúdo original: textos, imagens, códigos e simulações a partir de modelos probabilísticos treinados em larga escala. Além dos avanços tecnológicos, são discutidos os impactos sociais, educacionais e profissionais decorrentes dessa transformação, bem como os desafios éticos, regulatórios e epistemológicos

associados à autonomia crescente desses sistemas. Por fim, o artigo aponta direções para pesquisas futuras, destacando a necessidade de abordagens interdisciplinares que conciliem inovação tecnológica, responsabilidade ética e uso crítico da inteligência artificial em diferentes contextos.

PALAVRAS-CHAVE: Inteligência Artificial, Inteligência Artificial Clássica, Aprendizado Profundo, Inteligência Artificial Generativa, Ética e Regulação da Inteligência Artificial

FROM CLASSICAL ARTIFICIAL INTELLIGENCE TO GENERATIVE ARTIFICIAL INTELLIGENCE: AN EVOLUTIONARY ANALYSIS

ABSTRACT: This article presents an evolutionary analysis of Artificial Intelligence (AI), examining its trajectory from classical paradigms grounded in symbolic systems, logical rules, and deterministic models to the contemporary emergence of deep neural networks and generative models. Initially, classical AI is discussed from a historical and conceptual perspective, highlighting its contributions, computational limitations, and reliance on explicitly formalized knowledge. Subsequently, the transition toward statistical and connectionist approaches is analysed, driven by increases in computational power, the large-scale availability of data, and the development of deep learning algorithms. The study emphasizes the emergence of generative artificial intelligence as a new paradigm, characterized by the ability to produce original content such as texts, images, code, and simulations based on large-scale probabilistic models. Beyond technological advances, the paper discusses the social, educational, and professional impacts resulting from this transformation, as well as the ethical, regulatory, and epistemological challenges associated with the growing autonomy of these systems. Finally, the article outlines directions for future research, highlighting the need for interdisciplinary approaches that reconcile technological innovation, ethical responsibility, and the critical use of artificial intelligence across different contexts.

KEYWORDS: Artificial Intelligence, Classical Artificial Intelligence, Deep Learning, Generative Artificial Intelligence, Ethics and Regulation of Artificial Intelligence

INTRODUÇÃO

A Inteligência Artificial (IA), enquanto campo de estudo, percorreu um caminho notável, evoluindo de sistemas que rigidamente seguiam lógicas predefinidas para entidades capazes de gerar textos, imagens e soluções com um grau surpreendente de originalidade. Essa jornada representa mais do que um avanço técnico; é uma redefinição fundamental da relação entre humano e máquina no domínio da

cognição e da criatividade. Inicialmente fundamentada na hipótese de que “todo aspecto da aprendizagem ou qualquer outra característica da inteligência pode, em princípio, ser descrito com tanta precisão que se pode fazer uma máquina para simulá-lo” (MINSKY, 1969, p. v), a IA buscou, por décadas, capturar a inteligência via representação simbólica explícita. Contudo, o paradigma contemporâneo se apoia em uma abordagem radicalmente distinta, baseada em dados e correlações estatísticas em larga escala. Como observam Goodfellow, Bengio e Courville (2016), “o aprendizado profundo permite que o computador construção de conceitos complexos a partir de conceitos mais simples”, um processo que acontece de forma distribuída e muitas vezes não interpretável, diferindo profundamente da transparência lógica buscada pelos pioneiros. Este artigo visa mapear essa trajetória evolutiva, explorando as fundamentações da IA clássica, a revolução do *deep learning* e o advento dos modelos generativos, refletindo sobre seus resultados e implicações.

A Inteligência Artificial (IA) constitui um campo de pesquisa interdisciplinar cuja evolução histórica reflete mudanças significativas nos paradigmas de modelagem da cognição e do aprendizado computacional. Desde suas origens, associadas a sistemas formais baseados em lógica simbólica e regras explicitamente definidas, até o estágio atual, caracterizado por modelos estatísticos de grande escala capazes de gerar textos, imagens e outros artefatos digitais, a IA passou por transformações metodológicas que alteraram substancialmente seus pressupostos teóricos e suas aplicações.

Nos primeiros desenvolvimentos da área, predominou a concepção de que processos cognitivos poderiam ser representados por meio de estruturas simbólicas manipuláveis computacionalmente. Essa perspectiva fundamenta-se na hipótese de que aspectos da inteligência humana poderiam ser descritos de forma suficientemente precisa para permitir sua simulação por máquinas (MINSKY, 1969). A partir desse pressuposto, consolidaram-se abordagens como sistemas especialistas, motores de inferência lógica e bases de conhecimento, que obtiveram resultados em domínios bem delimitados, mas apresentaram limitações relacionadas à escalabilidade, à adaptação a novos contextos e à dependência de conhecimento previamente formalizado (RUSSELL; NORVIG, 2021).

A partir do final do século XX, observa-se uma transição progressiva para abordagens estatísticas e conexionistas, impulsionada pelo aumento da capacidade computacional, pela ampliação do acesso a grandes volumes de dados e pelo desenvolvimento de algoritmos de aprendizado de máquina. Nesse contexto, o aprendizado profundo (*deep learning*) consolidou-se como um paradigma dominante, permitindo a construção automática de representações hierárquicas a partir dos dados, sem a necessidade de especificação explícita de regras simbólicas. Conforme descrito por Goodfellow, Bengio e Courville (2016), essas arquiteturas aprendem múltiplos níveis de abstração por meio de redes neurais profundas treinadas por otimização numérica.

A consolidação desse paradigma conduziu, mais recentemente, ao desenvolvimento de modelos de inteligência artificial generativa, baseados em arquiteturas neuronais de larga escala, como os modelos do tipo *transformer*. Esses sistemas diferenciam-se das abordagens tradicionais por sua capacidade de modelar distribuições probabilísticas complexas e gerar novas instâncias de dados que preservam regularidades estatísticas observadas durante o treinamento (Brown *et al.*, 2020). Tais modelos ampliaram o escopo das aplicações da IA incluindo tarefas de geração de linguagem natural, síntese de imagens, produção de código e simulações computacionais (Bommasani *et al.*, 2021).

A adoção crescente desses sistemas introduz questões relevantes de natureza técnica, ética e epistemológica, relacionadas à interoperabilidade dos modelos, à confiabilidade das saídas geradas, à presença de vieses nos dados e às implicações do uso desses sistemas em contextos educacionais, científicos e profissionais. Diante desse cenário, este artigo propõe uma análise da trajetória evolutiva da inteligência artificial, articulando os fundamentos da IA simbólica, a transição para o aprendizado profundo e o surgimento dos modelos generativos, com o objetivo de contextualizar seus impactos e limites à luz das transformações recentes do campo, em consonância com o escopo do texto anexo

DA INTELIGÊNCIA ARTIFICIAL CLÁSSICA À INTELIGÊNCIA ARTIFICIAL GENERATIVA: EVOLUÇÃO CONCEITUAL E TECNOLÓGICA

A Inteligência Artificial (IA) clássica, também denominada IA simbólica ou baseada em regras, constituiu o paradigma dominante desde o surgimento formal do campo, na década de 1950, até aproximadamente o final dos anos 1980. Esse modelo fundamentava-se na hipótese de que os processos cognitivos humanos poderiam ser formalizados por meio da manipulação explícita de símbolos, estruturas lógicas e regras dedutivas bem definidas. Nessa perspectiva, a inteligência era compreendida como um processo essencialmente simbólico, sequencial e lógico, no qual o conhecimento humano deveria ser previamente codificado em sistemas computacionais (Russell; Norvig, 2021).

Os sistemas especialistas representam a principal materialização desse paradigma. Aplicações como o MYCIN, desenvolvido para apoio ao diagnóstico médico, baseavam-se em extensas bases de conhecimento compostas por regras do tipo SE-ENTÃO, associadas a mecanismos de inferência lógica. Linguagens como LISP e Prolog tornaram-se ferramentas centrais nesse contexto, por sua capacidade de manipular símbolos e expressões lógicas de forma eficiente (Nilsson, 2010). Apesar dos avanços obtidos em domínios altamente estruturados, a IA simbólica revelou limitações significativas quando aplicada a ambientes complexos, incertos ou ruidosos.

Entre os principais entraves desse paradigma destacam-se a dependência de conhecimento explicitamente formalizado, a baixa escalabilidade dos sistemas baseados em regras e a dificuldade de lidar com ambiguidade, senso comum e aprendizado contínuo. Essas limitações ficaram conhecidas na literatura como o problema da representação do conhecimento, evidenciando a dificuldade de traduzir o conhecimento humano tácito e contextual em estruturas formais rígidas (Mccarthy, 1989). Como consequência, a IA simbólica mostrou-se pouco adaptável a problemas do mundo real, nos quais os dados são incompletos, dinâmicos e de alta dimensionalidade.

Um avanço conceitual relevante nesse período intermediário foi a introdução da Lógica *Fuzzy*, proposta por Lotfi A. Zadeh no final da década de 1960. Diferentemente da lógica clássica, fundamentada em valores binários de verdade, a Lógica *Fuzzy* permite a representação formal de graus de pertinência, possibilitando a modelagem computacional de conceitos vagos, imprecisos ou linguisticamente definidos (Zadeh, 1965).

Em seu trabalho seminal, Zadeh (1968) argumenta que muitos fenômenos do mundo real não podem ser adequadamente descritos por categorias rígidas, sendo mais apropriado empregar estruturas matemáticas capazes de lidar com transições graduais entre estados. Essa abordagem rompe com a rigidez da lógica booleana e introduz um formalismo compatível com o raciocínio aproximado humano.

A Lógica *Fuzzy* desempenhou papel fundamental na ampliação do escopo da Inteligência Artificial clássica, sobretudo em sistemas de controle, tomada de decisão e inferência baseada em conhecimento linguístico. Do ponto de vista epistemológico, esse paradigma representa uma inflexão importante ao reconhecer explicitamente os limites da precisão formal na representação do conhecimento, antecipando debates contemporâneos sobre incerteza, explicabilidade e confiança em sistemas inteligentes. Embora distinta dos modelos estatísticos atuais, a Lógica *Fuzzy* estabelece uma ponte conceitual entre a IA simbólica tradicional e abordagens posteriores baseadas em aproximação e aprendizado a partir de dados.

A transição para um novo paradigma ocorreu com o ressurgimento das Redes Neurais Artificiais (RNAs) e do Aprendizado de Máquina (*Machine Learning*), impulsionados por avanços simultâneos em poder computacional, disponibilidade massiva de dados e desenvolvimento de algoritmos de otimização eficientes. Diferentemente da abordagem simbólica, os modelos conexionistas abandonam a programação explícita de regras e passam a aprender representações internas diretamente a partir dos dados, por meio de processos estatísticos e iterativos (HAYKIN, 2009).

O desenvolvimento e a popularização do algoritmo de retropropagação do erro (*backpropagation*) permitiram o treinamento eficiente de redes neurais com múltiplas camadas ocultas, dando origem ao paradigma do aprendizado profundo (*deep learning*). Essas arquiteturas demonstraram elevada capacidade de extração automática de características hierárquicas e abstratas, alcançando desempenhos superiores em tarefas de percepção, como reconhecimento de imagens, processamento de linguagem natural e reconhecimento de fala áreas nas quais a IA simbólica apresentava desempenho limitado (Goodfellow; Bengio; Courville, 2016).

A evolução desse paradigma culmina, até o momento, na consolidação da Inteligência Artificial Generativa, que representa uma mudança qualitativa no escopo das aplicações de IA. Enquanto os modelos discriminativos tradicionais concentram-se em tarefas de classificação e regressão, os modelos generativos buscam aprender a distribuição probabilística subjacente aos dados, permitindo a síntese de novas amostras estatisticamente plausíveis. Esse avanço amplia significativamente o papel da IA, que passa a atuar não apenas como ferramenta analítica, mas também como agente de produção de conteúdo (BROWN *et al.*, 2020).

Arquiteturas como as Redes Adversariais Generativas (GANs), baseadas na competição entre uma rede geradora e uma rede discriminadora, e os Transformadores, fundamentados em mecanismos de atenção capazes de modelar dependências de longo alcance, constituem os principais pilares técnicos desse novo paradigma (Goodfellow *et al.*, 2014; Vaswani *et al.*, 2017). A partir dessas bases, emergem os Modelos de Linguagem de Grande Escala (LLMs), como o GPT, bem como modelos multimodais voltados à geração de imagens, áudio e vídeos.

Esses sistemas são caracterizados pelo pré-treinamento em grandes volumes de dados não estruturados e pela capacidade de adaptação a múltiplas tarefas por meio de instruções em linguagem natural (*prompts*). Tal característica redefine a interação humano-máquina, deslocando o foco da programação tradicional para a engenharia de *prompts* e para o uso contextual da IA como ferramenta cognitiva ampliada. No entanto, esse avanço também introduz desafios relevantes de natureza ética, epistemológica e regulatória, especialmente no que se refere à confiabilidade, à transparência e aos impactos sociais da autonomia crescente desses sistemas (Floridi *et al.*, 2018).

RESULTADOS E DISCUSSÕES

A evolução recente da Inteligência Artificial (IA), em especial com a consolidação dos modelos generativos, produziu impactos mensuráveis e de caráter estrutural em múltiplos domínios sociais, científicos e econômicos. A chamada IA Generativa ampliou significativamente o acesso a ferramentas avançadas de produção de

conteúdo, viabilizando a geração automatizada de textos, imagens, vídeos, áudios e códigos por usuários não especialistas. Esse processo tem sido descrito na literatura como um fenômeno de democratização cognitiva, no qual capacidades antes restritas a profissionais altamente qualificados passam a ser mediadas por sistemas computacionais treinados em larga escala (Bommasani *et al.*, 2021)

No campo científico, esses sistemas têm contribuído para a aceleração de descobertas, especialmente em áreas como o desenvolvimento de fármacos, a engenharia de materiais e a modelagem molecular, ao permitir a exploração eficiente de espaços de solução complexos e de alta dimensionalidade (JUMPER *et al.*, 2021). Na indústria, a IA generativa vem sendo empregada na otimização de processos de design, simulação e prototipagem, reduzindo ciclos de desenvolvimento e custos operacionais, além de favorecer abordagens baseadas em cocriação humano-máquina (Goodfellow; Bengio; Courville, 2016). No cotidiano, observa-se uma transformação profunda na interação humano-computador, marcada pela difusão de assistentes conversacionais baseados em grandes modelos de linguagem, que incorporam capacidades de compreensão contextual, adaptação discursiva e geração de respostas em linguagem natural (BROWN *et al.*, 2020).

Entretanto, os benefícios associados a esse avanço tecnológico coexistem com desafios relevantes de natureza técnica, ética, jurídica e social. Entre eles, destacam-se a amplificação de desinformação em larga escala, a produção de *deepfakes*, a reprodução e intensificação de vieses sociais presentes nos dados de treinamento, bem como controvérsias relacionadas à propriedade intelectual, autoria e uso indevido de conteúdos protegidos (Floridi *et al.*, 2018). Adicionalmente, o impacto da IA generativa sobre os mercados de trabalho criativos e cognitivos suscita debates sobre reconfiguração de competências, deslocamento ocupacional e necessidade de políticas públicas de adaptação e regulação.

Um aspecto central desse debate reside na limitada interpretabilidade dos modelos generativos contemporâneos. A lógica interna desses sistemas, frequentemente caracterizada como uma “caixa-preta”, contrasta com a transparência intencional que marcou os paradigmas clássicos da IA simbólica e baseada em regras. Essa opacidade introduz novos dilemas relacionados à explicabilidade, responsabilização e confiança nos sistemas automatizados, exigindo o desenvolvimento de marcos regulatórios, metodologias de auditoria algorítmica e abordagens de IA explicável (XAI) capazes de mitigar riscos sem comprometer os avanços alcançados (Russell; Norvig, 2021).

DIMENSÕES ÉTICAS, EPISTEMOLÓGICAS E LIMITES CIENTÍFICOS DA INTELIGÊNCIA ARTIFICIAL GENERATIVA

A consolidação da Inteligência Artificial Generativa introduz desafios que extrapolam o domínio estritamente tecnológico, exigindo uma análise crítica de seus fundamentos epistemológicos, éticos e científicos. Diferentemente dos paradigmas clássicos da Inteligência Artificial, nos quais os sistemas computacionais operavam como instrumentos determinísticos ou probabilísticos claramente delimitados, os modelos generativos contemporâneos passam a atuar como mediadores ativos na produção, organização e reformulação de informações, influenciando processos de interpretação e decisão humana.

Do ponto de vista epistemológico, esses sistemas tensionam concepções tradicionais de conhecimento científico baseadas em causalidade explícita, transparência metodológica e verificabilidade. Os modelos generativos não constroem explicações a partir de princípios formais ou leis causais explicitamente representadas, mas inferem regularidades estatísticas em grandes volumes de dados. Como resultado, produzem saídas semanticamente coerentes e contextualmente adequadas, ainda que desprovidas de justificativas causais acessíveis ao usuário humano. Essa característica reforça a distinção entre *capacidade preditiva* e *explicação científica*, impondo limites claros à interpretação epistemológica do conhecimento gerado (Floridi et al., 2018).

A opacidade estrutural desses modelos, frequentemente caracterizada como o problema da “caixa-preta”, constitui um limite científico relevante, sobretudo em domínios nos quais explicabilidade, auditabilidade e responsabilização são requisitos centrais. Em contraste com a IA simbólica clássica, que privilegiava representações lógicas transparentes, a IA generativa prioriza desempenho empírico e generalização estatística, deslocando o foco do controle explícito para a validação a posteriori dos resultados (Russell; Norvig, 2021).

Esses limites tornam-se ainda mais evidentes quando se analisa o papel da interação humano-máquina na mediação do comportamento dos modelos generativos, especialmente por meio da engenharia de *prompt*, discutida a seguir.

Engenharia de Prompt: Mediação Cognitiva, Limites Metodológicos e Responsabilidade Ética

A engenharia de *prompt* emerge como um elemento metodológico central na utilização de modelos de Inteligência Artificial Generativa, desempenhando papel distinto da programação tradicional. Enquanto abordagens clássicas de desenvolvimento de sistemas computacionais se baseiam na especificação explícita

de algoritmos, regras e estruturas de controle, a engenharia de *prompt* consiste na formulação de instruções em linguagem natural que condicionam, de maneira indireta, o comportamento do modelo generativo.

Do ponto de vista científico, a engenharia de *prompt* pode ser compreendida como uma forma de mediação cognitiva, na qual o usuário humano atua como agente regulador do espaço de respostas possíveis do modelo. Essa mediação não altera a arquitetura interna nem os parâmetros do sistema, mas influencia a ativação de padrões latentes aprendidos durante o treinamento. Assim, o comportamento do modelo passa a depender não apenas de sua estrutura estatística, mas também da qualidade semântica, contextual e normativa das instruções fornecidas (Brown *et al.*, 2020).

O desenvolvimento de um bom *prompt* constitui um elemento central para a interação eficaz entre usuários humanos e sistemas de inteligência artificial generativa, pois atua como um mediador cognitivo que condiciona a qualidade, a precisão e os limites epistemológicos das respostas produzidas. Um *prompt* bem elaborado deve definir claramente o objetivo da tarefa, fornecer contexto adequado, explicitar o formato esperado da saída e impor restrições que reduzam ambiguidades, como limites de extensão, estilo linguístico e normas técnicas adotadas. Além disso, a atribuição de um papel específico ao modelo e a organização das instruções em etapas contribuem para maior controle metodológico e previsibilidade dos resultados. Sob essa perspectiva, a engenharia de *prompt* não se restringe a um procedimento operacional, mas configura-se como uma prática técnico-científica alinhada a princípios de explicabilidade, rigor conceitual e responsabilidade no uso da IA, especialmente em contextos educacionais e acadêmicos. Essa compreensão dialoga com estudos recentes desenvolvidos por Mendonça e colaboradores, que analisam criticamente o uso de tecnologias digitais e de inteligência artificial na educação, enfatizando a necessidade de mediação pedagógica consciente e metodologicamente fundamentada (Mendonça *et al.*, 2023).

As IA Generativas geram texto para texto, texto para imagem ou vice versa, texto para gerar ou melhorar códigos computacionais, como por exemplo a transformação de uma imagem do autor principal e seu filho de forma lúdica em mostra a figura 1, feita através da ferramenta de geração ou modificação do *Gemini Google*® (Google LLC, 2026). Ressalta-se que *prompts* escrito em inglês tem maior probabilidade de sucesso em problemas de maior complexidade, devido a base de conhecimento é de aproximadamente 80% a 90% em inglês

Contudo, essa característica evidencia limites metodológicos importantes. A dependência da formulação dos *prompts* introduz variabilidade significativa nos resultados, dificultando a reprodutibilidade e a padronização de experimentos

científicos baseados em modelos generativos. Pequenas variações linguísticas ou contextuais nas instruções podem produzir respostas substancialmente distintas, o que impõe desafios à validação empírica e à comparação sistemática de resultados. Dessa forma, a engenharia de *prompt* não pode ser considerada um mecanismo de controle determinístico, mas sim uma prática probabilística sujeita a incertezas estruturais.

Sob a perspectiva epistemológica, a engenharia de *prompt* não resolve o problema da explicabilidade dos modelos generativos. Ainda que instruções bem definidas possam aumentar a coerência e a adequação contextual das respostas, elas não fornecem acesso direto aos mecanismos internos de inferência do sistema. Assim, a engenharia de *prompt* atua como um instrumento de orientação pragmática, e não como uma estratégia de explicação científica no sentido estrito (Goodfellow; Bengio; Courville, 2016).

O ajuste sequencial de um *prompt* consiste em um processo iterativo de refinamento no qual a resposta gerada por um modelo de linguagem é analisada criticamente e utilizada como base para a reformulação da instrução subsequente, incorporando maior contexto, restrições, exemplos ou critérios explícitos de avaliação. Nesse ciclo contínuo de feedback, o usuário identifica ambiguidades, lacunas informacionais ou desvios conceituais presentes na resposta e os traduz em ajustes progressivos no *prompt*, promovendo uma convergência gradual para resultados mais precisos, consistentes e alinhados ao objetivo pretendido. Essa estratégia, amplamente discutida na literatura recente sobre engenharia de *prompts*, permite explorar de forma sistemática as capacidades dos modelos generativos sem alterar seus parâmetros internos, funcionando como um mecanismo externo de controle semântico e metodológico sobre o processo de geração textual (CHEN *et al.*, 2024).



Figura 1 – Uso lúdico da ferramenta de edição de imagem do Gemini.

Do ponto de vista ético, a centralidade da engenharia de *prompt* desloca parte significativa da responsabilidade pelo conteúdo gerado para o usuário humano. Instruções ambíguas, incompletas ou orientadas por objetivos questionáveis podem induzir o modelo à geração de informações enviesadas, imprecisas ou normativamente problemáticas. Nesse sentido, o uso responsável da IA generativa exige diretrizes normativas para a formulação de *prompts*, incluindo: a delimitação explícita do escopo da tarefa, a restrição quanto à geração de inferências não verificáveis e a validação crítica dos resultados antes de sua aplicação prática (Bommasani *et al.*, 2021).

Além disso, a engenharia de *prompt* evidencia uma limitação estrutural da IA generativa: a transferência de parte do controle cognitivo e ético do sistema para a interação humano-máquina. Essa constatação reforça a necessidade de abordagens complementares, como técnicas de Inteligência Artificial Explicável (XAI) e modelos híbridos neuro-simbólicos, capazes de oferecer maior transparência, rastreabilidade e controle normativo sobre os sistemas generativos (Floridi *et al.*, 2018).

A Figura 2 ilustra a relação entre o usuário humano, a engenharia de *prompt* e o modelo de Inteligência Artificial Generativa, destacando os principais limites epistemológicos associados a essa interação. O usuário atua como agente formulador de instruções em linguagem natural (*prompt*), que funcionam como mecanismo de mediação cognitiva entre a intenção humana e o espaço probabilístico de respostas do modelo.

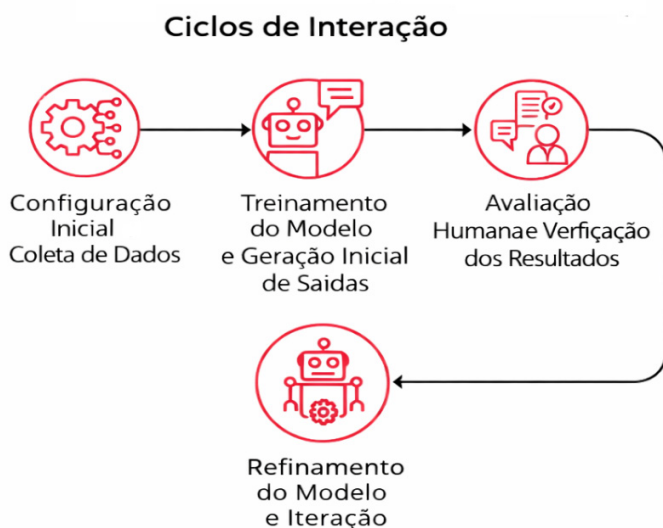


Figura 2 – Relação entre o usuário, a engenharia de *prompt* e o modelo de I.A.G.

O modelo generativo, por sua vez, opera como um sistema estatístico de alta complexidade, caracterizado por baixa interpretabilidade interna (“caixa-preta”), o que restringe a rastreabilidade causal e a explicação científica das saídas produzidas. Os limites epistemológicos emergem precisamente nessa interface, onde desempenho empírico, controle indireto e responsabilidade ética coexistem sem garantir transparência ou explicabilidade plena.

E finalmente uma visão global do trabalho é apresentada na figura 3, por meio de um infográfico gerado pelo *NotebookLM* (Google LLC, 2026).

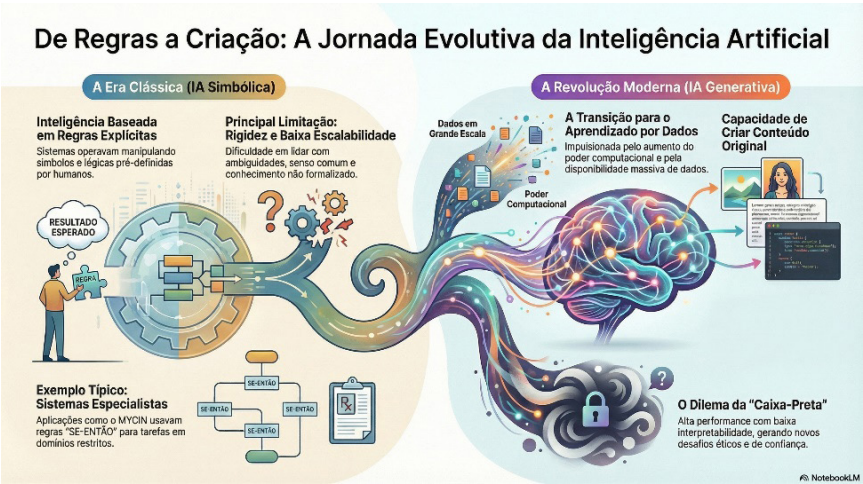


Figura 3 – Infográfico “A era clássica” VS “Revolução moderna”.

CONCLUSÃO E TRABALHOS FUTUROS

O artigo apresenta uma síntese crítica da evolução histórica e conceitual da Inteligência Artificial, evidenciando a transição dos paradigmas clássicos, fundamentados em sistemas simbólicos e regras explícitas, para abordagens estatísticas e conexionistas, culminando na emergência da Inteligência Artificial Generativa. Argumenta-se que essa transformação ultrapassa ganhos incrementais de desempenho computacional, configurando uma mudança estrutural nos pressupostos teóricos, metodológicos e epistemológicos do campo. Enquanto a IA clássica privilegiava transparência, controle e previsibilidade, os modelos contemporâneos baseados em aprendizado profundo deslocam o foco para a aprendizagem automática a partir de grandes volumes de dados, ampliando significativamente o escopo de aplicações, mas introduzindo desafios relacionados à opacidade, aos vieses

e à confiabilidade. O texto destaca a IA Generativa como um fenômeno de impacto transversal nos domínios científico, educacional, industrial e social, cujos benefícios coexistem com limites técnicos, éticos e regulatórios. Nesse contexto, a engenharia de prompt é analisada como um elemento metodológico central na mediação entre usuários e sistemas generativos, funcionando como mecanismo de orientação probabilística que amplia a utilidade prática dos modelos, ao mesmo tempo em que explicita seus limites epistemológicos. Conclui-se que a compreensão crítica da trajetória da IA é essencial para fundamentar seu uso responsável, apontando como direções futuras pesquisas empíricas comparativas, o desenvolvimento de abordagens híbridas, o avanço em IA explicável, auditoria algorítmica e governança de dados, bem como o fortalecimento de perspectivas interdisciplinares voltadas à regulação, formação crítica e sustentabilidade do uso dessas tecnologias.

AGRADECIMENTO

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

REFERÊNCIAS

BOMMASANI, Rishi; HUDSON, Drew A.; ADLER, Ehsan. *On the opportunities and risks of foundation models*. Stanford: Center for Research on Foundation Models, 2021.

BROWN, Tom B.; MANN, Benjamin; RYDER, Nick; SUBBIAH, Melanie; KAPLAN, Jared; DHARIWAL, Prafulla; NEELAKANTAN, Arvind; et al. *Language models are few-shot learners*. *Advances in Neural Information Processing Systems*, v. 33, p. 1877–1901, 2020.

CHEN, Banghao; ZHANG, Zhaofeng; LANGRENÉ, Nicolas; ZHU, Shengxin. Unleashing the potential of prompt engineering for large language models. *Journal of Artificial Intelligence Research*, v. 78, p. 1–35, 2024.

FLORIDI, Luciano; COWLS, Josh; KING, Thomas C.; et al. *AI4People — an ethical framework for a good AI society*. *Minds and Machines*, v. 28, n. 4, p. 689–707, 2018.

GOODFELLOW, Ian J.; BENGIO, Yoshua; COURVILLE, Aaron C. *Deep Learning*. Cambridge (MA): MIT Press, 2016.

GOODFELLOW, Ian J.; POUGET-ABADIE, Jean; MIRZA, Mehdi; XU, Bing; WAHLSTROM, David; OZAI, Sherjil; COURVILLE, Aaron C.; BENGIO, Yoshua. *Generative adversarial nets*. In: *Advances in Neural Information Processing Systems*, v. 27, p. 2672–2680, 2014.

GOOGLE LLC. *Gemini*. Disponível em: <https://gemini.google.com/>. Acesso em: 08 jan. 2026.

GOOGLE LLC. *NotebookLM*. Disponível em: <https://notebooklm.google/>. Acesso em: 08 jan. 2026.

HAYKIN, Simon S. *Neural Networks and Learning Machines*. 3. ed. New York: Pearson, 2009.

JUMPER, John; EVANS, Richard; PRITZEL, Alexander; GREEN, Tim; FIGURNOV, Michael; RONNEBERGER, Olaf; TUNYASUVANUKOOL, Kuan; BATES, Russell; ŽÍDEK, Adam; POTAPENKO, Alexander; BRIDGLAND, Alexander; MEYER, Clemens; KOHL, Sebastian A. A.; BALLARD, Andrew J.; COWIE, Alex; ROMERA-PAREDES, Bernardino; NIKOLOV, Stanislav; JAIN, Rishub; ADLER, Jonathan;

BLACKWELL, Philip; VIELLE, Curtis; LIEBSCHNER, Daniel; BAKER, David; HASSABIS, Demis. *Highly accurate protein structure prediction with AlphaFold*. *Nature*, v. 596, n. 7873, p. 583–589, 2021. DOI: 10.1038/s41586-021-03819-2.

MCCARTHY, John; THOMASON, Richmond (org.). *Artificial intelligence, logic and formalizing common sense*. Dordrecht: Springer, 1989.

MENDONÇA, Márcio; PEREIRA, João Luiz; CARVALHO, Patrícia; SANTOS, Larissa; SILVA, Ana Beatriz; SOUZA, Marcos Vinícius. *Tecnologias digitais e inteligência artificial na educação: limites, potencialidades e mediações pedagógicas*. Ponta Grossa: Atena Editora, 2023.

MINSKY, Marvin; MCCARTHY, John; SHANNON, Claude E. (orgs.). *Semantic Information Processing*. Cambridge (MA): MIT Press, 1969.

NILSSON, Nils J. *The Quest for Artificial Intelligence*. Cambridge: Cambridge University Press, 2010.

QUANSER Inc. *Inverted Pendulum System Documentation*. Markham: Quanser Inc., 2024.

RUSSELL, Stuart; NORVIG, Peter. *Artificial Intelligence: A Modern Approach*. 4. ed. Hoboken (NJ): Pearson, 2021.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki; USZKOREIT, Jakob; JONES, Llion; GOMEZ, Aidan N.; KAISER, Łukasz; POLOSUKHIN, Illia. *Attention is all you need*. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach (CA): Curran Associates, 2017. p. 5998–6008.

ZADEH, Lotfi A. *Fuzzy sets*. *Information and Control*, v. 8, n. 3, p. 338–353, 1965.

ZADEH, Lotfi A. *Fuzzy algorithms*. *Information and Control*, v. 12, n. 2, p. 94–102, 1968.