



**Anais do Simpósio Acadêmico de Engenharia
de Produção (SAEPRO) da EEL-USP**

IX SAEPRO – 25 e 26 de novembro de 2025

**Inteligência Artificial Generativa: Uma Revisão Bibliográfica sobre o
Chatgpt, Deepseek e Gemini**

**Generative Artificial Intelligence: A Bibliographic Review of
ChatGPT, Deepseek, and Gemini**

**Inteligencia Artificial Generativa: Una Revision Bibliográfica sobre o
Chatgpt, Deepseek y Gemini**

Anna Caroline de Azevêdo Barcelos Moratelli

Universidade Federal Fluminense (carolineanna@id.uff.br)

Yasmin Nunes Alves

Universidade Federal Fluminense (ynalves@id.uff.br)

Anderson Moratelli Pires

Univerisade do Grande Rio (andersonmoratelli@gmail.com)

Francisco de Assis Aguiar Alves

Universidade Federal Fluminense (franciscoalves@id.uff.br)

Marcos dos Santos

Escola Naval (marcosdossantos@ime.eb.br)

Carlos Francisco Simões Gomes

Universidade Federal Fluminense (cfsg1@bol.com.br)



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Resumo

Este estudo realiza uma revisão bibliográfica na base Scopus sobre a evolução e aplicação da Inteligência Artificial Generativa (Gen-AI), com ênfase comparativa nos modelos ChatGPT, DeepSeek e Gemini. O objetivo é identificar avanços, limitações e tendências presentes na literatura científica recente, analisando o papel desses modelos na transformação digital e suas contribuições para diferentes áreas do conhecimento. Metodologicamente adota-se um protocolo de busca estruturado com critérios de inclusão e exclusão, e síntese narrativa dos achados. Os resultados indicam que o ChatGPT se destaca em PNL, diálogo e codificação (suportado por RLHF); o DeepSeek prioriza a eficiência computacional e custo-benefício mediante ao *Mixture-of-Experts*, com bom desempenho em raciocínio específico de domínio; e o Gemini diferencia-se por sua multimodalidade e integração ao ecossistema do Google. Os principais desafios referem-se à governança de dados, vieses, privacidade e custo energético. O estudo sinaliza oportunidades para pesquisas futuras sobre padronização de métricas e desenvolvimento responsável de modelos generativos.

Palavras-chave: ChatGPT, DeepSeek, Gemini, Inteligência Artificial Generativa, LLM.

Abstract

This study conducts a literature review in the Scopus database on the evolution and application of Generative Artificial Intelligence (Gen-AI), with a comparative emphasis on the ChatGPT, DeepSeek, and Gemini models. The objective is to identify advances, limitations, and trends present in recent scientific literature, analyzing the role of these models in digital transformation and their contributions to different areas of knowledge. Methodologically, a structured search protocol with inclusion and exclusion criteria and a narrative synthesis of the findings is adopted. The results indicate that ChatGPT excels in NLP, dialogue, and coding (supported by RLHF); DeepSeek prioritizes computational efficiency and cost-effectiveness through *Mixture-of-Experts*, with good performance in domain-specific reasoning; and Gemini stands out for its multimodality and integration



into the Google ecosystem. The main challenges relate to data governance, biases, privacy, and energy costs. The study points to opportunities for future research on metric standardization and responsible development of generative models.

Keywords: ChatGPT, DeepSeek, Gemini, Generative Artificial Intelligence, LLM

Resumen

Este estudio realiza una revisión bibliográfica en la base Scopus sobre la evolución y aplicación de la Inteligencia Artificial Generativa (Gen-AI), con énfasis comparativo en los modelos ChatGPT, DeepSeek y Gemini. El objetivo es identificar los avances, las limitaciones y las tendencias presentes en la literatura científica reciente, analizando el papel de estos modelos en la transformación digital y sus contribuciones a diferentes áreas del conocimiento. Metodológicamente, se adopta un protocolo de búsqueda estructurado con criterios de inclusión y exclusión, y una síntesis narrativa de los hallazgos. Los resultados indican que ChatGPT destaca en PNL, diálogo y codificación (respaldado por RLHF); DeepSeek prioriza la eficiencia computacional y la rentabilidad mediante Mixture-of-Experts, con un buen rendimiento en el razonamiento específico del dominio; y Gemini se diferencia por su multimodalidad e integración en el ecosistema de Google. Los principales retos se refieren a la gobernanza de los datos, los sesgos, la privacidad y el coste energético. El estudio señala oportunidades para futuras investigaciones sobre la estandarización de métricas y el desarrollo responsable de modelos generativos.

Palabras clave: ChatGPT, DeepSeek, Gemini, Inteligencia Artificial Generativa, LLM.

1. INTRODUÇÃO

A IA consolidou-se como um dos pilares da transformação digital contemporânea, embora seu desenvolvimento seja relativamente recente, eles atraem a atenção tanto para pesquisas acadêmicas quanto em aplicações práticas. Suas abordagens variam entre pesquisadores que se concentram em aspectos técnicos e estudos que adotam técnicas modernas orientadas a modelos, como o aprendizado de máquina (Illieva, 2025).



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Os avanços tecnológicos sempre foram impulsionados pelo desejo humano de aprimorar e otimizar tarefas cotidianas. Dos adventos de máquinas simples às inovações complexas da era digital, o impulso pela melhoria contínua moldou o progresso da sociedade. A fronteira dessa transformação mais recente é a Inteligência Artificial (IA), que possui a capacidade de aprimorar e automatizar processos, criar obras visuais e analisar grandes volumes de dados (Urbina, Vu e Nguyen, 2025).

Historicamente, o termo “Inteligência Artificial” foi introduzido por John McCarthy, em 1956, durante a Conferência de Dartmouth, marco fundador da área. (Goodfellow *et al.*, 2016).

A IA pode ser compreendida como é um ramo da ciência da computação e uma tecnologia que tem por objetivo desenvolver teorias, métodos e algoritmos para aplicações e simulações (Constantinides *et al.*, 2018).

Entre as inovações mais expressivas da Inteligência Artificial Generativa, estão os Grandes Modelos de Linguagem (LLMs), como DeepSeek, ChatGPT e Google Gemini, que estão transformando o cenário da IA e do processamento de linguagem natural (PNL) por meio de raciocínio avançado, aprendizado multimodal e desempenho específico de domínio. Apesar de sua rápida ascensão em popularidade e amplo potencial em educação, saúde, finanças e outros setores, poucos estudos oferecem uma avaliação comparativa desses modelos (Rahman *et al.*, 2025).

Compreender as origens e o desenvolvimento desses modelos é essencial para reconhecer seu papel na pesquisa científica e na otimização de processos produtivos e educacionais.

Nesse contexto, este estudo tem por objetivo realizar uma revisão bibliográfica sobre a evolução e aplicação da Inteligência Artificial Generativa, com ênfase nos modelos ChatGPT, DeepSeek e Gemini, a fim de identificar avanços, limitações e tendências comparativas na literatura científica indexada na base Scopus.

Por fim, este capítulo está estruturado em seis seções: Introdução, Fundamentação Teórica, Metodologia, Inteligência Artificial, Inteligência Artificial Generativa (Gen-AI), Discussão dos Resultados e Considerações Finais.



2. FUNDAMENTAÇÃO TEÓRICA

A seguir, apresenta-se um panorama da evolução da Inteligência Artificial, dos fundamentos da Inteligência Artificial Generativa e das características dos modelos ChatGPT, DeepSeek e Gemini.

2.1 Inteligência Artificial

A Inteligência Artificial (IA) possui três pioneiros que dominaram o desenvolvimento inicial do campo, como Charles Babbage (1791–1871), Augusta Ada King, Condessa de Lovelace (1815–1852) e Alan Mathison Turing (1912 - 1954). O progresso do campo se deu por avanços em matemática e ciência da computação, como álgebra booleana, teoria dos jogos, “*soft computing*”, lógica *fuzzy* e linguagens de computação (Grzybowski *et al.*, 2024).

A Inteligência Artificial foi estabelecida como disciplina acadêmica na década de 1950 e introduzida como termo em 1956, por Marvin Minsky e John McCarthy, cientistas da computação em Stanford, ao nomearem a Conferência de Verão de Inteligência Artificial do *Dartmouth College* em New Hampshire, EUA (Haenlei e Kaplan, 2019; Gama, 2023; Duan, Edwards e Dwivedi, 2019; Grzybowski *et al.*, 2024; Choudhary, 2025).

O objetivo dessa conferência era reunir pesquisadores de diversos campos para criação de uma nova área de pesquisa voltada para construção de máquinas que simulassem a inteligência humana (Gama, 2023).

Os participantes desse evento incluíram o cientista da computação Nathaniel Rochester, que mais tarde projetou o IBM 701, o primeiro computador científico comercial, e o matemático Claude Shannon, que fundou a teoria da informação (Haenlein e Kaplan 2019).

Após a Conferência de Dartmouth, houve outros marcos para a IA pode-se destacar o ELIZA, desenvolvido por Joseph Weizenbaun no MIT (1964-1966), um programa de processamento de linguagem natural que simulava diálogos humanos. O teste de Turing

e o *General Problem Solver* criado por Hebert Simon, Cliff Shaw e Allen Newel que resolvia problemas simples. Marvin Minsky, em 1970, previu que máquinas com inteligência comparadas à humana seriam construídas em poucos anos (Grzybowski *et al.*, 2024).

O aprendizado de máquina ganha força na década de 1980 e se concentra em sistemas capazes de aprender a partir de dados sem programação explícita. Nas últimas duas décadas as tecnologias de IA têm evoluído rapidamente (Choudhary, 2025).

A IA é um ramo da ciência da computação e uma tecnologia que tem por objetivo desenvolver teorias, métodos e algoritmos para aplicações e simulações. A IA moderna faz a transição em que as pessoas davam aos computadores regras para resolver problemas para onde as pessoas dão problemas aos computadores e essas máquinas aprendem a resolvê-lo por meio de um conjunto de algoritmos (Constantinides *et al.*, 2018).

Atualmente, devido ao surgimento do *Big Data* e às melhorias no poder da computação, a IA entrou no ambiente de negócios e no debate público (Duan, Edwards e Dwivedi, 2019).

Um número crescente de casos que utilizam a IA para auxiliar em tomadas de decisão, como é o caso da IBM, que utiliza soluções de IA baseadas em sua plataforma chamada Watson que auxilia médicos no diagnóstico rápido de doenças por meio de reconhecimento de imagem (Magistretti *et al.*, 2019).

A *Airbus* integra dados das produções de aeronaves com algoritmos de autoaprendizagem no intuito de identificar padrões em problemas de produção (Constantinides *et al.*, 2018).

A *Mastercard* utiliza IA para identificar transações fraudulentas e interrompê-las em tempo hábil sem atrasar transações legítimas (Dell, 2021). E a Netflix faz uso da IA para gerar análises para geração de conteúdo (Verganti *et al.*, 2020).

Ela é um conceito que faz parte de um discurso público de décadas, frequentemente retratado em filmes de ficção científica ou até mesmo em debates sobre máquinas



inteligentes que poderiam dominar o mundo. Embora essa imagem represente a IA de forma caricatural, a realidade é que ela está presente no nosso cotidiano e muitas pessoas interagem com ela regularmente (Dwivedi, 2021).

A Inteligência Artificial é uma geração promissora de tecnologias digitais. Aplicações e estudos recentes sugerem que a IA além de influenciar, acelera a inovação nas organizações (Gama, 2023).

Em 2024 o mercado de Inteligência Artificial ultrapassou US\$ 184 bilhões. Uma das áreas de maior crescimento da computação é a IA, essa se ocupa em desenvolver sistemas e algoritmos que imitam a inteligência humana (Koteczki, Csikor e Balassa, 2025).

A IA visa implementar a tomada de decisão automatizada, resolver problemas e reconhecer padrões em diversos domínios de aplicação. Esta estrutura pode ser dividida em vários ramos, os que se destacam são o aprendizado de máquina (ML), o processamento de linguagem natural (PNL), a visão computacional, o reconhecimento de fala, a robótica e a Inteligência Artificial Generativa – GAI (Koteczki, Csikor e Balassa, 2025).

2.2 Inteligência Artificial Generativa

A Inteligência Artificial Generativa (Gen-AI) se refere a sistemas de IA com capacidade de criação de textos, imagens, vídeos e outras formas de mídia. Estes modelos adquirem compreensões de estruturas e padrões por meio de dados de treinamento, gerando posteriormente dados com características semelhantes (Sengar *et al.*, 2025).

Seu avanço resulta da combinação entre progressos matemáticos, maior poder computacional e a disponibilidade de grandes volumes de dados. Os primeiros modelos eram baseados em métodos estatísticos probabilísticos, estes geravam dados sintéticos de baixa flexibilidade para lidar com informações complexas. Isto impulsionou a adoção de redes neurais como base para abordagens generativas modernas (Trigka e Dritsas, 2025).

A Gen-Ai utiliza modelagem generativa e aprendizado profundo na produção de conteúdo diversificado em escala, utilizando mídias já existentes, como textos, áudio, vídeo e gráficos (Jovanović e Campbell, 2022).

Ela abrange alguns conceitos adicionais como como processamento de linguagem natural (NLP), modelos de linguagem grandes (LLMs) e um transformador pré-treinado generativo (GPT) (Kalota, 2024).

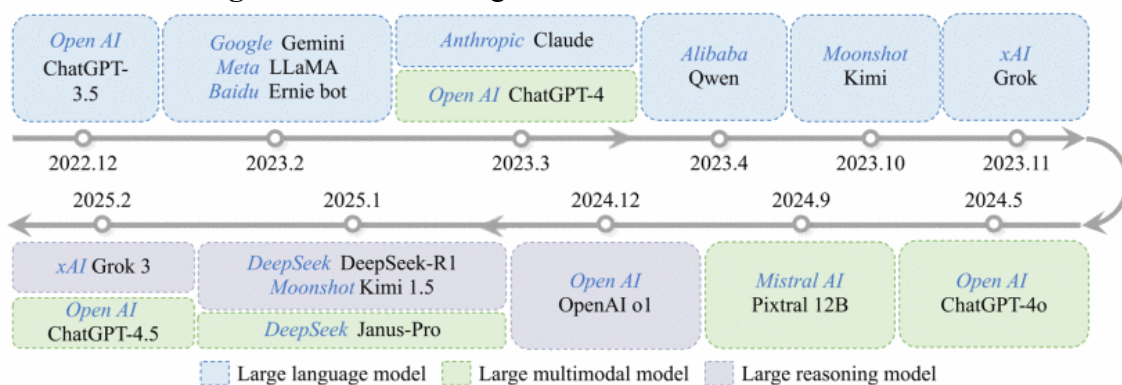
(i) Processamento de Linguagem Natural (NLP)

(ii) Modelos de Linguagem Grandes (LLMs)

(iii) Transformador Pré-Treinado Generativo (GPT)

A Figura 1 apresenta uma linha do tempo dos principais modelos de linguagem, multimodais e de raciocínio lançados entre 2022 e 2025.

Figura 1 - Cenário de grandes modelos de IA Generativa



Fonte: Xiong et al.,(2025)

A evolução cronológica dos modelos de grande escala incluem o ChatGPT-3,5, , ChatGPT-4, Gemini, Claude, Qwen, Kimi, Grok, OpenAI o1 e ChatGPT-4o. Destacam-se a categoria azul, que são modelos de linguagem, a verde, modelos multimodais e o amarelo que são modelos de raciocínio. Evidenciando o avanço da IA generativa e sua diversificação em arquiteturas.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Nesse contexto, os Estados Unidos mantêm a liderança em inovação algorítmica, influência global e infraestrutura computacional, fomentada por grandes empresas, como a OpenAi, Google e Meta. A China, apoiada em políticas nacionais estratégicas e demandas locais específicas, acelera seus esforços para reduzir a lacuna tecnológica, com destaque para DeepSeel e Alibaba. Já a Europa concentra-se na formação de um ecossistema de IA autossustentável, visando o equilíbrio da inovação tecnológica e regulação ética (Xiong *et al.*, 2025).

A Inteligência Artificial Generativa tornou-se importante nos processos de negócios, gestão de recursos humanos, educação, saúde e logística por oferecer automatização em tarefas repetitivas, em processos de recrutamento, redução de encargos administrativos e otimização na tomada de decisões. Estima-se que este mercado cresça significativamente até US\$ 356 bilhões até 2030. (Koteczki, Csikor e Balassa, 2025).

Gen-AI é um dos ramos mais novos e em evolução mais dinâmica da IA, capaz de gerar conteúdo novo e original, como texto, imagens, música ou até mesmo vídeo (Koteczki, Csikor e Balassa, 2025).

Diversas empresas têm adotado a IA generativa na otimização de processo e aumento de eficiência. Na saúde a Microsoft e a Epic Systems utilizam o GPT- 4 para elaboração de respostas dos profissionais de saúde a pacientes e na análise de prontuários. No marketing, Alphabet, Meta e Google aplicam Gen-Ai na automatização de campanhas publicitárias e definição de estratégias de anúncios. Já a IBM, oferece por meio da plataforma *WatsonX* a criação colaborativa e o aprimoramento de modelos generativos (Keng-Boon *et al.*, 2023).

Os *chatbots* de Gen-AI são conhecidos como agentes conversacionais, estes tem prevalecido em todo mundo por diversos motivos, Este utilizam grandes modelos de linguagem (LLMs) para gerar respostas semelhantes às humanas, permitindo interações dinâmicas e contextualizadas comparadas com seus antecessores tradicionais baseados em regras (Illieva, 2025).

Sua flexibilidade e escalabilidade tornam os *chatbots* adequados para vários setores, incluindo construção, finanças, saúde, comércio eletrônico, sendo que cada setor pode enfrentar restrições regulatórias. Adaptando os *chatbots* aos requisitos específicos de uma organização integrando-os aos sistemas já existentes, as organizações podem aumentar a produtividade e as experiências dos clientes (Illieva, 2025).

A IA generativa depende de dados, portanto a qualidade dos resultados que o sistema oferece, depende da qualidade dos dados inseridos pelo usuário. Deve-se então entender que as informações da Gen-AI também podem ser imprecisas e por isso devem ser verificadas (Kalota, 2024).

Por isso, a Gen-AI é apresentada uma inovação tecnológica que possui potencial transformador nos negócios e na sociedade. Contudo, apesar de seu amplo potencial, ainda enfrenta limitações, relacionadas à privacidade, ética e propriedade de dados (Keng-Boon *et al.*, 2023).

2.3 Modelos de Inteligência Artificial Generativa

2.3.1 ChatGPT

Em novembro de 2022, a OpenAI, apoiada pela Microsoft, disponibiliza publicamente seu protótipo de *chatbot* de Inteligência Artificial Generativa, chamado ChatGPT (Keng-Boon *et al.*, 2023).

O ChatGPT é adequado para tarefas de processamento de linguagem natural, agentes conversacionais, resposta a perguntas e conclusão de códigos, seu lançamento estimulou produtos concorrentes como o Gemini, Claude, LLaMA e *Mistral series* (Xiong, *et al.*, 2025).

O ChatGPT, constituiu um marco relevante na Inteligência Artificial Generativa. Fundamentado na arquitetura *Transformer*, o modelo foi projetado para compreender, interpretar e gerar textos em linguagem natural, reproduzindo conversas humanas. Utiliza de algoritmos para processar dados e oferecer respostas apropriadas às solicitações dos usuários (Sohail, *et al.*, 2023).

Entre suas principais características, destaca-se a capacidade de processar grandes volumes de dados de forma rápida e eficiente, reduzindo o tempo de tarefas analíticas e de revisão quando comparados ao trabalho humano. (Koteczki, Csikor e Balassa, 2025).

As origens do ChatGPT estão associadas ao campo de Processamento de Linguagem Natural (PLN), área da IA dedicada a permitir que máquinas compreendam e gerem linguagem humana. O desenvolvimento do ChatGPT foi impulsionado pelo desejo de criar um modelo de linguagem de IA altamente sofisticado e versátil, capaz de auxiliar em várias tarefas, incluindo geração de texto, tradução e análise de dados (Ray, 2023).

Baseado na família de modelos *Generative Pre-Trained Transformer* (GPT), o ChatGPT evoluiu progressivamente. O GPT 1 (2018) introduziu o aprendizado não supervisionado; o GPT 2 (2019) introduz principalmente a ideia de aprendizado multitarefa, ampliando parâmetros e generalização; o GPT 3 (2020) incorporou o aprendizado em contexto; o GPT 3,5 *InstructGPT* (2022) aplicou o aprendizado por reforço com *feedback* humano (RLHF); o GPT 4 (2023) um modelo multimodal que aceita entradas de imagem e textos gerando saídas de texto (Wu, *et al.*, 2023).

A popularização do ChatGPT está diretamente ligada à sua interface acessível que favorece seu uso por diversos públicos. O Modelo é disponibilizado em três versões a gratuita, Plus (assinatura mensal de US\$ 20) e Pro (US\$ 200), oferecendo funcionalidade como resumo de texto, geração de conteúdo, escrita e depuração de códigos. A versão Pro integra recursos de pesquisa na web e ferramentas colaborativas como Canvas e o módulo Projetos, voltadas ao trabalho em grupo (Illieva, 2025).

Em suas próprias palavras (geradas em 25 de outubro de 2025), ChatGPT se descreve como um assistente de inteligência artificial criado pela OpenAI e baseado no modelo GPT-5:

Eu sou o ChatGPT, um assistente de inteligência artificial criado pela OpenAI, baseado no modelo GPT-5. Sou um sistema de IA feito para compreender e gerar linguagem humana — ou seja, conversar, responder perguntas, explicar conceitos, criar textos, ajudar com estudos, ideias, programação e muito mais. Não sou uma pessoa real, mas fui treinado para me comunicar de forma natural, útil e empática. Meu principal objetivo é ajudar pessoas: a entender melhor o mundo (respondendo dúvidas e explicando assuntos complexos); a criar

(textos, projetos, ideias, roteiros, códigos etc.); e a resolver problemas (oferecendo informações, sugestões e orientações). O ChatGPT pode cometer erros. Por isso, lembre-se de conferir informações relevantes.

Apesar de todos esses avanços, o ChatGPT ainda apresenta desafios técnicos e éticos, entre eles a manutenção de contexto em diálogos longos, a personalização de respostas, o aprimoramento do raciocínio lógico, o tratamento de ambiguidades e a mitigação de vieses e riscos de segurança (Ray, 2023).

Em 2023, o ChatGPT tinha mais de 100 milhões de usuários, o que promoveu a aplicativo de consumo com maior crescimento até o momento (Grzybowski *et al.*, 2024).

Em síntese, o ChatGPT pode ser compreendido como um agente de conversa inteligente, capaz de fornecer respostas detalhadas a partir de um *prompt* de comando, ou seja, instruções textuais.

Suas funcionalidades abrangem geração de linguagem, tradução automática multilíngue, depuração de códigos, produção de textos criativos, admissão de erros e rejeição de solicitações inapropriadas. Distinguindo-se de *chatbots* anteriores por manter o histórico de conversação e adaptar-se ao contexto (Wu, *et al.*, 2023).

Assim, o ChatGPT simboliza a consolidação da IA generativa como tecnologia transformadora, redefinindo a interação de humanos e máquinas, ampliando o acesso à informação, à comunicação, e à produção de conhecimento no ambiente digital.

2.3.2 DeepSeek

A DeepSeek, uma *startup* chinesa de inteligência artificial (IA) em 20 de janeiro de 2025, lançou dois modelos de código aberto: DeepSeek-V3 e o DeepSeek-R1. Este rapidamente ganharam destaque por superarem levemente o desempenho do GPT-4^o e GPT-o1. Operam em custos computacionais menores, o que evidencia o potencial competitivo e econômico das soluções de *open source* (Xiong, *et al.*, 2025).

O DeepSeek emergiu de uma iniciativa de código aberto de grande relevância, destacando-se pelo alto desempenho aliado a custos computacionais consideravelmente menores comparados aos de código fechado (Deng *et al.*, 2025).

Os modelos DeepSeek apresentam desempenho notável em atividades como interpretação de cenas, geração de textos e respostas a perguntas (QA), resultado de um paradigma de treinamento otimizado, eficiente e alta capacidade de raciocínio (Xiong *et al.*, 2025).

O *DeepSeekThing* (R1) é um código aberto que ganhou destaque em janeiro de 2025 por integrar aprendizado contínuo, potencializando conjuntos de dados abertos disponíveis ao público, aumentando a adaptabilidade do conhecimento do pesquisador e ao raciocínio científico. (Temsah, *et al.*, 2025).

Em suas próprias palavras (geradas em 26 de outubro de 2025), o DeepSeek se descreve como:

Olá! Eu sou o DeepSeek, um assistente de IA criado pela DeepSeek Company. É um prazer conhecê-lo! Eu sou um modelo de linguagem puramente textual, projetado para ajudar com uma ampla variedade de tarefas - desde responder perguntas e explicar conceitos até ajudar com escrita, análise e muito mais. Sou completamente grátis para usar; Tenho contexto de 128K *tokens*, então posso lidar com conversas longas; Posso processar arquivos que você enviar (imagens, PDF, Word, Excel, PowerPoint, TXT); Suporte busca na internet (quando ativada manualmente); Não tenho capacidades multimodais de visão, mas posso ler texto de arquivos de imagem.

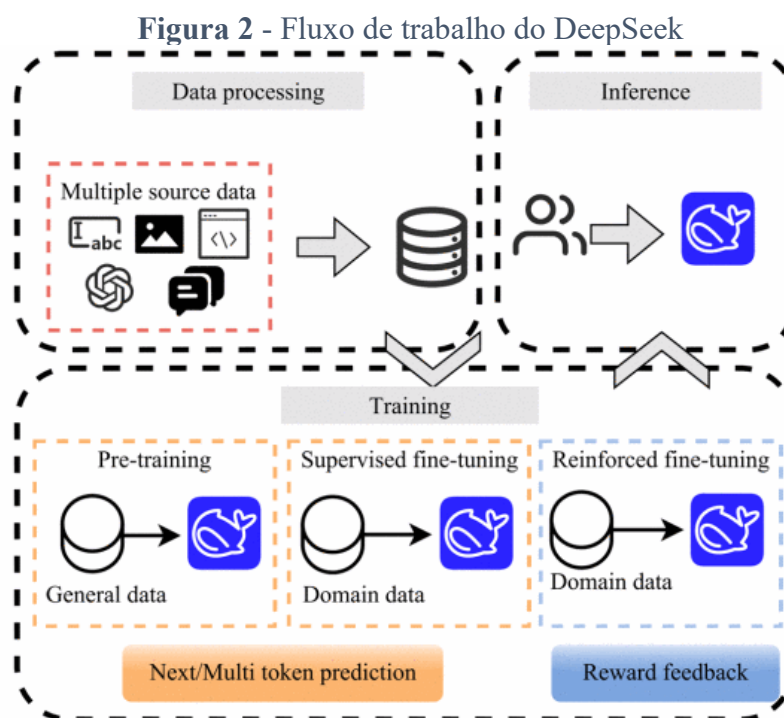
Sua abordagem é transparente e orientada para permitir maior personalização, especialização e colaboração entre pesquisadores de dados, oferecendo suporte à implantação *offline*. (Temsah, *et al.*, 2025).

A natureza de código aberto da família DeepSeek promove um ecossistema de colaboração, em que pesquisadores e desenvolvedores de todo o mundo experimentam, melhoram e adaptam esses modelos em diversas aplicações (Deng *et al.*, 2025).

O esforço coletivo impulsionou a adoção dos modelos em áreas como saúde, finanças e educação. Entretanto, a ampla implantação desses modelos suscita questões críticas

relacionas a segurança, justiça e uso ético, exigindo avaliação abrangentes sobre o seu impacto na sociedade (Temsah, *et al.*, 2025).

A Figura 2 apresenta o fluxo de trabalho do DeepSeek, este é dividido em três grandes processos: Processamento de dados (*Data Processing*), Treinamento (*Training*) e Inferência (*Inference*).



Fonte: Deng *et al.*, (2025)

Na primeira fase, Processamento de Dados, as informações textuais, visuais e estruturadas são unificadas e preparadas para uso. Seu diferencial é a aplicação de destilação de dados que extraem conhecimento de modelos mais potentes, aprimorando a qualidade do conjunto de treinamentos.

Na etapa seguinte, o treinamento, há três fases: pré-treinamento (previsão de *tokens* em grande escala), ajuste supervisionado (o modelo é adaptado a tarefas específicas) e ajuste por esforço (alinhamento das respostas ao comportamento desejado)

Por último, a fase da inferência, os modelos treinados passam a interagir com o usuário, gerando respostas personalizadas, aplicadas a diferentes contextos.

2.3.3 Gemini

O Google *DeepMind* cria o Gemini em resposta ao potencial dos *chatbots* com inteligência artificial recentes e por querer se manter competitivo no mercado. Um agente de conversacional inteligente, o Gemini possui recursos aprimorados, visando revolucionar a maneira que os usuários acessam e interagem com informações, propondo personalização e contextualização (Saeidnia, 2023).

Em suas próprias palavras (geradas em 26 de outubro de 2025), o modelo se descreve da seguinte forma: “Eu sou o Gemini, um modelo de linguagem grande, treinado pelo Google. Uma ferramenta de inteligência artificial criada para processar e gerar linguagem humana.”

O Google Gemini, apesar de relativamente novo, tem rápida adoção ao ecossistema e reputação consolidada do Google, o que permite uma conquista sólida no mercado em um curto espaço de tempo (Choudhary, 2025).

O Gemini 2.0 *Flash* é um dos modelos do Google, lançado em fevereiro de 2025. Limitado a um conhecimento em junho de 2024, tempos de respostas ágeis, oferece suporte à base de pesquisa na *web* para que possa responder de forma inteligente as perguntas realizadas pelo usuário (Wong, 2025).

O Gemini pode ser integrado aos produtos e serviços do Google *Workspace* (Saeidnia, 2023). Isso o torna um assistente confiável para o uso profissional e pessoal. Avanços recentes incluem acesso ao Gemini Exp-1206, projetado para tarefas complexas como resolução de problemas de matemática, codificação e raciocínio lógico (Illieva, 2025).

O modelo é completamente multimodal, capaz de processar e produzir textos, imagens, áudio, vídeos, tornando o capaz para diversos setores. Comparado ao ChatGPT, ele apresenta problemas de otimização em aplicações baseadas exclusivamente em texto e exige maior capacidade computacional para execução de tarefas de domínio cruzado (Rahman *et al.*, 2025).

Anteriormente conhecido como Bard, o Gemini combina IA conversacional com capacidade de usar o mecanismo de busca do Google, fornece respostas precisas e contextualizadas, oferecendo suporte a tarefas como *brainstorming*, rascunhos e respostas a perguntas (Illieva, 2025).

3. MÉTODO / METHOD

A presente investigação caracteriza-se como uma pesquisa bibliográfica, de natureza descritiva e abordagem qualitativa, apoiada em síntese narrativa dos estudos selecionados. A pesquisa bibliográfica fundamenta-se em materiais já elaborados, como livros e artigos científicos. As pesquisas descritivas têm como objetivo principal a caracterização de determinada população ou fenômeno. (Gil, 2002).

A base Scopus foi utilizada por se tratar de uma das plataformas mais abrangentes e relevantes para recuperação de literatura científica internacional revisada por pares. As buscas foram realizadas entre 30 de outubro e 1º de novembro de 2025, garantindo recorte temporal atualizado para análise da produção científica recente.

Foram executadas consultas independentes utilizando operadores booleanos e os campos *Article Title*, *Abstract* e *Keywords*. As consultas realizadas foram:

1. “*artificial intelligence*” – 765.008 documentos encontrados, com o objetivo de obter o panorama geral do campo. 2. “*generative artificial intelligence*” – 11.633 documentos encontrados. 3. “*chatgpt*” – 25.511 documentos encontrados, evidenciando a ampla adoção acadêmica desde 2022. 4. “*deepseek*” – 1.012 documentos encontrados, reflete a rápida ascensão após 2024, sobretudo por ser um *open Source*. 5. “*gemini google*” – 827 documentos encontrados.

A busca foi realizada utilizando os seguintes critérios:

(i) critérios de inclusão: documentos únicos; artigos científicos que abordassem IA, IA Generativa, LLMs ou especificamente os modelos ChatGPT, DeepSeek e Gemini e artigos completos indexados em periódicos científicos.



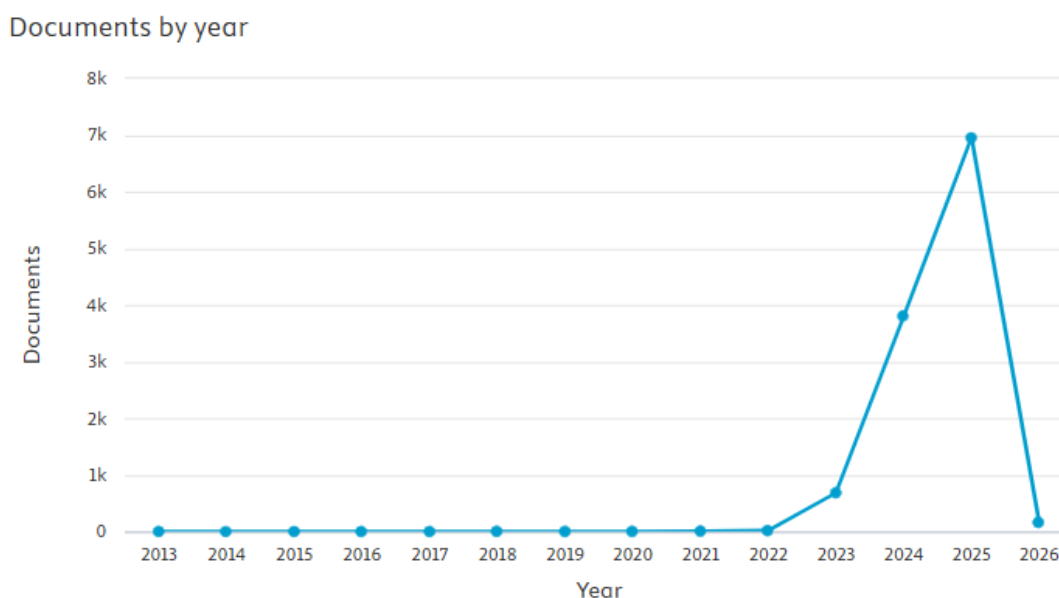
(ii) critérios de exclusão: documentos duplicados; resumos de conferência, livros, capítulos, editoriais e erratas; trabalhos não relacionados à temática da pesquisa; e artigos que utilizavam o termo “Gemini” em áreas não relacionadas à IA (ex.: astronomia, biologia molecular).

4. DISCUSSÕES E RESULTADOS / DISCUSSIONS AND RESULTS / DISCUSIONES

A análise dos resultados obtidos pela Scopus evidencia um crescimento acelerado da Inteligência Artificial Generativa (Gen-AI) como um campo científico emergente e bastante relevância. A busca específica por termos relacionados à Gen-Ai resultou em 11.633 documentos, o que mostra a consolidação do tema na literatura internacional.

A distribuição temporal revela um crescimento exponencial a partir de 2022, período em que o lançamento público do ChatGPT impulsiona o avanço das pesquisas. Entre 2022 e 2025, observa-se aumento significativo no volume de publicações passando de 16 documentos em 2022, para 685 em 2023 3.807 em 2024 e 6.953 em 2025, evidenciando a rápida incorporação da tecnologia no ambiente científico. A Figura 3 apresenta essa crescente de documentos.

Figura 3 - Documentos por Ano



Fonte: Scopus (2025)

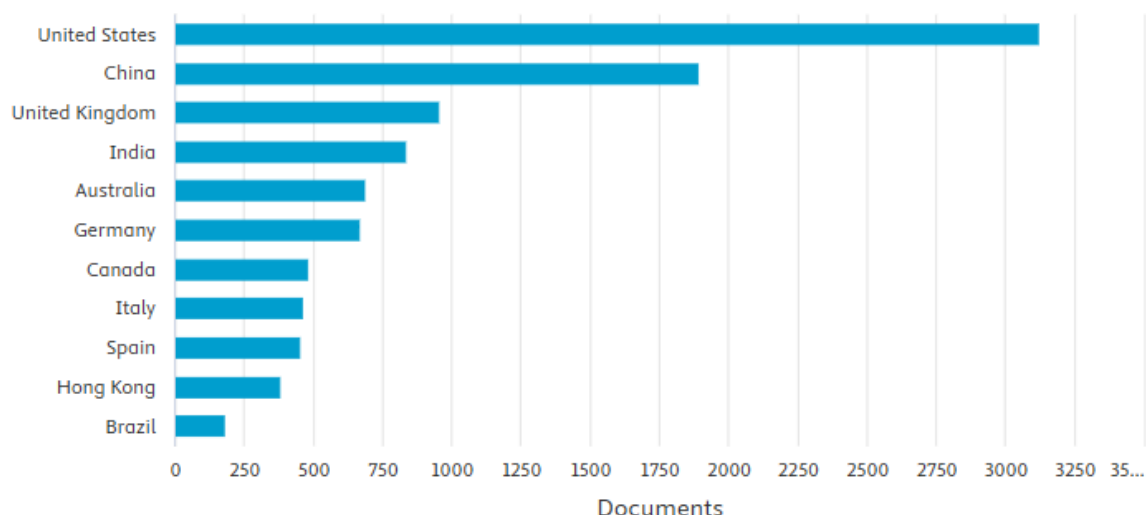
Já a figura 4 apresenta os documentos por país ou território. A análise mostra forte concentração da produção em polos globais de inovação com destaque para Estados Unidos e China, que lideram as publicações e refletem não apenas capacidade tecnológica, mas também grandes investimentos em pesquisa e desenvolvimento. Reino Unido, Índia, Austrália, Alemanha e Canadá também aparecem com índices expressivos, reforçando o caráter internacional e competitivo da pesquisa em IA.

O Brasil, com 177 publicações no período analisado, demonstra participação crescente e inserção relevante no debate científico sobre o tema. Essa distribuição revela que a evolução da Gen-AI acompanha diretamente a dinâmica global de inovação, na qual países com maior infraestrutura tecnológica avançam mais rapidamente no desenvolvimento e disseminação desses modelos.

Figura 4 - Documentos por País ou Território

Documents by country or territory

Compare the document counts for up to 15 countries/territories.



Fonte: Scopus (2025)

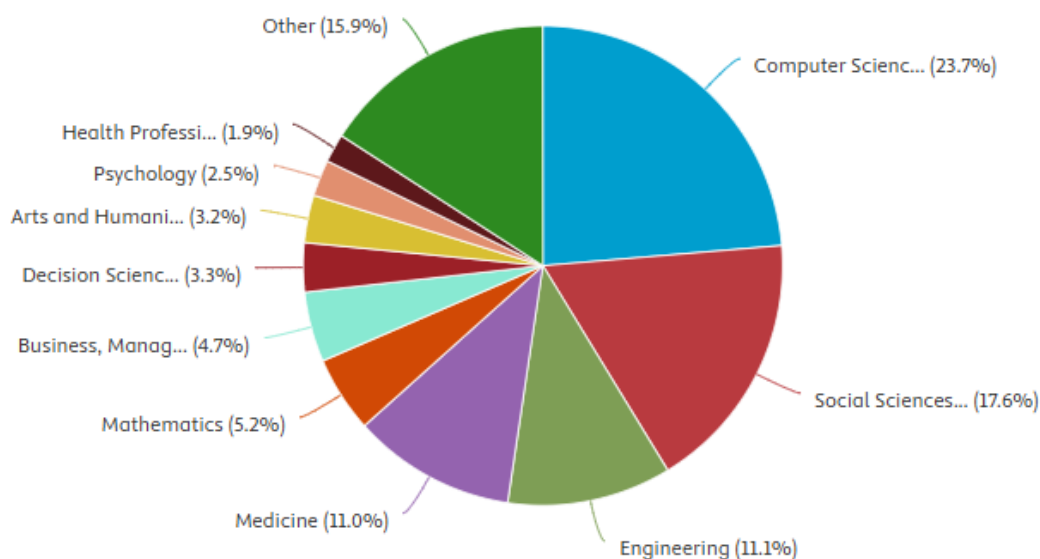
Em relação às áreas do conhecimento evidenciadas na Figura 5, os dados evidenciam a natureza transversal da Gen-AI. A maior parte das publicações está concentrada na área de Ciência da Computação, responsável por 23,7% dos estudos, seguida pelas Ciências Sociais (17,6%), Engenharia (11,1%), Medicina (11,0%), Matemática (5,2%) e Business & Management (4,7%).

Esses números reforçam que a IA generativa extrapola o campo técnico ao se expandir rapidamente por áreas aplicadas, educacionais, sociais e biomédicas, consolidando-se como tecnologia de impacto organizacional, cognitivo e social.

A presença expressiva de pesquisas em Saúde, Ciências Sociais e Administração aponta para a maturidade e a multidisciplinaridade da abordagem, que abrange desde aplicações clínicas e diagnósticas até estudos sobre comportamento, educação, governança e mercado.

Figura 5 - Documentos por Área

Documents by subject area



Fonte: Scopus (2025)

A partir desse panorama global, torna-se possível relacionar a produção científica às características específicas dos três modelos analisados: ChatGPT, DeepSeek e Gemini 2.5 Flash.

O ChatGPT aparece na literatura por seu equilíbrio entre profundidade analítica e interatividade, possui uma janela de contexto estendida em até 1 milhão de *tokens* e diversas funções, tais como, navegação, execução de código, geração de imagens e personalização via GPTs. A combinação de arquitetura densa com aprendizado por reforço com *feedback* humano (RLHF) que permite diálogos contextuais mais longos e coerentes, tornando-o ideal para aplicações em educação, programação e pesquisa. Seus índices superiores em desempenho matemático (88) e codificação (52) reforçam essa liderança no raciocínio lógico e computacional.

O DeepSeek-V3, mostra alta escalabilidade e eficiência, por utilizar o modelo *Mixture-of-Experts* (MoE) para ativar parâmetros de forma seletiva durante a inferência. Essa arquitetura possibilita baixo custo computacional e forte desempenho em raciocínio lógico e tarefas de domínio específico, como aplicações médicas e financeiras. Seus índices elevados de raciocínio lógico (90,8%) explicam o uso crescente em ambientes que demandam tomada de decisão estruturada e análise complexa.

O Gemini 2.5 Flash diferencia-se por sua multimodalidade, integrando texto, imagem, áudio e vídeo de maneira fluida. Essa capacidade o torna mais adequado para ambientes corporativos e de pesquisa, especialmente em tarefas que envolvem raciocínio complexo e grandes volumes de dados, com suporte direto às ferramentas do Google *Workspace*. Os resultados encontrados na literatura sobre multimodalidade e integração de dados convergem com o perfil do Gemini, reforçando sua adequação para aplicações empresariais e institucionais.

Do ponto de vista comparativo, observa-se que os três modelos não competem na mesma direção, mas se complementam dentro do ecossistema generativo: o ChatGPT domina tarefas linguísticas, analíticas e de codificação; o DeepSeek-V3 se destaca em eficiência e raciocínio especializado; e o Gemini lidera em multimodalidade e integração corporativa. Essa complementaridade é refletida na Tabela 1.

Tabela 1 - Comparação entre ChatGPT 5 (Plus), DeepSeek-V3 e Gemini (2,5 Flash)

Recurso	ChatGPT 5 (Plus)	DeepSeek-V3	Gemini (2.5 Flash)
Arquitetura	Estrutura de transformador densa, com todos os parâmetros habilitados durante a inferência. Garante desempenho robusto em uma ampla gama de atividades de processamento de linguagem natural (PLN)	Arquitetura <i>Mixture of Experts</i> (MoE), que aumenta a eficiência computacional e reduz significativamente os custos de treinamento	Arquitetura multimodal baseada em <i>transformers</i> capaz de processar e gerar texto, imagens, áudio e vídeo. Este projeto permite a integração suave de diversos formatos de dados



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Descrição	Inclui navegação na <i>web</i> , execução de código, geração de imagens e GPTs personalizados. Utiliza RLHF, para melhorar a coerência e a fluidez das respostas	Adota o mecanismo MoE, ativando apenas os parâmetros necessários para tarefas específicas, reduzindo custos e otimizando desempenho	Focado em processamento multimodal, integração com serviços Google e raciocínio avançado. Realiza inferência com múltiplas entradas e saídas
Vantagens	Versátil em tarefas de criação de conteúdo, assistência em codificação e análise de dados	Elevada eficiência e raciocínio em tarefas específicas de domínio, com desempenho escalável e <i>open source</i>	Forte raciocínio multimodal, integração nativa com o Google <i>Workspace</i> e alta velocidade de processamento
Conjunto de dados	40% Texto da <i>web</i> ; 20% Código; 20% Conversações; 10% Livros; 10% Enciclopédias multilíngues	30% Texto da <i>web</i> ; 25% Dados específicos de domínio (médico, jurídico, financeiro); 15% Artigos acadêmicos; 15% Código; 10% Diálogos; 5% Pares imagem-texto	30% Dados da <i>web</i> ; 20% Código; 20% Pares imagem-texto; 15% Dados de domínio; 10% Dados com curadoria; 5% Diversos
Índice de Matemática	88	82	87
Raciocínio Lógico	88,7%	90,8%	90,0%
Pontos Fortes	PNL versátil, forte em diálogo e codificação	Raciocínio específico de domínio, escalável e de código aberto	Criação avançada de conteúdo multimodal
Tokens	200 mil a 1 milhão	128 mil	1.048.576
Preço	assinaturas Plus US\$ 20/mês	Gratuito	Gratuito

Fonte: Adaptado de (Illieva, 2025; Rahman *et al.*, 2025; Deng *et al.*, 2025)

Em síntese, os resultados revelam que a Inteligência Artificial Generativa é um campo em rápida expansão, fortemente influenciado por avanços tecnológicos recentes e por investimentos globais em pesquisa. O crescimento expressivo das publicações, a diversidade de áreas envolvidas e a distribuição geográfica concentrada demonstram o potencial transformador da tecnologia.

A análise comparativa dos modelos ChatGPT 5 (Plus), DeepSeek-V3 e Gemini 2.5 Flash confirma que cada arquitetura responde a necessidades distintas, oferecendo soluções específicas que vão desde o processamento de linguagem natural até a multimodalidade integrada. A literatura também aponta desafios recorrentes, tais como viés algorítmico, privacidade, governança de dados e custo computacional, reforçando a necessidade de desenvolvimento responsável e ético dos modelos generativos.

5. CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo apresentar um panorama crítico da Inteligência Artificial (IA) e da Inteligência Artificial Generativa (Gen-AI), destacando sua evolução, maturidade tecnológica e os principais setores que vêm incorporando essas soluções. A revisão bibliográfica demonstrou que modelos generativos como ChatGPT, DeepSeek e Gemini representam um marco na automação cognitiva, combinando raciocínio avançado, aprendizado multimodal e especialização por domínio. Esses avanços têm impactado diretamente áreas como educação, saúde, finanças, operações e gestão do conhecimento.

A análise comparativa evidencia que o ChatGPT se consolida como referência em Processamento de Linguagem Natural, diálogo e codificação, potencializado por uma ampla janela de contexto e pelo uso de RLHF. O DeepSeek, por sua vez, prioriza eficiência computacional e escalabilidade, apresentando forte desempenho em raciocínio lógico e aplicações específicas de domínio graças à arquitetura MoE. Já o Gemini diferencia-se pela capacidade multimodal e pela integração nativa com ecossistemas corporativos, o que o torna especialmente adequado a ambientes de negócios e projetos de análise avançada de dados.

Embora representem um avanço significativo, esses modelos dependem diretamente da qualidade dos dados utilizados, tanto no treinamento quanto nas informações fornecidas pelo usuário. Consequentemente, os resultados gerados podem apresentar imprecisões e devem ser avaliados criticamente por especialistas, especialmente em contextos sensíveis. Além disso, persistem desafios relacionados à privacidade, ética, governança e propriedade de dados, apontando para a necessidade de protocolos robustos de uso responsável.

Para trabalhos futuros recomenda-se o aprofundamento de estudos empíricos voltados à comparação sistemática de diferentes modelos generativos, utilizando métricas padronizadas de desempenho, confiabilidade e impacto social. Investigações dessa natureza podem contribuir para consolidar diretrizes técnicas e éticas que sustentem o desenvolvimento seguro, transparente e responsável da Inteligência Artificial Generativa.

REFERÊNCIAS / REFERENCES / REFERENCIAS

CHOUDHARY, T. "Political Bias in Large Language Models: A Comparative Analysis of ChatGPT-4, Perplexity, Google Gemini, and Claude," in **IEEE Access**, vol. 13, pp. 11341-11379, 2025, doi: 10.1109/ACCESS.2024.3523764.

CONSTANTINIDES, PANOS, OLA HENFRIDSSON, AND GEOFFREY G. PARKER. 2018. "Introduction—Platforms and Infrastructures in the Digital Age." *Information Systems Research* **29**(2): 381–400. <https://doi.org.ez24.periodicos.capes.gov.br/10.1287/isre.2018.0794>.

DELL. 2021. **From Analytics to AI**. Disponível em: <<https://www.delltechnologies.com/asset/en-my/products/ready-solutions/briefs-summaries/from-analytics-to-ai-brochure.pdf>>. Acesso em: 25 de outubro de 2025.

DENG, L. "Artificial Intelligence in the Rising Wave of Deep Learning: The Historical Path and Future Outlook [Perspectives]," in **IEEE Signal Processing Magazine**, vol. 35, no. 1, pp. 180-177, Jan. 2018, doi: 10.1109/MSP.2017.2762725.

DENG *et al.*, "Exploring DeepSeek: A Survey on Advances, Applications, Challenges and Future Directions," in **IEEE/CAA Journal of Automatica Sinica**, vol. 12, no. 5, pp. 872-893, May 2025, doi: 10.1109/JAS.2025.125498.

DUAN, Y., EDWARDS, J. S., DWIVEDI, Y, W. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda, **International Journal of Information Management**, Volume 48, 2019, Pages 63-71, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>.

DWIVEDI, Y, K. Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy, **International Journal of Information Management**, Volume 57, 2021, 101994, ISSN 0268-4012, <https://doi.org/10.1016/j.ijinfomgt.2019.08.002>.

GAMA, F., MAGISTRETTI, S. Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. **Journal of product innovation management**, Volume 42, Issue1, <https://doi-org.ez24.periodicos.capes.gov.br/10.1111/jpim.12698>

GIL, A. C. **Como elaborar projetos de pesquisa**. [s.l.] Editora Atlas SA, 2002.

GRZYBOWSKI, A., *et al.*, A History of Artificial Intelligence, **Clinics in Dermatology**, Volume 42, Issue 3, 2024, Pages 221-229, ISSN 0738-081X, <https://doi.org/10.1016/j.clindermatol.2023.12.016>.

HAENLEIN, M., KAPLAN, A. (2019). A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence. **California Management Review**, 61(4), 5-14. <https://doi-org.ez24.periodicos.capes.gov.br/10.1177/0008125619864925>

ILIEVA, G. 2025. “Extension of Interval-Valued Hesitant Fermatean Fuzzy TOPSIS for Evaluating and Benchmarking of Generative AI Chatbots”. **Electronics (Switzerland)** 14 (3). <https://doi.org/10.3390/electronics14030555>.

JOVANOVIĆ, M., E CAMPBELL, M. "Generative Artificial Intelligence: Trends and Prospects," in **Computer**, vol. 55, no. 10, pp. 107-112, Oct. 2022, doi: 10.1109/MC.2022.3192720.

KALOTA F. A Primer on Generative Artificial Intelligence. **Education Sciences**. 2024; 14(2):172. <https://doi.org/10.3390/educsci14020172>

KOTECZKI, R., CSIKOR, D. & BALASSA, B.E. The role of generative AI in improving the sustainability and efficiency of HR recruitment process. **Discov Sustain** 6, 601 (2025). <https://doi-org.ez24.periodicos.capes.gov.br/10.1007/s43621-025-01484-3>.

MAGISTRETTI, STEFANO, LORENZO ARDITO, ANTONIO MESSENI PETRUZZELLI. 2021. “Framing the Microfoundations of Design Thinking as a Dynamic Capability for Innovation: Reconciling Theory and Practice.” **Journal of**



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

Product Innovation Management 38: 645–667. <https://doi-org.ez24.periodicos.capes.gov.br/10.1111/jpim.12586>.

RAHMAN, A. *et al.* Comparative analysis based on DeepSeek, ChatGPT, and Google Gemini: Features, techniques, performance, future prospects, **Systems and Soft Computing**, Volume 7, 2025, 200396, ISSN 2772-9419, <https://doi.org/10.1016/j.sasc.2025.200396>.

RAY, P. P. ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, **Internet of Things and Cyber-Physical Systems**, Volume 3, 2023, Pages 121-154, ISSN 2667-3452, <https://doi.org/10.1016/j.iotcps.2023.04.003>.

SAEIDNIA, H. R. (2023;), "Welcome to the Gemini era: Google DeepMind and the information industry". **Library Hi Tech News**, Vol. ahead-of-print No. ahead-of-print. <https://doi-org.ez24.periodicos.capes.gov.br/10.1108/LHTN-12-2023-0214>

SENGAR, S.S., HASAN, A.B., KUMAR, S. *et al.* Generative artificial intelligence: a systematic review and applications. **Multimed Tools Appl** 84, 23661–23700 (2025). <https://doi-org.ez24.periodicos.capes.gov.br/10.1007/s11042-024-20016-1>

SOHAIL, S. S. *et al.*, Decoding ChatGPT: A taxonomy of existing research, current challenges, and possible future directions, **Journal of King Saud University - Computer and Information Sciences**, Volume 35, Issue 8, 2023, 101675, ISSN 1319-1578, <https://doi.org/10.1016/j.jksuci.2023.101675>.

SRIVASTAVA, P.R., H.K. SINGH, S. SAKSHI, J.Z. ZHANG, Q. LI. 2024. "Identifying Alternative Options for Chatbots With Multi-Criteria Decision-Making: A Comparative Study". **Journal of Database Management** 35 (1).

TEMSAH, K. A., ALHASAN, I., ALTAMIMI, A. JAMAL, A. AL-EYADHY, K. H. MALKI, AND M.-H. TEMSAH, "DeepSeek in healthcare: Revealing opportunities and steering challenges of a new open-source artificial intelligence frontier," **Cureus**, vol. 17, p. 2, Feb. 2025.

TRIGKA, M., DRITSAS, E. "The Evolution of Generative AI: Trends and Applications," in **IEEE Access**, vol. 13, pp. 98504-98529, 2025, doi: 10.1109/ACCESS.2025.3574660.

URBINA, J. T., VU, P. D., NGUYEN, M. V. Disability Ethics and Education in the Age of Artificial Intelligence: Identifying Ability Bias in ChatGPT and Gemini, **Archives of Physical Medicine and Rehabilitation**, Volume 106, Issue 1, 2025, Pages 14-19, ISSN 0003-9993, <https://doi.org/10.1016/j.apmr.2024.08.014>.



Anais do Simpósio Acadêmico de Engenharia de Produção (SAEPRO) da EEL-USP

IX SAEPRO – 25 e 26 de novembro de 2025

XIONG, L., *et al.*, "DeepSeek: Paradigm Shifts and Technical Evolution in Large AI Models," in **IEEE/CAA Journal of Automatica Sinica**, vol. 12, no. 5, pp. 841-858, May 2025, doi: 10.1109/JAS.2025.125495.

WONG, J. Speculative futures of education: utopian and dystopian scenarios envisioned by Chatgpt, Gemini, and Deepseek. **Discov Educ** 4, 261 (2025). <https://doi-org.ez24.periodicos.capes.gov.br/10.1007/s44217-025-00692-3>

WU T. *et al.*, "A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development," in **IEEE/CAA Journal of Automatica Sinica**, vol. 10, no. 5, pp. 1122-1136, May 2023, doi: 10.1109/JAS.2023.123618.