

**\*Eduardo Motta Ferreira da Silva**

Graduando em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia de Mogi das Cruzes, Mogi das Cruzes, SP, Brasil.

ORCID: 0009-0003-6408-4071

E-mail:

edumotta2005@gmail.com

**\*\*Gastão Victor dos Santos Gonçalves**

Graduando em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia de Mogi das Cruzes, Mogi das Cruzes, SP, Brasil.

ORCID: 0009-0002-5941-8674

E-mail: gastaovictor@outlook.com

**\*\*\*Rita de Cássia Muniz de Lima**

Graduanda em Análise e Desenvolvimento de Sistemas pela Faculdade de Tecnologia de Mogi das Cruzes, Mogi das Cruzes, SP, Brasil.

ORCID: 0009-0003-2422-691X

E-mail: ritamuniz1995@gmail.com

## INTELIGÊNCIAS ARTIFICIAIS GENERATIVAS DE IMAGEM: estudo comparativo entre ferramentas de geração text-to-image

DOI: [doi.org/10.5281/zenodo.22712420](https://doi.org/10.5281/zenodo.22712420)

### RESUMO

Este trabalho analisa o comportamento de diferentes modelos de Inteligência Artificial generativa na criação de imagens a partir de prompts textuais. Três prompts distintos foram aplicados a três sistemas diferentes de geração de imagens, e os resultados foram comparados com base em critérios como fidelidade ao prompt, ausência de distorções, coerência visual, proporção e realismo. A análise identificou pontos fortes e limitações de cada modelo, oferecendo uma visão crítica sobre sua capacidade de traduzir linguagem natural em representações visuais precisas.

**Palavras-Chave:** IA; Geração de Imagem; Prompt.

### GENERATIVE IMAGE ARTIFICIAL INTELLIGENCES: a comparative study of text-to-image generation tools

### ABSTRACT

This work analyzes the behavior of different generative Artificial Intelligence models in creating images from textual prompts. Three distinct prompts were applied to three different image generation systems, and the results were compared based on criteria such as prompt fidelity, absence of distortions, visual coherence, proportion, and realism. The analysis identified strengths and limitations of each model, providing a critical perspective on their ability to translate natural language into precise visual representations.

**Keywords:** AI; Image Generation; Prompt.

### INTELIGENCIAS ARTIFICIALES GENERATIVAS DE IMÁGENES: estudio comparativo entre herramientas de generación text-to-image

### RESUMEN

Este trabajo analiza el comportamiento de diferentes modelos de Inteligencia Artificial generativa en la creación de imágenes a partir de prompts textuales. Tres prompts distintos fueron aplicados a tres sistemas diferentes de generación de imágenes, y los resultados fueron comparados con base en criterios como fidelidad al prompt, ausencia de distorsiones, coherencia visual, proporción y realismo. El análisis identificó fortalezas y limitaciones de cada modelo, ofreciendo una perspectiva crítica sobre su capacidad de traducir lenguaje natural en representaciones visuales precisas.

**Palabras clave:** IA; Generación de Imágenes; Prompt.

## INTRODUÇÃO

A tecnologia se encontra em vários lugares do cotidiano da sociedade atual. Com o avanço tecnológico, não tardou para que surgissem as Inteligências Artificiais (IAs), capazes de gerar respostas que simulam o comportamento humano. Entre suas diversas aplicações, há a geração de imagens através de um texto disponibilizado pelo usuário.

Com a chegada de diversas IAs capazes de gerar imagens disponibilizadas ao público, é de se esperar que cada uma possua uma forma diferente de gerar ou de compreender o que precisa ser gerado. Logo, essa situação traz o questionamento: como as Inteligências Artificiais diferem entre si na geração e compreensão de imagens através de textos?

Com isso, esse artigo busca trazer as diferenças e as semelhanças na forma de geração de imagem entre as principais IAs de geração de imagem, a fim de analisar o modo de pensamento artificial para moldagem da imagem e o quão condizente a intenção original do usuário se mostrou no resultado final.

## 2 FUNDAMENTAÇÃO TEÓRICA

A inteligência artificial (IA) pode ser explicada como uma tecnologia capaz de simular a inteligência humana, como o raciocínio e a criatividade (Stryker; Kavlakoglu, 2024). Apesar de ter ganhado visibilidade recente com IAs generativas como o ChatGPT (2022), seu histórico é mais antigo: segundo Teixeira (2019), começou a surgir de forma prática "após a Segunda Guerra Mundial (de 1945 em diante)". O termo "Inteligência Artificial", porém, só foi utilizado em 1956 pelo cientista John McCarthy, mesmo ano da criação do primeiro software de IA em execução. A partir dos anos 2000, com a era do big data e da computação em nuvem, a tecnologia expandiu-se, permitindo gerenciar volumes crescentes de dados para treinar modelos (Stryker; Kavlakoglu, 2024), até chegar às versões atuais com pensamento profundo e criação de imagens.

Para gerar uma imagem, a IA precisa de uma técnica que interprete e construa a forma desejada. Entre as mais conhecidas estão GANs, CNNs e RNNs (Lima, 2024 apud Russel; Norvig, 2013). A Rede Adversária Generativa (GAN) funciona por meio de redes geradoras e discriminadoras: a geradora cria um novo objeto a partir dos dados recebidos, enquanto a discriminadora separa resultados realistas dos sintéticos, aperfeiçoando o aprendizado do gerador (Varughese, [s.d]). A Rede Neural Convolucional (CNN) atua por camadas, convolucional (cálculo e identificação de padrões), de agrupamento (redução de resolução e complexidade) e totalmente conectada (classificação) (Lima, 2024). Já a Rede Neural Recorrente (RNN) é treinada com dados sequenciais, buscando informações anteriores para formular respostas com coerência; atualmente está em desuso relativo, mas ainda é utilizada (Stryker, [s.d]).

Recentemente surgiram modelos potentes de geração de imagem a partir de texto (text-to-

image), baseados em difusão e condicionamento semântico. Três se destacam: o DALL·E (OpenAI), ferramenta proprietária com alta fidelidade ao prompt e boa performance em cenas complexas, embora limitada no controle fino de detalhes; o MidJourney, voltado à produção artística e forte apelo estético, que por vezes "interpreta artisticamente" o prompt em detrimento do realismo; e o Stable Diffusion, modelo open source com alto grau de personalização (seeds, negative prompts, fine-tuning), porém mais exigente em conhecimento técnico e recursos. Comparativos recentes indicam que nenhuma ferramenta domina todos os aspectos: o DALL·E tende a ser melhor em fidelidade, o MidJourney em estética e o Stable Diffusion em customização e custo.

A comparação justa dessas ferramentas pode ser organizada em três categorias de critérios, sintetizadas na Tabela 1.

| <b>Categoria</b>        | <b>Exemplos de critérios</b>                                                                                               |
|-------------------------|----------------------------------------------------------------------------------------------------------------------------|
| Automáticos / objetivos | Métricas de similaridade (FID, CLIP); fidelidade ao prompt (TIFA); generalizabilidade (VLEU).                              |
| Humanos / subjetivos    | Estética; clareza e coerência dos elementos; coerência visual (proporção, perspectiva, luz, artefatos); preferência geral. |
| Práticos / usabilidade  | Facilidade de uso; latência; custo; flexibilidade de parâmetros; licenciamento; escalabilidade.                            |

Fonte: Os autores.

É importante reconhecer trade-offs: um modelo com alta fidelidade automática pode falhar em estética ou exigir muitos recursos. A avaliação humana costuma complementar as métricas automáticas para capturar nuances que algoritmos não percebem.

### **3 MATERIAIS E MÉTODOS**

Foram utilizados como base teórica artigos científicos, livros e publicações especializadas que abordam os temas de inteligência artificial generativa, redes neurais difusionais e modelos de aprendizado profundo aplicados à geração de imagens. As fontes foram obtidas por meio de plataformas acadêmicas como Google Scholar, além de bibliotecas digitais públicas e universitárias.

Embora modelos como DALL·E, MidJourney e Stable Diffusion sejam amplamente referenciados na literatura especializada, este estudo optou por avaliar ferramentas de acesso gratuito e irrestrito ao público geral, sendo selecionadas três plataformas de geração de imagem por IA — ChatGPT (OpenAI), Copilot (Microsoft) e Gemini (Google) — por considerá-las mais representativas do uso cotidiano, possuírem planos gratuitos e serem acessíveis para fins de comparação prática.

O experimento consistiu na formulação e envio de um conjunto padronizado de prompts (descrições textuais) a cada uma das inteligências artificiais selecionadas. Para a análise das imagens

geradas, foram estabelecidos três parâmetros de avaliação:

Parâmetro 1 – Fidelidade ao prompt: verificar se todos os elementos descritos no prompt estão presentes na imagem gerada.

Parâmetro 2 – Ausência de distorções: verificar se objetos e personagens não apresentam deformações em relação ao padrão real (como uma moto com apenas uma roda ou um ser humano com três braços).

Parâmetro 3 – Consistência de estilo: verificar se a IA manteve o estilo solicitado sem mesclar estilos distintos (por exemplo, gerar em cartoon uma imagem descrita para ser realista).

Os prompts foram definidos com base em critérios de neutralidade, clareza e diversidade temática, de modo a avaliar aspectos distintos das ferramentas, como realismo, coerência visual e fidelidade à descrição.

Cada IA recebeu os mesmos prompts e produziu imagens sob as mesmas condições de contexto, buscando assegurar comparabilidade entre os resultados. As imagens geradas foram analisadas qualitativamente por meio de critérios visuais, considerando elementos como composição, detalhamento, proporção e uso de cores.

A definição dos prompts seguiu uma amostragem intencional, própria de estudos qualitativos exploratórios, voltada a abranger cenários com diferentes níveis de complexidade e natureza, e não a representar estatisticamente o universo de gerações possíveis. Estudos comparativos semelhantes seguem a mesma lógica: Muhr et al. (2024) compararam três modelos de geração de imagens a partir de dez prompts para identificar e classificar diferenças e erros entre as ferramentas. Nessa direção, foram selecionados três prompts com propósitos distintos: o primeiro voltado à representação de múltiplos elementos, relações espaciais e movimento; o segundo, à interpretação de um contexto histórico-cultural associado a um estilo artístico específico; e o terceiro, à composição cinematográfica, iluminação e características visuais. Essa diversidade permitiu avaliar fidelidade ao prompt, ausência de distorções e consistência de estilo em cenários hipotéticos variados.

Os três prompts utilizados foram os seguintes:

Prompt 1 - Aja como um pintor profissional e gere uma imagem de um homem adulto com estatura baixa utilizando roupa formal andando de moto em alta velocidade por uma avenida movimentada de uma cidade de interior com o pôr do sol ao horizonte com um céu alaranjado. Ele está utilizando fones de ouvido e ao fundo da imagem é possível ver um cachorro caramelo lambendo um sorvete de morango que caiu no chão.

Prompt 2 - Baseado no barroco português, faça uma releitura em imagem de como foi realizado o plebiscito de 1993, onde a população brasileira votou entre a volta da república democrática ou da monarquia constitucional ou parlamentar, utilizando elementos da época.

Prompt 3 - Faça um retrato cinematográfico em estilo neo-noir de uma mulher jovem,

fotógrafa amadora, com cabelo curto azul-escuro, vestindo jaqueta de couro preta e tênis brancos, correndo numa estação de metrô abandonada com vagões cobertos de grafite, luzes de néon roxo refletindo no chão molhado, alta resolução, profundidade de campo rasa.

Os resultados obtidos foram então comparados e discutidos à luz da literatura, com o objetivo de identificar diferenças de estilo, capacidade interpretativa e limitações técnicas entre as inteligências artificiais analisadas.

Para a geração das imagens, foi necessário realizar login nas plataformas das ferramentas utilizadas. Com o objetivo de evitar qualquer influência do histórico de uso, como conversas e gerações de imagens anteriores, foi criado um endereço de e-mail temporário e um cadastro novo em cada plataforma. Adicionalmente, cada imagem foi gerada em uma sessão de conversa independente, garantindo que os resultados não fossem afetados por interações prévias.

#### **4 RESULTADOS E DISCUSSÃO**

A primeira ferramenta analisada (IA1) foi o ChatGPT, desenvolvido pela OpenAI, utilizado em seu plano gratuito, na versão GPT-5.1 disponível à época da coleta.

A imagem 1 demonstrou boa fidelidade ao prompt, com a maioria dos elementos descritos presentes e sem distorções relevantes em relação ao padrão visual esperado. No entanto, foram identificadas duas divergências: o cachorro foi posicionado no primeiro plano da imagem, contrariando a instrução de que deveria aparecer ao fundo; e o motociclista não transmite visualmente a sensação de alta velocidade descrita no prompt.

A imagem 2 mostrou-se coerente com o prompt, apresentando a palavra "plebiscito" grafada corretamente e contrastes de luz e sombra característicos do estilo barroco. Contudo, foram identificadas distorções anatômicas sutis, como a presença de um sexto dedo na mão direita do homem posicionado ao lado da urna de votação, além de uma leve deformação no pote de tinta nanquim. Merece destaque também a ausência de elementos que remetessem ao contexto brasileiro do plebiscito, o que compromete parcialmente a fidelidade à proposta original.

A imagem 3 foi a que mais fielmente reproduziu os elementos do prompt entre as geradas pela IA1, respeitando os contrastes, as cores e a atmosfera descritos. A única inconsistência identificada foi uma distorção anatômica: a personagem apresenta um sexto dedo na mão direita, com a qual segura a câmera.

**Imagem 1 - Imagem gerada pela IA1 baseada no prompt 1.**



Fonte: ChatGPT.

**Imagem 2 - Imagem gerada pela IA1 baseada no prompt 2.**



Fonte: ChatGPT.

Imagem 3 - Imagem gerada pela IA1 baseada no prompt 3.



Fonte: ChatGPT.

Em resumo, houve os seguintes resultados:

Tabela 2: Análise de parâmetros e prompts da IA1.

|                    | Prompt 1                                          | Prompt 2                                          | Prompt 3                              |
|--------------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------|
| <b>Parâmetro 1</b> | Maioria dos elementos descritos foram respeitados | Maioria dos elementos descritos foram respeitados | Elementos descritos foram respeitados |
| <b>Parâmetro 2</b> | Objetos e personagens condizentes com a realidade | Objetos e personagens levemente distorcidos       | Personagem levemente distorcido       |
| <b>Parâmetro 3</b> | Manteve consistência em toda a imagem             | Manteve consistência em toda a imagem             | Manteve consistência em toda a imagem |

Fonte: os autores.

A segunda ferramenta analisada (IA2) foi o Copilot, desenvolvido pela Microsoft, igualmente utilizado em seu plano gratuito.

A imagem 4 mostrou-se quase integralmente coerente com o prompt, mantendo consistência no estilo de pintura e nas proporções dos objetos e personagens. A principal divergência identificada foi a ausência de movimento na avenida: ao contrário do descrito, a via foi representada vazia. Observou-se também uma descontinuidade no traçado da rua, porém, por existirem configurações viárias semelhantes em diferentes partes do mundo, esse elemento não foi classificado como distorção.

A imagem 5 apresentou riqueza de detalhes, com a palavra "plebiscito" grafada corretamente e a inclusão de elementos que contextualizam historicamente a cena, como a bandeira brasileira atual, o ano do plebiscito e o brasão da família imperial de Orleans e Bragança, referências que permitem ao observador identificar imediatamente o evento retratado. Os contrastes de luz e sombra e os

ornamentos são condizentes com o estilo barroco solicitado, e, embora as vestimentas difiram das representadas na imagem 2, também são compatíveis com o período histórico evocado.

A imagem 6 foi a que obteve melhor desempenho geral entre todas as analisadas, respeitando integralmente os elementos do prompt, mantendo proporções adequadas e preservando a coerência de estilo ao longo de toda a composição.

**Imagem 4 - Imagem gerada pela IA2 baseada no prompt 1.**



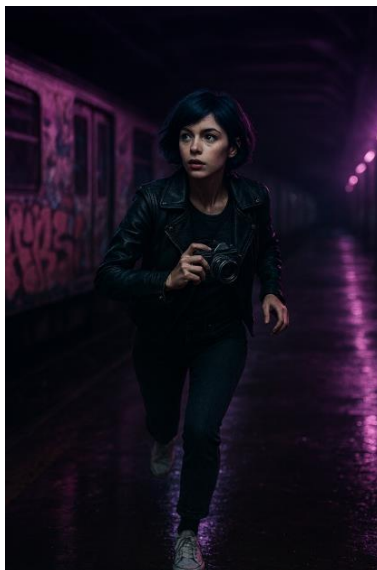
Fonte: Copilot.

**Imagem 5 - Imagem gerada por IA2 baseada no prompt 2.**



Fonte: Copilot.

Imagem 6 - Imagem gerada por IA2 baseada no prompt 3.



Fonte: Copilot.

Tabela 3: Análise de parâmetros e prompts da IA2.

|             | Prompt 1                                          | Prompt 2                                          | Prompt 3                                          |
|-------------|---------------------------------------------------|---------------------------------------------------|---------------------------------------------------|
| Parâmetro 1 | Maioria dos elementos descritos foram respeitados | Elementos descritos foram respeitados             | Elementos descritos foram respeitados             |
| Parâmetro 2 | Objetos e personagens condizentes com a realidade | Objetos e personagens condizentes com a realidade | Objetos e personagens condizentes com a realidade |
| Parâmetro 3 | Manteve consistência em toda a imagem             | Manteve consistência em toda a imagem             | Manteve a consistência em toda a imagem           |

Fonte: os autores.

A terceira ferramenta analisada (IA3) foi o Gemini, desenvolvido pela Google, utilizado em sua versão gratuita.

**Imagem 7 - Imagem gerada pela IA3 baseada no prompt 1.**



Fonte: Gemini.

A imagem 7 foi a que apresentou maior número de divergências em relação ao Prompt 1. Entre os problemas identificados estão: o motociclista aparece trafegando pela calçada em vez da avenida; o personagem não transmite estatura baixa nem sensação de alta velocidade; pedestres aparecem caminhando no meio da via; e o cachorro está posicionado no primeiro plano, contrariando a descrição. Ainda assim, a imagem manteve proporções condizentes com a realidade e preservou consistência de estilo ao longo da composição, sendo esses os únicos parâmetros satisfeitos.

**Imagem 8 - Imagem gerada por IA3 baseada no prompt 2.**



Fonte: Gemini.

A imagem 8 também apresentou desempenho insatisfatório em relação ao prompt. Não há elementos que remetam ao contexto do plebiscito brasileiro de 1993, e o estilo adotado pela ferramenta aproxima-se do Renascimento, com traços de simetria e arquitetura greco-romana, em vez do barroco português solicitado. Foram identificadas ainda distorções anatômicas, como personagens com quatro dedos nas mãos e feições faciais deformadas, caracterizando falha tanto no Parâmetro 1 quanto no Parâmetro 2.

**Imagem 9 - Imagem gerada por IA3 baseada no prompt 3.**



Fonte: Gemini.

A imagem 9 foi a que melhor atendeu ao prompt entre as três geradas pelo Gemini. Os elementos descritos foram representados com boa fidelidade, com a única divergência sendo a predominância de luzes azuis e rosas em vez das luzes roxas especificadas. A personagem e os demais elementos da composição não apresentaram distorções anatômicas ou visuais perceptíveis, satisfazendo plenamente os Parâmetros 2 e 3.

**Tabela 4: Análise de parâmetros e prompts da IA3**

|                    | <b>Prompt 1</b>                              | <b>Prompt 2</b>                              | <b>Prompt 3</b>                                   |
|--------------------|----------------------------------------------|----------------------------------------------|---------------------------------------------------|
| <b>Parâmetro 1</b> | Poucos elementos descritos foram respeitados | Poucos elementos descritos foram respeitados | Maioria dos elementos foram respeitados           |
| <b>Parâmetro 2</b> | Objetos e personagens levemente distorcidos  | Objetos e personagens levemente distorcidos  | Objetos e personagens condizentes com a realidade |
| <b>Parâmetro 3</b> | Manteve consistência em toda a imagem        | Não manteve consistência na imagem           | Manteve consistência em toda a imagem             |

Fonte: os autores.

## 5 CONCLUSÃO E PERSPECTIVAS

Este estudo comparou o desempenho de três ferramentas de geração de imagens por inteligência artificial, ChatGPT (OpenAI), Copilot (Microsoft) e Gemini (Google), a partir de três prompts padronizados, avaliados segundo critérios de fidelidade ao prompt, ausência de distorções e consistência de estilo.

Os resultados evidenciaram diferenças significativas entre as ferramentas. O Copilot (IA2) apresentou o desempenho mais consistente entre as três IAs analisadas, respeitando integralmente os elementos descritos nos três prompts e mantendo proporções e estilo sem distorções perceptíveis, com destaque para a imagem gerada a partir do Prompt 3, considerada de execução próxima ao ideal. O ChatGPT (IA1) demonstrou boa fidelidade geral aos prompts, com pequenas falhas pontuais, como o posicionamento incorreto do cachorro no Prompt 1 e distorções anatômicas sutis nas imagens 2 e 3, mas manteve consistência de estilo em todos os casos. O Gemini (IA3) foi o que apresentou maior dificuldade em traduzir os prompts com precisão, especialmente nos Prompts 1 e 2, nos quais poucos elementos foram respeitados e houve inconsistência de estilo no segundo caso, embora tenha obtido resultado satisfatório no Prompt 3.

Esses achados são coerentes com o que a literatura aponta: nenhuma ferramenta domina todos os aspectos simultaneamente, e o desempenho varia conforme a natureza da tarefa. Prompts com maior carga descritiva e elementos culturais específicos, como o Prompt 2, revelaram as maiores discrepâncias entre as ferramentas, sugerindo que a capacidade de interpretação contextual ainda representa um desafio relevante para os modelos atuais.

Como limitações, este estudo utilizou exclusivamente as versões gratuitas das ferramentas, o que pode restringir a capacidade de geração em relação às versões pagas. Além disso, a avaliação foi conduzida de forma qualitativa e por um número reduzido de avaliadores, sem o uso de métricas automáticas como FID ou CLIP, o que limita a generalização dos resultados. Investigações futuras poderiam ampliar o conjunto de prompts, incluir versões pagas das ferramentas, incorporar métricas objetivas de avaliação e realizar validação com júri mais amplo.

Por fim, recomenda-se que usuários que busquem alta fidelidade descritiva priorizem o Copilot ou o ChatGPT em detrimento do Gemini para tarefas com prompts detalhados, e que, independentemente da ferramenta escolhida, as imagens geradas sejam submetidas à revisão humana antes de qualquer uso profissional ou acadêmico, dado que todas as ferramentas analisadas produziram ao menos algum tipo de distorção ou omissão.

Os autores declaram que as ferramentas de IA generativa (ChatGPT, Copilot e Gemini) foram utilizadas exclusivamente como objeto experimental de análise neste estudo, não tendo sido empregadas na redação, revisão textual ou geração de conteúdo do manuscrito.

## REFERÊNCIAS

CAO, J. et al. **VLEU: a Method for Automatic Evaluation for Generalizability of Text-to-Image Models**. Disponível em: <<https://arxiv.org/abs/2409.14704>>. Acesso em: 27 set. 2025.

GOOGLE. **Gemini**. Disponível em: <<https://gemini.google.com/app>>.

HU, Y. et al. **TIFA: Accurate and Interpretable Text-to-Image Faithfulness Evaluation with Question Answering**. Disponível em: <<https://arxiv.org/abs/2303.11897>>.

LIMA, Marcos Vinícius de Moura. **Imagens Geradas Por Inteligência Artificial: Técnicas e Exemplos**. 2024. 44f. **Trabalho de Conclusão de Curso - Graduação** (Redes de Computadores) - Universidade Estadual de Goiás, Pires do Rio.

MICROSOFT COPILOT. **Microsoft Copilot**. Disponível em: <<https://copilot.microsoft.com/>>. OPENAI. **ChatGPT**. Disponível em: <<https://chatgpt.com/>>.

SCHWARTZ, R. et al. The Ethics of AI-Generated Art: Ownership and Authorship in the Age of Generative Models. **AI & Society**, Springer, p. 1-12, London, UK, 2023.

STRYKER, C. **O que é uma rede neural recorrente (RNN)?** Disponível em: <<https://www.ibm.com/br-pt/think/topics/recurrent-neural-networks>>. Acesso em: 5 out. 2025.

STRYKER, C.; KAVLAKOGLU, E. **O que é inteligência artificial (IA)?** Disponível em: <<https://www.ibm.com/br-pt/think/topics/artificial-intelligence>>. Acesso em: 25 set. 2025.

TEIXEIRA, João. **O que é inteligência artificial**. E-galáxia, 2019.

VARUGHESE, J. **Generative adversarial networks**. Disponível em: <<https://www.ibm.com/think/topics/generative-adversarial-networks>>.

HESSEL, Jack et al. **CLIPScore: A Reference-free Evaluation Metric for Image Captioning**. 1 jan. 2021. Disponível em: <<https://aclanthology.org/2021.emnlp-main.595/>>

JAYASUMANA, Sadeep et al. **Rethinking FID: Towards a Better Evaluation Metric for Image Generation** **arXiv.org**. [S.l.: S.n.]. Disponível em: <<https://arxiv.org/abs/2401.09603>>. Acesso em: 8 ago 2026.

WILES, Olivia et al. **Revisiting Text-to-Image Evaluation with Gecko: On Metrics, Prompts, and Human Ratings**. Disponível em: <<https://arxiv.org/abs/2404.16820>>. Acesso em: 8 ago. 2026.

MUHR, Paula et al. Evaluating Text-to-Image Generated Photorealistic Images of Human Anatomy. **Cureus**, 21 nov. 2024. Disponível em: <<https://www.cureus.com/articles/308456>>