

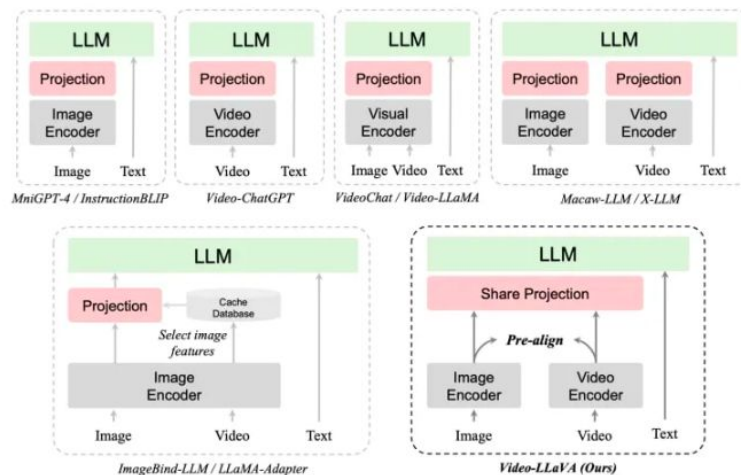
## **ETAPA 9**

# **Geração de imagens**

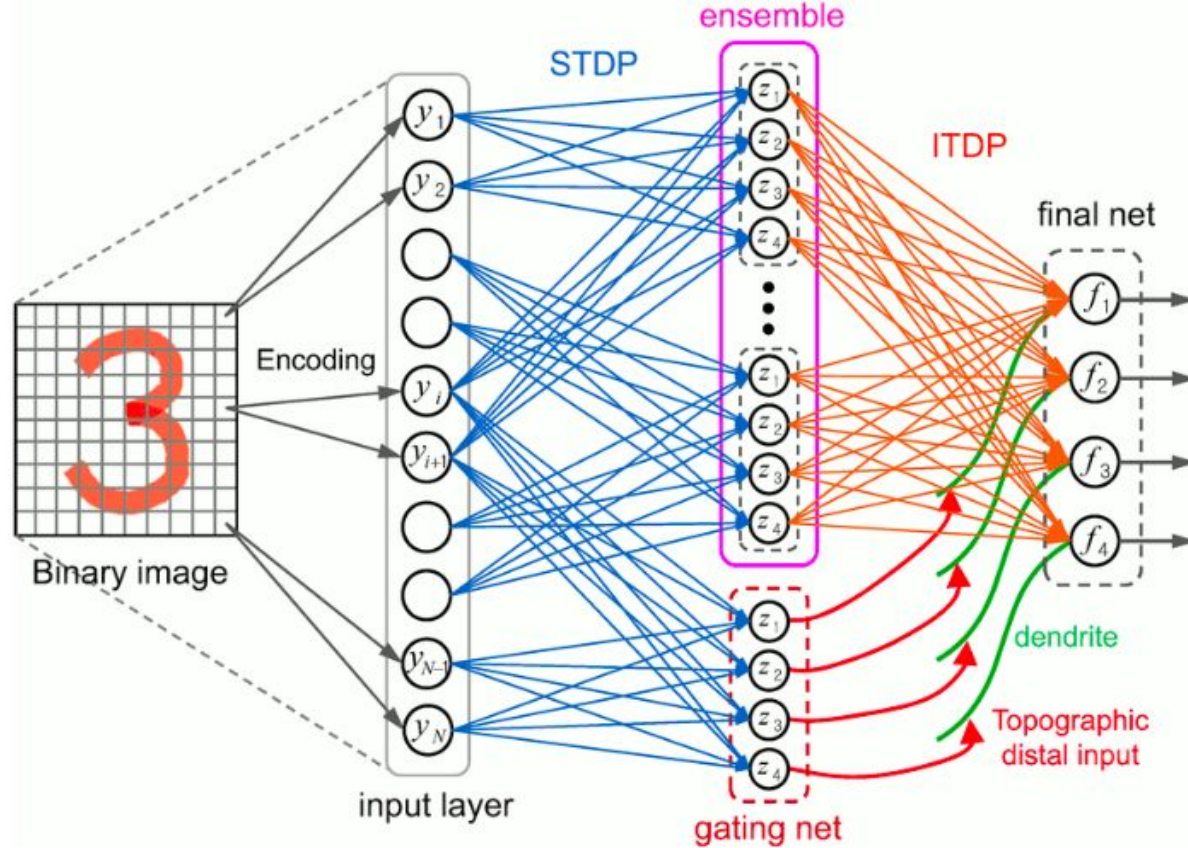
# IA Generativa com Imagens

Assim como LLMs, podemos aplicar as arquiteturas e conceitos de processamento de texto em imagens. Dentre os principais modelos, dois destacam-se por suas abordagens distintas no campo da geração de imagens.

- **Modelos de difusão** trabalham a partir de ruídos aleatórios, reconstruindo progressivamente uma imagem. Reconhecida por sua alta qualidade em detalhes e suavidade (arte e fotografia)
- **Transformadores** utilizam 'tokens' para prever a próxima parte da imagem com base no contexto fornecido. Destacam-se em tarefas que exigem consistência semântica (legendas visuais).



**Figure 1. Comparing Different LVLm Paradigms.** Video-LLaVA aligns images and videos before projection, allowing LLM to learn from a unified visual representation and endowing LLM with the ability to comprehend both images and videos simultaneously.

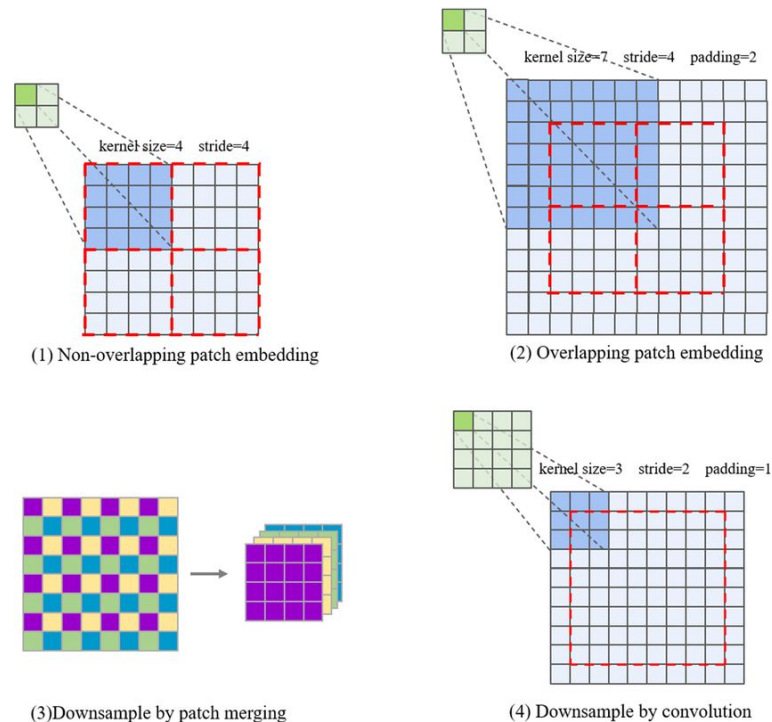


SEM-ITDP ensemble network architecture. The STDP connections, which projects from the selected input neurons to each WTA circuit, together with the WTA circuits constitute the SEM ensemble. The ITDP connections have the same connectivity as the logical ITDP model. All of the ensemble, gating and final output networks use the same SEM circuit model.

# Patches de Imagens

**Patches** são a unidade de processamento em modelos de visão, como o Vision Transformer (ViT). Uma imagem é dividida em pequenos blocos de pixels (patches), que são processados de forma semelhante a tokens de texto.

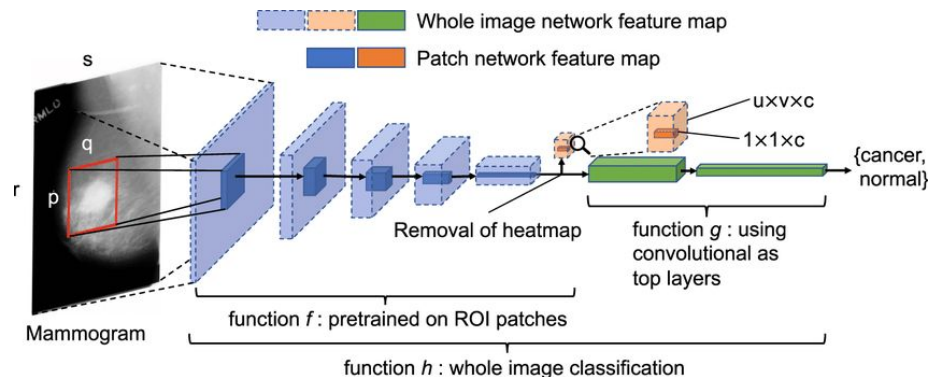
**Divisão de Imagens:** Por exemplo, uma imagem de 256x256 pixels pode ser dividida em patches de 16x16 pixels, resultando em 256 patches.



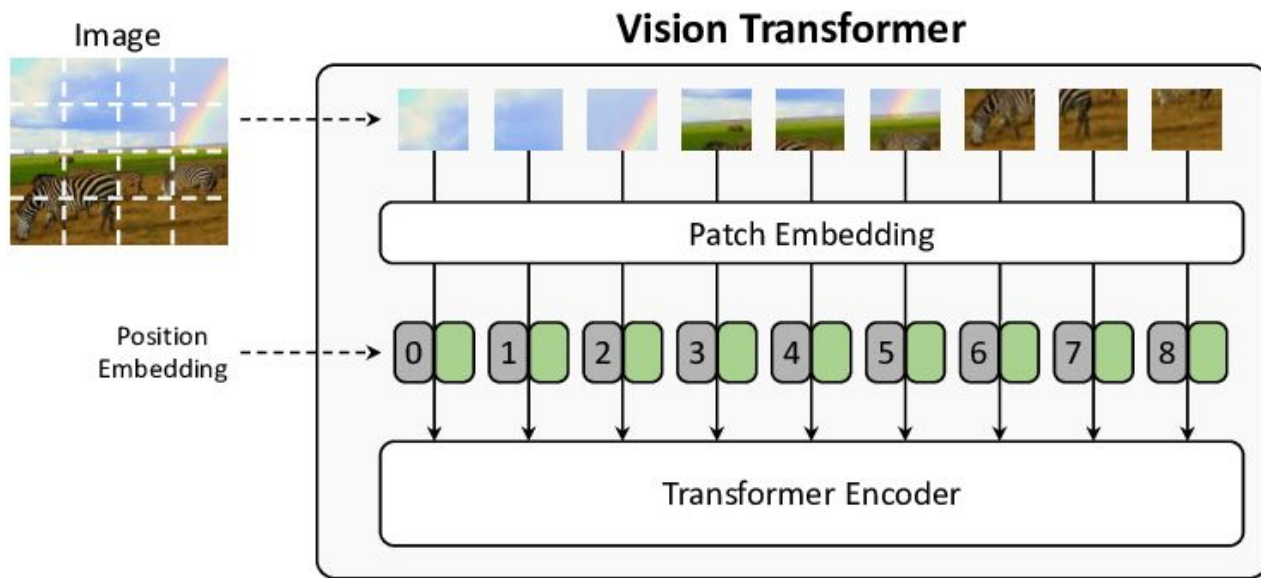
# Patches de Imagens

**Semelhança com Tokens:** assim como tokens capturam o contexto no texto, patches capturam relações locais e globais dentro de uma imagem. O modelo aprende como diferentes partes da imagem se relacionam, melhorando a compreensão de estruturas visuais.

**Codificação Posicional:** como patches não têm uma ordem natural como tokens de texto, cada patch recebe uma codificação posicional para manter a estrutura espacial da imagem.



# Patches de Imagens com Transformers





# Desenvolver prompts eficazes para gerar imagens no Stable Diffusion (SD)

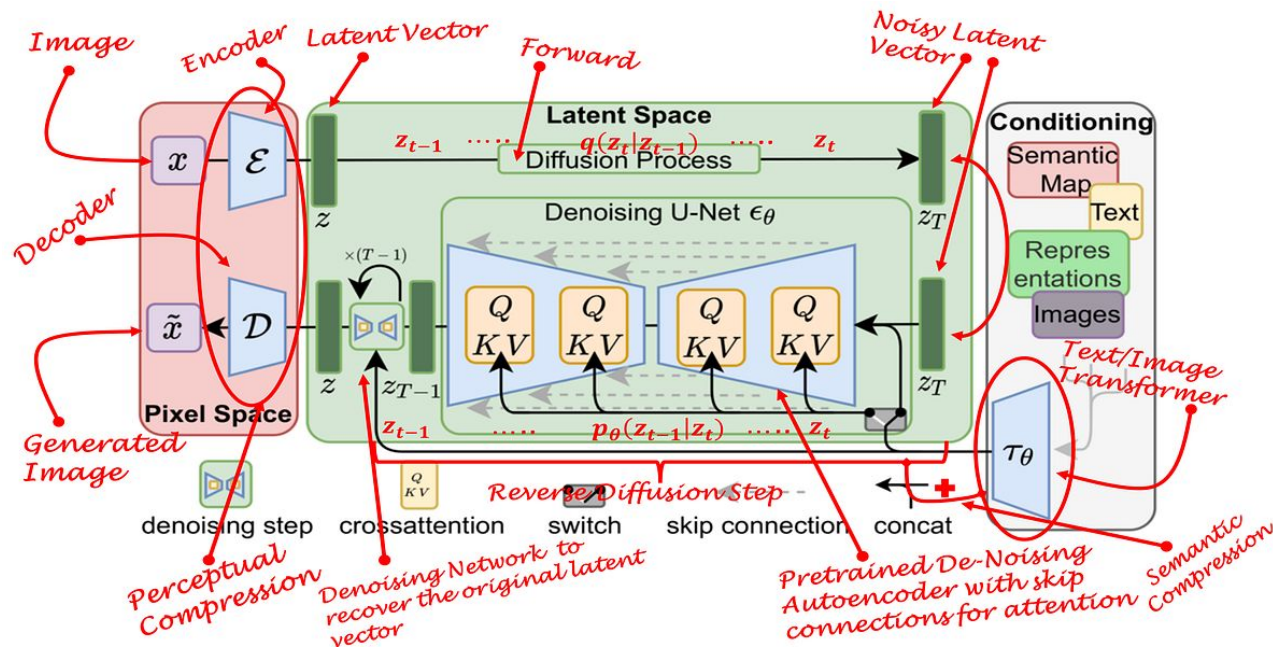
O Stable Diffusion é um modelo de geração de imagens open-source baseado em processos de difusão latente, que transforma ruído aleatório em imagens claras e detalhadas. É treinado para reconstruir imagens progressivamente, refinando detalhes a partir de ruído até formar o resultado final.

**Open-source:** Disponível para personalização e adaptação a aplicações específicas.

**Desempenho eficiente:** Capaz de gerar imagens de alta qualidade em hardware de consumo, como GPUs modernas.

**Modularidade:** Suporte para integração com extensões, como ControlNet e LoRA, que expandem suas funcionalidades.

# Patches de Imagens com Difusão





# Patches de Imagens com Difusão

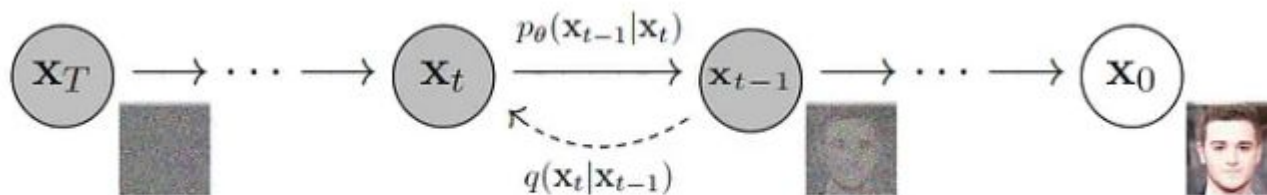
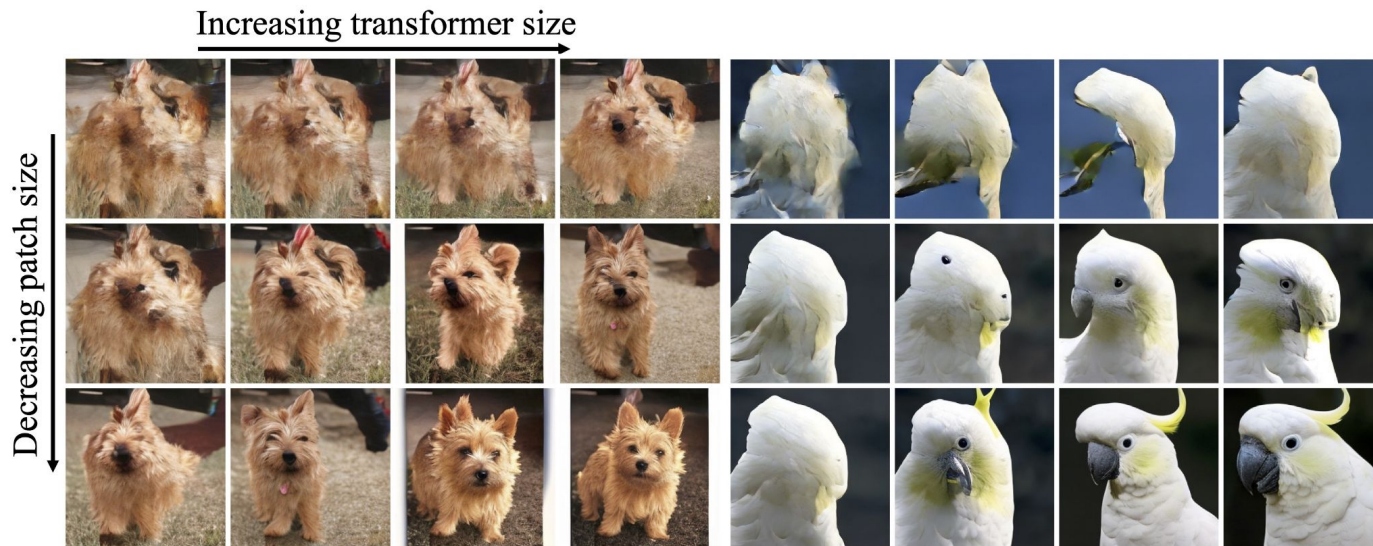


Figure 2: A typical Diffusion Model Process (Source: [1])

# Patches de Imagens com Transformers



# Desenvolver prompts eficazes para gerar imagens no Stable Diffusion (SD)

Estratégias para controle fino de detalhes:

**Combinação de estilos:** Misture referências como "pintura a óleo" com "realismo fotográfico".

**Especificidade:** Inclua descrições claras de luz, textura e composição.

**Ajustes técnicos:** Parâmetros como CFG Scale (força do prompt) e steps (detalhamento).

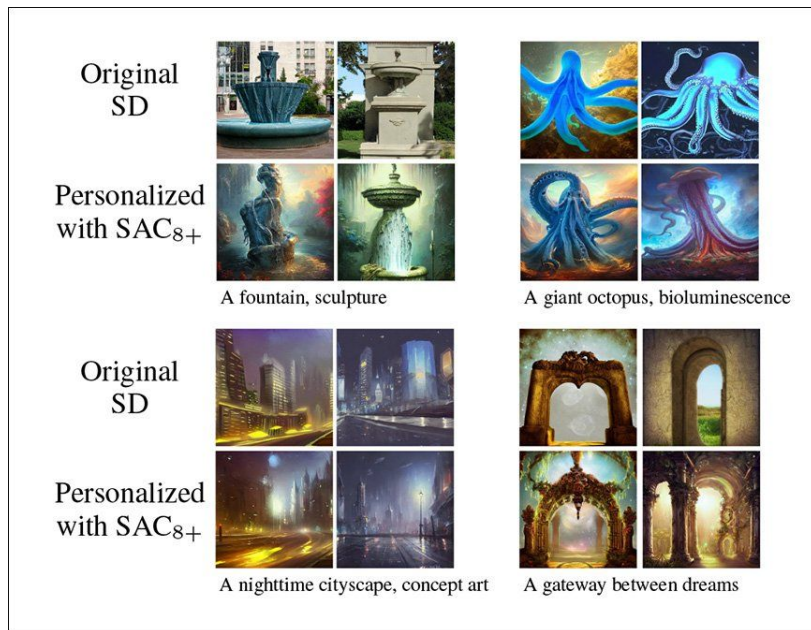


# Desenvolver prompts eficazes para gerar imagens no Stable Diffusion (SD)

A geração guiada no Stable Diffusion permite usar imagens de referência ou mapas predefinidos para influenciar o resultado, garantindo maior consistência e controle artístico.

## Métodos de geração guiada

- **Imagem como referência:** O modelo utiliza características de uma imagem base para criar variações estilizadas.
- **Mapas de guia:** Esboços ou mapas de profundidade orientam a composição final.





# Desenvolver prompts eficazes para gerar imagens no Stable Diffusion (SD)

## Parâmetros

### **Prompt**

Entrada de texto que descreve a imagem desejada. Seja específico para obter resultados detalhados. Inclua adjetivos, objetos, estilos e contexto (ex.: "Uma paisagem futurista com luzes de neon à noite").

### **Prompt Negativo**

Entrada de texto que especifica elementos a serem evitados na imagem gerada. Use para remover artefatos ou características indesejadas (ex.: "borrado, distorcido, texto").

### **Strength (Força) (img2img)**

Controla o quanto a saída adere à imagem de entrada. Intervalo: 0.1 (mínima alteração) a 1.0 (máxima alteração). Use valores baixos para preservar os detalhes originais. Valores altos para transformações criativas.

### **Guidance Scale (Escala de Orientação)**

Ajusta o peso da influência do prompt na geração. Tipicamente 7-15. Valores baixos produzem saídas mais criativas. Valores altos geram imagens mais alinhadas ao prompt.

### **Máscara e Imagem (inpainting)**

Uma imagem binária indicando as áreas a editar (branco) e preservar (preto).

### **Sampling Steps (Passos de Amostragem)**

Número de iterações para refinar a imagem, comumente 20-50. Valores mais altos melhoram a qualidade, mas aumentam o tempo de processamento.

### **Scheduler/Sampler (Agendador/Amostrador)**

Define a estratégia de remoção de ruído. Exemplos: DDIM, Euler, LMS, etc. Experimente diferentes amostradores para obter resultados variados.



# Construir Aplicações com Geração de Imagens com IA

- **Imagens de Personagens:** utilizar LLM para combinar as imagens dos personagens com as suas sumarizações.
- **Personagens de Realidades Alternativas:** utilizar IA para combinar as imagens dos personagens com outras imagens.

**Geração de imagens**  
Exercícios de geração de imagens testando o efeito de cada parâmetro.

- Prompt Negativo
- Guidance Scale
- Sampling Steps
- Scheduler/Sampler

**Adaptação de imagens**  
Exercícios usando imagens, ou rascunhos, como base para a geração de novas imagens:

- Imagem inicial
- Inpainting / Outpainting

# DALL-E

O DALL-E (OpenAI) é um modelo de geração de imagens que transforma descrições textuais em representações visuais, utilizando uma combinação de técnicas de aprendizado profundo (transformadores) para criar imagens originais baseadas em prompts fornecidos pelo usuário.

- **Geração criativa:** Capaz de criar imagens artísticas ou conceituais com base em descrições.
- **Edição de imagens:** Alterações detalhadas em partes específicas de uma imagem.
- **Exploração visual:** Geração de múltiplas variações de uma mesma ideia.

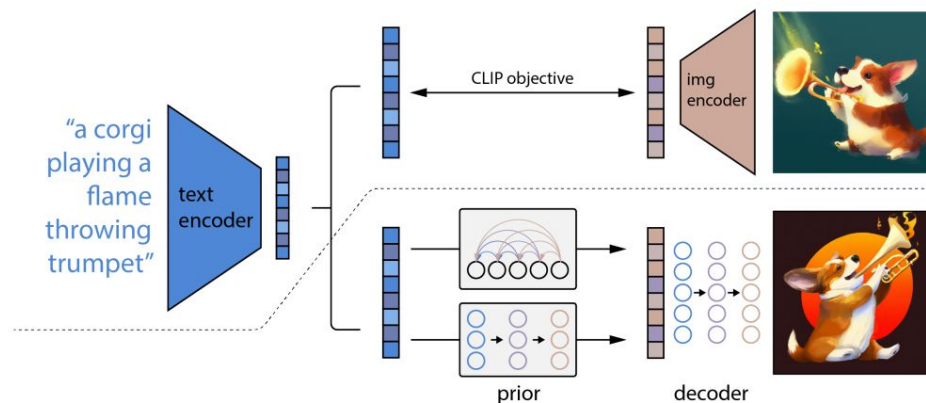


Figure 2: A high-level overview of unCLIP. Above the dotted line, we depict the CLIP training process, through which we learn a joint representation space for text and images. Below the dotted line, we depict our text-to-image generation process: a CLIP text embedding is first fed to an autoregressive or diffusion prior to produce an image embedding, and then this embedding is used to condition a diffusion decoder which produces a final image. Note that the CLIP model is frozen during training of the prior and decoder.



# DALL-E

## Diretrizes Avançadas para Prompts no DALL-E

Para obter imagens de alta qualidade, é necessário dominar técnicas avançadas de controle de estilo, luz, composição e detalhes específicos. Esses aspectos são definidos diretamente nos prompts, influenciando a precisão e a estética dos resultados.

Técnicas para refinar prompts:

- Use adjetivos específicos para descrever emoções ou tons (e.g., "melancólico", "vibrante").
- Faça iterações, ajustando o prompt com base no resultado inicial.

### Estilo visual

Defina claramente o estilo, como "arte abstrata", "fotografia", ou "realismo mágico".

### Iluminação

Especifique o tipo de luz, como "iluminação suave", "luz do pôr do sol" ou "contraste dramático".

### Composição

Inclua informações sobre o enquadramento, como "plano médio", "vista aérea" ou "ângulo baixo".

### Detalhes adicionais

Adicione elementos como "textura de óleo", "fundo nebuloso" ou "ambiente futurista".





# Utilizar técnicas de prompt para criação de imagens no MidJourney

MidJourney é uma ferramenta inovadora de geração de imagens baseada em IA, focada em estética artística e designs criativos.

Diferenciais:

- Produz imagens com detalhes altamente refinados e um toque artístico único.
- Oferece suporte direto para variações estilísticas específicas, como surrealismo, realismo e estilos de arte clássica.
- Foco em usuários criativos, como artistas, designers e profissionais de marketing.

## **MidJourney**

Excelente para projetos artísticos e abstratos, com foco em texturas e nuances visuais.

## **DALL·E:**

Mais versátil, ideal para aplicações técnicas ou conceituais, como diagramas ou imagens fotográficas.



# Utilizar técnicas de prompt para criação de imagens no MidJourney

No MidJourney, a estrutura do prompt é essencial para guiar o modelo na criação de imagens que reflitam os desejos artísticos do usuário. Os prompts podem incluir descrições detalhadas e ajustes de estilo e textura para produzir resultados únicos.

- Descrição principal: Defina o objeto ou cena central ("uma floresta assombrada").
- Estilo artístico: Especifique o tipo de arte ("em estilo aquarela", "arte digital futurista").
- Texturas e detalhes: Inclua elementos visuais, como "iluminação suave", "detalhes metálicos".
- Parâmetros adicionais: Ajuste a qualidade (resolução) e proporções (16:9).

**Prompt:** "Um castelo flutuante em estilo gótico, céu nublado, iluminação dramática, texturas detalhadas, arte digital."



# Comparação entre DALL·E, MidJourney e Stable Diffusion

Os modelos DALL·E, MidJourney e Stable Diffusion possuem características distintas que os tornam mais adequados para diferentes aplicações.

## DALL·E

Versatilidade em interpretação de prompts; útil para uso técnico e empresarial.

Estilo visual menos refinado em comparação com o MidJourney.

Aplicações técnicas ou negócios

## MidJourney

Foco em qualidade artística, ideal para designs criativos e estéticos.

Menor precisão para descrições técnicas detalhadas.

Criatividade e arte.

## StableDiffusion

Próprio para soluções open-source, altamente personalizável, funciona localmente.

Configuração técnica mais complexa; pode demandar ajustes detalhados.

Projetos personalizáveis ou locais

## Problemas Conhecidos - Mãos e olhos

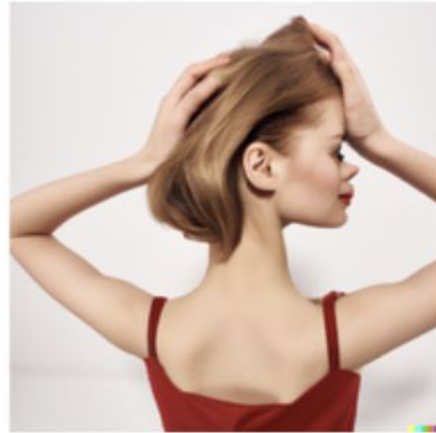
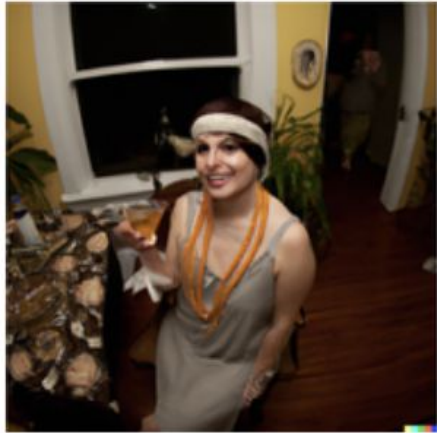
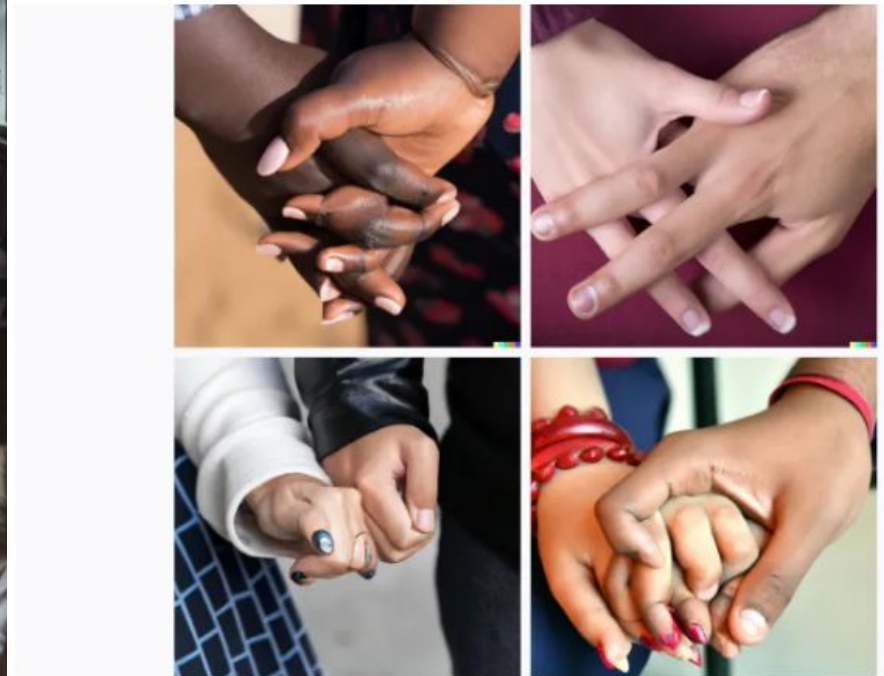


Figure 7-6. Deformed hands and eyes



Dall-E was an early AI leader, but hands are not its thing. | Dall-E prompted by Alan Truly / Digital Trends

Figure 7-9. Women eating salads and laughing







closeup of a hand usin scissors to cut shiny auburn hair

88

Create



Images



"Compassionate manager" by Stable Diffusion.

Images



"Manager" by Stable Diffusion.



MIDJOURNEY



DALL-E 2



STABLEDIFFUSION



beatles lego set, catalogue photograph



# Classificação de imagens

Pernas ou Salsichas?



Preguiça ou Croissant au Chocolat?



Chihuahua ou Muffin?





# Construir Aplicações com Geração de Imagens com IA

## **dataprep.py:**

- **Imagens do Episódio:** utilizar o LLM e os sumários dos episódios para criar imagens ilustrativas do episódio através de meta-prompting.
- **Imagens da Temporada:** utilizar técnica de self-ask prompting para instruir o LLM a criar uma imagem ilustrativa da temporada com algum 5 estilos diferentes.

## **dashboard.py:**

- **Assistente de Marketing Digital:** utilizar prompt-chaining para combinar a criação de textos de campanhas com imagens para publicação em redes sociais.
- **Visualização das imagens de sumários:** usar chain-of-thoughts prompting para alterar o código já criado para a aba *Episódios* e *Temporadas*.