A Prefeitura de Belo Horizonte apresenta:



CURSO DE INTELIGÊNCIA ARTIFICIAL APLICADA ÀS ARTES

Acessível em Libras

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Ai é nós! [livro eletrônico] : curso de inteligência artificial aplicada às artes / Raphael Silva Ferreria ... [et al.]. -- 1. ed. -- Belo Horizonte : Ed. dos Autores, 2026.
PDF

Outros autores: Ana Luiza Pedrosa Camilo, Nelson Rodrigues Pombo Júnior, Italo Cardoso Travenzoli
ISBN 978-65-02-50759-9

1. Arte digital 2. Artes 3. Inteligência artificial I. Ferreria, Raphael Silva. II. Camilo, Ana Luiza Pedrosa. III. Pombo Júnior, Nelson Rodrigues.
IV. Travenzoli, Italo Cardoso.

26-416026.0 CDD-700

Índices para catálogo sistemático:

1. Inteligência artificial : Artes 700

Maria Alice Ferreira - Bibliotecária - CRB-8/7964

FICHA TÉCNICA

Instrutores/conteudistas:
Ana Luiza Pedroza Camilo; Nelson Rodrigues Pombo Junior; Ítalo Cardoso Travenzoli; Raphael Ferreira;

Coordenação pedagógica:
Fernanda Carvalho

Design instrucional:
Raphael Ferreira

Projeto gráfico e diagramação:
Fernanda Carvalho e Julia Lechi

SUMÁRIO

TEXTO 6

AULA 1 6
1 - Introdução 6
1.1 Principais movimentos literários 9
1.2 Gêneros literários 10
2 - Pesquisa artística 11
3 - IAs generativas e produção textual 11
3.1 Interfaces para utilização de Grandes Modelos de Linguagem (LLMs) 11
3.2 Modelos de linguagem 12
3.3 Modelos abertos 13
3.4 Parâmetros 13
4 - Engenharia de prompts 15
4.1 Modelo de como criar um bom prompt para textos generativos 15
4.2 Prompts aplicados à geração de textos poéticos 16
4.3 Personalização de chatbots 16
AULA 2 17
5 - Aplicação do projeto e pesquisa artística 17
6 - Pós-produção 17
6.1 Revisão ortográfica e gramatical em Modelos de Linguagem 17
6.2 Editores 17
6.3 Tipografia 19
6.4 Peças gráficas 19
REFERÊNCIAS 19

IMAGEM 21

AULA 1 21
1 - Introdução 21
2 - A Imagem digital 22
2.1 Pixel 22
2.2 Imagem Raster 23
2.3 Imagem Vetorial 23
2.4 Código-aberto (open-source) 24
3 - Ferramentas de geração e edição de imagens digitais 25
3.1 Modelos código aberto de Inteligência Artificial 25
3.2 Interfaces gráficas para rodar o Stable Diffusion 25
3.2.1 ComfyUI 26
3.2.2 Automatic1111 26
3.2.3 Serviços para rodar as interfaces na nuvem 27
3.2.4 Alternativas para rodar o Stable Diffusion e demais IAs 27
3.2.5 Modelos de IA para edição de imagens 27
3.3 Programas de edição de imagem 28
3.3.1 Programas de edição de imagens rasterizadas 29
3.3.2 Programas de edição de imagens vetoriais 29
4 - Pesquisa artística em criação de imagens com IAs generativas 29
AULA 2 32
5 - Geração de imagens por inteligência artificial 32
5.1 Engenharia de prompts 32
5.1.2 Como criar um bom prompt 33
5.1.3 Prompts negativos 34
5.1.4 Gerador de prompts 34
6 - Stable Diffusion 35
6.1 Tarefas (tasks) 35
6.2 Parâmetros 35
7 - Detalhes sobre a interface ComfyUI 37
7.1 - Como gerar imagens no ComfyUI 38
8 - Pós-produção 42
8.1 Upscaling 42
8.2 Padrões de cores 43
REFERÊNCIAS 43

ÁUDIO **46**

AULA 1 ..... **46**
1 - INTRODUÇÃO **46**
2 - UM BREVE PANORAMA DAS IAS NO MUNDO DOS SONS **46**
3 - SOM E ARTE **48**
3.1 Tipos de artes do som **49**
4 - Pesquisa artística **50**
5 - Som e áudio **51**
5.1 Som **51**
5.1.1 Características do som **52**
5.2 Áudio **53**
5.2.1 Analógico x digital **53**
6 - Produção de áudio **54**
6.1 Processos de Produção **54**
6.1.1 Sampleamento (Amostragem) **55**
6.1.2 Síntese sonora **55**
6.2 Efeitos sonoros **55**
6.3 Softwares de produção de áudio **56**
6.4 Softwares para produção de áudio generativo **56**
AULA 2 ..... **58**
7 - MODELOS DE LINGUAGEM **58**
7.1 Funções (tasks) **58**
7.2 Parâmetros **59**
8 - ENGENHARIA DE PROMPTS **60**
8.1 Prompts postivos **60**
8.2 Prompts negativos **60**
8.3 Gerador de prompts **61**
9 - PÓS-PRODUÇÃO **61**
9.1 Pós-produção em softwares de IA generativa **62**
9.2 Formatos de áudio **63**
REFERÊNCIAS **64**

VÍDEO **65**

AULA 1 ..... **65**
1 - INTRODUÇÃO **65**
2 - UM BREVE HISTÓRICO DAS IAS NO MUNDO DOS VÍDEOS **65**
3 - AS ARTES DO VÍDEO **66**
4 - PESQUISA ARTÍSTICA **67**
5 - LINGUAGEM AUDIOVISUAL **68**
AULA 2 ..... **70**
6 - MODELOS DE LINGUAGEM **70**
6.1 Funções (tasks) **71**
6.2 Parâmetros **72**
7 - ENGENHARIA DE PROMPTS **73**
7.1 Prompts postivos **75**
7.2 Prompts negativos **75**
7.3 Gerador de prompts **75**
8 - PRODUZINDO CONTEÚDO AUDIOVISUAL COM IAS GENERATIVAS **76**
8.1 Formatos de arquivos de vídeo **78**
REFERÊNCIAS **79**

TEXTO

Competência:

Utilizar de maneira contextualizada os fundamentos da criação de textos poéticos generativos com Inteligência Artificial (IA) por meio de softwares livres e modelos abertos.

Resultados de aprendizagem:

Aplicar demonstrativamente os fundamentos da criação de textos generativos com IAs por meio de softwares livres de modo crítico e propositivo.

AULA 1

VIDEOAULAS DISPONÍVEIS NA ABA "MÓDULO 2 - AULA 1"

1 - Introdução

Neste módulo, vamos aprender a criar textos com ajuda da Inteligência Artificial (IA), usando apenas ferramentas gratuitas e abertas, ou seja, que qualquer pessoa pode usar e adaptar. O objetivo é que você entenda como essas tecnologias funcionam, conheça algumas referências históricas da literatura e aplique esses conhecimentos na sua prática artística.

Luis Carlos Ribeiro Kasprzykowski em sua dissertação "Sistema de Geração de Texto Para Performance", propôs a criação de um sistema capaz de gerar textos automaticamente para serem usados em apresentações ao vivo. A ideia do artista, pesquisador e professor universitário nas áreas de artes, multimídia e tecnologias digitais, era investigar como a tecnologia pode colaborar com a criação artística, integrando som, palavra e performance em tempo real.

Segundo Kasprzykowski (2020), usar computadores para criar textos não é algo novo. Já nos anos 1950, artistas e programadores começaram a construir máquinas que escrevem frases parecendo humanas.

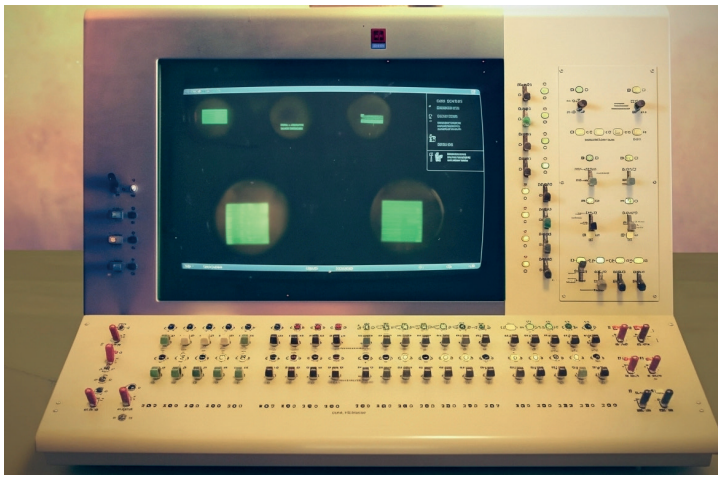
Um exemplo disso é o projeto "Love Letters", criado em 1953 por Christopher Strachey (Figura 1) *e 2).* Ele fez um programa que usava um computador antigo chamado Ferranti Mark 1 para gerar cartas de amor automáticas.

O programa funcionava assim: ele tinha uma estrutura de frases pronta, como por exemplo:

"Você é meu/minha [adjetivo] [substantivo]. Meu/minha [adjetivo] [substantivo] [verbo] seu/sua [adjetivo] [substantivo]."

Depois, o computador escolhia palavras aleatórias de uma lista para preencher os espaços.

Figura 1 – Ferranti Mark I e o projeto Love Letters



Fonte: Imagem do autor (criação com IA generativa)

Figura 2 – Programadores gerando cartas de amor no Ferranti Mark I

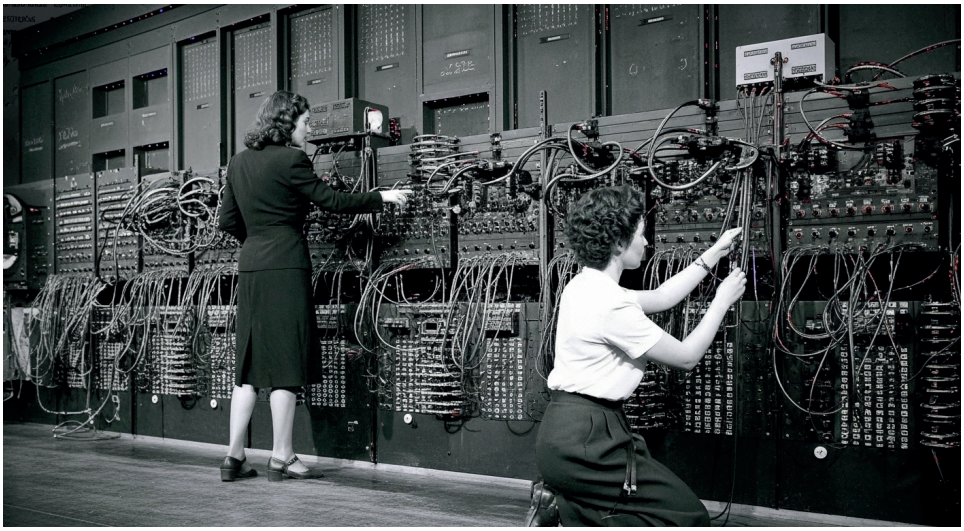


Fonte: Imagem do autor (criação com IA generativa)

Assim, o computador conseguia criar milhões de combinações diferentes, mesmo com um vocabulário pequeno. Aqui é importante ter em mente a distinção: diferentemente de IAs que apenas classificam, identificam ou analisam, como o caso citado da Ferranti Mark 1, a IA generativa, que vamos discutir nesta disciplina é capaz de produzir dados novos com base no que aprendeu durante o treinamento. Ela não copia exatamente o que viu — em vez disso, ela recria com variações, combinando padrões de maneira nova.

A (Figura 3) mostra como era o primeiro computador eletrônico da história. O ENIAC foi encomendado pelo Exército dos Estados Unidos da América (EUA) para realizar cálculos balísticos, ou seja, para calcular trajetórias de projéteis.

Figura 3 - Programadoras operam o primeiro computador eletrônico da história, o ENIAC.



Fonte: Imagem do autor (criação com IA generativa)

As primeiras programadoras do ENIAC foram mulheres matemáticas brilhantes, incluindo: Jean Jennings Bartik; Kay McNulty; Betty Holberton; Ruth Teitelbaum; Marlyn Wescoff; Frances Bilas *(Figura 4).*

Figura 4 – As primeiras programadoras do ENIAC



Fonte: Imagem do autor (recriação com IA generativa)

Com o tempo, os computadores ficaram menores, mais baratos e começaram a chegar na casa das pessoas, principalmente nas décadas de 1970 e 1980. Isso fez com que o interesse em usar a computação para a escrita criativa aumentasse bastante.

Foi nessa época que surgiram novas formas de criar e ler textos no computador, como por exemplo:

- **Hipertexto -** textos com links que levam a outras partes ou páginas (tipo os links da internet).
- **Literatura interativa -** onde o leitor pode escolher caminhos diferentes na históriadentro das possibilidades do digital.
- **Poesia digital -** que mistura palavras, imagens e até sons criados por meio computador. Nos anos 1990, esse movimento cresceu ainda mais. Foi quando a Internet começou a se popularizar e o texto escrito era a principal forma de comunicação online, já que os vídeos e imagens ainda eram difíceis de carregar com as conexões lentas da época. Esse período foi muito importante porque marcou início daquilo que a gente chama atualmente de cibercultura, ou seja, a cultura criada a partir do uso da tecnologia e da internet.

Hoje, com o avanço das IAs, qualquer pessoa pode experimentar, por exemplo, a criação de poesia, prosa e outros textos a partir de comandos simples. Mais à frente falaremos um pouco sobre o *GPT4all* e *Google AIStudio* e sua proposta de avanço da IA, antes disso, vamos começar conversando sobre literatura.

Você sabe o que queremos dizer quando falamos de literatura?

De acordo com o dicionário Michaelis, podemos definir literatura como a arte de compor escritos, em prosa ou em verso, de acordo com determinados princípios teóricos ou práticos. E o que será que isso tem a ver com o digital e as IAs? Vamos continuar conversando...

1.1 Principais movimentos literários

Para entender melhor como podemos usar isso de forma criativa na nossa própria produção artística, é importante olhar também para a história da literatura. Muitos movimentos literários anteriores ao advento do digital, desafiaram as formas tradicionais de escrever, brincando com as palavras, as estruturas e os sentidos - de um jeito que conversa muito com o que hoje vemos nas criações feitas com inteligência artificial. São eles:

- **Vanguardas europeias (1910–1930) -** Movimentos como Futurismo, Dadá e Cubismo que quebraram regras tradicionais, usando fragmentos, colagens e experimentação na arte e na escrita.
- **Surrealismo (década de 1920) -** Buscava expressar o mundo dos sonhos e do inconsciente, usando a escrita automática: escrever sem pensar para liberar a criatividade.
- **Concretismo (décadas de 1950–1960) -** Movimento brasileiro que valorizava a forma visual das palavras no papel, criando poemas que são também imagens.
- **Oulipo (década de 1960) -** Grupo de escritores que criavam textos usando regras e restrições específicas, como escrever sem usar certas letras.
- **Poesia digital (anos 1990 até hoje) -** Poemas feitos para e por meio digital, que podem ter imagens, sons e interação com o leitor, explorando novas formas de criar e ler poesia.

1.2 Gêneros literários



Na literatura tradicional, aprendemos que os textos costumam se organizar em três grandes gêneros - lírico, épico e dramático - cada um com características próprias que ajudam a entender o tipo de história ou sentimento que está sendo contado.

O gênero lírico é aquele que expressa sentimentos, emoções e percepções subjetivas do eu. Um exemplo famoso é o *Soneto de Fidelidade*, de Vinicius de Moraes, no qual o poeta fala sobre o amor de forma intensa e emocional: "*Que não seja imortal, posto que é chama / Mas que seja infinito enquanto dure."* Nesse caso, o foco não está em contar uma história, mas sim em transmitir uma emoção.

Já o **gênero épico ou narrativo** tem como principal característica o relato de ações de personagens em determinado tempo e espaço. Um bom exemplo é o livro Capitães da Areia, de Jorge Amado, que conta a história de um grupo de meninos de rua que vive em um trapiche abandonado em Salvador: "Pedro Bala e sua turma dormem no trapiche abandonado. São meninos de rua de Salvador, vivendo aventuras e violências num Brasil desigual."

Por fim, o **gênero dramático** é voltado para a representação teatral. Os textos são construídos principalmente por meio de diálogos entre personagens e são feitos para serem encenados. Um exemplo marcante é a peça O Auto da Compadecida, de Ariano Suassuna.

Em uma de suas cenas, os personagens João Grilo e Chicó trocam falas como:

JOÃO GRILO: *"Eu não sou homem de roer corda."*
CHICÓ: *"Mas tenho medo, João!"*

Observe que neste tipo de texto, não há narrador, a ação acontece por meio das falas e gestos dos personagens.

Esses três gêneros ajudam a entender as diferentes formas de se contar, sentir ou representar algo por meio da linguagem literária. Porém, com o avanço das tecnologias digitais e das IAs, novos gêneros literários começaram a surgir, que não cabem tão facilmente nas categorias antigas.

Você já deve ter visto - e até compartilhado - um meme por aí, né? Pois é, os memes são um exemplo de como a internet criou novos gêneros de comunicação, que misturam humor, crítica e mensagens diretas, muitas vezes em poucas palavras ou imagens. Outro exemplo são as *web art*, como uma de suas principais características o uso ativo e crítico da própria rede como meio e matéria-prima artística e do *storytelling* automatizado, onde as histórias são criadas em tempo real por

IAs. Tudo isso mostra como o mundo digital muda a forma de se comunicar, criar e se expressar.

2 - Pesquisa artística

A **pesquisa artística** é o **fio condutor** que dá consistência e coerência ao trabalho de um artista. Ao longo dos próximos módulos, você vai desenvolver sua própria criação poética, e para isso, é fundamental investigar como os movimentos e gêneros literários podem servir como ponto de partida para suas experimentações. A cada etapa, você vai construir um repertório criativo, relacionando referências e explorando linguagens que alimentem a sua produção autoral.

Por exemplo: o **Surrealismo** nos convida a deixar a imaginação correr solta, como num sonho. Ele valoriza o que vem do nosso **inconsciente**, aquelas ideias inesperadas que surgem quando não estamos tentando controlar tudo. Já o **Concretismo** olha para o texto como uma imagem: **as palavras viram forma**, som e espaço na página — como se o poema também pudesse ser desenhado. O movimento **Oulipo**, por outro lado, mostra que criar com **limites** também é poderoso. Usar regras, como escrever um texto sem a letra "e" ou rimar de formas diferentes, pode fazer a criatividade sair por outros caminhos novos e surpreendentes. Enquanto isso, os **gêneros literários tradicionais**, como o **lírico**, ajudam a gente a falar dos nossos sentimentos, daquilo que toca o coração. O **épico e o dramático** podem virar histórias em vídeo, teatro, música ou até quadrinhos — é só usar a criatividade para reinventar essas formas.

Não se prenda à ideia de que o texto precisa ter começo, meio e fim. Você pode escrever em lista, em fluxo de consciência, como se fosse uma carta, uma legenda de rede social, um sonho, uma manchete de jornal, uma partitura. Brinque com o formato. Um texto artístico muitas vezes surge quando a gente muda o jeito de olhar para as coisas do dia a dia. Aquilo que parece comum ou sem graça pode ganhar um novo sentido quando olhamos com mais maginação ou sensibilidade. Enxergar o cotidiano com um olhar mais poético é uma forma poderosa de criar.

3 - IAs generativas e produção textual

Nos últimos anos, as **inteligências artificiais (IAs) generativas** revolucionaram a forma como interagimos com a tecnologia. Para a produção textual, essas ferramentas são como ter um **assistente superinteligente** que pode te ajudar a **escrever, resumir, criar roteiros, e até mesmo gerar poemas e letras de música.** Elas funcionam processando grandes volumes de dados de texto para aprender padrões, o que lhes permite criar conteúdo novo e original. É importante entender que as IAs não substituem a criatividade humana, mas sim a amplificam, oferecendo novas possibilidades e agilizando o processo de criação. Se você quer explorar essa tecnologia na prática, existem diversas ferramentas disponíveis.

3.1 Interfaces para utilização de Grandes Modelos de Linguagem (LLMs)

Que tal conhecer o **ChatGPT,** um dos mais populares chatbots desenvolvidos pela OpenAI, ou o **GPT4All,** uma alternativa de código aberto que você pode usar de forma mais flexível e offline, sem

necessitar de conexão com a internet? Essas interfaces vão te ajudar a organizar ideias, experimentar formas e aprofundar a pesquisa artística que estamos desenvolvendo neste módulo.

Essas ferramentas serão apresentadas com mais detalhes no vídeo que compõe este módulo. No vídeo, você vai entender como e quando usá-las, de acordo com os caminhos que você escolheu até aqui: seja para explorar um movimento literário que chamou sua atenção, seja para experimentar um gênero ou linguagem que despertou sua curiosidade.

Fique atento(a), pois essas ferramentas serão úteis nos momentos em que você precisar dar forma às suas ideias, testar possibilidades ou registrar o processo criativo.

Dica

Como rodar IAs localmente no seu celular:

Se você se interessa por rodar IAs localmente e offline, sem a necessidade de conexão com a internet, o **Google Edge Gallery** permite usar modelos de IA diretamente no seu smartphone. Ele pode ser baixado de forma gratuita diretamente da loja de aplicativos oficial da Google (Play Store) ou Apple (App Store). Lá você poderá escolher de forma intuitiva os modelos de linguagem que deseja rodar.



LINKS ÚTEIS

Chat GPT (O Chat-bot proprietário mais popular)
https://chatgpt.com/

Google AI Studio (Nele você poderá rodar modelos livres como o Gemma)
https://aistudio.google.com/prompts/new_chat

Google AI Edge Gallery (Rode IAs localmente, offline em seu celular)
https://play.google.com/store/apps/details?id=com.google.ai.edge.gallery&pcampaignid=web_share&pli=1

GPT4All (Rode diversos modelos de IAs localmente em seu computador ou celular)
https://www.nomic.ai/gpt4all

3.2 Modelos de linguagem

Agora que você já entendeu melhor como interagir com ferramentas de IA, vamos dar um passo além: entender o que são os Modelos de Linguagem ou **LLMs – Large Language Models.** Modelos de linguagem são sistemas de inteligência artificial treinados para entender, gerar e interagir com textos em linguagem natural (como o português, inglês, etc.). Eles aprendem a partir de grandes volumes de texto e conseguem prever a próxima palavra ou frase com base no que já foi dito,

por isso conseguem responder perguntas, escrever textos, resumir ideias, traduzir, criar poemas, entre outras tarefas.

Existem modelos especializados para diferentes objetivos (Tasks). Vamos conhecer!

- **Text-to-Text (Texto para Texto) -** Esse é o mais comum. Você escreve uma pergunta ou um comando e recebe como resposta um texto. Exemplo: ChatGPT, Google Gemini, LLaMA.

- **Image-to-Text (Imagem para Texto) -** A IA analisa uma imagem e descreve em texto o que está vendo. Super útil em acessibilidade ou para descrever imagens artísticas.

- **Voice-to-Text (Voz para Texto) -** Aqui a IA transforma o áudio da sua voz em texto escrito. Pense nos apps de transcrição automática, como o Whisper da OpenAI.

Ou seja, para o nosso caso, você como criador(a), precisa entender qual ferramenta usar dependendo da sua necessidade.

3.3 Modelos abertos

Modelos de linguagem abertos (ou open source) são modelos cujo código, arquitetura ou pesos de treinamento são **disponibilizados publicamente**. Isso permite que qualquer pessoa (pesquisador, desenvolvedor ou artista) acesse, estude, modifique e use o modelo em seus próprios projetos. Eles são diferentes dos **modelos fechados** que são **controlados por empresas** e só podem ser usados via API (interface que conecta aplicativos) ou interface web, com acesso limitado.

Vejamos exemplos:

- **LLaMA (Meta) -** Modelo com licença específica, que permite uso acadêmico e experimental.
- **Gemma (Google) -** Modelo de código aberto, com licença Apache 2.0, ótimo para testes de IA locais. IA local é quando a inteligência artificial funciona direto no seu aparelho — como no computador ou celular — sem precisar de internet ou de enviar dados para servidores na nuvem (ou seja, em outros lugares pela internet). Isso traz mais rapidez, porque tudo é feito ali mesmo, e também mais privacidade, já que suas informações não são compartilhadas com outras empresas ou sistemas online.
- **Mistral -** Outro modelo open source muito usado por quem trabalha com processamento de linguagem. Processamento de linguagem é o jeito como computadores e máquinas aprendem a entender, interpretar e até produzir textos e falas, como se “falassem a nossa língua”. É o que permite, por exemplo, que você converse com assistentes virtuais, traduza textos automaticamente ou receba sugestões de palavras no celular enquanto digita. É quando a tecnologia aprende a lidar com a linguagem humana — escrita ou falada — para facilitar a comunicação entre pessoas e máquinas.

3.4 Parâmetros

Parâmetros são os **valores ajustáveis** que um modelo de linguagem aprende durante seu treinamento. Eles definem como o modelo processa e gera texto, ou seja, como ele “pensa”. Em termos

técnicos, um modelo de linguagem é uma rede neural com milhões ou bilhões de conexões entre neurônios artificiais. Cada conexão tem um “peso” e isso é um parâmetro.

Quantos parâmetros um modelo pode ter? Mais que bilhões. E isso pode significar por exemplo: um modelo com 7 bilhões de parâmetros (7B) já pode escrever bem, traduzir e responder perguntas básicas. Enquanto um com 70B ou mais pode gerar textos longos, coerentes, com múltiplos estilos e realizar tarefas complexas (como resolver problemas matemáticos ou entender ironia).

Agora, **atenção aos termos técnicos** que vão te acompanhar sempre que for trabalhar com geração de texto, mas nesse caso, os **parâmetros** são definidos no código de programação. **São ajustes que você poderá alterar e que influenciarão os resultados obtidos**.

Então temos:

- **Tokens -** São como pedacinhos de texto. Toda resposta da IA é formada por um número de tokens. Saber disso te ajuda a controlar o tamanho e custo da geração. Os modelos têm limites de tokens para entrada + saída. Alguns podem ter limite de tokens por interação.
- **Temperatura -** Controla o nível de aleatoriedade nas respostas. Temperatura baixa (ex.: 0,2) = respostas mais objetivas. Alta (ex.: 0,8) = respostas mais criativas e inesperadas.
- **Top-k e Top-p -** Controlam como um modelo de linguagem escolhe as palavras na hora de gerar texto. Isso porque quando a IA está escrevendo, ela não escolhe apenas a palavra mais provável, ela considera várias opções possíveis.

4 - Engenharia de prompts

Criar um bom texto com IA depende de saber **dar o comando certo**: isso é **engenharia de prompt**!

Modelos de linguagem (como o GPT-4) não “pensam” como seres humanos. Eles respondem com base em padrões de linguagem aprendidos em grandes volumes de texto. Por isso, a forma como você pergunta muda totalmente o que a IA responde.

Elementos que ajudam a construir um bom prompt:

1. **Clareza do objetivo -** O que você quer que a IA faça? (Escrever? Reformular? Traduzir? Criar?)
2. **Estilo ou forma desejada -** Texto curto ou longo? Poema, carta, roteiro?
3. **Tom ou emoção -** Triste, engraçado, neutro, irônico?
4. **Referências ou contexto -** Inspirado em algum autor? Usando um vocabulário específico?
5. **Restrições ou estruturas -** Número de palavras, tipo de rima, sem usar certas letras, etc.

Para artistas e pesquisadores, a engenharia de prompts pode ser parte do processo de criação:

- Criar poemas com respostas da IA;
- Subverter instruções intencionalmente;
- Dialogar com a máquina como se ela fosse personagem, curadora, oráculo ou editora;
- Misturar respostas humanas e artificiais na composição final.

4.1 Modelo de como criar um bom prompt para textos generativos

Complete a tabela a seguir e experimente os prompts! (Use uma folha de papel ou um editor de textos para rascunho e experimente possibilidades de prompts, baseadas na tabela).

Como usar essa tabela:

- Escolha um tipo de criação que lhe interesse.
- Combine o tema, a forma e o tom sugeridos ou crie suas próprias variações.
- Monte um prompt claro e específico para usar em ferramentas de IA generativa ou na sua criação manual.
- Experimente e ajuste o prompt para obter diferentes resultados.

Tipo de criação	Tema sugerido	Forma / Estrutura	Tom / Estilo	Exemplo de prompt
Poesia		4 estrofes curtas		
Narrativa curta	Futuro, tecnologia	1-2 páginas	Melancólico, ficção científica	
Descrição de cena				Descreva uma cidade futurista onde todas as ruas são espelhos, criando uma atmosfera misteriosa.
Texto experimental	Texto experimental Labirinto, palavra	Microconto, sem letra “e”	Angustiado, acelerado	Escreva um microconto sem usar a letra “e” sobre uma pessoa presa num labirinto de palavras.
Spoken Word	Internet, pressa	Texto curto para 1 minuto		
Mistura de estilos			Dramático, irônico	
Diálogo ficcional				

4.2 Prompts aplicados à geração de textos poéticos

Ao delimitar o escopo da criação, o prompt atua como uma semente poética, direcionando o modelo a selecionar vocabulário, ritmo e imagens que se alinham à sensibilidade desejada, resultando em um texto que possui maior profundidade e coerência artística.

4.3 Personalização de chatbots

Personalização de chatbots é o processo de **ajustar a linguagem, o tom, o comportamento e as respostas do chatbot para que ele se conecte melhor com um público específico ou atenda às necessidades particulares.** Isso torna a interação mais natural, eficaz e satisfatória para quem está usando. É importante porque um chatbot genérico pode parecer frio, distante ou pouco útil. Quando personalizado, ele cria uma experiência mais acolhedora, compreende melhor as dúvidas do usuário e responde de forma mais relevante, aumentando a eficiência do atendimento e a satisfação do público.

- **Defina o estilo do chatbot -** Decida qual será a identidade do seu chatbot. Ele será um escritor clássico e romântico, um poeta moderno e direto, ou um autor de ficção científica, um parceiro de composição? Essa definição garante que todos os textos gerados tenham coerência e estilo. Um estilo consistente é um fator importante para a identidade artística.
- **Escolha um foco temático específico -** Não peça para o bot escrever sobre tudo. Escolha um tema principal onde ele será especialista, como: a natureza, o sentimento de nostalgia, a vida urbana, ou a mitologia. Ao restringir o assunto, você permite que o bot desenvolva um vocabulário rico e aprofundado naquele nicho. Isso resulta em descrições mais criativas e originais.
- **Use textos de referência -** O treinamento do bot deve ser feito com obras pertinentes e de qualidade. Suba arquivos relacionados ao tema. A qualidade do texto que ele lê é a qualidade do texto que ele irá produzir.
- **Tenha um antídoto para alucinações -** Crie um plano claro para quando o bot não entender o que você pediu e criar respostas erradas, alucinadas. Oriente o bot a perguntar informações sobre o contexto e referências sobre assunto, a fim de evitar enganos. Como os modelos são pré-treinados, trocar de modelo pode ser a solução quando a alucinação se manter.

Obs: Os erros mesmos quando indesejados, podem ser boas fontes de inspiração e vários artistas trabalham justamente com o potencial das falhas.

VIDEOAULAS DISPONÍVEIS
NA ABA ”MÓDULO 2 - AULA 2”

5 - Aplicação do projeto e pesquisa artística

Agora é com você! Pegue o início da sua pesquisa artística que começou na Aula 1 e construa um chatbot criativo para te ajudar no processo. Pode ser um bot assistente que sugira títulos, inicie versos ou até que critique suas produções para te ajudar a melhorá-las.

6 - Pós-produção

Chegou a hora de **finalizar** o que foi criado! Afinal, o texto não termina quando a IA te entrega uma resposta. Aqui vamos aprender como transformar o que você produziu em algo pronto para ser mostrado ao mundo.

A pós-produção envolve etapas de **edição e finalização** do arquivo trabalhado em formatos que possam ser utilizados em linguagens específicas e de forma profissional. Ela está diretamente relacionada com a forma com a qual você irá expor sua obra.

6.1 Revisão ortográfica e gramatical em Modelos de Linguagem

Quando você chegar na sua primeira versão textual, você pode ainda pedir para a própria IA revisar o texto. Teste prompts como aplicando o que já foi falado na aula anterior.

Exemplo: “Revise meu texto corrigindo erros de ortografia e gramática, mas sem mudar o estilo de escrita.”

Aplique o conhecimento de diferentes formas e a favor da sua qualidade de produção.

6.2 Editores

Se o seu texto vai virar arte visual (um pôster, um post, uma peça gráfica), editores gratuitos podem ser seus aliados.

- **GIMP** (https://www.gimp.org/) **-** Editor de imagens vetoriais e bitmap, perfeito para manipulação de fotos. Open source e funciona no windows, macos e Linux.
- **Photopea -** Funciona direto no navegador e é bem parecido com o Photoshop. Nele você consegue aplicar efeitos de texto incríveis, usar camadas e exportar em alta qualidade. Você consegue rodar em notebooks, desktops e dispositivos móveis.
- **PicsArt -** parecido com o GIMP, porém versão mobile, gratuito.
- **Canva -** Proprietário, multifuncional e intuitivo, é uma interface completa para te auxiliar desde a criação até a finalização dos arquivos. Oferece planos gratuitos com limitações de uso e planos

pagos com bom custo benefício. Oferece acesso livre para estudantes, professores e OSCs. Pode ser utilizado no navegador ou via app instalado em dispositivos móveis.

- **Adobe express -** Concorrente direto do Canva de acesso gratuito e funcionalidades similares. Pode ser utilizado via app ou diretamente no navegador.

Mas também existem os pagos como CorelDraw, Illustrator, Photoshop, Lightroom (versão mobile também).

6.3 Tipografia



A **tipografia é a arte e a técnica de organizar letras e palavras de forma visual**. Ela começa na antiguidade com inscrições em pedra e manuscritos feitos à mão, passando pela grande revolução de Gutenberg no século XV, com a criação da prensa móvel, momento em que a tipografia se torna ferramenta de reprodução em massa.

Durante os séculos seguintes, surgem estilos clássicos como as serifadas romanas e as góticas, que dominam livros, jornais e documentos. No século XIX, com a Revolução Industrial, novas tecnologias e demandas comerciais incentivam o surgimento de tipos mais ousados e decorativos de letras. Já no século XX, o modernismo traz uma busca por clareza e funcionalidade, com fontes como Helvetica e Futura ganhando destaque.

Com a chegada do computador e do design digital, a tipografia se torna ainda mais acessível e dinâmica. Inclusive: a tipografia continua evoluindo, agora com o uso de inteligência artificial na criação de novos tipos e formas de interação textual.

Escolher a **fonte certa dá personalidade ao seu texto**. Mas o que isso quer dizer? A tipografia — ou seja, o estilo, o formato e a aparência das letras — influencia diretamente na forma como a mensagem é percebida. Uma fonte pode parecer séria, divertida, moderna ou até misteriosa. Assim como escolhemos cores e imagens para criar um clima, escolher a letra certa também ajuda a contar melhor sua história.

No vídeo 03 deste modulo para conhecer um pouco do site **Google Fonts** e aprender a baixar e aplicar a fonte em uma imagem/peça gráfica.

Atenção: Sempre use fontes com licença aberta, principalmente se o seu projeto for público ou comercial. Tome cuidado com sites paralelos que não oferecem segurança no download de arquivos.

6.4 Peças gráficas

Quando falamos em **peças gráficas**, estamos nos referindo a **qualquer material visual criado para comunicar algo**, seja uma arte para redes sociais, um cartaz impresso, um convite digital, um banner ou até uma capa de livro. Cada uma dessas peças tem características técnicas que precisam ser adequadas ao seu meio de exibição: digital ou impresso.

A **composição** é como os elementos gráficos são organizados dentro da peça. Uma boa composição equilibra forma e função, guiando o olhar e facilitando a leitura da mensagem. Vejamos os principais elementos:

- **Hierarquia -** O que deve chamar mais atenção? Título, imagem, botão, chamada?
- **Alinhamento -** Textos e imagens organizados de forma limpa e coerente.
- **Contraste -** Diferenças de cor, forma ou tamanho para destacar o que importa.
- **Espaçamento (respiro) -** Evitar poluição visual e permitir que o conteúdo “respire”.
- **Tipografia -** Escolher fontes adequadas ao tom e à função da peça (moderna, clássica, divertida, séria).

O formato final importa!

Veja algumas noções que você precisa conhecer:

Característica	**Web**	**Impressão**
Resolução da imagem	72 DPI	300 DPI
Cores da imagem	RGB	CMYK
Formato de Arquivo	PNG, JPEG, GIF	PDF, TIFF, EPS
Objetivo da composição	Atrair, comunicar rápido	Imprimir com clareza e impacto
Cuidados principais	Leveza, adaptação à tela	Fidelidade de cor e nitidez

REFERÊNCIAS

FRANZÃO, Luana. Do ENIAC ao notebook: confira a evolução dos computadores nas últimas décadas. CNN Brasil, São Paulo, 23 fev. 2021. Disponível em: https://www.cnnbrasil.com.br/tecnologia/do-eniac-ao-notebook-confira-a-evolucao-dos-computadores-nas-ultimas-decadas/. Acesso em: 23 jun. 2025.

GIMP. Disponível em: https://www.gimp.org. Acesso em: 15/06/2025

Google. Gemini AI Studio. 2024.

Hugging Face. LLM Tools. 2024.

KASPRZYKOWSKI, Luís Carlos Ribeiro. Sistema de geração automática de texto para performance. Dissertação (Mestrado em Multimédia) – Universidade do Porto, Porto, 2020. 63 f.

Meta AI. LLaMA Models. 2024.

OpenAI. ChatGPT. 2024.

Photopea. Disponível em: https://www.photopea.com

Picsart Photo Editor. [Aplicativo mobile]. Versão 24.0.1. São Francisco: Picsart Inc., 2024. Disponível em: https://picsart.com. Acesso em: 30 jun. 2025.

AMADO, Jorge. Capitães da Areia. São Paulo: Companhia das Letras, 2008.

IMAGEM

Competência:

Utilizar de maneira contextualizada os fundamentos da criação de imagens generativas com IAs por meio de softwares livres.

Resultados de aprendizagem:

Aplicar demonstrativamente os fundamentos da criação de imagens generativas com IAs por meio de softwares livres de modo crítico e propositivo.

AULA 1

VIDEOAULAS DISPONÍVEIS NA ABA "MÓDULO 3 - AULA 1"

1 - Introdução

O objetivo desta disciplina é que você possa se expressar através de imagens digitais, utilizando as ferramentas mais simples e sofisticadas que estão disponíveis de graça. Hoje temos ao nosso dispor ferramentas que nos permitem gerar ou trabalhar imagens mesmo sem saber desenhar, pintar ou fotografar; ou sem que tenhamos lápis, pincéis e câmeras. Estamos falando de ferramentas como Inteligência Artificial Generativa, ou seja, softwares que geram conteúdos diversos, incluindo imagens; bem como softwares que permitam editar e trabalhar os resultados obtidos através destas ferramentas.

Pretendemos que você:

- Consiga conceber a ideia ou conceito daquilo que você quer expressar através destas ferramentas;
- Compreenda os caminhos que você pode seguir para realizar a sua ideia;
- Saiba os fundamentos da imagem digital, para então conseguir fazer o melhor uso das ferramentas disponíveis;
- Compreenda como obter imagens através de ferramentas de Inteligência Artificial generativa, sabendo como dar comandos que delimitem o que você quer - e que não quer - nas imagens;
- Consiga trabalhar as imagens geradas através de ferramentas de edição, de modo que você personalize os resultados de acordo com suas necessidades.

Para começar, precisamos entender os fundamentos da imagem digital, já que não trabalharemos como materiais físicos, como lápis, pincéis, tintas, papel e telas, ou seja, aqueles materiais que compramos na papelaria para desenhar ou pintar.

Precisamos compreender o que é o pixel, que é a unidade básica da imagem digital, com a qual todas as ferramentas que utilizaremos trabalharão.
Além disso, podemos ter dois tipos principais de imagem digital, que são as imagens raster e as imagens vetoriais.

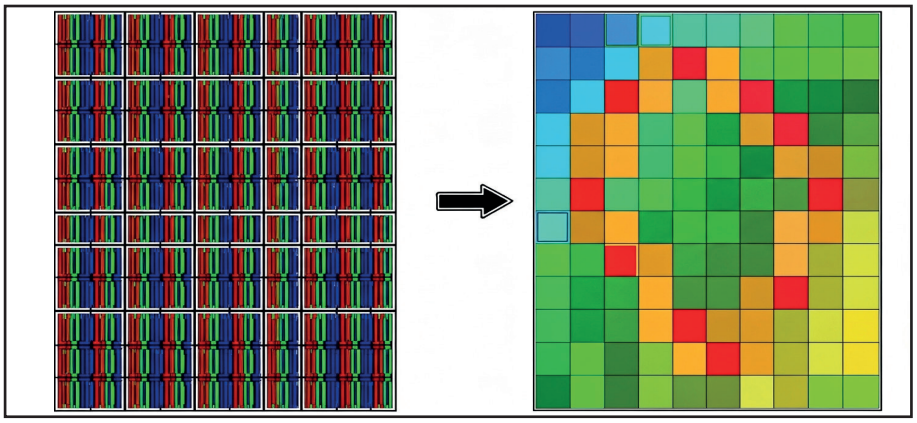
2 - A Imagem digital

2.1 Pixel

Um **pixel** (abreviação de picture element) é a menor unidade de uma imagem digital. Ele representa um único ponto em uma grade que compõe toda a imagem. Cada pixel possui uma cor específica, definida por valores numéricos que indicam a quantidade de vermelho, verde e azul (no caso do modelo de cor RGB, usado pela maioria das telas).

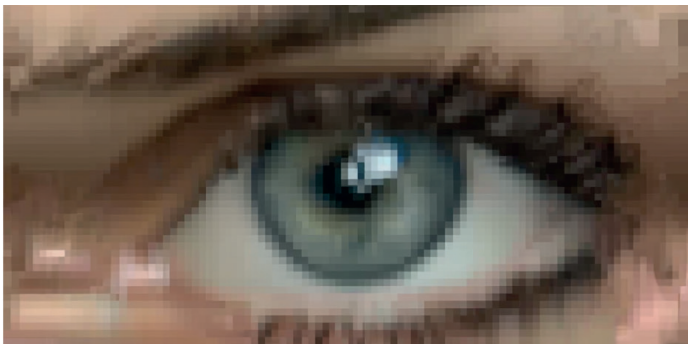
Quando olhamos uma imagem digital normalmente, os pixels não são perceptíveis individualmente. Mas ao dar bastante zoom, é possível ver que a imagem é formada por pequenos quadrados coloridos, que são os pixels.

Figura 1 - Os pixels RGB organizados em uma série de linhas e colunas.



Fonte: Do autor. Imagem criada por IA

Figura 2 - Imagem de baixa resolução, na qual é possível identificar os pixels individuais.



Fonte: Do autor. Imagem criada por IA

2.2 Imagem Raster

Uma imagem **raster** (ou bitmap, que vem de mapa de bits) é composta por uma grade de **pixels**, onde cada pixel representa um ponto específico com uma cor própria. A edição raster envolve manipular diretamente esses pixels.

Figura 3- Imagem raster



Fonte: Do autor.

Por exemplo, ao pintar com um pincel digital sobre uma tela branca, o que está acontecendo é a udança da cor dos pixels naquela área específica. Ao dar zoom em uma imagem raster, é possível ver claramente os quadrados coloridos (os pixels) que formam a imagem.

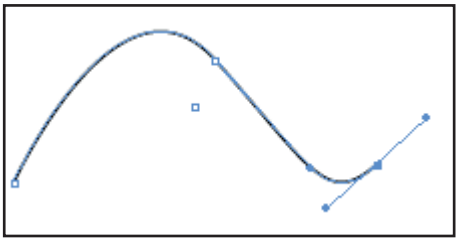
Uma imagem raster é baseada em pixels, em que cada pixel tem informações de cor.

A edição raster é pixel a pixel e, quando ampliada, a imagem perde qualidade e fica “pixelada”. É o tipo de imagem usada em fotos e pinturas digitais.

2.3 Imagem Vetorial

Uma **imagem vetorial** é uma imagem baseada em expressões matemáticas, e independente de pixels. Ao contrário das imagens raster, que são formadas por uma grade de pontos (pixels), as imagens vetoriais são criadas por caminhos definidos por pontos chamados **“nós”**, localizados em coordenadas nos **eixos x e y**.

Figura 4- Imagem vetorial



Fonte: Do autor.

Esses caminhos podem formar formas como retângulos, círculos, curvas, linhas etc. Quando você redimensiona uma imagem vetorial (aumenta ou diminui), o computador recalcula os caminhos matematicamente, o que permite que a imagem seja redimensionada sem perda de qualidade — ou seja, ela nunca fica pixelada.

Quando você redimensiona ou move esses pontos, o computador calcula e redesenha o caminho e exibe a nova forma.

Agora que temos esta noção inicial sobre as imagens digitais, podemos então relacionar quais ferramentas temos à nossa disposição. Este curso busca trabalhar com ferramentas que não precisamos pagar para usá-las, o que significa dizer que trabalharemos com ferramentas de código aberto (open source) ou então que são gratuitas, apesar de não terem código aberto. Mas o que significa ser código aberto?

2.4 Código-aberto (open-source)



Quando nos referimos a um software de "**código-aberto" (ou open source)**, significa que estamos trabalhando não só com um programa gratuito, mas um software cujo código-fonte está disponível para toda comunidade. Por código-fonte, entendemos todas as instruções que fazem um programa funcionar, tal como como uma receita de bolo, que os programadores escrevem em linguagem de programação.

Isto permite que seja possível ver como este programa foi feito; que ele seja modificado, adicionando ou mudando funções e que ele seja compartilhado sem que se pague por isso.
Para que um programa seja considerado verdadeiramente "código aberto", ele precisa seguir algumas regras bem claras. Essas regras garantem que o programa seja livre para todos usarem, sem discriminação, sem cobrança e sem limitar o que você pode fazer com ele.

Você pode usar o programa para estudar, trabalhar, criar arte, dar aulas ou fazer pesquisas; pode mudar o programa e criar uma nova versão com seu nome; pode usar junto com outros programas, mesmo que eles não sejam abertos.

Muita coisa que usamos hoje em tecnologia só existe graças ao movimento de software livre, já que o compartilhamento do conhecimento ajuda a construir um mundo mais colaborativo.

Agora que sabemos o que são ferramentas de código aberto, podemos listar os programas que temos disponíveis para gerar ou editar imagens digitais.

3 - Ferramentas de geração e edição de imagens digitais

Para facilitar nosso aprendizado, podemos dividir de um lado os programas de geração de imagem por Inteligência Artificial, e, por outro, os programas para edição de imagens digitais. Além disso, como há dois tipos principais de imagens digitais - as raster e as vetoriais - temos programas específicos para trabalhar com cada um destes tipos de imagem.

3.1 Modelos código aberto de Inteligência Artificial

Veremos neste tópico que para obtermos **imagens** por **Inteligência Artificial** será necessário:

a) um modelo de geração de imagens
b) interfaces gráficas que permitem utilizar este modelo de forma intuitiva

O modelo que utilizaremos é o **Stable Diffusion**, já que dentre as ferramentas de código totalmente aberto, esta é considerada a principal delas, cujo código fonte é disponibilizado publicamente pela Stability AI no Github, o que possibilita sua instalação e customização por desenvolvedores, artistas e pesquisadores.

Stable Diffusion é um modelo de inteligência artificial generativa desenvolvido para transformar textos em imagens. Ele pertence à categoria de modelos text-to-image (texto para imagem) e utiliza redes neurais avançadas para interpretar descrições textuais e gerar imagens coerentes com o conteúdo escrito.

O Stable Diffusion, no entanto, não possui uma interface gráfica oficial — ele é, essencialmente, um modelo que funciona por meio de código. Para torná-lo mais acessível a usuários que não têm familiaridade com programação, surgiram diversas interfaces gráficas que funcionam como “camadas visuais” sobre o modelo base. Essas interfaces permitem que o usuário interaja com o Stable Diffusion por meio de botões, menus e painéis intuitivos, facilitando o processo de geração de imagens.

Além do **Stable Diffusion**, temos diversos modelos de IAs que geram imagens. Entre os modelos proprietários, uma boa opção é o **Midjourney**, **Leonardo**, e os modelos **Imagen** e **Nano Banana** da Google. Os modelos Google podem ser testados no site Google AI Studio (https://aistudio.google.com/prompts/new_image) que permite a criação de cotas diárias de imagens. Já entre os modelos abertos, destacam-se o **Flux**, o **Qwen**.

Existem várias formas para testá-los. Você pode baixar cada modelo e instalar localmente em sua máquina através do site do Hugging Face e carregá-los no ComfyUI. Outra opção é testá-los nos próprios sites das plataformas com cotas e créditos gratuitos, ou assinar um planos de acordo com suas necessidades.

3.2 Interfaces gráficas para rodar o Stable Diffusion

Para facilitar o uso do Stable Diffusion, surgiram interfaces gráficas como o ComfyUI e Automatic1111, que são opções de código aberto populares e robustas. Ambas são disponibilizadas para download e instalação para rodar localmente, ou seja, utilizando os recursos do computador - como memória RAM e VRAM - onde foram instalados.

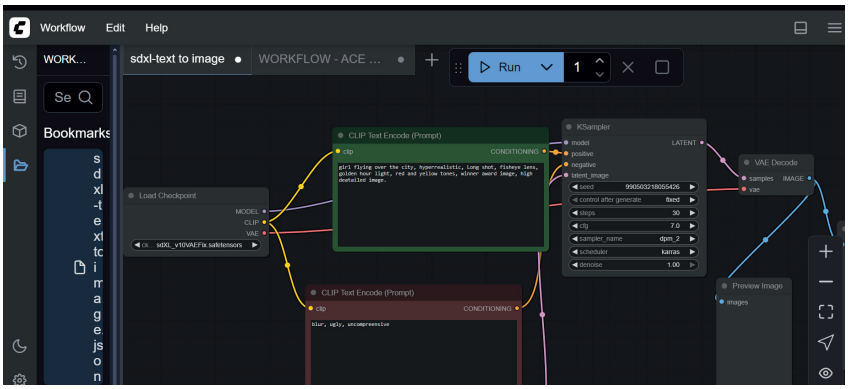
A maioria destas interfaces demandam uma especificação de hardware com memória RAM suficiente (geralmente maior que 8GB) e precisam de e placas gráficas (GPUs) com no mínimo 6GB de VRAM, o que pode ser um empecilho para rodar em computadores mais antigos ou com configurações modestas.

Como nem todos possuem computadores potentes o suficiente para rodar estas interfaces e modelos de IA, veremos que há a possibilidade de rodá-las em serviços na nuvem, que são, contudo serviços pagos que se assemelham a um aluguel remoto de um computador equipado adequadamente para rodar estas interfaces e gerar imagens com o Stable Diffusion.

3.2.1 ComfyUI

O **ComfyUI** (https://www.comfy.org/), que se diferencia por seu sistema baseado em nós (node-based), ou seja, o usuário pode montar fluxos personalizados conectando blocos que representam funções, como entrada de texto, carregamento de modelo, geração de imagem, aplicação de filtros, etc. Essa estrutura modular permite uma maior experimentação e flexibilidade, sendo especialmente útil para projetos mais complexos ou técnicos.

Figura 5 - Interface do ComfyUI



Fonte: Do autor.

3.2.2 Automatic1111

O **Automatic1111** (https://github.com/AUTOMATIC1111/stable-diffusion-webui) é uma interface web também open source que roda localmente no computador e oferece uma série de recursos adicionais, como seleção de modelos, ajuste fino de parâmetros, upscaling, entre outras. Todos os ajustes são acessados de forma visual, através de campos de texto, caixas de seleção e sliders, sem a necessidade de codificação.

A interface ComfyUI será abordada mais aprofundadamente nas aulas, mas encorajamos que vocês testem a interface Automatic1111 para descobrir com qual você mais se adapta.

3.2.3 Serviços para rodar as interfaces na nuvem

Para aqueles que não desejam ou não podem instalar esses recursos localmente, há alternativas baseadas em **nuvem**. Um exemplo é o **Google Colab**, que oferece notebooks prontos para uso do Stable Diffusion. O Google Colab, por exemplo, hospeda notebooks prontos que rodam o modelo em servidores temporários, bastando ao usuário executar comandos já configurados. A execução de notebooks funciona de graça, mas com severas limitações, sendo necessário pagar para obter melhores resultados.

Há alternativas pagas e mais práticas como o **Runpod** (https://www.runpod.io/), **QuickPod** (https://quickpod.io/), **Novita** (https://novita.ai/) e **Nordy** (https://nordy.ai) que apresentam fluxos prontos para usar do **Automatic1111** e **ComfyUI**.

3.2.4 Alternativas para rodar o Stable Diffusion e demais modelos de IA

Outra alternativa prática são os espaços interativos do **Hugging Face Spaces**, onde diversos modelos de IA, incluindo variações do Stable Diffusion, podem ser testados diretamente no navegador. Essa opção não exige instalação nem configuração prévia, sendo ideal para quem busca praticidade e acesso imediato a ferramentas de IA generativa, ela também oferec acesso às demais IAs citadas anteriormente.

Outras opções são usar os sites oficiais das prórpias plataformas, que oferecem planos gratuitos com algumas limitações de uso e planos de assinatura para diversos tipos de usuários.

Exemplos:

Qwen-Image (IA open source multimodal criado pela Alibaba) - https://quen.chat

Modelos “Imagen” da família Google (Modelos proprietários) - https://labs.google/fx/pt/tools/image-fx
https://aistudio.google.com/prompts/new_image



3.2.5 Modelos de IA para edição de imagens

Para quem busca controle avançado na edição de imagens, o Stable Diffusion é a referência open source, permitindo manipulações detalhadas através de comandos de texto (prompts) ou ferramentas complexas como ControlNets e LoRAs. No entanto, usar o Stable Diffusion dessa forma exige

um conhecimento técnico considerável e tempo para ajustes. É aqui que entram IAs mais recentes, como o **Nano Banana** e o **Qwen**, que simplificam tarefas que antes eram complexas: elas realizam edições intuitivas, como trocar roupas ou misturar rostos de maneira consistente, exigindo do usuário apenas comandos de texto simples, transformando o que era um processo técnico em uma experiência muito mais direta e fácil.

Nano Banana: O Mestre do Realismo Rápido (Proprietário - Google)

O Nano Banana é, na verdade, um apelido para o modelo Gemini 2.5 Flash Image do Google. Pense nele como uma ferramenta profissional de edição de imagem controlada por uma empresa (proprietária), onde você não vê o "segredo" do código. Sua maior força é a consistência facial e o realismo. Ele pode pegar duas ou mais fotos e misturá-las em uma nova cena de forma convincente, como se fosse uma foto real, mantendo os rostos dos personagens intactos. É ideal para edições complexas e rápidas, como mudar a roupa de uma pessoa ou o fundo de uma foto com um comando de texto simples, buscando agilidade e alta qualidade sem complicações técnicas.

Qwen: O Especialista em Texto e Flexibilidade (Open Source - Alibaba)

O Qwen-Image-Edit é o modelo de edição de imagens do Alibaba. Ele é open source (código aberto), o que significa que o código é público, e qualquer pessoa ou empresa pode usá-lo, modificá-lo e adaptá-lo. Ele se destaca na renderização de texto dentro das imagens. Se você precisa que a IA escreva palavras, frases ou logotipos em uma imagem, ele faz isso de maneira muito mais limpa e legível que muitos concorrentes. É melhor para desenvolvedores e usuários avançados que querem personalizar a ferramenta ou integrá-la a outros softwares (como o ComfyUI), valorizando a flexibilidade e o controle total sobre o resultado final.

Em resumo, se você quer velocidade e realismo fácil, use o Nano Banana. Se você precisa de controle, código aberto e excelência em texto, o Qwen é a escolha ideal e tudo por meio de promtps.

3.3 Programas de edição de imagem

Falaremos de alguns softwares que permitem criar ou editar imagens digitalmente, porém, a criação e edição é feita de forma "manual", através de procedimentos análogos aos processos físicos de criação de imagem, como desenhar com pincel, colorir, recortar partes da imagem, estampar, etc. Ou seja, a imagem é criada do zero ou então depende de uma imagem existente para que seja transformada.

Focaremos no uso destes softwares para esta segunda opção, ou seja, transformar uma imagem já existente, que, no caso, será a imagem gerada por meio de Inteligência Artificial. Utilizaremos estes softwares de edição em situações como a) a imagem gerada não sai exatamente da forma como desejamos, ou b) quando queremos adicionar elementos que não estão presentes na imagem ou c) quando queremos compor a imagem gerada com outras imagens, sejam imagens capturadas do mundo real ou outras imagens geradas por IA.

É importante saber que existem ferramentas diferentes para trabalhar os diferentes tipos de imagem que falamos, ou seja, imagens raster ou imagens vetoriais. É o que veremos a seguir.

3.3.1 Programas de edição de imagens rasterizadas

Aqui estão algumas ótimas opções gratuitas e de código aberto para quem busca alternativas aos programas de edição de imagens mais conhecidos.Para imagens raster (fotos, ilustrações, etc.), esses programas são ótimas alternativas ao Adobe Photoshop.

- **GIMP** (https://www.gimp.org/) - É a opção open source mais completa e versátil para edição geral de imagens, com uma ampla variedade de ferramentas.
- **Krita** (https://krita.org/pt-pt/) - Perfeito para quem se dedica à arte digital e animação, com funcionalidades específicas para pintura e desenho.
- **Photopea** (https://www.photopea.com/) - Uma solução poderosa que roda diretamente no seu navegador, sem precisar instalar nada. É ótimo para acesso rápido a um editor.
- **Pinta** (https://www.pinta-project.com/) - Ideal para quem busca algo simples e leve, focado em tarefas de edição mais rápidas e descomplicadas.
- **MyPaint -** Com uma interface limpa e fluida, é uma excelente escolha para quem quer se concentrar na pintura digital, sem distrações.

3.3.2 Programas de edição de imagens vetoriais

Para imagens vetoriais (logos, ícones, ilustrações técnicas, etc.), esses programas substituem o Adobe Illustrator e Corel Draw.

- **Inkscape** (https://inkscape.org/pt-br/) - Uma ferramenta completa e robusta, ideal para trabalhos vetoriais profissionais e complexos.
- **Vectr** (https://vectr.com/) - Mais simples e leve, permite colaboração em tempo real e pode ser usado no navegador ou como aplicativo.
- **SVGEdit (**https://svgedit.netlify.app/editor/index.html) - Uma opção 100% online, perfeita para fazer edições rápidas e pontuais em arquivos SVG.

4. Pesquisa artística em criação de imagens com IAs generativas

Apresentamos aqui dicas para construção inicial de uma pesquisa artística com IA generativa e softwares de edição de imagem rasterizadas, já pensando no que será colocado em prática na aula seguinte, com a experimentação da interface ComfyUI e uso de algum software para edição e imagens raster de sua preferência.

1. Conceba uma ideia ou conceito expressivo

Pense: o que você quer comunicar ou explorar por meio da imagem?

Pode ser algo pessoal (emoções, lembranças), social (temas como identidade, meio ambiente, tecnologia) ou imaginativo (fábulas, mundos alternativos, cenas surreais).

DICA

Escreva uma frase-síntese da sua ideia, como se fosse o título ou a "alma" da obra.

2. Escolha uma abordagem visual coerente com sua ideia

Isso significa pensar como sua ideia poderia ser traduzida visualmente.

Exemplo:

se for um tema nostálgico, talvez um estilo vintage; se for algo distópico, talvez um visual cyberpunk ou sombrio.

3. Pesquise inspirações visuais

Busque imagens, filmes, estilos de arte, artistas, movimentos artísticos ou produções visuais que se relacionem com sua ideia.

Registre 3 ou mais referências visuais que possam influenciar a composição e estética da sua imagem.

4. Delimite o estilo estético desejado

Alguns estilos que podem ser usados como referência nos prompts:

- **Realismo fotográfico**
- **Anime**
- **Arte digital futurista**
- **Pintura impressionista**
- **Estética vaporwave**
- **Arte surrealista**
- **Quadrinhos ou cartoon**

É possível misturar estilos! Exemplo: "personagem anime em uma pintura no estilo impressionista do século XVIII".

5. Planeje os elementos principais da cena

Liste os elementos que você quer que apareçam na imagem: personagens, cenários, objetos, clima, cores predominantes.

Exemplo:

"Uma mulher negra com cabelos trançados, de olhos fechados, flutuando no céu noturno, com planetas coloridos ao fundo e nuvens em formato de animais".

6. Escreva um primeiro rascunho de prompt (comando para a IA)

Use frases curtas e objetivas, com palavras-chave visuais.
Inclua o que você quer e também o que não quer.

Exemplo de prompt:

uma floresta onírica cheia de árvores brilhantes e lanternas flutuantes, no estilo anime, iluminação suave, 8K, sem texto, sem desfoque

7. Escolha uma estratégia de execução

Estabeleça como será feito o fluxo de criação:

Vai gerar só uma imagem por IA e retocar em softwares de edição?
Vai gerar várias variações e depois misturar em softwares de edição?
Vai criar colagens com fotos reais + imagens de IA?

Defina o ponto de partida e as ferramentas que pretende usar em cada etapa.

Essas etapas iniciais ajudam a transformar uma ideia vaga em um projeto artístico planejado, com intencionalidade estética e argumentativa — mesmo antes de gerar qualquer imagem.

VIDEOAULAS DISPONÍVEIS
NA ABA ”MÓDULO 3 - AULA 2”

5 - Geração de imagens por inteligência artificial

Agora aprofundaremos nosso conhecimento sobre as ferramentas de geração de imagens por Inteligência Artificial. Abordaremos a **engenharia de prompts**, ou seja, os comandos que damos às interfaces para delimitar como é a imagem que esperamos que seja gerada. Trataremos mais a fundo a interface gráfica **ComfyUI** - para o modelo **Stable Difusion** e seus **parâmetros** principais.

5.1 Engenharia de prompts

Um **prompt** nada mais é do que uma **descrição textual que orienta o modelo de inteligência artificial** sobre o que deve ser criado. É um **comando** por texto que instrui um modelo de inteligência artificial sobre o que gerar. No caso do Stable Diffusion, é a descrição escrita que orienta a criação da imagem desejada.

Engenharia de prompts é a prática de elaborar comandos textuais (prompts) **de forma estratégica** para obter resultados mais precisos, coerentes e criativos ao interagir com modelos de inteligência artificial generativa.

Esses modelos funcionam com base na entrada de texto fornecida pelo usuário, e a forma como esse texto é construído influencia diretamente a qualidade, o estilo e a fidelidade da imagem gerada.

Na prática, a engenharia de prompts envolve **selecionar cuidadosamente palavras-chave, ajustar a ordem dos termos, incluir descrições detalhadas e até combinar estilos artísticos, referências culturais, técnicas de iluminação, cores e composições visuais.** O objetivo é guiar o modelo para produzir uma imagem que atenda ao que o usuário deseja, minimizando ambiguidades e evitando interpretações inesperadas.

Por exemplo, um prompt genérico como “um gato” pode gerar resultados muito variados, enquanto um prompt mais elaborado, como “um gato preto sentado em uma poltrona vermelha, estilo realista, luz suave, fundo desfocado”, oferece ao modelo instruções muito mais claras e específicas.

Além disso, a engenharia de prompts não é uma ciência exata, mas sim uma habilidade que exige prática, testes e refinamento. Muitos usuários desenvolvem seus próprios estilos e estratégias, testando variações e analisando os efeitos de diferentes construções textuais. Com o tempo, é possível entender melhor como o modelo responde a certos termos e como pequenas alterações no prompt podem gerar mudanças significativas no resultado final.

Também é possível se inspirar em prompts feitos por outros artistas com base no resultado obtido, ou mesmo usar geradores de prompts, o que trataremos mais adiante.

5.1.2 Como criar um bom prompt

Para criar um bom prompt é essencial compreender que a qualidade da imagem gerada depende diretamente da forma como a solicitação é escrita. Quanto mais detalhado e claro for esse comando, mais fiel será o resultado àquilo que o usuário deseja.

O primeiro passo para elaborar um bom prompt é **definir o tema central da imagem**, ou seja, o que será representado. Em seguida, é importante **incluir detalhes visuais relevantes**, como cores, texturas, roupas, expressões faciais, poses, cenário, estilo artístico e iluminação.

A organização das palavras é outro ponto a se considerar com atenção. A ordem dos elementos pode influenciar o resultado, e certos termos técnicos ajudam a refinar a estética da imagem, como “realistic lighting” (iluminação realista), “ultra-detailed” (ultra detalhado), “cinematic angle” (ângulo cinematográfico), entre outros. Além disso, o uso de referências estilísticas, como “no estilo de Van Gogh” ou “arte cyberpunk”, contribui para definir o tom visual da produção.

Para criar um bom prompt no Stable Diffusion, é importante ser **claro, específico e descritivo**. Um bom prompt deve informar ao modelo exatamente o que você deseja ver na imagem. Para isso, siga estas dicas:

- **Use substantivos principais -** descreva o tema ou objeto central da imagem (ex: uma floresta, uma mulher medieval, um robô futurista).
- **Adicione detalhes visuais -** inclua elementos como cores, expressões, roupas, ambiente, iluminação, ângulo da câmera, etc.

Exemplo:

“Uma mulher de cabelo azul, usando kimono vermelho, em uma rua japonesa à noite, iluminação de neon.”

- **Indique estilo artístico (se desejar)** - você pode sugerir estilos como realista, pintura a óleo, anime, arte digital, estilo Van Gogh, etc.
- **Use adjetivos com moderação -** eles ajudam a qualificar, mas o excesso pode confundir o modelo.
- **Evite ambiguidades** - quanto mais direto e objetivo for o seu texto, mais próximo o resultado será do que você imaginou.

Exemplo de bom prompt:

“Um castelo medieval cercado por névoa ao amanhecer, estilo pintura digital, cores frias, montanhas ao fundo, luz suave, perspectiva frontal.”

Quanto mais você pratica, melhor entende como o modelo responde aos seus comandos. Por isso, testar variações e observar os resultados faz parte do processo criativo.

5.1.3 Prompts negativos

Prompts negativos são comandos utilizados para informar ao modelo de inteligência artificial o que se deseja **evitar** na imagem gerada. Ou , enquanto o prompt principal descreve o que deve aparecer, o prompt negativo específica elementos que não devem ser incluídos.

Essa técnica é muito útil para refinar os resultados e corrigir problemas recorrentes nas imagens, como distorções, objetos fora de lugar, estilos indesejados ou erros anatômicos. Por exemplo, se ao gerar uma imagem de um rosto o modelo costuma apresentar olhos deformados ou múltiplos dedos, é possível inserir no prompt negativo termos como “deformed eyes, extra fingers, bad anatomy, low quality” para reduzir ou eliminar esses problemas.

O uso de prompts negativos melhora significativamente a qualidade visual da imagem final, tornando o processo mais controlado e profissional. Eles funcionam como um **filtro,** limitando o que o modelo deve considerar na geração da imagem, e são especialmente eficazes quando combinados com prompts bem construídos e detalhados.

5.1.4 Gerador de prompts

Para quem busca ajuda para criar prompts mais eficazes para o Stable Diffusion, existem diversas ferramentas gratuitas e de código aberto que simplificam a tarefa. Elas funcionam como assistentes na arte da engenharia de prompts, oferecendo recursos que vão desde a construção de comandos detalhados até a busca por inspiração em vastas galerias de arte.

Em **2025**, os **geradores de prompts gratuitos ou de código aberto** (open source) mais populares para o uso com **Stable Diffusion** são:

- **Promptomania** (https://promptomania.com/) - totalmente gratuito e oferece uma interface interativa baseada em menus. Ele permite que o usuário selecione características como estilo artístico, iluminação, enquadramento, personagens e até prompts negativos.

- **Lexica Prompt Builder** (https://lexica.art/) - uma ferramenta gratuita que funciona como um banco de dados de imagens geradas com Stable Diffusion. Ela permite pesquisar por palavras-chave e visualizar os prompts utilizados em cada imagem, servindo como fonte de inspiração e aprendizado. Embora não seja open source, o uso é totalmente gratuito e sem necessidade de cadastro.

- **Gerador de prompts da Hugging Face** (https://huggingface.co/spaces/gokaygokay/FLUX--Prompt-Generator) - Gratuito e com código aberto ele gera automaticamente sugestões de prompts com base em temas diversos e pode ser utilizado diretamente no navegador, sem instalação.

- **PromptHero Generator** (https://prompthero.com/) - oferece acesso gratuito à sua base de dados de prompts e imagens, embora partes da plataforma sejam pagas. Ele permite a exploração de comandos organizados por estilo, modelo ou artista.

6 - Stable Diffusion

Como introduzido anteriormente, o **Stable Diffusion** é um modelo de inteligência artificial generativa desenvolvido para transformar **textos em imagens** nos mais diversos estilos, dependendo da forma como o prompt e os parâmetros da interface gráfica são estruturados.

Entre os **estilos** mais comuns e mencionados no material estão: **realismo fotográfico, pintura digital, arte conceitual, anime e mangá, ilustrações em 3D,** além de estilos inspirados em artistas clássicos e contemporâneos, como **Van Gogh** ou estilos específicos como **cyberpunk, barroco, steampunk** e **cartoon**. Essa variedade é possível graças à flexibilidade do modelo e à ampla base de dados com a qual foi treinado.

Outro aspecto importante é a modularidade do Stable Diffusion, que permite ao usuário combinar diferentes ferramentas, extensões e scripts — como LoRA, embeddings, e ControlNet — para refinar os resultados. Além disso, o modelo funciona localmente, o que garante maior privacidade e independência da nuvem. Ele também oferece suporte a tarefas como text-to-image, image-to--image, inpainting e outpainting. Falaremos mais sobre as diversas tarefas (ou tasks); parâmetros; ControlNet e LoRA mais adiante.

Licença:

Quanto à licença, o Stable Diffusion é distribuído sob a CreativeML Open RAIL-M License, uma licença open source que permite o uso, modificação e redistribuição gratuita do modelo, inclusive para fins comerciais, desde que o uso esteja em conformidade com princípios éticos.

6.1 Tarefas (tasks)

No contexto do Stable Diffusion, o termo **“tasks”** (ou “tarefas”, em português) refere-se a **funções específicas** ou processos que podem ser executados utilizando modelos de inteligência artificial. Essas tasks representam diferentes tipos de aplicações ou objetivos dentro do uso do modelo, não se limitando apenas à geração de imagens a partir de texto.

As tasks mais comuns associadas ao Stable Diffusion incluem:

- **Text-to-Image -** a mais popular, na qual o modelo gera uma imagem a partir de uma descrição textual (prompt).
- **Image-to-Image (img2img) -** gera uma nova imagem a partir de outra imagem, podendo manter ou modificar seu estilo, composição ou detalhes.
- **Inpainting -** permite editar ou completar uma imagem original, preenchendo áreas ausentes ou apagadas com base em um novo prompt.
- **Outpainting -** amplia uma imagem para além de seus limites originais, criando cenários contínuos a partir da borda.
- **Style transfer / remix -** recria uma imagem em um novo estilo artístico, mantendo a estrutura básica do conteúdo original.

Essas tasks são como “modos de uso” do modelo, e cada uma delas pode ser ativada ou configura

da por meio de interfaces gráficas como Automatic1111 ou ComfyUI. Algumas tasks são executadas com ferramentas complementares ou modelos adaptados, mas todas partem da arquitetura base do Stable Diffusion. **Como a proposta do curso é uma introdução aos modelos de IA, focaremos na task text-to-image.**

6.2 Parâmetros

Os **parâmetros** do **Stable Diffusion** são os **ajustes técnicos** que o usuário pode configurar para influenciar o resultado final da imagem gerada. Entender esses parâmetros é essencial para controlar melhor a qualidade, o estilo e a coerência da imagem produzida. A experimentação e o ajuste fino são partes fundamentais do processo de gerar imagens com mais controle, coerência e qualidade.

A seguir, estão os principais parâmetros deste modelo:

- **Prompt -** É o texto que descreve o que deve ser gerado. Pode conter palavras, frases, descrições visuais e referências estilísticas.
- **Negative Prompt -** Campo para inserir o que não deve aparecer na imagem (ex: “extra fingers, blurry, low quality”).
- **Steps (ou Sampling Steps) -** Define quantas etapas de refinamento a imagem vai passar durante a geração. Quanto maior o número de steps (ex: 20, 30 ou 50), maior tende a ser a qualidade e a fidelidade do resultado — embora também aumente o tempo de geração.
- **Seed -** É o número que define o ponto de partida do processo aleatório. Usar o mesmo seed com o mesmo prompt geralmente gera a mesma imagem, o que é útil para reprodutibilidade.
- **CFG Scale (Classifier-Free Guidance Scale)** - Controla o quanto o modelo deve seguir fielmente o prompt. Um valor baixo (ex: 3) dá mais liberdade ao modelo; um valor alto (ex: 12) faz com que a imagem siga o prompt mais rígidamente. O valor padrão recomendado costuma ser entre 7 e 10.
- **Resolution (ou Tamanho da Imagem) -** Define a largura e altura da imagem gerada, geralmente em pixels (ex: 512x512, 768x768). Resoluções maiores exigem mais memória da GPU.
- **Aspect ratio (proporção) -** é a relação entre a largura e a altura que define a forma e o enquadramento visual da imagem, como 1:1 (quadrada), 16:9 (horizontal) ou 9:16 (vertical).
- **Sampler -** É o algoritmo de amostragem usado no processo de difusão. Cada sampler (como Euler, DPM++, Heun, etc.) tem características próprias. Alguns geram imagens mais suaves, outros mais detalhadas. O curso recomenda testar diferentes samplers conforme o estilo desejado.
- **Model Checkpoint -** Refere-se ao modelo principal carregado na interface. Pode ser o modelo padrão (como Stable Diffusion 1.5, 2.1, SDXL) ou um modelo personalizado (como DreamShaper, Anything, Realistic Vision, etc.).

Muitos dos parâmetros utilizados no modelo Stable Diffusion também podem ser observados em outros modelos de criação de imagens generativas. São conceitos fundamentais que se repetem em diversos modelos. Entre eles estão os parâmetros de prompts; prompts negativos; seed; resolution; checkpoint. Explore outros modelos de criação de imagem generativa e observe as semelhanças de parâmetros.

7 - Detalhes sobre a interface ComfyUI

A ComfyUI é uma interface altamente flexível e visual para o Stable Diffusion, destacando-se por sua arquitetura baseada em nós (node-based). Essa lógica modular oferece controle total e visualização precisa de cada fase do processo, pois permite montar fluxos personalizados para diferentes tarefas de forma clara e adaptável.

Vantagens e Recursos Avançados

A interface é compatível com todos os recursos avançados do Stable Diffusion, sendo muito procurada por profissionais que buscam precisão. Embora a curva de aprendizado inicial seja um pouco maior, sua estrutura permite salvar, reutilizar e compartilhar fluxos de trabalho completos. Por ser open source e gratuita, ela garante maior controle sobre a privacidade e o desempenho, com a comunidade constantemente expandindo suas funcionalidades através de pacotes de nós personalizados.

Principais blocos (nós) da ComfyUI

Prompt Text (CLIP Text Encode) - É o nó onde o usuário escreve o prompt principal (o que deseja que apareça na imagem) e, se desejar, o prompt negativo (o que quer evitar). Esses textos são processados e transformados em vetores de linguagem que o modelo entende.

Load Checkpoint - Esse nó carrega o modelo base que será usado para gerar a imagem, como Stable Diffusion 1.5, 2.1 ou SDXL. Também é possível carregar modelos customizados, como DreamShaper, Realistic Vision, entre outros.

Sampler - Controla o algoritmo de amostragem, ou seja, como o modelo vai interpretar o ruído inicial e transformá-lo na imagem final. É onde se define o número de steps, o tipo de sampler (Euler, DPM++, etc.), e o CFG Scale (grau de fidelidade ao prompt).

Latent Image ou Empty Latent Image - Cria a imagem “em branco” (latente), geralmente com dimensões padrão (ex: 512x512). É o ponto de partida para o processo de difusão.

KSampler (ou outros tipos de Sampler) - Executa o processo de geração com base nos vetores de texto, no ruído inicial e nas configurações anteriores. É o “coração” da geração.

VAE Decode -Transforma a imagem do espaço latente em uma imagem visível e final, que pode ser exibida na tela ou salva no computador.

Save Image / Preview Image - Permite visualizar o resultado final, salvá-lo ou compará-lo com outras variações.

Além desses blocos básicos, a ComfyUI permite incluir nós extras para LoRA, ControlNet, Refiner, inpainting, img2img, upscaling e muitos outros efeitos. Cada nó tem configurações próprias e pode ser combinado com outros, o que torna a interface extremamente personalizável e poderosa, mas nesse curso não aprofundaremos nesses pontos.

7.1 - Como gerar imagens no ComfyUI

A interface do ComfyUI já abrirá já com uma série de blocos pré-montados, cada um é direcionado para determinado tipo de tarefa. Outra possibilidade é começar um novo worflow do zero, algo que pode ser desafiador para usuários iniciantes. Como este curso tem um caráter introdutório, focaremos em um worfflow que executa a task text-to-mage, ou seja, que cria imagens a partir de prompts de texto.

Para utilizar o **ComfyUI gratuitamente na nuvem**, uma possibilidade é acessar a plataforma Nordy.ai que até o momento disponibiliza créditos para o plano free. Você poderá acessar a tarefa task-to image no seguinte **link: https://nordy.ai/workflows/66b0bd05e8c05e0008c608d9**

Caso sua máquina tenha uma configuração robusta com uma placa de vídeo dedicada de 16GB ou mais, recomendamos i**nstalar o ComfyUI gratuitamente em seu dispositivo**. Você poderá fazer isso gratuitamente através do site **www.comfy.org**.



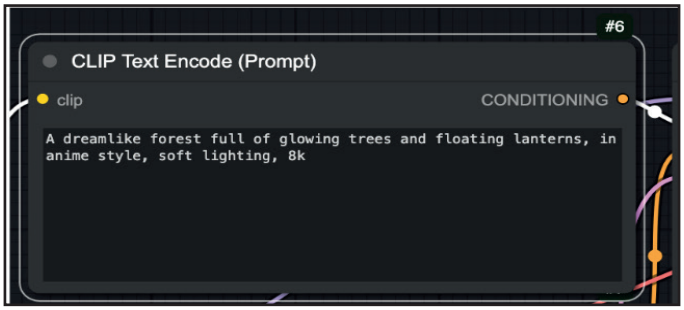
Agora é hora de praticar.
Bora pra cima!

1 - Dê zoom no quadro “CLIP text encode (Prompt)” que está ligado no nó “positive” (figura 6) do quadro KSampler. É neste quadro que você inserirá o prompt principal, sem os prompts negativos.

2 - Insira o prompt que você criou na aula passada sem o prompt negativo.

Importante - Como foi dito, alguns modelos foram treinados em inglês, então eles compreendem melhor os prompts neste idioma. Traduza seu prompt com o Google Tradutor e cole o prompt traduzido neste quadro.

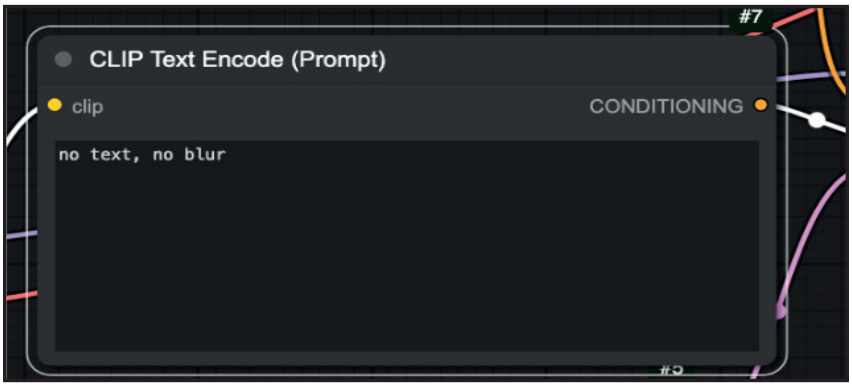
Figura 6 - Quadro para o prompt positivo.



Fonte: do autor.

3 - Agora dê zoom no quadro “CLIP text encode (Prompt)” que está ligado no nó “negative” (figura 7) do quadro KSampler. Note que também utilizei o exemplo da aula passada “sem texto, sem desfoque”, também traduzido para o inglês.

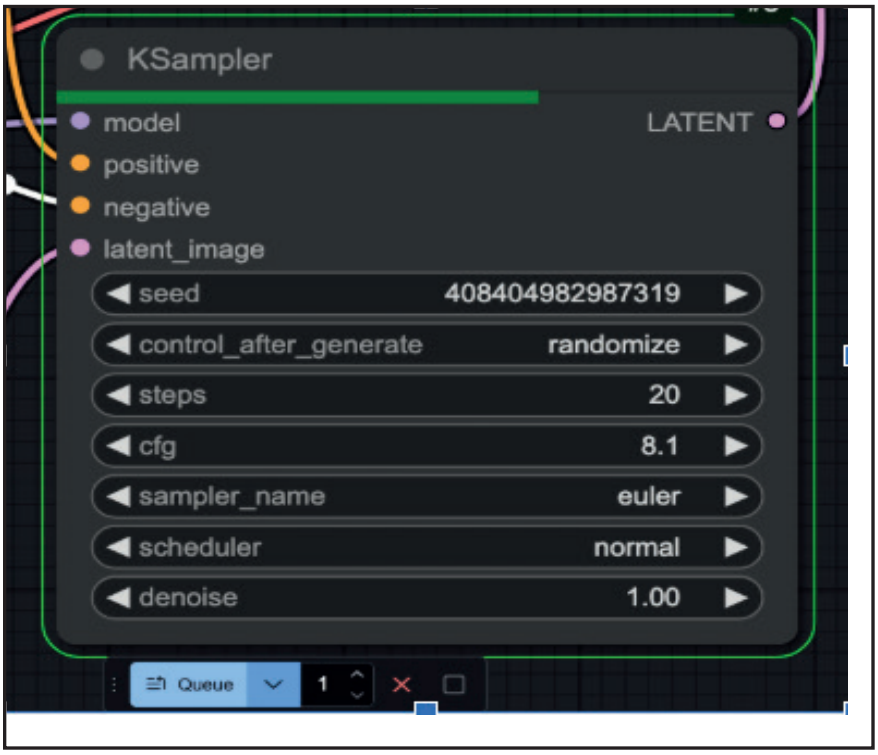
Figura 7 - Quadro para o prompt negativo.



Fonte: do autor.

4 - Clique no botão “Queue” (Em algumas interfaces poderá estar “RUN”) no canto inferior central da tela. Você poderá acompanhar a barra de progresso verde no bloco “KSampler” demonstrando que sua imagem está sendo gerada.

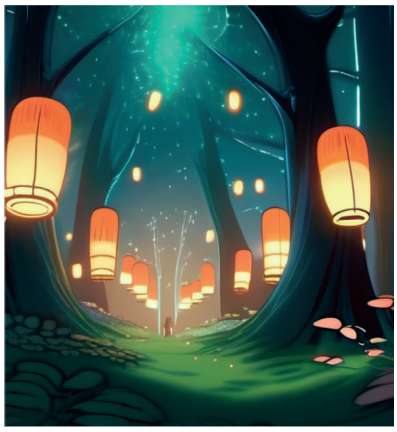
Figura 8 - Barra de progresso verde na parte superior do bloco KSampler.



Fonte: Do autor.

E pronto! Sua primeira figura foi gerada (figura 9).

Figura 9 - Imagem gerada pelo ComfyUI usando Stable Diffusion.



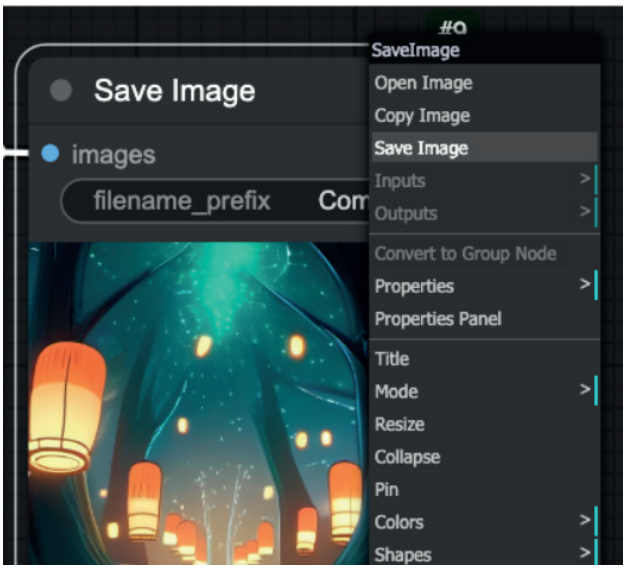
Fonte: do autor. Imagem gerada por IA

DICA

Clique seguidas vezes no botão “Queue” e note que uma nova imagem aparecerá, geralmente muito diferente da anterior, mesmo com os mesmos prompts. Um novo seed será gerado, transformando os resultados.

5 - Para salvar a imagem em seu computador, clique com o botão direito sobre a imagem no bloco “Save Image” (figura 10).

Figura 10 - Como salvar a imagem gerada pelo ConfiUI usando Stable Diffusion.

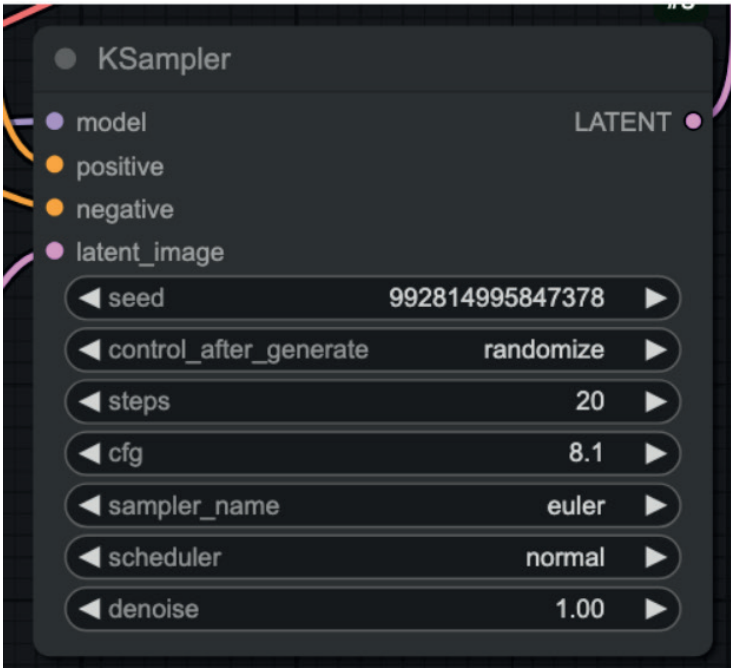


Fonte: do autor.

DICA

Retome a apresentação dos parâmetros do ComfyUI e expore livremente os parâmetros no quadro KSampler (figura 11).

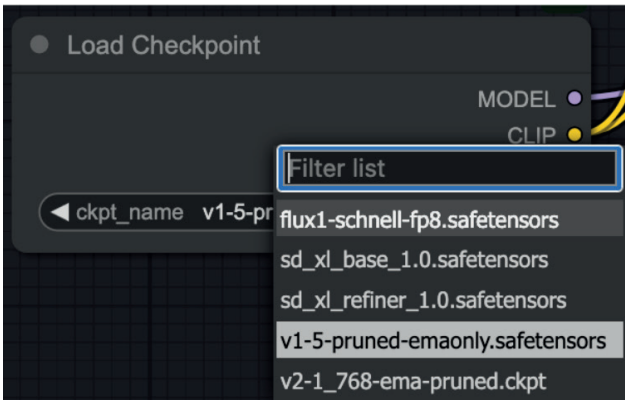
Figura 11 - Quadro KSampler e seus parâmetros.



Fonte: do autor.

6- No quadro “Load Checkpoint” (figura 12), explore outros checkpoints e veja o que resulta na imagem gerada. Lembre-se de sempre apertar o botão “Queue” para gerar uma nova imagem depois de mudar os modelos utilizados e demais parâmetros. No caso, foi utilizado o checkpoint de nome “v1-5-pruned-emaonly.safetensors”. Trocando de checkpont, você troca de modelo de IA utilizado que esteja instalado no dispositivo em utilização.

Figura 12 - Checkpoints no quadro Load Checkpoint.

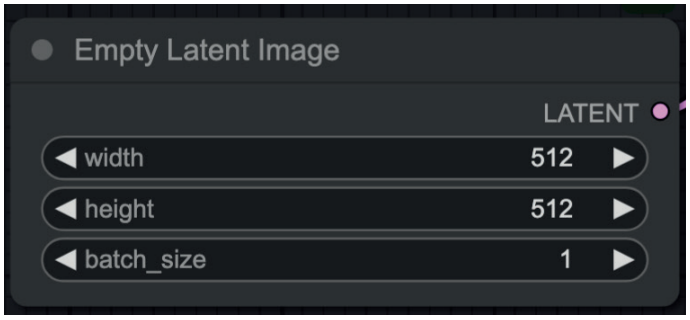


Fonte: do autor.

Lembre-se de sempre apertar o botão “Queue” para gerar uma nova imagem depois de mudar os parâmetros.

6 - Altere o tamanho da imagem no quadro Empty Latent Image (figura 13) modificando os parâmetros “width” (largura) e “height” (altura). Explore também o parâmetro “batch size”. Lembre-se de sempre apertar o botão “Queue” para gerar uma nova imagem depois de mudar os parâmetros.

Figura 13 - Quadro KSampler e seus parâmetros.



Fonte: Elaboração própria.

8 - Agora experimente outros prompts positivos e negativos em seus respectivos quadros, vendo qual imagem mais se aproxima da ideia que você planejou.

8 - Pós-produção

A pós-produção é o processo de **edição e aprimoramento da imagem** após sua criação inicial pela Inteligência Artificial. A imagem gerada pela IA é apenas o ponto de partida; a pós-produção garante que o resultado final tenha mais qualidade, fidelidade estética e usabilidade para projetos profissionais. Este processo pode ser feito em editores externos (como GIMP ou Krita) para ajustes de cor, contraste, remoção de imperfeições ou com os prórprios modelos de IA.

Você pode alterar imagens diretamente com prompts por meio de modelos de IAs como o **Nano Banana - Gemini 2.5 - flash** (https://aistudio.google.com/prompts/new_chat?model=gemini-2.5-flash-image-preview) ou **Qwen** (https://chat.qwen.ai/). A partir de tasks **image-to-image,** você poderá fazer o upload de uma imagem e alterá-la com prompts como **“troque a cor de determinado objeto”,** ou “**retire o personagem x”**. Essas são possibilidades que funcionam bem nesses modelos. Para tarefas mais complexas recomendamos a utilização de editores de imagem.

8.1 Upscaling

Uma das principais ferramentas de pós-produção é o **Upscaling**, que serve para ampliar imagens pequenas (ex: 512x512 pixels) para resoluções maiores (ex: 1024x1024 ou mais). O modelo gera imagens menores devido a limitações de hardware, mas o **upscaling**, usando métodos de IA, não apenas estica a imagem, mas também reconstrói detalhes e melhora a nitidez. Isso é essencial para aplicações em design, impressão e ilustração digital.
Alguns sites que oferecem upscale gratuito:

Nordy
Você poderá rodar um workflow no ComfyUI e fazer o upscale de sua imagem no navegador.
https://nordy.ai/comfyui/?flow=673faccc3c269af1b336c66b

Upscayl
Você pode baixar e instalar em seu dispositivo ou rodar na nuvem com créditos de IA.
https://upscayl.org

8.2 Padrões de cores

Para a finalização do seu arquivo, é crucial entender os padrões de cores: **RGB** e **CMYK**. Toda **imagem gerada por IA é nativamente RGB (Red, Green, Blue)**, o padrão ideal para projetos que serão vistos em telas digitais (computador, celular). Se o seu projeto será **impresso** fisicamente, você deve sempre converter a imagem para o **padrão CMYK** (Cyan, Magenta, Yellow, Key/Black) antes de enviar para a gráfica.

REFERÊNCIAS

AUTOMATIC1111. Stable Diffusion Web UI. GitHub, [s. d.]. Disponível em: https://github.com/AUTOMATIC1111/stable-diffusion-webui. Acesso em: 2 ago. 2025.

CIVITAI. Community-trained models and LoRAs for Stable Diffusion. [S. l.], [s. d.]. Disponível em: https://civitai.com. Acesso em: 2 ago. 2025.

COMPARECAMP. Vectr review: pricing, pros, cons & features. [S. l.], [s. d.]. Disponível em: https://comparecamp.com/vectr-review-pricing-pros-cons-features/. Acesso em: 2 ago. 2025.

COMFYUI. Node-based UI for Stable Diffusion. GitHub, [s. d.]. Disponível em: https://github.com/comfyanonymous/ComfyUI. Acesso em: 2 ago. 2025.

CURSO DE DISEÑO GRÁFICO. What is Vectr: main advantages and features. [S. l.], [s. d.]. Disponível em: https://curso-diseno-grafico.org/en/en/what-is-vectr-main-advantages-and-features/. Acesso em: 2 ago. 2025.

FILEFORMAT. Imagens raster vs vetoriais – uma breve comparação. Blog FileFormat, [s. l.], 18 nov. 2021. Disponível em: https://blog.fileformat.com/pt/image/raster-vs-vector-images-a-brief-comparison/. Acesso em: 2 ago. 2025.

GIMP. GIMP documentation – online GPL and GIMP bibliography. [S. l.], [s. d.]. Disponível em: https://docs.gimp.org/2.6/en/bibliography.html#bibliography-online-gpl. Acesso em: 2 ago. 2025.

GIMP. What is GIMP?. The Knowledge Academy, [s. d.]. Disponível em: https://www.theknowledgeacademy.com/blog/what-is-gimp/. Acesso em: 2 ago. 2025.
GITHUB. SVG Edit discussions. [S. l.], [s. d.]. Disponível em: https://github.com/SVG-Edit/svgedit/discussions. Acesso em: 2 ago. 2025.

GREAT LEARNING. GIMP tutorial for beginners. [S. l.], [s. d.]. Disponível em: https://www.mygreatlearning.com/gimp/tutorials. Acesso em: 2 ago. 2025.

HUGGING FACE. Spaces – prompt generator. [S. l.], [s. d.]. Disponível em: https://huggingface.co/spaces. Acesso em: 2 ago. 2025.

INKSCAPE. About Inkscape. [S. l.], [s. d.]. Disponível em: https://inkscape.org/about/. Acesso em: 2 ago. 2025.

KRITA. Official manual – Krita 5.2. [S. l.], [s. d.]. Disponível em: https://docs.krita.org/en/index.html. Acesso em: 2 ago. 2025.

LAION. OpenCLIP. [S. l.], [s. d.]. Disponível em: https://github.com/mlfoundations/open_clip. Acesso em: 2 ago. 2025.

LEXICA. Lexica.art – AI-generated images and prompts. [S. l.], [s. d.]. Disponível em: https://lexica.art. Acesso em: 2 ago. 2025.

MAGE.SPACE. Free text-to-image AI generator. [S. l.], [s. d.]. Disponível em: https://www.mage.space. Acesso em: 2 ago. 2025.

NPMJS. Package: svgedit. [S. l.], [s. d.]. Disponível em: https://www.npmjs.com/package/svgedit. Acesso em: 2 ago. 2025.

OPENAI. CLIP: learning transferable visual models from natural language supervision. 2021. Disponível em: https://openai.com/research/clip. Acesso em: 2 ago. 2025.

OPENALTY. Krita – free digital painting software. [S. l.], [s. d.]. Disponível em: https://www.openaltly.com/software/krita. Acesso em: 2 ago. 2025.

OPEN SOURCE INITIATIVE. The open source definition. [S. l.], [s. d.]. Disponível em: https://opensource.org/osd. Acesso em: 3 ago. 2025.

PCMAG. MUCHMORE, Michael. Photopea review – a no-cost online photo editor that almost rivals Photoshop. 2 abr. 2024. Disponível em: https://uk.pcmag.com/photo-editing/151687/photopea. Acesso em: 2 ago. 2025.

PHOTOPEA. How to use Photopea for beginners. GeekChamp, [s. d.]. Disponível em: https://geekchamp.com/how-to-use-photopea-for-beginners-photopea-tutorial/. Acesso em: 2 ago. 2025.
PHOTOPEA. How to use Photopea: a comprehensive guide for beginners. AI Pure, 25 set. 2024. Disponível em: https://aipure.ai/articles/how-to-use-photopea-a-comprehensive-guide-for-beginners. Acesso em: 2 ago. 2025.

PHOTOPEA. Wikipedia – Photopea. [S. l.], 2024. Disponível em: https://en.wikipedia.org/wiki/Photopea. Acesso em: 2 ago. 2025.

PINTA. Official website. [S. l.], [s. d.]. Disponível em: https://www.pinta-project.com/. Acesso em: 2 ago. 2025.

PINTA. Openalty – Pinta overview. [S. l.], [s. d.]. Disponível em: https://www.openaltly.com/software/pinta. Acesso em: 2 ago. 2025.

PROMPT HERO. Prompt generator for Stable Diffusion and Midjourney. [S. l.], [s. d.]. Disponível em: https://prompthero.com/prompt-generator. Acesso em: 2 ago. 2025.

PROMPTOMANIA. AI art prompt builder. 2025. Disponível em: https://promptomania.com. Acesso em: 2 ago. 2025.

ROMBACH, Robin et al. High-resolution image synthesis with latent diffusion models. arXiv, 2022. Disponível em: https://arxiv.org/abs/2112.10752. Acesso em: 2 ago. 2025.

SELFMADE DESIGNER. Learn graphic design steps. [S. l.], [s. d.]. Disponível em: https://selfmadedesigner.com/learn-graphic-design-steps/#project. Acesso em: 2 ago. 2025.

SELFMADE DESIGNER. What is Inkscape?. [S. l.], [s. d.]. Disponível em: https://selfmadedesigner.com/what-is-inkscape/. Acesso em: 2 ago. 2025.

SELFMADE DESIGNER. Which graphic design software to learn?. [S. l.], [s. d.]. Disponível em: https://selfmadedesigner.com/graphic-design-software-to-learn/. Acesso em: 2 ago. 2025.

STABILITY AI. Stable Diffusion GitHub Repository. 2022. Disponível em: https://github.com/CompVis/stable-diffusion. Acesso em: 2 ago. 2025.

SVGEDIT. Editor online. [S. l.], [s. d.]. Disponível em: https://svgedit.netlify.app/editor/index.html. Acesso em: 2 ago. 2025.

TEACHER NETWORK. Learn MyPaint. [S. l.], [s. d.]. Disponível em: https://teacher-network.in/OER/index.php/Learn_MyPaint. Acesso em: 2 ago. 2025.

THE KNOWLEDGE ACADEMY. What is Inkscape?. [S. l.], [s. d.]. Disponível em: https://www.theknowledgeacademy.com/blog/what-is-inkscape/. Acesso em: 2 ago. 2025.

UMATECHNOLOGY. SVG Edit: create and edit SVG images online. [S. l.], [s. d.]. Disponível em: https://umatechnology.org/svg-edit-create-edit-svg-images-online/. Acesso em: 2 ago. 2025.

VECTR. Official website. [S. l.], [s. d.]. Disponível em: https://vectr.com. Acesso em: 2 ago. 2025.
WIKIPEDIA. MyPaint. [S. l.], [s. d.]. Disponível em: https://en.wikipedia.org/wiki/MyPaint. Acesso em: 2 ago. 2025.

ÁUDIO

Competência:

Utilizar de maneira contextualizada os fundamentos da criação de áudios generativos com IAs, contemplando ferramentas proprietárias, ferramentas abertas e softwares livres..

Resultados de aprendizagem:

Aplicar demonstrativamente os fundamentos da da criação de áudios com IAs generativas de modo crítico e propositivo.

AULA 1

VIDEOAULAS DISPONÍVEIS NA ABA "MÓDULO 4 - AULA 1"

1 - INTRODUÇÃO

Neste módulo, daremos um rolê pelo mundo dos sons. Aprenderemos a criar faixas de áudio completas, desde o seu planejamento inicial e pesquisa até a finalização, utilizando inteligências artificiais (IAs) generativas e softwares livres.

Passaremos pela história das IAs no universo sonoro e veremos os princípios básicos e termos fundamentais para a compreensão do trabalho com áudio. Apresentaremos as diversas possibilidades de utilização do áudio em obras de arte de diferentes linguagens.

Exploraremos os principais efeitos para uso criativo em suas faixas e visitaremos modelos de IA generativa Open Source, além de softwares de edição que integram algoritmos avançados ao seu fluxo de trabalho. O objetivo é que você entenda como essas ferramentas podem auxiliar de forma colaborativa no processo de criação em áudio.

2 - UM BREVE PANORAMA DAS IAS NO MUNDO DOS SONS

A ideia de a música ser feita por máquinas não é nova! Sempre criamos ferramentas para nos ajudarem a expandir nossa capacidade musical, desde o violino e o piano até o simples sino dos ventos (Figura 1), que faz música sozinho com a força do vento. No passado, até compositores como Mozart usavam um "jogo de dados" para criar novas melodias de forma aleatória, um método

ligados ao mundo da música e do som, fator em comum com a computação.

Figura 1 – Sino dos ventos



Fonte: Imagem do autor (criação com IA generativa)

Depois de um tempo, a partir da década de 1940, artistas da Música Concreta expandiram as possibilidades da música, usando sons gravados do dia a dia (como o barulho de uma caneca ou de uma serra) como a principal matéria-prima para a criação. Essas técnicas, hoje são comuns em contexos diversos, como trilhas sonoras para obras audiovisuais, ou mesmo em videos virais de internet, no qual artistas tocam instrumentos inusitados para se expressar.

Na década de 1990, a popularização da computação digital e a música eletrônica, ampliaram esse universo, permitindo criar sons e ritmos totalmente novos diretamente no computador. Muitas polêmicas ainda dividem músicos puristas que prezam pelo som orgânico, derivado do virtuosismo do músico e os que abraçam as tecnologias digitais em suas produções e usam os computadores como colaboradores. Seja em performances de artistas da música eletrônica (Figura 1) com várias parafernalhas, como no show do tiozinho do churrasco (Figura 2) tocando seu teclado com um acompanhamento automático, diferentes tipos de IAs estão presentes nos processos.

Figura 2 – Performance de música eletrônica



Fonte: Imagem do autor (criação com IA generativa)

Figura 3 – Performance do “tio dos teclados”



Fonte: Imagem do autor (criação com IA generativa)

E não para por aí. A partir dos anos 2000, as inteligências artificiais além de possibilitar a criação de sons e músicas diretamente do computador ou de um celular, estão presentes nas recomendações de músicas de estilos semelhantes que recebemos ao utilizar plataformas de streaming, ou mesmo na identificação de voz presentes em assistentes virtuais, na tradução de discursos em tempo real, na criação de locuções para vídeos e muito mais.

As tecnologias que permitiram esses avanços no mundo dos sons são as **redes neurais, modelo computacional inspirado no cérebro humano, formado por camadas de “neurônios” interconectados que processam dados.** Ao receber uma grande quantidade de informações, a rede neural aprende a identificar padrões complexos, como harmonias, ritmos e tons.

Com o avanço das redes neurais especialmente a partir de 2010, se popularizam as inteligências artificiais generativas, que permitem que hoje estejam disponíveis para qualquer usuário, Interfaces como Udio, Suno AI e Eleven Labs, programas que conseguem aprender o estilo de milhares de músicas e gerar faixas completas, com letra e voz, de forma incrivelmente realista a partir de poucos cliques e prompts de texto.

Apesar de a IA assustar alguns artistas e makers, devemos pensar que as tecnologias estão envolvidas em todas as atividades do trabalho humano, desde ferramentas rudimentares até as IAs generativas que abordaremos nesse módulo.

3 - SOM E ARTE

O som é algo que faz parte de nossas vidas desde o útero materno. Mesmo pessoas com deficiências auditivas podem sentir as vibrações sonoras em ressonância com seu corpo. Além de um fenômeno físico e biológico o som é cultural e perpassa diversas linguagens artísticas. Nossa maneira de falar, os movimentos do corpo, o meio ambiente, emfim tudo produz algum tipo de som.

E esse contexto universal serve de matéria prima para as diversas potencialidades artísticas das sonoridades.

Quando falamos em arte e som, de primeira já nos surge a música, mas ela é só uma de suas inúmeras linguagens. Vamos conhecer algumas delas e ampliar o nosso repertório!

3.1 Tipos de artes do som



O som está relacionado à diversas linguagens, seja na **música,** nas **artes sonoras**, ou mesmo no **design sonoro** de obras audiovisuais. Todas elas se encontram nas fronteiras da **arte, ciência e tecnologia.** Vamos conhecer alguns dos tipos de artes do som:

Música

Música é a arte de organizar sons e silêncios ao longo do tempo, utilizando-os para criar uma forma de expressão estética, comunicativa ou lúdica. Dentro da música temos várias vertentes que contemplam desde a musica clássica à popular. Na música, alguns gêneros são mais dedicados ao entretenimento, já outros estão mais conectados com a pesquisa em arte ciência e tecnologia. Confira alguns gêneros mais experimentais:

Música Eletroacústica - Mistura sons eletrónicos (sintetizadores) com sons acústicos gravados. **Ex:** Morton Subotnick - *Silver Apples of the Moon* (primeira composição eletrônica de longa duração a ser lançada por uma gravadora clássica)
Musica concreta - Usa sons do dia a dia (ruídos de trem, conversas) e os manipula para criar uma composição. **Ex:** Pierre Schaeffer - *Étude aux chemins de fer* (composta por uma montagem e manipulação de gravações de ruídos de locomotivas e trens.)

Arte Sonora

A arte sonora surge da interdisciplinaridade entre a música e demais linguagens artísticas, como as artes visuais. O que conhecemos hoje como arte contemporânea é fruto de de diversos movimentos de vanguarda que expandiram as fronteiras do que se entende por arte. Grupos como o Fluxus, os Futuristas e muitos outros tornaram possível com suas ações o surgimento do que hoje chamamos de artes sonoras, isto é, uma arte derivada dos sons, mas sem a necessidade de melodia ou harmonia. Confira alguns tipos de arte sonora:

- Perfomance - Explora as possibilidades do corpo e podem se basear nas potencialidades sonoras do mesmo. Ex: Inês Terra - Língua Fora (A artista usa o próprio corpo como uma fonte sonora, explorando o som produzido pela língua em relação a outros ruídos e ao silêncio. A obra é uma performance que transforma o som mais íntimo em uma experiência artística e corporal.)
- Instalação Sonora - Obra de arte criada para um espaço específico, onde o som é o elemento central da experiência. Ex: Doug Aitken - Sonic Pavilion (A obra é um pavilhão circular de vidro no Inhotim que, através de microfones instalados em um poço de mais de 200 metros de profundidade, capta e amplifica em tempo real o som da Terra.)
- Paisagens Sonoras (Soundscapes)
- Paisagens Sonoras (Soundscapes) - Registros de ambientes sonoros para capturar a “identidade” de um local. Ex: R. Murray Schafer - The Vancouver Soundscape (gravação documental de 1973, dirigida por R. Murray Schafer, que catalogou os sons naturais, urbanos e as assinaturas sonoras da cidade de Vancouver)
- Poesia sonora - Poesia que foca na sonoridade das palavras em vez do significado **Ex:** Paulo Bruscky - Poazia (o artista explora as potencialidades sonoras de um sal de frutas efervescente)

4 - Pesquisa artística

Se você acha que pesquisa é coisa de cientista de jaleco, pense de novo. **Pesquisar em artes é uma aventura de criação e descoberta, e o laboratório pode ser seu ateliê, seu computador, a rua ou sua própria mente**. Vamos direto ao ponto para você entender como funciona.

A pesquisa artística, como já discutimos, é um trânsito entre o fazer e o pensar. Na arte sonora, seu “fazer” é a escuta, a gravação, a manipulação de frequências, timbres e espacialidades. Seu “pensar” é sobre o que esse som significa, como ele afeta nossa percepção do espaço, do tempo e da emoção.

Agora, insira a IA generativa nesse fluxo. Ela não é apenas uma ferramenta, como um microfone ou um software de edição. Ela é um agente intermediário, uma parceira de diálogo com sua própria lógica, seus próprios vieses e sua própria “escuta” algorítmica.

Pesquisar em arte é como ser um detetive da sua própria criatividade. Em vez de apenas criar algo bonito, você investiga uma ideia ou pergunta através do seu processo artístico. Seus experimentos, seus rabiscos, suas gravações e até seus erros não são falhas, mas sim as pistas e descobertas da sua investigação. Não existe uma fórmula pronta: você inventa seu próprio jeito de explorar, seja com tinta, com sons de objetos do dia a dia ou com qualquer outro material. A arte se torna uma aventura para entender o mundo, onde o caminho que você percorre ao criar é tão importante quanto a obra de arte final.

Hoje, você pode levar essa aventura para o mundo digital usando Inteligências Artificiais generativas como parceiras criativas. Para pesquisar com sons, por exemplo, você não precisa de um estúdio caro; seu laboratório se torna uma conversa com a IA. Você a provoca com textos (prompts) para que ela imagine e crie paisagens sonoras, ritmos estranhos ou timbres que ninguém nunca ouviu. A grande sacada é que os **"erros" e as respostas inesperadas da IA são suas maiores descobertas, revelando a lógica de uma "mente" não-humana.** Assim, sua pesquisa se transforma em uma exploração de novos universos sonoros.

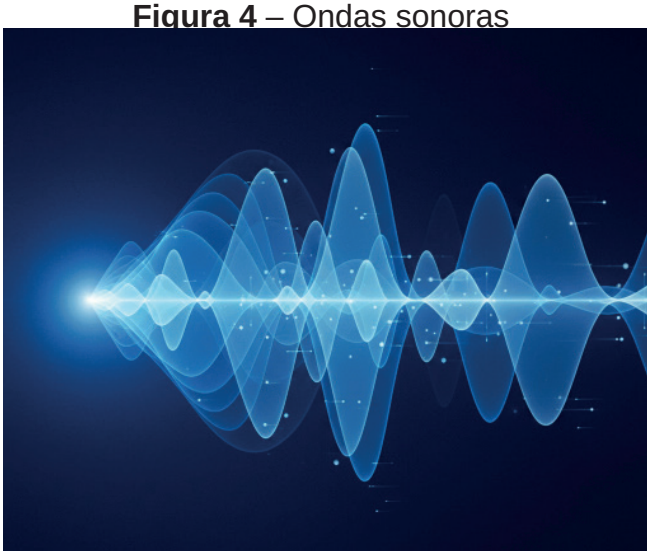
5 - Som e áudio

Estamos nos referindo a som e áudio, mas afinal, não é tudo a mesma coisa? Nada disso, apesar de usarmos essas palavras com frequência, cada uma tem o seu lugar. Vamos entender quais são as diferenças e principais características de cada uma delas.

5.1 Som

Som é uma onda de energia mecânica, longitudinal e tridimensional que se propaga através de um meio físico (figura 4).

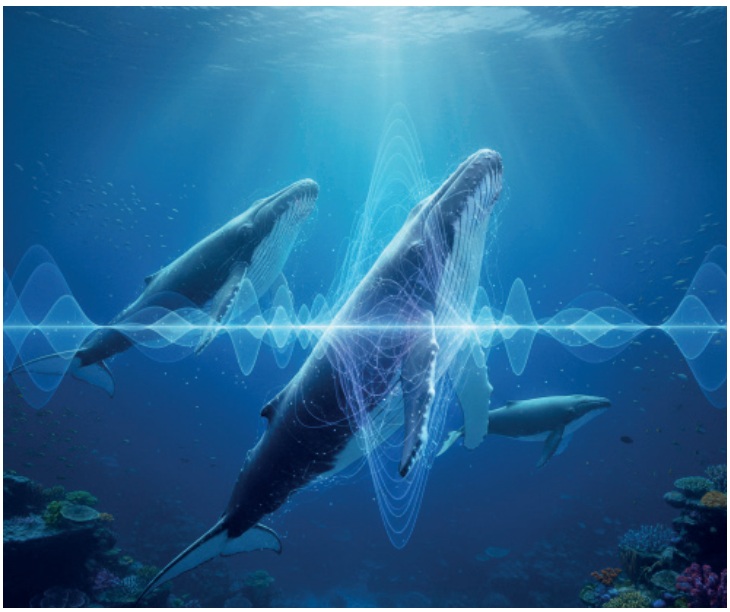
Figura 4 – Ondas sonoras



Fonte: Imagem do autor (criação com IA generativa)

Como exemplo de meio físico temos o ar, a água, ou um material sólido. Essa onda é gerada por uma vibração que empurra as moléculas de ar próximas a ela, criando diferenças de pressão que resultam no som. Um exemplo do som se propagando na água é o canto das baleias (Figura 5).

Figura 5 – Representação visual do canto das baleias



Fonte: Imagem do autor (criação com IA generativa)

Resumindo, ele é o resultado de uma vibração que se propaga e é percebida no nosso corpo, especialmente nos ouvidos. Mas também sentimos as vibrações do som em outros tecidos, como quando sentimos o pancadão bater no estômago.

5.1.1 Características do som

E quais são as principais características do som? Vamos ver agora, cada uma delas!

- Intensidade - A intensidade do som se refere à força das vibrações, ou seja, amplitude da onda sonora. Quando falamos para o DJ aumentar o som, estamos nos referindo à intensidade, que cresce ao girarmos o botão de volume.
- Altura (Frequência) - A frequência é o tamanho da onda. Frequências baixas (aqui nos referimos ao tamanho da onda e não à intensidade) caracterizam sons graves, isto é a voz de um locutor de rádio, um rugido de um leão, um trovão, um bumbo batendo. Já os sons altos, ou agudos são os sons da voz de uma criança, de uma mulher, de um apito, de pássaros cantando, de um violino. As frequências médias, ficam entre as altas e baixas, como um violão, um violoncelo ou a voz de um adolescente.
- **Timbre -** Timbre é a "**cara**", a "**identidade"** do som. É a característica que nos permite distinguir um som do outro. Quando tocamos uma nota musical, por exemplo um "lá" que vibra em 440hz, esses mesmos 440hz soarão de forma diferente em um piano, um saxofone, uma guitarra ou uma voz cantando. Conseguimos identificar os instrumentos por conta de seu timbre.

5.2 Áudio

O áudio é a materialização do som em um dispositivo (Figura. É o som que foi capturado, processado, armazenado ou transmitido por meios eletrônicos. Enquanto o som é a vibração física que se propaga, o áudio é a representação elétrica ou digital dessa vibração.

Figura 6 – Gravação de uma faixa de áudio em estúdio



Fonte: Imagem do autor (criação com IA generativa)

5.2.1 Analógico x digital

O áudio pode ser analógico ou digital e por mais que para os nativos digitais essa conversa não faça muito sentido, para músicos puristas, hipsters e admiradores do disco de vinil são fieis seguidores da sonoridade analógica.

Mas afinal, qual é a diferença entre o analógico e digital? O áudio **analógico** é a representação do som como um **sinal elétrico contínuo**, que varia de forma **idêntica à onda sonora original.** Presente em mídias como fitas cassete e vinis, ele é conhecido por sua sonoridade **“quente” e natural**, mas também é mais vulnerável a ruídos e à degradação ao longo do tempo. É considerado um som mais “vivo”, mais orgânico, apesar desta ser uma concepção bem subjetiva.

Já o áudio **digital** é a **representação do som usando código binário** (zeros e uns). O som é amostrado e transformado em dados numéricos, sendo o formato de áudio presente em **CDs, MP3s e serviços de streaming**. Suas principais vantagens são a reprodução precisa, a **resistência a ruídos** e a f**acilidade para edição e armazenamento**. Sim, o digital revolucionou a produção de áudio e se hoje estamos aqui estudando audio e IAs é graças ao áudio digital.

Antes do áudio digital, gravar algum registro fonográfico era para poucos. Com a popularização do áudio digital, qualquer um com um computador e uma ideia na cabeça passou a ter a chance de produzir suas próprias faixas. Hoje, com as IAs generativas, a produção de áudio se tornou ainda mais simples. Com um punhado de palavras e uma ideia na cabeça, nem o computador é mais necessário, qualquer um pode criar suas canções com seus celulares ou qualquer outro dispositivo digital que se conecte à internet.

6 - Produção de áudio

A **produção de áudio** é a arte e a t**écnica de capturar, criar, editar e manipular sons para gerar um produto final**, como uma música, um podcast, uma trilha sonora para cinema ou um efeito sonoro. É o processo que transforma ideias sonoras em realidade, seja por meio de gravações ou da criação de sons do zero.

O áudio produzido pode ter diversas aplicações que integram o campo das artes.

- Trilhas Sonoras - Criadas para filmes, séries, videogames ou peças de teatro, elas ajudam a contar a história e a evocar emoções no público. As trilhas sonoras de filmes, séries e jogos são formadas por três elementos principais que são a trilha musical, que inclui a música, diálogos e efeitos Sonoros.

DICA

Você pode usar as IAS generativas para criar efeitos diversos. O Stable Audio é uma boa ferramenta open source para tal finalidade.

- Produção Musical - O foco é a criação de canções para artistas, bandas ou playlists. Envolve desde a gravação dos instrumentos e vocais até a mixagem e masterização final. A produção musical é um processo de cinco etapas que são a composição e arranjo (Criação da melodia, letra e estrutura da música); gravação (captura dos sons de instrumentos e vocais); edição (limpeza das gravações, corrigindo erros e organizando as faixas); mixagem (Equilíbrio de todas as faixas e aplicação de efeitos para criar um som coeso) e masterização (polimento final para otimizar o som da música em qualquer sistema de áudio).

DICA

As IAs generativas são usadas na produção musical para compor novas melodias, criar arranjos complexos gerar batidas originais e sintetizar timbres únicos com modelos abertos como Magenta e proprietários como o Suno são bons nessas tarefas.

6.1 Processos de Produção

A forma como o som é criado é uma das distinções mais importantes na produção de áudio. As duas principais abordagens são o sampleamento e a síntese sonora e cada um dos processos tem suas peculiaridades, apesar de muitas vezes se misturarem.

6.1.1 Sampleamento (Amostragem)

O **sampleamento** é o processo de **gravar um som já existente para usá-lo na produção**. Um gravador digital captura um trecho de áudio (o “sample”), que pode ser uma nota de piano, uma batida de bateria, ou até um vocal. Depois, esse arquivo de áudio pode ser reproduzido, manipulado e usado como um instrumento em uma composição.

6.1.2 Síntese sonora

A **síntese sonora** é o processo de **criar um som do zero usando circuitos eletrônicos ou software**. Em vez de gravar, você constrói a onda sonora a partir de formas de onda básicas (como ondas senoidais, quadradas ou dente de serra) e manipula parâmetros como a frequência, a modulação e o timbre.

6.2 Efeitos sonoros

O som pode ser transformado para soar como se estivesse debaixo d’água, em um templo ou como um remix de DJ. Essas transformações são feitas por efeitos sonoros (FX), que se baseiam em três matrizes principais: Filtragem de Pente (Comb Filtering), Reverberação (Reverb) e Saturação (Saturation).

Filtragem de Pente (Comb Filtering)

A Filtragem de Pente ocorre quando dois sons idênticos são tocados com um pequeno atraso no tempo. Mesmo um atraso mínimo entre o som original e sua cópia já “muda a cara” do áudio. Os principais efeitos gerados por essa matriz são:

- **Delay -** Cria repetições do som, variando de um leve efeito de profundidade a um eco claramente audível, muito usado para dar ritmo extra.
- **Chorus -** Simula um coro (múltiplas fontes sonoras) cantando a mesma nota, adicionando corpo e profundidade ao som, comum em guitarras new wave (anos 80) e vocais.
- **Flanger -** Gera um som metalizado e cíclico, parecido com um avião passando. É dramático e cria texturas psicodélicas.
- **Phaser -** Parecido com o Flanger, cria um som robotizado e psicodélico ao manipular a fase (picos e vales) da onda sonora.

Reverberação e Saturação

Essas matrizes definem a espacialidade e a intensidade do som.

- **Reverberação (Reverb) -** É a ambientação do som, mostrando como ele reflete no ambiente. O reverb permite que o ouvinte perceba se o som está em um quarto, em um hall ou em uma igreja, dando dimensão ao áudio. O eco é um tipo de reverb longo.
- **Saturação -** É um aumento de sinal que adiciona calor, brilho e presença ao som, sendo uma forma sutil de distorção. Quando levada ao extremo, causa a distorção que faz o som parecer “rasgado”, como as guitarras de rock ou o som de um megafone.

Os efeitos sonoros são cruciais para a IA generativa. Eles podem ser incluídos nos prompts (comandos de texto) para que os modelos criem músicas com ambiências específicas (ex: “melodia com reverb de catedral”) ou aplicados na pós-produção para finalizar e editar o áudio gerado pela IA.

6.3 Softwares de produção de áudio

Para criar e produzir áudio, usamos softwares chamados DAWs (Digital Audio Workstations). Elas são como um estúdio de gravação virtual, permitindo gravar, editar e finalizar áudios de forma profissional.

Exemplos de DAWs

- Proprietárias (Pagas) - Pro Tools (padrão da indústria), Ableton Live (popular na música eletrônica) e FL Studio (muito usado no Hip Hop).
- Livres (Gratuitas) - Audacity (https://www.audacityteam.org/)

O Poder dos Plug-ins

As DAWs se tornam ainda mais poderosas com a instalação de aplicativos extras de terceiros, chamados Plug-ins.Eles funcionam como “acessórios” para expandir as funcionalidades da DAW. **Os Tipos de Plug-ins** Podem ser **efeitos clássicos** (como reverb e delay) ou **instrumentos virtuais** (simuladores de piano, amplificadores de guitarra ou baterias eletrônicas).

Há muitas empresas que produzem plug-ins de alta qualidade, como Native Instruments e Universal Audio (pagos), mas também há ótimas opções gratuitas, como as da Analog Obsession (https://analogobsession.com).



6.4 Softwares para produção de áudio generativo

O mercado da produção musical e audiovisual está sendo remodelado por ferramentas de inteligência artificial generativa. Confira algumas das plataformas que oferecem diversos serviços ao usuário.

Ferramentas proprietárias

Os modelos mais conhecidos, como **Suno AI** e **Udio**, são proprietários (pagos e de código fechado) e se destacam por criar músicas completas, com letra, voz e instrumentação, a partir de um simples comando de texto. A conveniência é alta, mas o usuário não tem controle sobre o

produção de audiobooks e vocais gerados por IA. Quem não se lembra daquela clássica voz irônica dos memes de vídeo que circulam nas redes sociais? É o Adam, um avatar de voz amplamente utilizado na plataforma ElevenLabs. Todas as plataformas oferecem planos gratuitos com algumas limitações, além de assinaturas com pacotes específicos de benefícios.

Ferramentas open source

Para ter controle total sobre a criação, a interface **ComfyUI** (já citada no módulo 3 deste curso) é uma ferramenta **open source (**código aberto) que usa um sistema de blocos interconectados (nós) para personalizar cada etapa do fluxo de trabalho da IA generativa de áudio, imagem e vídeo.Nele você poderá rodar modelos abertos diretamente em seu computador. Para os que desejam utilizá-lo diretamente no navegador, uma possibilidade é a computação em nuvem, como na plataforma runpod.

Para quem busca experimentação gratuita, o mercado também oferece boas opções, mesmo que não sejam de código aberto. O aplicativo **Speechma** é uma alternativa para a síntese de voz gratuita, permitindo a criação de áudio vocal sem custos. As plataformas **Labs** e **AI Studio** da Google também oferecem a possibilidade de criar locuções e faixas musicais diretamente no navegador.

Por fim, a comunidade de IA se apoia em plataformas de teste acessíveis. O **Hugging Face Spaces** e o **Google Colab** são ambientes online que permitem que qualquer pessoa use e experimente diversos modelos de IA de áudio (muitos deles open source) diretamente no navegador. Isso elimina a necessidade de instalar programas complexos ou ter um computador potente, tornando a exploração da tecnologia acessível a todos.

LINKS ÚTEIS

Suno AI - https://www.suno.com
Udio - https://www.udio.com
ElevenLabs - https://www.elevenlabs.io
Speechma - https://www.speechma.com
Google Labs - https://labs.google/fx/pt/tools/music-fx
Google AIStudio - https://aistudio.google.com/generate-speech
ComfyUI - https://www.comfy.org
Rundpod - https://www.runpod.io/

VIDEOAULAS DISPONÍVEIS
NA ABA ”MÓDULO 4 - AULA 2”

7 - MODELOS DE LINGUAGEM

A mágica da geração musical por IA se baseia em duas arquiteturas principais: os modelos T**ransformer** e os modelos de **difusão** (Diffuser). Os Transformers são essenciais para entender e gerar a estrutura sequencial e a linguagem musical (como notas e harmonias coesas). Já os modelos de Difusão (como o Diffrhytm) se concentram em criar o áudio final com alta fidelidade e realismo, eliminando ruídos e produzindo detalhes sonoros profissionais. É a combinação dessas tecnologias que permite que a IA não apenas componha, mas produza faixas prontas.

Para quem busca mais liberdade e controle, o movimento open source (código aberto) oferece alternativas. Projetos como o **Magenta (do Google), Stable Audio, AceStep, Diffrythm, Yue AI, Kokoro TTS** são alguns exemplos de modelos que podem ser utilizados gratuitamente e localmente por meio de interfaces como o ComfyUI. Os modelos também podem ser baixados e testados na plataforma Huggin Face.

7.1 Funções (tasks)

Confira os quatro principais tipos de modelos de IA para áudio generativo e suas funções (tasks):

1. Geradores de voz (Text-to-speech - TTS) - Comumente chamada de Text-to-Speech (TTS), esta IA transforma qualquer texto que você digita em fala de computador, com uma voz que parece bem humana. Ela é a base para assistentes virtuais como Siri e Alexa, narrações de vídeos e também é usada em ferramentas de acessibilidade para ajudar pessoas com deficiência visual a “ler” textos.

2. Geradores de música (Text-to-music) - Esses modelos são conhecidos como Text-to-Music (quando criam música a partir de uma descrição em texto) ou MIDI-to-Audio (quando convertem dados musicais em áudio). Sua função é compor músicas, melodias, batidas e até arranjos completos, tudo do zero. Servem para ajudar DJs e produtores a criar novas faixas, gerar trilhas sonoras para jogos e vídeos, e são perfeitos para quem quer experimentar novos estilos musicais.

3. Geradores de efeitos sonoros (Text-to-Audio) - Chamados de Text-to-audio (um termo mais geral para qualquer áudio gerado por texto), esses modelos criam sons específicos que você descreve, como o barulho de chuva, um trovão ou o som de uma porta fechando. São ideais para criar efeitos sonoros em jogos, filmes, animações ou para adicionar sons interessantes em podcasts e outras produções de áudio.

4. Transferência de estilo de áudio (Audio-to-Audio) - Essa IA pega o “jeito” de um som (como a

7.2 Parâmetros

Os parâmetros são o treinamento da IA (o que ela aprendeu) e as opções de ajuste que você tem para guiá-la. Eles são a chave para conseguir um áudio de qualidade e de acordo com os objetivos planejados.

A IA de áudio é como um estúdio de música superpotente, e os parâmetros são os botões que você usa para dizer exatamente como quer o som. Eles controlam a “receita” final. Ajustar os parâmetros é como ser o produtor musical. Você guia a IA (o estúdio) para criar exatamente o que você tem em mente.

Principais parâmetros de IAs de áudio:

- Prompt - É o comando principal. Sua descrição detalhada do conteúdo a ser gerado (gênero, instrumentos, emoção, atmosfera).
- Duração - É o tamanho do áudio. Define o tempo exato (em segundos) que a faixa terá.
- CFG Scale (Guidance Scale) - É o nível de obediência da IA.Controla o quanto a IA deve seguir seu prompt. Valor alto para rigidez; valor baixo para mais criatividade.
- Taxa de Amostragem (Sample Rate) - É a Qualidade HD do Som. Determina a fidelidade técnica. Taxas mais altas geram áudio mais detalhado e profissional, mas exigem mais tempo de processamento.
- Seed - É o Código de reprodução. Um número que, se repetido com os mesmos comandos, garante que o resultado será exatamente o mesmo. Essencial para refinar uma música que você gostou.
- Passos de Geração (Generation Steps) - É o esforço de refinamento da IA. Controla o número de etapas para aprimorar o áudio. Mais passos significam um som mais polido e detalhado, embora demore mais.

LINKS ÚTEIS

Stable Audio - https://huggingface.co/stabilityai/stable-audio-open-1.0
AceStep - https://huggingface.co/spaces/ACE-Step/ACE-Step
Diffrythm - https://diffrhythm.com/ai-music-generator
Kokoro TTS - https://huggingface.co/spaces/ACE-Step/ACE-Step

8 - ENGENHARIA DE PROMPTS

A engenharia de prompts é a **arte de dar as instruções certas para a inteligência artificial**, garantindo que o áudio gerado seja exatamente o que você tem em mente. É como se você fosse um diretor de som, e a IA, a sua equipe de produção. Em vez de simplesmente dizer “faça um som”, você precisa ser o mais detalhado possível, seja para criar uma música, um efeito sonoro, uma locução, uma trilha sonora, foley ou uma arte sonora.

Os elementos de um bom prompt para áudio generativo são como a receita para uma criação sonora precisa. Comece com o **conteúdo principal** (música, locução, foley), **defina a atmosfera** ou **emoção** (alegre, tensa), e, dependendo do tipo de áudio, **inclua detalhes específicos**. Para músicas, especifique gênero, instrumentação e estrutura. Para locuções, defina as características da voz, a intenção e as pausas. E para foley ou paisagens sonoras, descreva o contexto, as ações e os materiais. Combinar esses elementos resulta em um prompt detalhado que guia a IA para gerar exatamente o som que você imaginou.

8.1 Prompts postivos

- **Músicas e Efeitos Sonoros -** Comunique o tipo de áudio, gênero, emoção e qualidade de produção para criar faixas ou efeitos sonoros precisos. **Ex**: “Música de rock clássico, instrumental, com guitarra elétrica distorcida, bateria com groove e baixo marcante. O som deve ter uma atmosfera nostálgica e poderosa, soando como uma gravação de fita analógica dos anos 80.”
- **Locuções e Narrações (Voz) -** Descreva a identidade completa da voz (idade, tom, sotaque) e a intenção da fala, usando pontuação para guiar o ritmo. **Ex:** Voz feminina, jovem (20 e poucos anos), com tom animado e entusiasmado. A voz deve estar narrando uma propaganda de bebida energética. O texto é: ‘Prepare-se! (pausa dramática) O limite é o céu!’”
- **Trilhas Sonoras (BGM) e Efeitos (Foley) -** Para trilhas, defina a cena e o clímax; para foley, descreva a ação, o material e o contexto do som que precisa ser criado. **Ex:** “Crie uma trilha sonora de suspense crescente para uma cena noturna na floresta. A música deve usar cordas lentas e notas de piano isoladas. Adicione o foley de galhos secos quebrando sob passos leves a cada 5 segundos.”
- **Artes Sonoras e Paisagens Sonoras -** Crie um ambiente sonoro descrevendo o local, os elementos específicos e a atmosfera emocional que devem compor a paisagem. **Ex:** “Crie uma paisagem sonora urbana e caótica, com o som de buzinas misturadas com sussurros de pessoas em uma língua desconhecida. O ambiente deve evocar um sentimento de solidão em meio à multidão.”

8.2 Prompts negativos

Assim como na criação de imagens, os **prompts negativos** são essenciais para refinar a geração de áudio. Eles servem para dizer ao modelo de IA o **que não incluir na saída final,** ajudando a remover ruídos indesejados, instrumentos que não se encaixam ou atmosferas que não combinam.

Alguns modelos de áudio generativo não tem um campo de prompts negativos em separado. Caso isso ocorra, você pode incorporá-los no prompt principal usando palavras como “sem” ou “exceto”.

Exemplo: “Música de rock clássico, com riff de guitarra pesado, bateria energética, sem solo de piano, sem vocais masculinos.”

8.3 Gerador de prompts

Para facilitar os processo de criação de prompts, podemos utilizar geradores de prompts, aplicativos criados para produzir comandos textuais para modelos de IAs generativas. Temos alguns modelos proprietários como o Suno Prompt, o gerador do ChatGPT bem como podemos criar geradores de prompts em LLMs como o Gemini, e Qwen.

Gerador de prompt online

https://sunoprompt.com/
https://chatgpt.com/g/g-y2J4bAO8A-ai-music-prompt-generator
https://poe.com/StableAudioGenerator

DICA

Você também pode usar o gerador de prompts que criamos.
Acesse pelo link:
https://chat.qwen.ai/s/deploy/869cbb6e-ce73-4078-87e5-c82c4de53130

9 - PÓS-PRODUÇÃO



A **pós-produção** de áudio é a etapa final e crucial na criação de qualquer material sonoro, seja para filmes, vídeos, podcasts ou músicas. É o processo de **refinar, aprimorar e montar todos os elementos de áudio para que o resultado final** seja coeso, claro e impactante.

Isso envolve uma série de tarefas, como a **edição**, onde se corta e organiza os sons; a **mixagem,** que balanceia os volumes e frequências para garantir que todos os elementos (diálogos, músicas, efeitos) sejam ouvidos com clareza; e a **masterização**, que otimiza o áudio final para diferentes plataformas, garantindo que ele soe consistente em qualquer dispositivo. Em essência, a pós-produção transforma a gravação bruta em um produto polido e profissional, pronto para o consumo público.

Um detalhe importante é que a pós-produção de áudio é algo trabalhoso e desafiador para iniciantes, pois além de um conhecimento específico na área de engenharia de áudio é necessário treinar os ouvidos e ter disponíveis equipamentos de alta fidelidade sonora, além de um ambiente propício como um home-studio ou um estúdio profissional. Portanto, neste módulo focaremos na pós-produção realizada com ferramentas de IAs específicas.

9.1 Pós-produção em softwares de IA generativa

A pós-produção em softwares com foco nas inteligências artificiais são possibilidades que democratizam a produção de áudio para o público geral. Modelos que utilizam de diversas tasks, como a análise de áudio, transferência de estilo entre outras funções trazem aos usuários a chance de fazer suas gravações e gerações soarem de forma profissional, ou o mais próximo desse ponto.

Aplicativos proprietários bem como os open source oferecem diferentes soluções e algoritmos que entregam serviços de mixagem, masterização, separação de instrumentos, entre diversas funções que você poderá descobrir navegando em cada uma das diversas plataformas que estão em constante atualização.

As plataformas de referência na pós-produção de áudio são **proprietárias** como a **Suno**, **Landr**, **Moises** e **BandLab** que funcionam por meio de assinaturas mensais ou anuais. Algumas delas oferecem soluções gratuitas para determinadas funções.

- **Para limpeza de voz -** Se você precisa limpar narrações ou vocais, o ElevenLabs Voice Isolator (https://elevenlabs.io/voice-isolator) é excelente. Ele remove ruídos de fundo e ecos, deixando o áudio da voz cristalino. O Elevenlabs pode ser baixado como app diretamente em dispositivos móveis nas lojas officiais de aplicativos.
- **Para mixagem -** A plataforma Moises (https://studio.moises.ai/library/) permite que você separe faixas individualmente de forma gratuita no plano free. Uma boa opção para utilizar em mixagens, alterando elementos de cada faixa individualmente.O Moises pode ser baixado como app diretamente em dispositivos móveis, nas lojas officiais de aplicativos.
- **Para masterização -** O BandLab Mastering (https://www.bandlab.com/mastering) é uma ferramenta fantástica e gratuita que faz a masterização do seu áudio. Ele ajusta o volume e o equilíbrio do som para que sua música ou gravação soe bem em qualquer lugar. O bandnlab pode ser baixado como app diretamente em dispositivos móveis nas lojas officiais de aplicativos.

No modelo **open source**, ou com licenças mais permissivas temos a **Songmastr**, o **Audacity** (que apesar de não ser uma ferramenta dedicada ao uso de IAs generativas conta com plugins que cumprem esta função como o **OpenVINO** entre vários outras empresas do mercado, além de modelos de linguagem que cumprem funções úteis em pós-produção como o **Music Mixing Style Transfer,** e o **SonicMaster.**

- **Para mixagens e restauração de áudio -** O SonicMaster é uma opção de modelo de linguagem que cumpre diversas funções. Você faz o upload da faixa e acrescenta seus obejtivos no prompt, como por exemplo “abaixe os vocais”, “retire os ecos”, “equilibre a gravação”, “aumente os agudos”, “reduza os graves”. Os comandos devem ser digitados em inglês. Você poderá usar

- o google tradutor ou uma LLM para te auxiliar. O modelo pode ser rodado localmente ou utilizado no navegador através da plataforma Hugging Face Spaces (https://huggingface.co/spaces/amaai-lab/SonicMaster)
- **Para masterização -** O Songmastr é uma opção Open Source que disponibiliza a masterização em cotas gratuitas ou por meio da compra de créditos de IA. Acesse pelo navegador, faça o upload do áudio a ser masterizado e o upload de um áudio de referência para que a mágica aconteça. Acesse pelo navegador através do link https://www.songmastr.com/.
- **Para rodar localmente -** O OpenVINO é um pacote de plug-ins que oferece diversas funções de podução e pós-produção de áudio. Com ele você pode fazer transcrições de áudio para texto, retirar ruídos de gravações, gerar músicas e tudo isso diretamente de dentro do Audacity, uma DAW open source multiplataformas. O OpenVino está disponível para Windows e MACOS e pode ser instalado por meio do link https://www.audacityteam.org/download/openvino/. (É necessário ter o Audacity instalado em sua máquina)

9.2 Formatos de áudio

Após os processos de edição, mixagem e masterização do áudio é hora de exportá-lo para a execução em dispositivos diversos. Para isso, contamos com diversos formatos de arquivo. Os formatos de áudio são a forma como o som é armazenado digitalmente, e cada um tem características distintas que afetam a qualidade, o tamanho do arquivo e a compatibilidade. A escolha do formato ideal depende do objetivo, seja para streaming, arquivamento ou edição profissional.

- **MP3 -** O MP3 é o formato mais popular e universalmente compatível. Sua principal característica é a compressão com perda de dados, que reduz drasticamente o tamanho do arquivo ao remover informações sonoras que o ouvido humano raramente detecta. Ideal para streaming e armazenamento em dispositivos móveis.
- **FLAC -** O FLAC é um formato de áudio sem perda de dados. Ele comprime o arquivo sem descartar nenhuma informação da gravação original, garantindo a qualidade perfeita do som. Ideal para arquivamento de coleções de música, para audiófilos e para distribuição de áudio de alta fidelidade.
- **WAV -** O WAV é um formato de áudio sem compressão, ou seja, ele não perde nenhuma informação, mantendo a qualidade máxima da gravação. É o padrão em estúdios de gravação, mas é inviável para streaming e uso cotidiano devido ao seu tamanho.

E chegamos ao fim deste módulo!

Bora produzir!

Assista aos vídeos das aulas para conferir na prática como gerar e finalizar suas obras.

REFERÊNCIAS

ALMEIDA, José Roberto P. de. Inteligência Artificial em Música (Parte 1). Musicologia, 2023. Disponível em: https://www.blogs.unicamp.br/musicologia/2023/07/17/inteligencia-artificial-em-musica-parte-1/. Acesso em: 5 out. 2025.

ALMEIDA, José Roberto P. de. John Chowning, o músico que "tropeçou" na síntese FM. Musicologia, 2019. Disponível em: https://www.blogs.unicamp.br/musicologia/2019/11/20/35/. Acesso em: 5 out. 2025.

AUDACITY TEAM. Audacity with OpenVINO AI Effects now on Mac. Audacity Blog, 2024. Disponível em: https://www.audacityteam.org/blog/openvino-mac/. Acesso em: 5 out. 2025.

DE NEGRI, Fernanda. Inteligência artificial e redes neurais. IPEA, 2019. Disponível em: https://www.ipea.gov.br/cts/pt/central-de-conteudo/artigos/artigos/106-inteligencia-artificial-e-redes-neurais. Acesso em: 5 out. 2025.

LUNA, Ianni Barros; CORREA, Antenor Ferreira. Arte sonora: atravessamentos entre som e imagem. In: SIMPÓSIO INTERNACIONAL DE INOVAÇÃO EM MÍDIAS INTERATIVAS, 5., 2018, Goiânia. Anais [...]. Goiânia: UFG, 2018. Disponível em: https://files.cercomp.ufg.br/weby/up/777/o/6_-_Ianni_Barros.pdf. Acesso em: 4 out. 2025.

GIBSON, David. The art of mixing: a visual guide to recording, engineering, and production. routledge, 2019.

RUSSELL, Aden. FM Synthesis: A Simple, No-BS Guide. EDMProd, 2024. Disponível em: https://www.edmprod.com/fm-synthesis/. Acesso em: 5 out. 2025.

RUSSELL, Aden. The Ultimate Guide to Subtractive Synthesis. EDMProd, 2024. Disponível em: https://www.edmprod.com/subtractive-synthesis/. Acesso em: 5 out. 2025.

TEIXEIRA, Marcos Almada; PALAZZO, L. A. M. O uso de redes neurais artificiais na composição de trilhas sonoras para jogos. In: CONGRESSO DA ASSOCIAÇÃO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM MÚSICA, 33., 2023, Rio de Janeiro. Anais [...]. Rio de Janeiro: UFRJ, 2023. Disponível em: https://anppom.org.br/anais/anaiscongresso_anppom_2023/papers/1628/public/1628-7842-1-PB.pdf. Acesso em: 5 out. 2025.

VÍDEO

Competência:

Utilizar de maneira contextualizada os fundamentos da criação de vídeos generativos com IAs por meio de softwares livres.

Resultados de aprendizagem:

Aplicar de forma demonstrativa os fundamentos da criação de vídeos generativos com IAs por meio de softwares livres, de modo crítico e propositivo.

AULA 1

VIDEOAULAS DISPONÍVEIS NA ABA "MÓDULO 5 - AULA 1"

1 - INTRODUÇÃO

Neste módulo, abordaremos ferramentas de IAs Generativas e softwares livres para a geração de vídeos artísticos, mas, além disso o pensamento crítico será nosso eixo fundamental. Como as ferramentas tecnológicas podem servir à expressão de nossas narrativas e identidades? Neste curso, a IA é o nosso objeto de estudo, mas a pergunta fundamental permanece: de que forma essa tecnologia impacta nossa capacidade de criar e, mais importante, quem controla as narrativas que ela ajuda a gerar? Com base no pensamento de Paulo Freire (1996), que nos ensina a ler o mundo para além das palavras, convido vocês a "ler" a IA não como uma caixa-preta mágica, mas como um sistema criado por pessoas e empresas, com seus próprios interesses e vieses. "Ensinar não é transferir conhecimento, mas criar as possibilidades para a sua própria produção ou a sua construção". Aqui, nossa "produção" será audiovisual, mas também será crítica. A geração de vídeos por IA, que, como aponta Ivana Bentes, remixa o imenso arquivo cultural existente, levanta questões éticas centrais sobre direitos autorais e a reprodução de preconceitos sociais.

Para trabalharmos de forma crítica, traçaremos um breve panorama da história das IAs na linguagem do vídeo e veremos os princípios básicos para a compreensão do seu funcionamento. Exploraremos o trabalho criativo com colaboração das IAs generativas e aboradaremos ferramentas livres, e modelos open source a fim de conquistar soberania e privacidade no fazer.

2 - UM BREVE HISTÓRICO DAS IAS NO MUNDO DOS VÍDEOS

A história da IA generativa de vídeo começa com a tecnologia de geração de imagens. Os primeiros modelos, como as GANs (Redes Adversariais Generativas), eram excelentes para criar fotos rea

listas, mas esbarravam no obstáculo fundamental para vídeos: a **consistência temporal**. A notável exceção foi sua eficácia nos **deepfakes** (troca de rostos), uma tarefa mais controlada que não exigia a criação de um mundo coerente do zero. Fora dessa aplicação específica, a incapacidade da IA de manter a coerência de objetos e cenários entre os quadros gerava vídeos curtos e instáveis, marcados por uma deformação visual constante conhecida como flickering.

A solução para a inconsistência veio com arquiteturas mais sofisticadas, como os modelos de difusão e os Transformers. Projetados especificamente para compreender sequências e contexto temporal, esses modelos permitem que a IA gere quadros que evoluem de forma lógica e coerente. Esse avanço tecnológico possibilitou ferramentas como o Gen-2 da Runway e, mais notavelmente, o Sora da OpenAI, que estabeleceu um novo padrão de qualidade com vídeos longos, estáveis e com uma física de mundo surpreendentemente realista. Contudo, esse salto qualitativo impôs um novo desafio: o gargalo de processamento. A criação de vídeos de alta fidelidade demanda um poder computacional extremo, tornando o custo e o acesso em larga escala as principais barreiras atuais da tecnologia.

3 - AS ARTES DO VÍDEO

Desde a sua popularização, o vídeo transcendeu a função de mero registro para se tornar um dos campos mais dinâmicos e versáteis da expressão artística. Utilizando a imagem em movimento como matéria-prima, artistas exploram diferentes formatos para construir narrativas, investigar a própria tecnologia ou criar experiências imersivas. Embora compartilhem a mesma base técnica, cada vertente das artes do vídeo possui linguagens, circuitos e propósitos radicalmente distintos.

Figura 1 - Projeção mapeada



Fonte: do autor - imagem gerada por IA

- **Cinema -** Utiliza a linguagem audiovisual para construir narrativas complexas e de grande apelo popular.
- **Videoarte -**Subverte as convenções tradicionais em favor da exploração conceitual e estética do próprio meio, como na obra pioneira “TV Buddh” de Nam June Paik.
- **Videoinstalação -** Expande o vídeo para o espaço tridimensional, criando ambientes imersivos onde o espectador interage com a obra, como na projeção dupla de “The Crossing” de Bill Viola.

- **Videoclipe -** Combina som e imagem de forma criativa para promover uma música, muitas vezes atingindo o status de curta-metragem artístico, como no icônico “Thriller” de Michael Jackson.
- **Videomapping -** No videomapping, a “linguagem do vídeo” refere-se à maneira como os elementos visuais e sonoros são projetados para criar uma narrativa ou experiência imersiva em superfícies não convencionais. É uma fusão de técnicas cinematográficas e design espacial.
- **Documentário -** Emprega a imagem em movimento para registrar e interpretar a realidade com um ponto de vista autoral, como no ensaio visual e sonoro de Godfrey Reggio em “Koyaanisqatsi”.

Figura 2 - Videoinstalações



Fonte: do autor - imagem gerada por IA

4 - PESQUISA ARTÍSTICA

O artista criador, deve agir como agentes críticos e transformadores. A Inteligência Artificial generativa não é uma força neutra; ela está moldando o futuro da arte, da cultura e da comunicação.

Dominar essas ferramentas significa também compreender suas implicações éticas. Significa questionar quem se beneficia com elas, como os dados são governados e qual o nosso papel em garantir que essa revolução tecnológica caminhe para um futuro mais justo, diverso e criativo. A jornada criativa com a IA é contínua, e cada projeto é uma oportunidade não apenas para explorar novas fronteiras estéticas, mas para afirmar nossa responsabilidade ética no mundo digital.

A pesquisa artística com IA não busca apenas criar imagens belas ou impactantes; ela busca usar essas imagens para dizer algo sobre o mundo que essa IA está criando.

Em resumo, a pesquisa artística no contexto da IA generativa transforma o ofício do artista. Ele deixa de ser apenas um criador de formas para se tornar um pesquisador crítico de sistemas, cujo ateliê é tanto um laboratório estético quanto um campo de investigação ética e social. A obra de

arte se torna a publicação visível dessa pesquisa complexa e fundamental para o nosso tempo.

A pesquisa está na arte de criar prompts, de misturar imagens, de forçar a IA a sair de sua zona de conforto e gerar resultados inesperados e poéticos. O artista não é um mero operador; ele é um “domador” que aprende os segredos e as manhas da ferramenta para dobrá-la à sua visão.

5 - LINGUAGEM AUDIOVISUAL

A linguagem audiovisual é o sistema de comunicação que constrói significados e emoções através da articulação entre elementos de imagem em movimento e som. Mais do que apenas registrar a realidade, essa linguagem utiliza um conjunto de códigos e técnicas para criar narrativas, atmosferas e discursos. Cada escolha, da forma como uma cena é enquadrada até a música que a acompanha, é um ato intencional que compõe a mensagem final da obra.

Confira os principais elementos da linguagem audiovisual:

- **Plano -** É a unidade básica do vídeo, definindo o que o espectador vê e a que distância. A escolha do plano direciona a atenção e estabelece a relação emocional com a cena. Os mais comuns são o **plano geral**, que mostra todo o cenário e contextualiza a ação; o **plano médio**, que enquadra os personagens da cintura para cima, ideal para diálogos; e o **close-up** (Primeiro Plano), que foca no rosto para revelar emoções e reações íntimas.

Figura 3 - Planos do audiovisual



Fonte: do autor - imagem gerada por IA

- **Ângulo -** Refere-se ao posicionamento vertical da câmera em relação ao que está sendo filmado, alterando a percepção do espectador. O **ângulo normal,** na altura dos olhos, cria uma visão neutra e realista. O **contra-plongée** (ângulo baixo/low camera), filmando de baixo para cima, engrandece o personagem, transmitindo poder e superioridade. Já o **plongée** (ângulo alto/high camera), de cima para baixo, diminui o personagem, sugerindo fragilidade ou opressão.

Figura 4 - Ângulos do audiovisual



Fonte: do autor - imagem gerada por IA

- **Movimento de Câmera -** A câmera pode se mover para adicionar dinamismo, seguir personagens ou revelar informações gradualmente. A **panorâmica** é o movimento de rotação da câmera em seu próprio eixo (horizontal ou vertical), usada para mostrar um cenário. O **travelling** é o deslocamento físico da câmera, que pode seguir um personagem ou se mover por um ambiente. O **zoom** *a*proxima ou afasta do assunto em cena**.** O uso da **câmera na mão** cria uma sensação de urgência e realismo, como em documentários ou cenas de ação.
- **Montagem (edição) -** É a "escrita" do filme, o processo de selecionar, combinar e ordenar os planos para construir o ritmo e o sentido da narrativa. Um corte seco cria uma transição direta e rápida. Uma **fusão** (dissolve) sobrepõe duas imagens, geralmente para indicar passagem de tempo. A **montagem paralela**, ou cross-cutting, intercala cenas que acontecem ao mesmo tempo em lugares diferentes, gerando suspense e conectando histórias.
- **Design de som (Sound design) -** Esta é a disciplina que cria todo o universo sonoro do filme. O trabalho é dividido em camadas que são mixadas no final para criar a experiência auditiva.
- **luminação -** A luz é usada para esculpir o espaço, criar atmosferas e direcionar o olhar. Uma iluminação de **alto contraste** (low key), com muitas sombras, é comum em filmes de suspense e noir para criar mistério. Já uma iluminação de **baixo contraste** (high key), clara e com poucas sombras, é típica de comédias e musicais, transmitindo otimismo e leveza.
- **Color Grading (Colorização / Gradação de cor) -** Este é o tratamento de cor realizado pelo colorista, em diálogo com o diretor de fotografia. O processo ocorre em duas fases: primeiro, a **correção de cor**, uma etapa técnica para balancear contraste e exposição, garantindo consistência visual. Em seguida, vem a **color grading**, a etapa criativa onde se aplica um "look" ou estilo, manipulando paletas de cores para criar uma atmosfera, evocar emoções e dar uma identidade visual única à obra.

VIDEOAULAS DISPONÍVEIS
NA ABA ”MÓDULO 5 - AULA 2”

6 - MODELOS DE LINGUAGEM

O ecossistema de IA generativa é hoje um campo de batalha entre duas filosofias: o código aberto (open source) e os sistemas proprietários.



Figura 5 - Ecossistemas aberto x jardins murados
Fonte: do autor - imagem gerada por IA

Modelos open source (Stable Diffusion, Kling, Wan) - promovem a transparência, a colaboração e a democratização. Permitem que a comunidade inspecione o código, identifique vieses e construa suas próprias ferramentas. Eles nos dão maior controle sobre nossos dados e processos.

Uma ferramenta poderosa é a interface **ComfyUI** que permite rodar diversos modelos de IAs de vídeo simultaneamente e com blocos interconectados que cumprem o fluxo de trabalho escolhido pelo usuário. Caso você tenha uma máquina que suporte, recomendamos seu uso, caso não, você pode usar a computação em nuvem, como em plataformas como a Runpod e Nordy.ai.

Modelos Proprietários (Veo, Sora, Runway, Pika) - funcionam como “jardins murados”. Oferecem conveniência, mas a um custo, a falta de transparência sobre como os modelos são treinados e como nossos dados são utilizados. Ao usar essas plataformas, estamos, em muitos casos, alimentando e aprimorando os sistemas dessas corporações.

A escolha de uma ferramenta não é apenas técnica, é também política. Optar por soluções de código aberto, sempre que possível, é um ato de fortalecimento da comunidade e de defesa da

Abaixo estão listados os links de plataformas de IAs open source, na qual você poderá testar ferramentas e produzir seus vídeos com créditos de IAs gratuitos nos planos gratuitos.

Também disponibilizamos o link do ComfyUI, caso você tenha uma máquina com uma configuração robusta e no mínimo 16GB disponíveis em uma placa de vídeo dedicada. Caso você não tenha um computador com esta configuração, você poderá rodar na nuvem. A plataforma Nordy disponibiliza créditos gratuitos para teste e se desejar você poderá comprar créditos de IA parar realizar suas gerações.

Por fim disponibilizamos os links para testar as plataformas proprietárias Google VEO e Runway, que disponibilizam cotas gratuitas para geração de vídeos.

LINKS ÚTEIS

Hunyan - https://github.com/Tencent-Hunyuan/HunyuanVideo
Wan - https://wan.video/
Wan (teste no Huggin Face Spaces) - https://huggingface.co/spaces/Wan-AI/Wan-2.2-5B
LTX (teste no Huggin Face Spaces) - https://huggingface.co/spaces/Lightricks/ltx-video-distilled

ComfyUI - https://www.comfy.org/

Nordy
Wan text-to-video + ComfyUI - https://nordy.ai/workflows/689593226aeb446a6d9fc9fb
Wan image-to-video + ComfyUI - https://nordy.ai/workflows/6895368512f06da6674c74b4

Google Veo 2 (No Google AI Studio) - https://aistudio.google.com/prompts/new_video
Google Veo 3 (No Google Lab) - https://labs.google/flow
Google Veo 3 (No Google Gemini) - https://gemini.google.com/
Runway - https://runwayml.com/

6.1 Funções (tasks)

Confira os principais tipos de modelos de IA para áudio generativo e suas funções (tasks):

Exploraremos a aplicação prática dos modelos de linguagem na geração de vídeo com Inteligência Artificial. O objetivo é que você experimente a criação de prompts e a utilização de diferentes tipos de modelos generativos: texto-para-vídeo, imagem-para-vídeo e áudio-para-vídeo. A prática é fundamental para entender as capacidades e as nuances de cada ferramenta.

- **Modelos de Texto-para-Vídeo (Text-to-Video) -** Os modelos

- **Modelos de Imagem-para-Vídeo (Image-to-Video) -** Os modelos de imagem-para-vídeo pegam uma imagem estática como base e a animam, adicionando movimento, efeitos ou transformações. Você pode fornecer um prompt adicional para guiar a animação.
- **Geradores de efeitos sonoros (Text-to-Audio) -** Chamados de Text-to-audio (um termo mais geral para qualquer áudio gerado por texto), esses modelos criam sons específicos que você descreve, como o barulho de chuva, um trovão ou o som de uma porta fechando. São ideais para criar efeitos sonoros em jogos, filmes, animações ou para adicionar sons interessantes em podcasts e outras produções de áudio (veja mais mo módulo 4 do curso).
- **Modelos de Áudio-para-Vídeo (Audio-to-Video) -** Modelos de áudio-para-vídeo geram ou sincronizam elementos visuais com base em um arquivo de áudio. Isso pode incluir a criação de visualizadores de música, sincronização labial para avatares falantes, ou a geração de cenas que correspondam ao ritmo e ao tom de uma trilha sonora.

6.2 Parâmetros

Os parâmetros em IAs de vídeo são os comandos e ajustes que você utiliza para guiar a inteligência artificial na criação de um clipe. Eles funcionam como as “instruções da receita” , dizendo à IA não apenas o que criar (o conteúdo), mas também como criar (o estilo, o movimento e os detalhes técnicos). Dominá-los é a chave para transformar uma ideia em um resultado visualmente impactante.

Principais parâmetros de IAs de vídeo:

- **Prompt de Texto (Text Prompt) -** É a instrução principal e mais importante. Trata-se de uma descrição detalhada da cena que você deseja criar, incluindo o sujeito (quem ou o quê), a ação (o que está acontecendo) e o estilo (a aparência visual, como “estilo Ghibli”, “fotorrealista”, “cyberpunk”).
- **Imagem de Referência (Image Prompt) -** Em vez de ou junto com o texto, você pode fornecer uma imagem. A IA usará essa imagem como ponto de partida para definir a composição, o estilo visual ou até mesmo a aparência de um personagem no vídeo gerado.
- **Vídeo de Referência (Video Prompt) - A**lgumas IAs permitem usar um vídeo como base. Isso é útil para aplicar um estilo a um vídeo existente (style transfer) ou para guiar o tipo de movimento e a estrutura da cena que você deseja no resultado final.
- **Prompt Negativo (Negative prompt)** - É uma lista de palavras-chave do que você não quer ver no seu vídeo. É extremamente útil para refinar a imagem, removendo elementos indesejados como “texto”, “desenho animado”, “mãos deformadas” ou “baixa qualidade”
- **Duração -** É o tamanho do vídeo. Define o tempo exato (em segundos).
- **CFG Scale (Guidance Scale) -** É o nível de obediência da IA.Controla o quanto a IA deve seguir seu prompt. Valor alto para rigidez; valor baixo para mais criatividade.
- **Aspect Ratio (Proporção da Tela) - D**efine o formato do vídeo. Os mais comuns são 16:9 (widescreen, para YouTube), 9:16 (vertical, para TikTok/Reels) e 1:1 (quadrado, para Instagram).
- **FPS (Frames Por Segundo) -** Define a fluidez do vídeo. Valores mais altos, como 24 ou 30 FPS, resultam em um movimento mais suave e cinematográfico, enquanto valores mais baixos podem criar um efeito de stop-motion ou de vídeo mais “seco”.
- **Seed -** É o Código de reprodução. Um número que, se repetido com os mesmos comandos, garante que o resultado será exatamente o mesmo. Essencial para refinar uma música que você gostou.

- **Câmera -** Permite que você dirija a “câmera virtual”. É possível especificar movimentos como pan (movimento horizontal), tilt (movimento vertical), zoom in/out (aproximação/afastamento) e definir o ângulo da câmera (visão de baixo, de cima, etc.).
- **Passos de Geração (Generation Steps) -** É o esforço de refinamento da IA. Controla o número de etapas para aprimorar o áudio. Mais passos significam um som mais polido e detalhado, embora demore mais.

7 - ENGENHARIA DE PROMPTS

A engenharia de prompts é o nosso meio de diálogo com a IA. É onde nossa intenção criativa se traduz em instrução. No entanto, essa “conversa” não é neutra. A forma como formulamos nossos pedidos pode ter consequências éticas diretas.

De acordo com Araújo e Junqueira (2025), analisando sobre a produção artística e musical, podemos ter a mesma abordagem com outros tipos de criação, pois

> quando tratamos do papel da IA generativa nas práticas artísticas juvenis, estamos focando em como os jovens articulam a IA generativa nos processos de apropriação, criação e produção artístico-musical, que vão desde a aprendizagem de um vocal extraído de uma música com ajuda da IA, perpassando pela criação de imagens a partir de prompts de comando em texto, até de vídeos. Em todos esses processos, a IA é coadjuvante em facilitar e tornar mais produtivas as práticas juvenis em questão (Araújo;Junqueira, 2025, s/p).

Ao criar um prompt, não estamos apenas descrevendo uma imagem; estamos fazendo escolhas. A decisão de incluir ou omitir certas palavras, de descrever personagens de uma determinada maneira ou de situar uma cena em um contexto específico, reflete nossa própria visão de mundo. É nossa responsabilidade, como criadores, usar essa ferramenta para construir representações justas e plurais.

Nesse sentido é importante entender a responsabilidade do criador diante do cenário que é da inteligência artificial.
Cada vídeo, texto ou imagem que geramos com IA nos coloca em uma posição de poder e responsabilidade.

É imperativo refletir sobre:

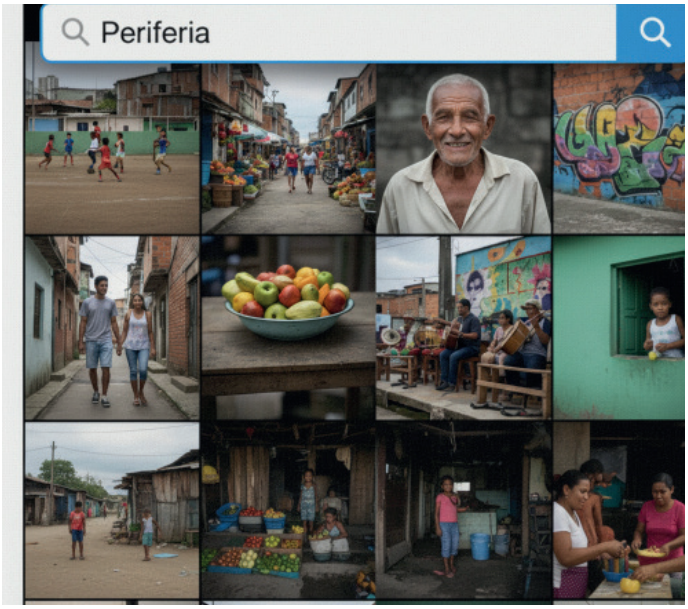
Originalidade vs. Curadoria - Qual é o nosso papel? Somos autores ou curadores de conteúdo gerado? A autoria reside na visão, nas escolhas e na intervenção humana que molda o processo. A IA é a ferramenta; a responsabilidade ética e criativa é nossa.

Uso Malicioso e Desinformação - A capacidade da IA de gerar conteúdo realista levanta a preocupação com os deepfakes, notícias falsas, enviezadas e a disseminação de desinformação.

Não que sejamos neutros, longe disso, mas é necessário não reforçar preconceitos inerentes da sociedade. A educação sobre o uso ético e responsável da IA é essencial para promover um am

biente digital saudável. Ao criar, devemos nos comprometer com a verdade e a autenticidade. Reflita sobre isso.

Figura 6: Busca de imagens de periferia



Fonte: do autor - imagem gerada por IA

Figura 7: Refinamento de prompt: Cultura vibrante na periferia



Fonte: do autor - imagem gerada por IA

As imagens acima mostram a palavra “Periferia” sendo digitada, e a IA gera imagens clichês de pobreza e abandono. Em seguida, o prompt é refinado para “Cultura vibrante na periferia, artistas

de rua, afrofuturismo", e as imagens se transformam em cenas de potência, criatividade e alegria. A animação ilustra como a intencionalidade no prompt pode guiar a IA para longe dos estereótipos

7.1 Prompts postivos



Para criar um prompt de vídeo eficaz, é fundamental estruturá-lo como um roteiro conciso, detalhando os elementos essenciais da cena. Isso exige descrever claramente o **assunto principal** (personagem ou objeto), a **ação** que ocorre, o **ambiente**, o **estilo visual** desejado (como cinematográfico ou animação), e as características de **iluminação** e **cor** que definem a atmosfera.

A elaboração de um bom prompt é um processo que combina criatividade com precisão, exigindo que o usuário seja específico, utilize **palavras-chave** fortes para guiar o tom, e **refine** o comando gradualmente com base nos resultados, transformando a comunicação com a IA em uma habilidade essencial.

Para aumentar o impacto visual, utilize palavras-chave de qualidade como: **cinematográfico, fotorrealista, 8K, detalhes intrincados, iluminação dramática e profundidade de campo**.

7.2 Prompts negativos

Como já vimos nos demais módulos, prompts negativos são instruções textuais que informam ao modelo de IA o que ele deve excluir ou evitar durante o processo de geração de conteúdo. Atuando como um filtro essencial, os prompts negativos aprimoram a geração de imagens por IA ao refinar a qualidade, controlar o estilo e aumentar a eficiência do processo, simplesmente ao direcionar a inteligência artificial para longe de resultados indesejados.

Para utilizar prompts negativos de forma eficaz, é essencial criar listas específicas e detalhadas do que deve ser evitado, ajustando-as continuamente através da experimentação para se adaptar às particularidades de cada modelo de IA.

Confira alguns exemplos de palavras chave para prompts negativos: **low quality, blurry, deformed, bad anatomy, watermark, text, signature, cropped, out of frame, noise, grainy.**

7.3 Gerador de prompts

DICA

Você também pode criar seu próprio gerador de prompts na LLM aberta Qwen. Basta abrir um novo chat, selecionar o modelo “Qwen-3-coder” e digitar seu prompt com o que você deseja acrescentar no seu gerador de prompts. Ela faz todo o trabalho e se algo não estiver correto é só você digitar que a IA corrige. Chegou no ponto que você deseja? Agora é só clicar no botão implementar e o Qwen disponibiliza um link pra você utilizar seu gerador de prompts personalizado.



8 - PRODUZINDO CONTEÚDO AUDIOVISUAL COM IAS GENERATIVAS



A ascensão das IAs generativas está redefinindo radicalmente o fluxo de trabalho da produção audiovisual, transformando cada etapa do processo. A tradicional captação de imagens com câmeras agora compete com a geração de cenas inteiras a partir de comandos de texto, enquanto o roteiro pode se desdobrar em argumentos mais flexíveis que guiam a máquina. Ferramentas como o storyboard são aceleradas, permitindo a visualização instantânea de planos complexos. Na pós-produção, a edição ganha novas possibilidades de transição e sequenciamento, e a finalização é otimizada com ferramentas inteligentes, fundindo criação e execução de maneiras inéditas.

Pré-Produção (planejamento e concepção)

Nesta etapa, a ideia é estruturada e visualmente planejada antes da criação efetiva. Para isso são necessários os seguintes elementos:

- **Roteiro (ou argumento) -** É a estrutura narrativa ou conceitual que serve como base para a criação, podendo ser um roteiro detalhado ou um simples argumento que, ao ser inserido na IA, se torna o guia central para a geração das cenas.
- **Storyboard -** Continua sendo a etapa de planejamento visual da obra, mas agora pode ser gerado ou prototipado de forma quase instantânea por IAs de imagem, permitindo visualizar a composição dos planos antes mesmo de criar os vídeos.

Produção (captação das imagens)

Aqui, o conceito se transforma em matéria-prima visual, com a IA atuando como a câmera, o cenário e os atores. É a hora em que o jogo acontece.

- **Captação de Imagens (geração)** - No fluxo de trabalho com IA, esta etapa deixa de ser o ato de filmar com uma câmera e passa a ser a criação de planos visuais através de prompts de texto, imagens ou vídeos em modelos generativos.

DICA

No campo da arte tudo é possível e a experimentação é sempre bem vinda. Explore possibilidades, misture imagens captadas com seu celular, com câmeras, fotos, geradores de imagens e vídeos, Explore sua criatividade.

Pós-Produção (refinamento e acabamento)

Com os clipes gerados, esta fase se concentra em montar a narrativa e dar o polimento profissional à obra. Existem IAs que oferecem soluções diretamente de sua própria plataforma e outras que podem ser usadas para funções específicas. Entra as open source as melhores opções são a **Wan, Hunyan, Seedance e Kling**. Os modelos proprietários contam com diversas plataformas como a **Sora, Runway, Google Flow** entre outras que trazem custos mais expressivos e todas as questões envolvidas ao modelo fechado.

- **Edição (Montagem) -** Nesta fase, os clipes gerados pela IA são sequenciados e ordenados para construir o ritmo e a narrativa final, um processo que também pode ser assistido por IAs para criar transições ou sugerir cortes.
- **Finalização -** Corresponde ao acabamento técnico e artístico do projeto, onde ferramentas de IA são cada vez mais usadas para aprimorar tarefas como a colorização (color grading), o design de som (sound design) e a melhoria da resolução (upscaling).

LINKS ÚTEIS

Upscaler gratuito para usar online - Tensor Pix
https://app.tensorpix.ai/videos
Upscaler gratuito para usar via APP - Capcut
https://www.capcut.com/pt-br/tools/ai-video-upscaler
Upscaler open source para rodar localmente no computador - Video 2x
https://github.com/k4yt3x/video2x

Editor de vídeos gratuito para desktops e smartphones - VN Editor
https://www.vlognow.me/

8.1 Formatos de arquivos de vídeo

Cada formato de arquivo de vídeo, como MP4, MOV ou MKV, funciona como um contêiner específico para imagem e som, e a escolha correta é o que garante que seu projeto será visualizado corretamente em qualquer tela.

Confira mais detalhes sobre cada um os formatos na lista abaixo:

MP4 - O formato mais universal e compatível, ideal para web, redes sociais e streaming por seu ótimo equilíbrio entre qualidade e tamanho de arquivo.
MOV - Padrão da Apple, comum em edição profissional (Final Cut, Premiere Pro), pois suporta vídeos de altíssima qualidade e transparência (canal alfa).
MKV - Um contêiner flexível e de código aberto, popular para filmes e séries em alta definição, pois consegue agrupar múltiplas faixas de áudio e legenda em um único arquivo.

Assista aos vídeos das aulas para conferir na prática como gerar e finalizar suas obras, utilizando de conhecimentos técnicos e pensamento crítico para promover o potencial transformador das artes.

Boas produções!

REFERÊNCIAS

ARAÚJO, Jose Magnaldo Moura; JUNQUEIRA, Eduardo Santos. MULTILETRAMENTOS E INTELIGÊNCIA ARTIFICIAL NAS PRÁTICAS ARTÍSTICAS JUVENIS: uma análise sobre a produção de vídeos remix. Revista Interinstitucional Artes de Educar, [S.L.], v. 11, n. 1, p. 99-118, 10 fev. 2025. Universidade de Estado do Rio de Janeiro. http://dx.doi.org/10.12957/riae.2024.86225.

“The Illustrated Transformer” por Jay Alammar: Um artigo visual que ajuda a entender, de forma mais intuitiva, como funcionam os modelos de linguagem que são a base de muitas IAs generativas.

BENTES, Ivana. Mídia-Arte: estéticas da comunicação e modelos teóricos. In: Anais do XXVIII Congresso Brasileiro de Ciências da Comunicação. Rio de Janeiro, 2005. Disponível em: http://www.portcom.intercom.org.br/pdfs/59581618554487303376021132904259551034.pdf. Acesso em: 9 ago. 2025.

BENTES, Ivana. Economia narrativa: do midiativismo aos influenciadores digitais. In: Interfaces do Midiativismo: do conceito à prática. 2018. Disponível em: https://interfacesdomidiativismo.wordpress.com/wp-content/uploads/2018/06/artigo-8.pdf. Acesso em: 9 ago. 2025

FREIRE, Paulo. Pedagogia da autonomia: saberes necessários à prática educativa. São Paulo: Paz e Terra, 1996.

SARZI-RIBEIRO, R. A. INTELIGÊNCIA ARTIFICIAL, ARTE E TECNOLOGIA: VISUALIDADES, AUDIOVISUALIDADES E SONORIDADES. Publicações, 2023. Disponível em: https://www.editorapublicar.com.br/ojs/index.php/publicacoes/article/view/797. Acesso em: 9 ago. 2025.

DUARTE,Roberto Dias. Guia Prático para Engenharia de Prompts em IA Generativa de Imagens. Disponível em: https://www.robertodiasduarte.com.br/guia-pratico-para-engenharia-de-prompts-em-ia-generativa-de-imagens/. Acesso em: 9 ago. 2025.

AI É NÓS!

Projeto 0272/2024 Este projeto foi realizado com recursos da Lei Municipal de Incentivo à Cultura de Belo Horizonte.

APOIO: ASSIMPO DISTÓPICA

REALIZAÇÃO: PROJETO REALIZADO PELA SOCIEDADE CIVIL COM RECURSOS ORIUNDOS DA POLÍTICA DE FOMENTO À CULTURA MUNICIPAL

INCENTIVO: LMIC LEI MUNICIPAL DE INCENTIVO À CULTURA — BELO HORIZONTE PREFEITURA