



Universidade Federal de Pernambuco  
Centro de Informática

Graduação em Engenharia da Computação

**Uma interface de redes generativas para  
publicitários**

Avner Augusto dos Anjos

Trabalho de Graduação

Recife  
27 de agosto de 2021

Universidade Federal de Pernambuco  
Centro de Informática

Avner Augusto dos Anjos

## **Uma interface de redes generativas para publicitários**

*Trabalho apresentado ao Programa de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação*

Orientador: *Prof. Dr. Filipe Carlos de Albuquerque Calegario*

Recife  
27 de agosto de 2021

*Dedico este trabalho a meus pais, que me "carregaram" até este momento, sem vocês nada disso seria possível.*

# Agradecimentos

Agradeço a todos os amigos que conheci durante minha graduação, alguns dos quais já fizeram ou ainda fazem parte da minha carreira profissional. Aprendi muito com vocês e espero que eu possa ter agregado em algo na jornada da graduação.

A todos os excelentes professores que me educaram e me auxiliaram na busca de conhecimento. Agradeço em especial ao Prof. Dr. Ariano Sarmiento pela mentoria em diversos momentos do curso e competições e ao Prof. Dr. Filipe Calegario por me orientar neste trabalho, pelas diversas reuniões que tivemos e por não me deixar desistir.

Por fim, gostaria de agradecer à minha família, meu pai - Carlos, minha mãe - Elizabete, minha irmã - Iohanah e minha noiva Camila, pelo suporte durante o curso, sei que não foi fácil. Muito Obrigado!

*"It's going to be legen... wait for it... DARY!"*  
—BARNEY STINSON (HIMYM)

# Resumo

Um dos grandes desafios de artistas da área de marketing é a criação de conteúdo que possa persuadir e ser mais efetivo que outros materiais concorrentes. Além disso, essa tarefa pode tomar muito tempo devido à grande quantidade de materiais existentes (o que dificulta a criação de um conteúdo original), tipos de design, layouts e etc. Recentemente, novos trabalhos sobre a utilização de redes adversárias generativas (GANs) para a geração automática de imagens surpreenderam ao apresentar imagens hiper-realistas com alta qualidade [RPG<sup>+</sup>21] e outras pesquisas utilizaram esses conceitos no ambiente de marketing de forma a criar o material final automaticamente [GJS<sup>+</sup>21]. Contudo, o nível de complexidade na utilização desse tipo de recurso para o publicitário ainda é muito alto. Este trabalho propõe e desenvolve a criação de um ambiente no Google Colaboratory com a intenção de facilitar o uso dos recursos de IA que ao invés de criar o material final, auxilia no processo de criação de publicitários ao gerar novas imagens. Os resultados apontam que apesar dos estudos recentes conseguirem a geração de material original, o foco de estudos futuros pode ser a avaliação da ferramenta para publicitários em processo de aprendizagem. Em relação a avaliação da ferramenta em si, as avaliações quantitativas e qualitativas, indicam que estamos no caminho certo no desenvolvimento da interface, mas que ainda há pontos de melhoria.

**Palavras-chave:** Marketing, GAN, Google Colab, Inteligência Artificial, propaganda, Ferramenta de apoio,

# Abstract

One of the biggest challenges for artists in the marketing field is creating content that can persuade and be more effective than other competing materials. Also, this task might take a lot of time due to the large amount of existing materials (which makes it difficult to create original content), design types, layouts and so on. Recently, new works on the use of generative adversarial networks (GANs) for the automatic generation of images surprised by presenting hyperrealistic images with high quality [RPG<sup>+</sup>21] and other researches used these concepts in the marketing environment in order to create the final material automatically [GJS<sup>+</sup>21]. However, the level of complexity in using this kind of resource for the advertiser is still very high. This work proposes and develops the creation of an environment in Google Colaborative with the intention of facilitating the use of AI resources that, instead of creating the final material, assists the creative process of advertisers by generating new images. The results show that, despite recent studies having managed to generate original material, the focus of future studies may be the evaluation of the tool for advertisers in the learning process. Regarding the evaluation of the tool itself, the quantitative and qualitative evaluations indicate that we are on the right path in the development of the interface, but that there are still areas for improvements.

**Keywords:** Marketing, GAN, Google Colab, Artificial Intelligence, Advertising, Assistant design tool

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação e Proposta	2
1.2	Organização do Trabalho	2
<b>2</b>	<b>Estado da arte</b>	<b>4</b>
2.1	Processo Criativo	4
2.2	Processamento de imagens	5
2.3	GANs	7
2.3.1	VQ-GAN	8
2.4	Trabalhos relacionados	9
2.4.1	Vinci	9
2.4.2	DALL-e	11
2.4.3	CLIP	12
<b>3</b>	<b>Proposta</b>	<b>13</b>
3.1	Caracterização do problema	13
3.2	Arquitetura	14
<b>4</b>	<b>Desenvolvimento</b>	<b>16</b>
4.1	Datasets	16
4.2	Plataforma	16
4.2.1	Interface	17
<b>5</b>	<b>Avaliação</b>	<b>19</b>
5.1	Métricas de avaliação	19
5.1.1	Usabilidade	19
5.1.2	Experiência do usuário	20
5.1.3	Análise qualitativa	20
5.2	Resultados e Discussão	21
5.3	Discussão	22
<b>6</b>	<b>Conclusão</b>	<b>24</b>
6.1	Trabalhos futuros	24



# Lista de Figuras

1.1	Da esquerda para a direita: a) Banner da luta de MMA entre um homem e uma mulher. b) Logo da McDonalds. c) Propaganda da RedBull.	1
2.1	Exemplos de anúncios gerados pelo <i>Vinci</i> .	5
2.2	Representação de uma imagem RGB	6
2.3	O símbolo $\Sigma$ representa o cálculo do soma ponderada, e o símbolo $\varphi$ representa a ativação função de processamento da soma ponderada e geração da saída do neurônio.)	7
2.4	Arquitetura e funcionamento de uma GAN.	8
2.5	Taxonomia das GANs.	9
2.6	Comparação entre propagandas gerados por: a) um designer humano, b) ferramenta Luban e c) Vinci.	10
2.7	Imagens de um bebê rabanete passeando com um cachorro gerado pelo DALL-e.	11
2.8	Imagens de um caracol feito de uma harpa gerado pelo DALL-e.	11
2.9	Arquitetura do CLIP.	12
3.1	Arquitetura da plataforma.	15
4.1	Entradas da rede.	18
5.1	Saídas geradas a partir do texto: “ <i>A man with wings drinking, RedBull.</i> ”	20
5.2	Saída gerada a partir do texto: “ <i>A man with wings drinking, RedBull.</i> ”, com a característica "artística".	21
5.3	Saída gerada a partir do texto: “ <i>An image with two people holding hands.</i> ”, com a característica "artística".	21
5.4	Resultados médios obtidos pelo User Experience Questionnaire.	22

# Lista de Tabelas

3.1	Tabela com o perfil dos entrevistados.	13
5.1	Tabela com o resultados do SUS.	22

# CAPÍTULO 1

## Introdução

Um ‘M’ em amarelo em um fundo vermelho. Adicionar o sabor de menta e espuma nas pastas de dentes. Um poster promovendo uma luta de MMA entre um homem e uma mulher. A frase “Red Bull gives you wiiings”. O que esses elementos têm em comum? Todos eles têm por objetivo influenciar um público alvo a alguma coisa. Seja dar a sensação de confiabilidade na marca e incentivar a ideia de fome. Ou então, trazer a sensação de que sua boca está mais limpa (mesmo o frescor não sendo fator influenciador nisso). Ou criar uma situação absurda de um luta entre homem e mulher para trazer à tona a discussão sobre violência doméstica. Ou o simples apelo de que o consumo de um produto vai dar ao usuário uma performance sobre humana.



**Figura 1.1** Da esquerda para a direita: a) Banner da luta de MMA entre um homem e uma mulher. b) Logo da McDonalds. c) Propaganda da RedBull.

**Fonte:** Retirado da internet

Um anúncio pode ser definido como “uma mensagem específica construída para informar, persuadir, promover, provocar ou motivar as pessoas em nome de uma marca, entidade ou causa” [Lan21]. Se por um lado, o aumento do número de conteúdo publicitário pode auxiliar no processo de ensino de novos profissionais da área, por outro é cada vez mais difícil ser criativo para ter ideias que consigam chamar mais atenção que outros ou ser mais atrativo que os competidores. Além disso, nem sempre um material inteiramente original será considerado de qualidade. Outro ponto é que essa tarefa criativa pode consumir muito tempo. Além disso, o trabalho do publicitário não se resume a escolher entre padrões previamente prontos: tem-

se o processo de criação, a sinergia entre componentes textuais, visuais e sonoros (quando aplicável), entre outros aspectos que devem ser levados em consideração.

Apesar de um anúncio ou propaganda poder ter vários elementos dentro do mesmo material, é inegável que elementos textuais e visuais são os mais importantes na criação desse conteúdo. Consequentemente, é preciso reconhecer que trabalhos recentes na área de geração de imagens utilizando redes adversárias generativas (GANs) [GPAM<sup>+</sup>14] têm alcançado excelentes resultados especialmente na criação de imagens ultrarealistas a partir de uma entrada textual [RPG<sup>+</sup>21].

Como mostrado em trabalhos recentes [RPG<sup>+</sup>21, ERO21, SC] o estado atual da síntese de imagens utilizando GANs tem sido constante fonte de surpresa, especialmente quando utilizado em conjunto com o CLIP [RKH<sup>+</sup>21], um modelo de rede neural que aprende conceitos visuais a partir de treinamento com linguagem natural. Pelos seus excelentes resultados na literatura e pelo fato de ser um técnica de *Zero-shot learning*, onde o modelo pode trabalhar com classes nunca antes vistas na fase de treinamento, a combinação GAN+CLIP foi escolhida para a utilização neste trabalho.

## 1.1 Motivação e Proposta

Diferentemente de tentar gerar o material final como é feito em [GJS<sup>+</sup>21] o problema investigado neste trabalho é o reflexo da utilização de imagens geradas por uma rede neural de aprendizagem de máquina profunda, ou *Deep Learning*, como ferramenta de auxílio no processo criativo dos profissionais de publicidade, usando uma interface que sirva como abstração na utilização de recursos de deep learning, mais especificamente a utilização de modelos GAN. Problemas como mapear o processo criativo e trabalhar a sinergia dos elementos indicam que manter o designer como principal ator do processo criativo e gerar ferramentas de apoio pode ser um bom direcionamento para estudos.

Para alcançar esse objetivo e validar a ideia, como proposto em [dO21] foi criada uma interface na plataforma Google Colaboratory que elimina a necessidade do profissional possuir altos recursos de hardware e abstrai do usuário final todo o processo de preparo e processamento de *datasets* além da implementação complexa dos modelos.

## 1.2 Organização do Trabalho

O presente trabalho está organizado da seguinte maneira:

- **Capítulo 2 - Estado da Arte:** Conceitos relacionados ao problema de geração de imagens utilizando GANs e como se dá o processo criativo no meio da publicidade. O capítulo apresenta ainda uma revisão da literatura e trabalhos relacionados.
- **Capítulo 3 - Proposta:** Este capítulo apresenta o processo de caracterização do problema a ser resolvido e a arquitetura da plataforma desenvolvida.

- **Capítulo 4 - Desenvolvimento:** Neste capítulo é mencionado os *datasets* utilizados para geração de múltiplas imagens e a descrição técnica da plataforma.
- **Capítulo 5 - Avaliação:** Aqui são apresentadas as métricas utilizadas na avaliação do trabalho assim como a discussão sobre os resultados obtidos.
- **Capítulo 6 - Conclusão:** Dificuldades, conclusões gerais, contribuições e possíveis trabalhos futuros são discutidos nesta seção.

## Estado da arte

Este capítulo apresenta a base teórica dos temas abordados ao longo deste trabalho. A Seção 2.1 descreve o processo criativo e o problemas enfrentados por publicitários. A seção 2.2 descreve o problema computacional na utilização de imagens, seu impacto na necessidade de hardware específico, além dos conceitos básicos sobre redes neurais. Na Seção 2.3, são apresentadas as redes adversariais generativas (GANs) em especial as VQ-GANs. A Seção 2.4 discorre sobre o estado da arte e ferramentas recentes criadas como propósito de auxiliar no processo criativo.

“Só há duas razões para se achar um lance: ou você **PRECISA** ou você conhece uma série de **PADRÕES** que vão ajudá-lo a encontrar tal lance.” — Jaacob Agaard (Grand Master Preparation)

### 2.1 Processo Criativo

A frase acima, apesar de bem aplicável no xadrez, nos leva a refletir sobre o impacto do preparo e da criação de métodos dentro do processo criativo para os profissionais da área de propaganda. A criatividade, como mostrado em [Pri14] e em [DW06], pode aumentar a eficácia e eficiência em até dez vezes e tem o segundo maior impacto na lucratividade da publicidade e no valor da marca a longo prazo. Dentro do contexto de propaganda, apesar de ser difícil haver um consenso de qual é a definição de criatividade, é fácil perceber que há uma distinção do conceito de criatividade dentro do contexto puramente artístico [Bel92].

Amabile [Ama96] argumenta que o processo criativo deve ser visto como um processo de cinco estágios e sugere um quinto estágio no processo criativo chamado de resultado. Ela sugere as seguintes etapas:

- **Apresentação do problema ou tarefa** (o problema pode vir de um estímulo interno ou externo).
- **Preparação** (coleta ou reativação de informações relevantes).
- **Geração de resposta** (busca na memória e ambiente imediato para responder ao problema).
- **Validação** de resposta (teste da possível resposta em relação aos critérios de conhecimento e avaliação).

- **Resultado** (decisão de progredir ou encerrar o processo). Este quinto estágio é uma fase de tomada de decisão. Esta fase reconhece que, se a solução do problema passar pela fase de validação, o processo é encerrado. Se a solução não for totalmente alcançada, as etapas anteriores do processo podem ser revisadas.

Turnbull e Wheeler [TW17] realizaram um estudo em agências de publicidade do Reino Unido e após uma série de entrevistas, estruturaram o processo criativo em 7 etapas que poderiam ser mapeadas nas 5 de Amabile, validando o modelo proposto.

Em [GJS<sup>+</sup>21] faz-se uma tentativa de automação do processo criativo na elaboração de anúncios. Apesar de apresentar *Vinci*, uma solução que cria de forma automática um produto final (anúncio), alguns desafios da área são listados pelos autores:

- **Q1:** Como formular o processo de design humano de uma forma que a máquina possa entender.
- **Q2:** Como selecionar elementos de design que correspondam tanto ao estilo semântico quanto visual.
- **Q3:** Como organizar elementos de design no plano final.

O resultado apresentado pelo *Vinci* gera anúncios completos que acertam na sinergia entre os elementos mas pecam no critério criatividade.



**Figura 2.1** Exemplos de anúncios gerados pelo *Vinci*.

**Fonte:** [GJS<sup>+</sup>21]

## 2.2 Processamento de imagens

Desde cedo o ser humano se vê cercado por imagens. O processo para o entendimento de objetos complexos, contudo, só se inicia entre os 8 meses e os dois anos de idade, que é quando

começamos a atribuir semântica ao que vemos. Traços viram curvas, começamos a entender cores e começamos a perceber profundidade [SOU20].

Segundo [SB11], uma imagem pode ser entendida como uma função bidimensional,  $f(x, y)$ , em que  $x$  e  $y$  são coordenadas espaciais (plano), e a amplitude de  $f$  em qualquer ponto é chamada de intensidade ou nível de cinza da imagem nesse ponto. Quando  $x$ ,  $y$  e os valores de intensidade de  $f$  são quantidades finitas e discretas, chamamos essa imagem de uma imagem digital. Sendo assim, as imagens digitais não trazem consigo nenhuma informação intrínseca, ou seja, computadores só enxergam uma matriz bidimensional de pixels, na qual cada pixel guarda uma informação sobre características como cor, luminosidade e saturação, dependendo do tipo de representação utilizado.



**Figura 2.2** Representação de uma imagem RGB

**Fonte:** Retirado de [SB11]

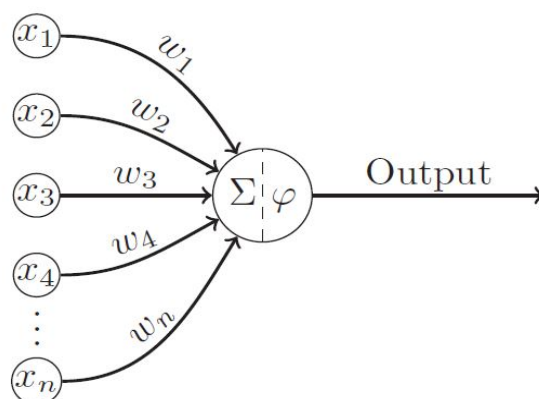
À área que manipula essas imagens digitais é dado o nome de Processamento de Imagens. Com a popularização das técnicas de Aprendizagem de Máquina, a área de processamento de imagens passa por uma verdadeira revolução. Segundo [Kel19], em Aprendizagem de Máquina usa-se algoritmos para escolher a melhor função, a partir de um conjunto de funções possíveis, que explica as relações entre características em um determinado conjunto de dados. Dentro dessa área, dois conceitos muito importantes são o de redes neurais e o de aprendizagem profunda.

Ainda por [Kel19] uma rede neural consiste em uma rede de unidades simples de processamento de informações, chamadas neurônios, onde as redes de aprendizado profundo são redes neurais que possuem muitas camadas ocultas de neurônios. Cada neurônio recebe várias entradas associadas com pesos, o qual são multiplicados. Se entendermos as camadas como várias matrizes, veremos que o processo é similar à multiplicação de matrizes.

De forma simplificada, quanto mais camadas temos em uma rede neural, ou seja, quanto maior a sua profundidade, maior o número de multiplicação de matrizes, aumentando assim a complexidade computacional do problema. Além disso, quanto maior o número de neurônios em cada camada, maior o número de pesos. Ou seja, maior o tamanho das matrizes a serem multiplicadas.

Sendo assim, devido às características de complexidade intrínseca na realização dos cál-





**Figura 2.3** O símbolo  $\Sigma$  representa o cálculo do soma ponderada, e o símbolo  $\phi$  representa a ativação função de processamento da soma ponderada e geração da saída do neurônio.)

**Fonte:** Retirado de [Kel19]

culos em redes neurais e processamento de imagens, é natural que para a realização dessas tarefas seja necessária a utilização de GPUs de alta performance a fim de alcançar eficiência, tornando-se fator impeditivo para alguns usuários finais, no caso deste trabalho, publicitários.

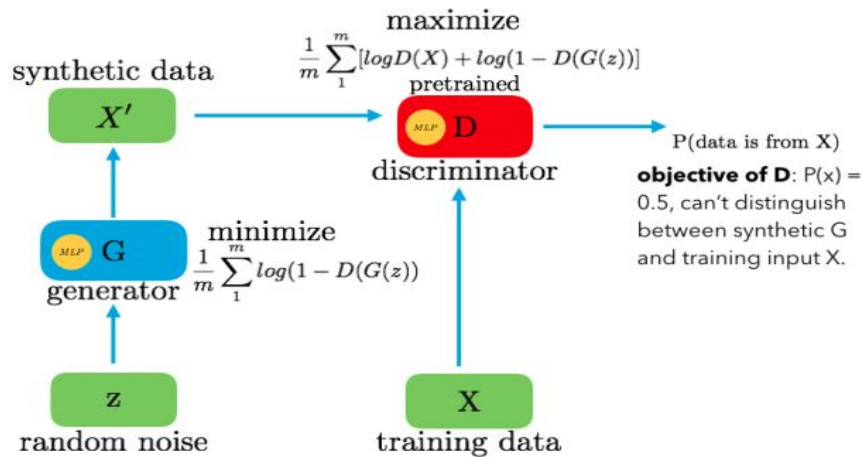
## 2.3 GANs

Primeiramente proposta por [GPAM<sup>+</sup>14], redes generativas adversárias têm por objetivo criar novos exemplos baseados no conjunto de dados usados para treinamento. Atualmente, pode ser aplicada em vários ramos de pesquisa como síntese de imagem, geração de texto [LSL18], processamento de imagem médica, colorização da imagem, tradução imagem à imagem, geração de arte e síntese de imagens [WCY<sup>+</sup>20], tendo esta última utilização especial sucesso devido ao poder de processamento de imagem das GANs.

Esta rede tenta solucionar o problema da aprendizagem sem a necessidade de um conjunto de dados que já possuam rotulação prévia (aprendizagem supervisionada), visto que modelos supervisionados têm maior utilização em problemas de regressão e classificação. Por outro lado, para solucionar problemas generativos, muito provavelmente não teremos um conjunto de dados pré-rotulado (aprendizagem não-supervisionada), o que levou ao aprofundamento no estudo da utilização de GANs.

Com recentes resultados surpreendentes, as GANs possuem este nome devido ao fato de possuírem dois agentes na sua estrutura, competindo entre si. Um dos agentes, o gerador G, aprende características de um conjunto de dados e gera um novo exemplar pertencente a este conjunto. O outro agente, um discriminador D, é um modelo discriminativo que aprende a identificar se uma imagem pertence ao modelo ou ao conjunto de dados. Nessa estrutura o gerador tenta criar imagens que enganem o discriminador em um sistema de competição min-max cíclica. O gerador não tem acesso direto a imagens reais - a única maneira de aprender

é por meio de sua interação com o discriminador [CWD<sup>+</sup>18]. O gerador G é treinado para confundir o discriminador e o discriminador é treinado para aumentar a acurácia na sua classificação. A rede geradora pode ser considerada, portanto, um mapeamento de algum espaço de representação, denominado espaço latente.



**Figura 2.4** Arquitetura e funcionamento de uma GAN.

**Fonte:** [GJS<sup>+</sup>21]

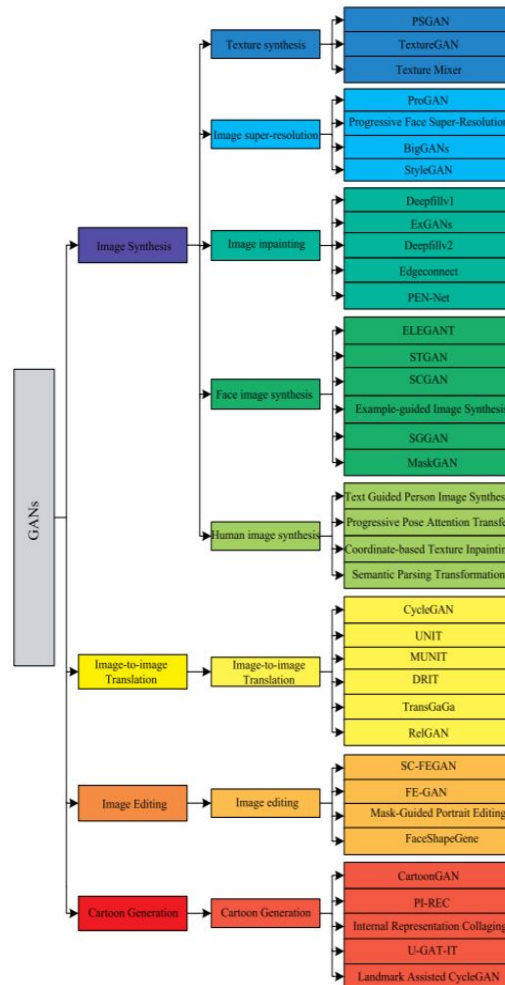
Devido à vasta gama de pesquisas destinadas a utilização de GANs em diferentes áreas, surgiram diversas aplicações baseadas em GANs como mostrado na figura 2.5

### 2.3.1 VQ-GAN

Em dezembro de 2020, Esser et al. propuseram a VQ-GAN [ERO21], uma variação da arquitetura GAN que aproveita as características das redes neurais convolucionais, CNNs, para permitir o uso de modelos do tipo *transformer* em imagens de alta qualidade.

Os transformers [VSP<sup>+</sup>17], modelos de deep learning que vêm ganhando bastante popularidade recentemente, utilizam uma técnica conhecida como mecanismo de atenção. Nesses modelos, uma sequência de dados de entrada é analisada e a importância de cada parte desses dados é computada. Por levar em consideração as interações entre cada dado e todos os outros daquela sequência, os transformers acabam sendo inviáveis para sequências muito longas, como é o caso em imagens de alta resolução.

Para contornar esse problema, as VQ-GANs usam uma estratégia de dois estágios, onde primeiramente incorpora-se os bias indutivos das CNNs para gerar um mapeamento que prioriza interações locais. Esse mapeamento é utilizado no segundo estágio, como entrada para a estrutura de transformer. Dessa forma, é possível utilizar transformers para modelar e sintetizar imagens de alta resolução.



**Figura 2.5** Taxonomia das GANs.

Fonte: [GJS<sup>+</sup>21]

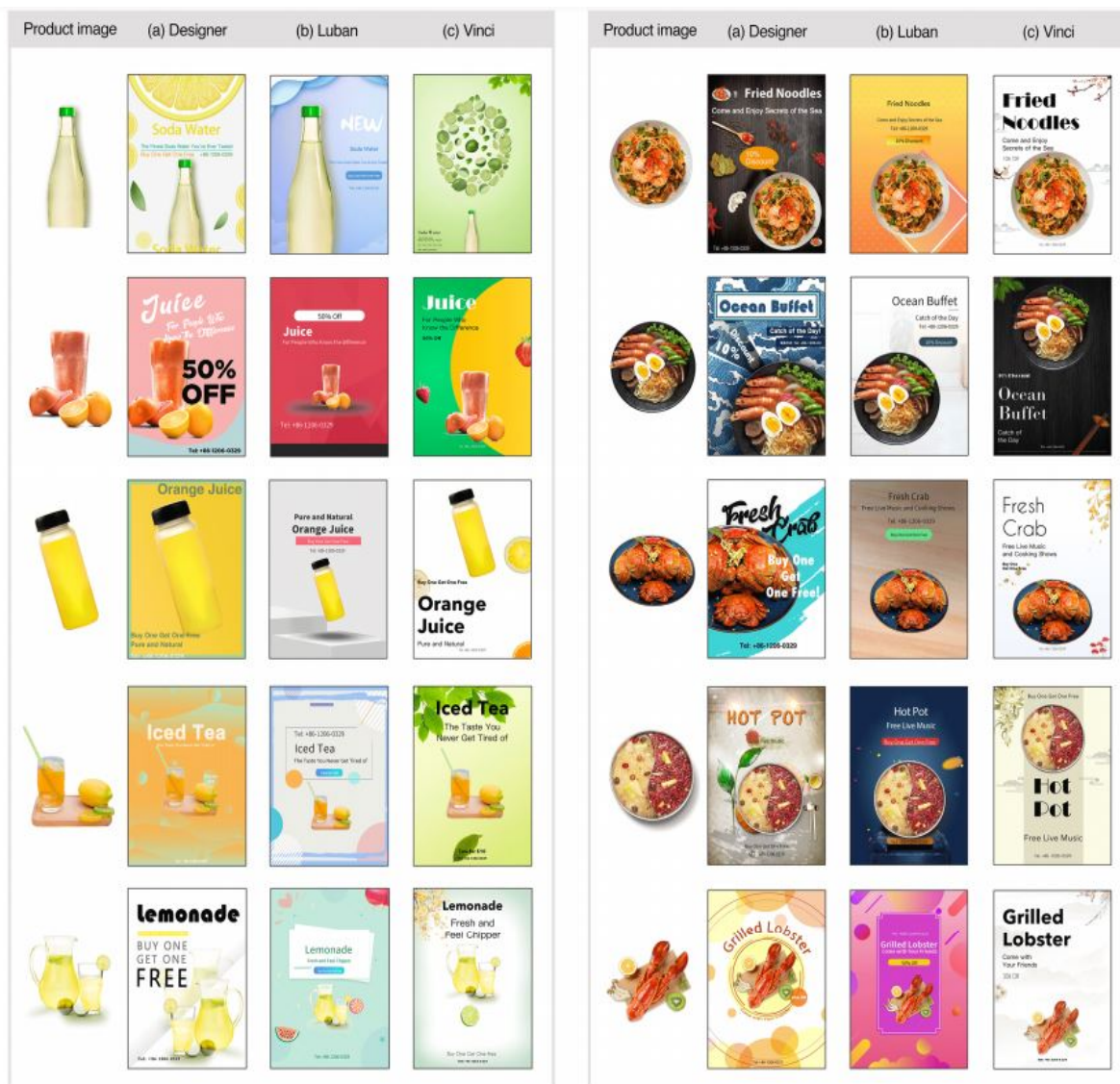
## 2.4 Trabalhos relacionados

Tendo como base a motivação de avaliar o reflexo da utilização de imagens sintetizadas no processo criativo e de inspiração de profissionais de propaganda e marketing, é necessário avaliar alguns trabalhos relacionados ao contexto a fim de tomar um direcionamento.

### 2.4.1 Vinci

Guo et al. propuseram recentemente *Vinci* [GJS<sup>+</sup>21], que consiste na primeira plataforma que usa técnicas de aprendizagem profunda capaz de sintetizar anúncios e a partir de feedback do usuário permitir pós-processamento do anúncio. Após uma série de entrevistas, foi proposta uma interface da plataforma que consiste em (i) uma caixa de entrada para imagem, no caso

o produto foco do anúncio, (ii) uma entrada textual para descrição ou frases que serão utilizadas no pôster, (iii) um painel de edição, onde são selecionados atributos como fonte do texto, organização dos elementos e etc. e (iv) uma lista com os posters gerados. Um dos principais desafios deste trabalho é manter a coesão e sinergia entre os elementos presentes no anúncio.



**Figura 2.6** Comparação entre propagandas gerados por: a) um designer humano, b) ferramenta Luban e c) Vinci.

**Fonte:** Retirado de [GJS<sup>+</sup>21].

Para geração dos posters, Vinci usa os conceitos de *Variational Autoencoders* apresentados em [KW13]. Após a criação da plataforma foi feito um desafio para criação de material que foi posteriormente comparado com poster gerados usando uma segunda plataforma de assistência ao desenvolvimento de propagandas, a Luban, e com posters criados por designers. Esta

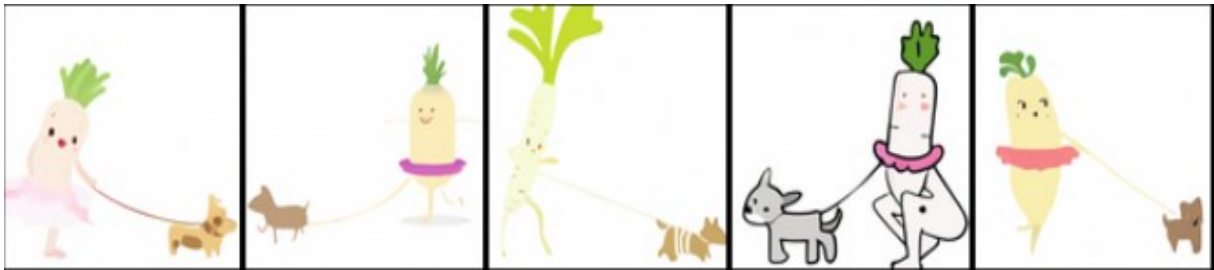
comparação pode ser vista na imagem 2.6

Os resultados mostraram que os pôsteres gerados pela Vinci foram mais bem avaliados no fator qualidade que os gerados pela Luban e obtiveram performance comparável aos gerados por designers humanos.

### 2.4.2 DALL-e

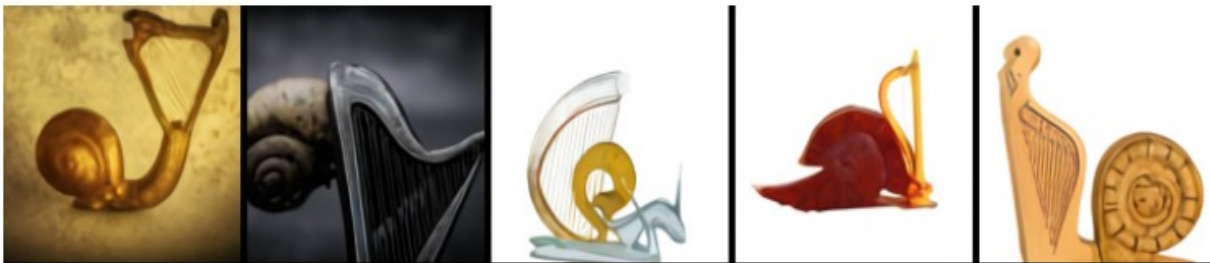
Outro trabalho que merece destaque é o proposto pela OpenAI, o Dalle [RPG<sup>+</sup>21], cujo nome é uma mistura da referência ao robô da pixar WALL-e e ao pintor Salvador Dali. Nele, uma entrada textual é processada dentro de uma rede neural de 12 bilhões de parâmetros, chamada de transformer, para gerar imagens a partir da interpretação do texto de entrada.

O trabalho que apresentou o DALL-E não divulgou completamente a arquitetura ou o código do modelo, contudo o modelo surpreendeu ao demonstrar os resultados obtidos a partir de entradas como “uma ilustração de um rabanete bebê em um tutu passeando com um cachorro” ou “um caracol feito de uma harpa”. Alguns exemplos de saídas para essas respectivas entradas podem ser vistas nas figuras 2.7 e 2.8.



**Figura 2.7** Imagens de um bebê rabanete passeando com um cachorro gerado pelo DALL-e.

**Fonte:** Adaptado da internet



**Figura 2.8** Imagens de um caracol feito de uma harpa gerado pelo DALL-e.

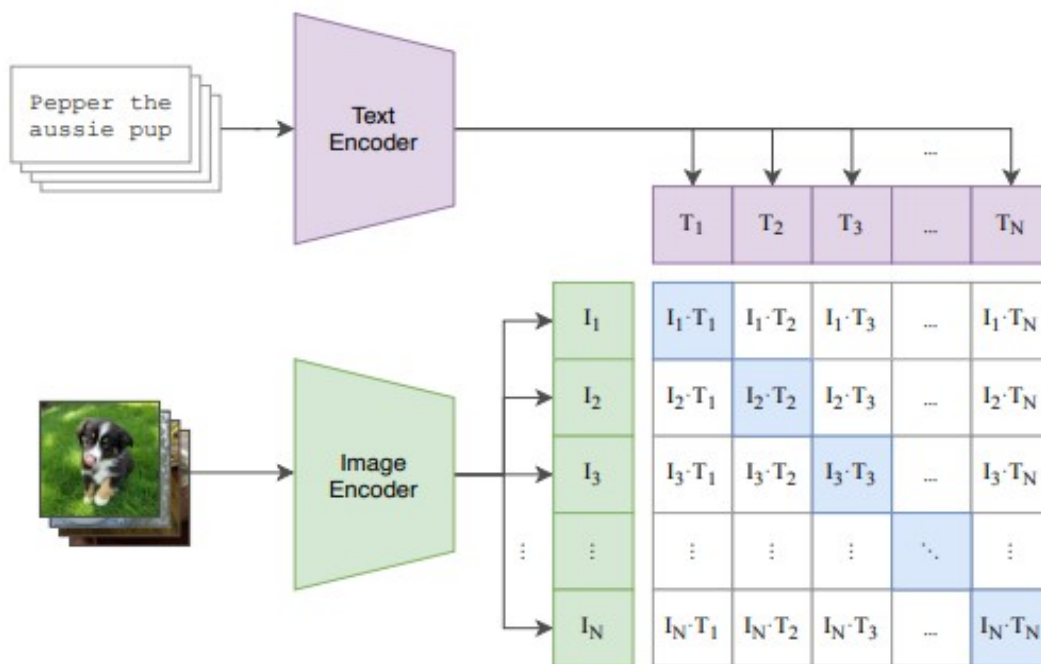
**Fonte:** Adaptado da internet

### 2.4.3 CLIP

Paralelamente ao DALL-E, a OpenAI, divulgou o trabalho denominado como CLIP acrônimo para *Contrastive Language-Image Pre-training*. O trabalho cria um modelo para ranquear quais imagens dentro de um conjunto retratam melhor uma determinada sentença ou entrada de texto. De uma forma mais abrangente, CLIP se propõe a resolver o problema: dados como entrada uma imagem e um texto, este par pertence ao mesmo conjunto?

CLIP é um modelo de *zero-shot* que tem por grande diferencial dos seus competidores o fato de ter sido treinado por um conjunto de imagens com curadoria de imagens rotuladas, como imagenet, CIFAR, wikiart e etc. As imagens utilizadas e as descrições dessas imagens foram extraídas da internet. Foram utilizados cerca de 400 milhões de pares texto-imagem. Isso evita o problema de por exemplo, numa imagem em que o elemento predominante fosse um cachorro, informações como a presença de outros elementos também presentes na imagem, como uma fonte, fossem perdidos.

A ideia do funcionamento do CLIP já era conhecida e consiste em passar tanto as entradas textuais como as visuais por encoders e maximizar a similaridade dos elementos da diagonal e minimizar a similaridade dos elementos de fora dela. Os resultados obtidos mostraram que essa abordagem pode ser competitiva com modelos supervisionados [RKH<sup>+</sup>21].



**Figura 2.9** Arquitetura do CLIP.

**Fonte:** Adaptado da internet<sup>1</sup>



## CAPÍTULO 3

# Proposta

Nas seções anteriores, foi discutido que o DALL-e, modelo no qual esse trabalho se inspira, foi desenvolvido pela OpenAI, que utilizava um transformer autoregressivo para geração de imagem a partir de texto, contudo pelo fato desde importante trabalho não ter divulgado a abordagem utilizada, como os pesos utilizados na rede ou código, o CLIP, trabalho também divulgado pela OpenAI, vem sendo foco de pesquisa de diversos membros da comunidade a fim de ser utilizado para rankear imagens geradas por outros modelos, como o *Big Sleep* e o *Deep Dream*. Tão logo quando lançado, foi utilizado como uma espécie de direcionador para os diferentes redes com os quais foi testado.

A abordagem utilizada no presente trabalho toma esta proposta como direcionamento e implementa uma arquitetura de VQ-GAN + CLIP[ali] para gerar imagens a partir de uma entrada textual e desenvolve uma interface que permita a publicitários a utilização de redes neurais, mais especificamente modelos adversarias generativos, sem a necessidade de entender os detalhes da tecnologia utilizada, nem a necessidade dos recursos de hardware que geralmente são necessários para a execução de tais modelos.

### 3.1 Caracterização do problema

Com o público-alvo definido, neste caso os publicitários, foi necessário uma pesquisa exploratória a fim de entender os principais pontos de deveriam ser levados em consideração no processo criativo. Tomando um formato de estudo de caso, foram entrevistados 4 publicitários, entre 22 e 28 anos, todos formados na área de publicidade e propaganda, tendo entre 1 e 5 anos de experiência no mercado, sendo duas mulheres um homem.

Nesta primeira etapa, a entrevista introduziu o conceito a ser estudado e buscou entender o processo criativo, avaliar se havia espaço para utilização de redes neurais dentro do contexto e identificar o que deveria ser implementado na plataforma para facilitar a criação de material.

**Tabela 3.1** Tabela com o perfil dos entrevistados.

Entrevistado	Sexto	Idade	Experiência
E1	F	22	6 meses
E2	F	26	1 ano
E3	M	28	1 ano e meio
E4	F	27	5 anos

Alguns relatos importantes foram:

E1 começou citando que *"Apesar de existir um processo criativo é preciso entender que a criatividade em si está no diferenciado contundo nem sempre tudo que é original será considerado bom."*

Do ponto de vista de E2 *"precisa-se entender o público-alvo e direcionar o processo criativo para que se possa facilitar o entendimento de onde se quer chegar."*

Para E3 *"É necessário entender qual tipo de material que pretende ser criado e qual seu objetivo, seja por exemplo engajar um público interno numa empresa a realizar determinada atividade ou incentivar a compra de produtos, o que é mais comum."*

Por fim E4 mencionou que *"Talvez a criação de uma ferramenta desse tipo auxiliasse mais quem tem pouca ou nenhuma experiência na área."*

Já sobre a possibilidade da utilização de uma rede neural para auxiliar no processo de criação E3 disse que *"É difícil prever a influência da utilização desse tipo de tecnologia visto que já existem algumas ferramentas online que não utilizam redes neurais mas geram um material final, apesar de serem materiais bem genéricos e uma montagem baseada em vários padrões já presentes dentro do sistema."* A partir de relatos como esse, fica claro que o foco deste trabalho deve ser a criação de uma ferramenta de assistência a publicitários e avaliar o reflexo da utilização do material gerado por ela no processo criativo.

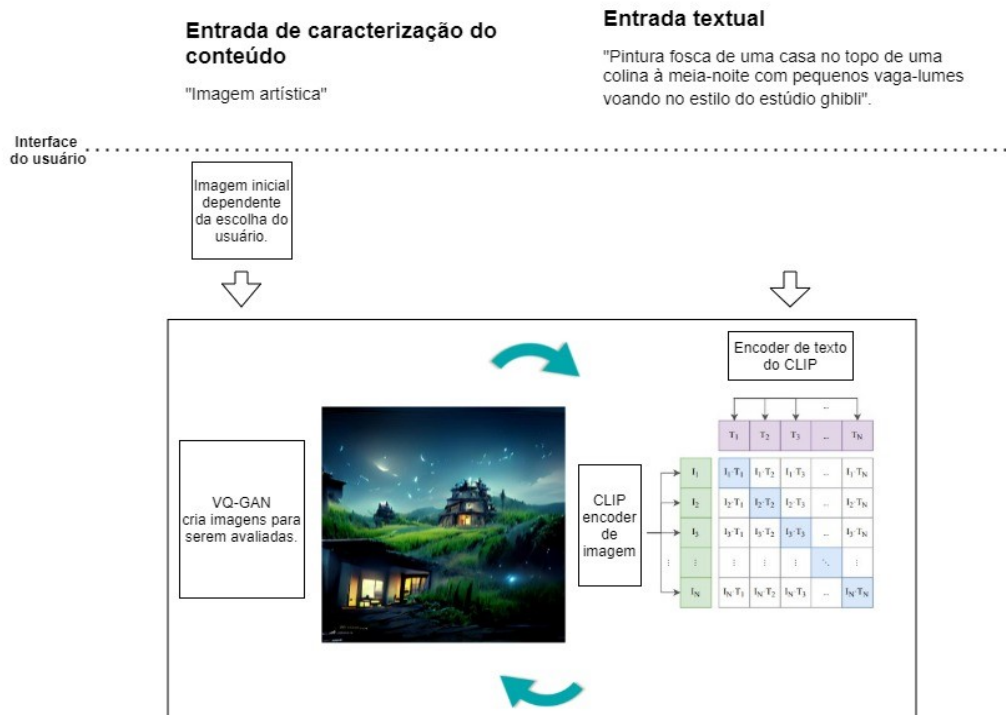
Apesar dos relatos mencionados acima não serem explicitamente iguais, todos os entrevistados mostraram entender que o público alvo e o objetivo da material que será criado modifica completamente as características do produto final e que nesse contexto devemos propiciar ao profissional a opção do mesmo prover entradas específicas, como uma entrada textual para descrição do que deve ser criado e opções sobre qual o objetivo do material.

## 3.2 Arquitetura

A partir das entrevistas realizadas na seção 3.1 e do conhecimento adquirido no estudo do estado da arte para geração de imagens, foi percebida a necessidade de estabelecer uma arquitetura que utilizasse uma VQ-GAN + CLIP como *core* da plataforma e uma interface que pudesse dar suporte no processo criativo. O detalhes da implementação dessa plataforma serão apresentados no capítulo 4.

Um diagrama com a arquitetura da plataforma é apresentado na figura 3.1. Ao usuário é dada a opção de dar três entradas, uma textual obrigatória, que descreve a imagem que deve ser gerada, uma entrada de múltipla escolha que define o principal objetivo da imagem final como engajamento ou aumento de faturamento e, caso deseje, uma entrada visual que serve como apoio para geração da imagem final. O modelo VQ-GAN é responsável por gerar imagens que são passadas como entrada para o encoder de imagens do CLIP, enquanto a entrada textual é passada para o enconder de texto do CLIP. Ao usuário é facultativo a entrada de objetivo e a imagem inicial, mas caso tenha optado por uma dessas alternativas, no primeiro caso será utilizado uma imagem pre-definida que tem como característica predominante demonstrar o objetivo escolhido e essa imagem será usada como imagem inicial do modelo, no segundo caso, a imagem dada como entrada é diretamente utilizada para gerar a imagem final. O processo de avaliação do CLIP é feito ciclicamente até número máximo pré-definido de iterações, que neste caso foi utilizado como 100, número com o qual o modelo já começava a apresentar imagens





**Figura 3.1** Arquitetura da plataforma.

**Fonte:** Autor

que já possuíam as características de convergência .

# Desenvolvimento

Este capítulo apresenta o desenvolvimento da plataforma e seus detalhes técnicos. A seção 4.1 descreve a escolha dos *datasets* utilizados e suas descrições e a seção 4.2 descreve a implementação de uma interface para profissionais de propaganda no Google Colaboratory.

### 4.1 Datasets

Três datasets foram utilizados para gerar as imagens: imagenet 1024, wikiart1024 e o sflickr. Os motivos principais para a escolha desses datasets foi o tamanho deles e o tempo de execução, que devido às limitações de hardware, não poderia ser muito grande para não afetar negativamente a experiência do usuário. Apresentado em ImageNet é uma base de dados de imagens organizado de acordo com a hierarquia WordNet, em que cada nó da hierarquia é representado por centenas de milhares de imagens[ima] é usada como referência na literatura para comparação de novos modelos de redes neurais. Já o wikiart é uma base de imagens gerada com obras de arte com mais de 250.000 itens de mais de 3.000 artistas. Por sua vez, o Flickr é referência padrão para a descrição de imagens baseada em frases. A necessidade da utilização de *datasets* diferentes se dá ao notar que precisamos oferecer uma diversidade imagens com características diferentes.

### 4.2 Plataforma

Contextualizado em um cenário onde se faz necessária a utilização de recursos de hardware como GPUs, este trabalho usou como alternativa à um servidor dedicado a plataforma desenvolvida pela Google para permitir o desenvolvimento de código em Python e acesso gratuito ao uso de GPUs, fatores atrativos aos desenvolvedores de redes neurais e analistas de dados.

Como mencionado no capítulo 3, este trabalho foi influenciado por um post da *Machine Learning at Berkeley (ML@B)* da Universidade da Califórnia que introduziu a arquitetura de VQ-GAN + CLIP. Tendo como ponto de partida o colab disponibilizado em [https://colab.research.google.com/drive/1ZAus\\_gn2RhTZWzOWUpPERNC0Q8OhZRTZ?usp=sharing](https://colab.research.google.com/drive/1ZAus_gn2RhTZWzOWUpPERNC0Q8OhZRTZ?usp=sharing), foi necessário algumas atualizações técnicas no código a fim de permitir o perfeito funcionamento da plataforma. Primeiramente, é importante citar que o código remove a necessidade do processo de treinamento do modelo, já que utiliza um *checkpoint* do estado da rede após o treinamento, isso ajuda a reduzir bastante o tempo total do processo. Contudo, alguns dos links de download dos checkpoints dos datasets estavam inválidos, sendo assim necessária a sua correção.

Dito isto, todo o código necessário para a ciclo básico de funcionalidade foi alterado para rodar com um único clique, preparando assim o ambiente para receber as entradas do usuário. Esse processo necessita ser executado a cada nova conexão a plataforma do Google Colab o que é um fator que influencia no tempo total do processo. O código que é considerado "ciclo básico" tem por função: instalar as bibliotecas Python necessárias para a execução da rede, clonar os repositórios do GitHub do *transformer* utilizado na rede e do CLIP, membros *core* da arquitetura desenvolvida. Além disso, baixa os *checkpoints* dos três *datasets* mencionados na seção 4.1.

Diferentemente do código original, algumas especificações técnicas foram abstraídas a fim de reduzir a complexidade de informações necessárias para a utilização da ferramenta como o número de *seeds*, o número de iterações e a escolha do *dataset*.

### 4.2.1 Interface

Como mencionado por E3 "*O processo de criação do material publicitário começa com um estudo sobre o contexto do produto, a partir daí é que se tem uma ideia de qual caminho se seguir...*", este relato nos faz refletir sobre como seria introduzido a intenção do publicitário na plataforma. O caminho mais natural é a adição de um *placeholder* para introdução de uma sentença que deve ser a descrição da imagem que ele gostaria de ver ou de algo que pudesse servir de inspiração.

Outro ponto a ser discutido é a possibilidade da geração de múltiplas imagens que pudessem servir de inspiração. Para atacar esse ponto, foi omitido do código original a escolha do *dataset* a ser utilizado pelo código, isto adicionado a execução da síntese de imagens utilizando o *checkpoints* da rede vários *datasets*. Assim, várias imagens com características diferentes, devido às características de cada conjunto de dados utilizado para treinar a rede, puderam ser criadas.

Como E2 mencionou na seção 3.1, o propósito do trabalho deve ser levado em consideração na criação do material, sendo assim, duas decisões importantes tiveram que ser tomadas: quanto inserção de uma imagem de entrada e quanto a adição de contextos internos a plataforma.

A primeira decisão diz respeito ao fato que a síntese de imagens parte a partir do ruído, ou seja, do ponto de vista humano a imagem não possui nenhum contexto semântico. Sendo assim, a imagem é criada do zero. Para dar uma alternativa ao usuário de inserir uma imagem e a partir de então essa imagem ser usada como ponto de partida. Isso se reflete em uma entrada textual que contém um caminho para uma imagem presente no Google Drive, infelizmente devido a limitações do Colab nenhuma implementação mais robusta pôde ser feita.

A segunda decisão diz respeito a propagandas com objetivos específicos, como: incentivar consumo e engajamento interno. Com esse objetivo, foi adicionada uma seleção na qual o usuário poderia escolher entre um dos objetivos citados acima. Essa opção permite a síntese de uma imagem que em vez tomar de tomar como início o ruído, vai usar uma imagem pré-definida pelo sistema que possui alguma característica que seja coesa com o objetivo selecionado. A escolha de apenas duas entradas facilita a validação da ideia de plataforma.

Para finalizar, é notável a diferença das características dos *datasets* e das imagens finais sintetizadas a partir de: i) um *dataset* com imagens reais ii) um *dataset* com imagens de obras artísticas. Para dar uma alternativa ao usuário de escolher se a obra teria detalhes mais abstratos,

foi adicionado uma *checkbox* que adicionaria o *dataset* WikiArt.

Um ponto positivo da utilização do Google Colab é possibilidade do uso dos recursos de widgets, que apesar de bastante simples, servem ao propósito do teste de conceito proposto.

A figura 4.1 mostra o widget que espera como entrada:

- Uma entrada textual(obrigatória), que descreve o conteúdo da imagem final a ser gerada.
- Um endereço do google Drive com uma imagem que será usada como base no processo de síntese da imagem.(opcional)
- Uma entrada via combo box, que define o propósito da propagada, seja incentivo ao consumo, motivacional ou artísticos. Esses três foram escolhidos na etapa de proposta a fim reduzir o número de possibilidades e a complexidade do problema.(opcional)

Para os itens 2 e 3, se escolhido uma entrada inicial ou um propósito, uma imagem inicial será dada como entrada para a rede, caso contrário, o processo de síntese se inicia a partir de ruído.



**Figura 4.1** Entradas da rede.

**Fonte:** Autor

O código final da plataforma pode ser encontrado disponível em: <https://colab.research.google.com/drive/1H1iKXOcbdLeQ10tP7lh9Cs-jQJEhcir0?usp=sharing>.

## CAPÍTULO 5

# Avaliação

Este capítulo apresenta a metodologia e discorre sobre resultados obtidos e está organizado da seguinte forma. A seção 5.1 apresenta as métricas utilizadas na avaliação do sistema proposto no capítulo 4 e a Seção 5.2 aponta os resultados obtidos após a avaliação dos usuários. Por fim, a seção 5.3 apresenta a discussão dos resultados, especialmente da etapa qualitativa.

### 5.1 Métricas de avaliação

Após a etapa de desenvolvimento todos os entrevistados foram submetidos a utilização da interface proposta no capítulo 4 a fim de avaliar o problema quantitativamente e qualitativamente. Segundo [?] os métodos podem ser definidos como:

- **Quantitativa:** analisa opiniões e informações através de números e de técnicas estatísticas (percentagem, média, desvio padrão, entre outros).
- **Qualitativa:** análise descritiva de dados, que tende ser feita individualmente, que não podem ser traduzidos em números.

Com o objetivo de gerar essas informações os entrevistados avaliaram a plataforma nos quesitos de experiência do usuário e usabilidade, além de responderem uma entrevista final sobre engajamento e dissertar subjetivamente sobre a plataforma e o material gerado.

#### 5.1.1 Usabilidade

Para avaliar a plataforma do quesito usabilidade foi utilizado o *System Usability Scale (SUS)*. Proposto em [B<sup>+</sup>96], o texto afirma que não ha medidas absolutas para usabilidade e que a usabilidade precisa ser definida com referencia a um contexto específico. O formulário do *SUS* traz uma série de afirmações com as quais o entrevistado deve responder em uma escala de 0 a 4 se discorda totalmente ou concorda totalmente. Após o preenchimento do formulário os itens 1,3,5,7 e 9 contribuem com a posição menos 1 e os itens 2,4,6,8 e 10 contribuem com 5 menos a posição na escala. Ao final, o valor total deve ser multiplicado por 2,5 para gerar o score final que varia entre 0 e 100. Para Jhon, a pontuação obtida indica que:

- **menor que 60:** Inaceitável.
- **60-69:** OK.
- **70-79:** Bom.

- **80-90:** Excelente.
- **maior que 90:** Melhor usabilidade possível.

Devido a alta utilização deste método para medir usabilidade, é considerado altamente confiável e válido.

### 5.1.2 Experiência do usuário

Já para a avaliação da plataforma quanto a experiência do usuário foi usado o software UEQ (*User Experience Questionnaire*). Em 2006 três experts da área de tecnologia da informação desenvolveram o UEQ. O formulário consiste em 26 pares de adjetivos, um antônimo do outro, no qual o entrevistado escolhe uma alternativa numa escala de 1 a 7. No site *www-ueq-online-org* é disponibilizado o formulário em 30 línguas além da ferramenta da análise de dados numa planilha de excel. A ferramenta classifica informações como: Atratividade, Perspicuidade, Eficiência, Confiabilidade, Estimulação e Novidade numa escala entre péssimo e excelente.

### 5.1.3 Análise qualitativa

Apesar de bem fundamentado numa análise quantitativa, a classificação da plataforma como uma ferramenta de auxílio no processo criativo traz consigo a necessidade de uma análise qualitativa, onde o usuário se sentisse mais aberto a expor sua opinião e citar por exemplos de melhoria da plataforma.

Por fim, foi feita uma análise de todos os dados coletados, de modo a avaliar a plataforma e apresentar os principais pontos de trabalho.

E1, por exemplo avaliou os resultados obtidos pela inserção dos parâmetros inseridos por ele como *"Muito abstratos, mas é interessante perceber que algumas características da marca foram adicionadas à imagem."*. A imagem 5.1 mostra a saída gerada pelo conjunto de parâmetros: i) texto: "A man with wings drinking, RedBull.", sem entrada inicial, "consumo" como objetivo e "imagens reais" como característica.



**Figura 5.1** Saídas geradas a partir do texto: "A man with wings drinking, RedBull."

**Fonte:** Autor

Para testar o conceito, o parametro "característica" foi mudado para artístico e o resultado foi a figura 5.3.



**Figura 5.2** Saída gerada a partir do texto: “*A man with wings drinking, RedBull.*”, com a característica “artística”.

**Fonte:** Autor

Para finalizar os exemplos, segue a saída a partir das entradas providas por E4. Os parâmetros foram os seguintes, texto: “an image with two people holding hands”, sem entrada inicial, “engajamento” como objetivo e “artístico” como característica. E4 conclui que “*é estranho mais ainda sim da pra ser utilizado dentro de um processo de 'brainstorm'*”.



**Figura 5.3** Saída gerada a partir do texto: “*An image with two people holding hands.*”, com a característica “artística”.

**Fonte:** Autor

Como contribuição final, o trabalho apresenta uma discussão sobre os resultados e propostas futuras de trabalhos na área.

## 5.2 Resultados e Discussão

Logo após a utilização da ferramenta os usuários foram submetidos aos testes de usabilidade e experiência do usuário. Para o teste SUS, de usabilidade, os resultados obtidos foram os mostrados na tabela 5.1

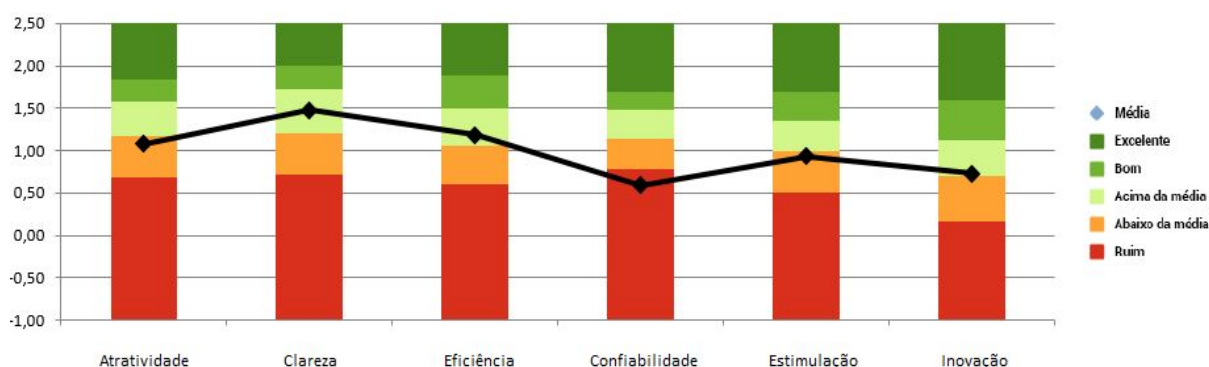
Os resultados de usabilidade mostram que apesar de simples, a interface da plataforma alcança um bom nível de satisfação durante seu uso, estando na média para sistemas de boa usabilidade. Contudo, alguns comentários foram feitos sugerindo uma funcionalidade *drag*

**Tabela 5.1** Tabela com o resultados do SUS.

Entrevistado	Pontuação	Resultado
E1	70	BOM
E2	72.5	BOM
E3	62.5	OK
E4	77.5	BOM

and drop para a inserção da imagem inicial.

Já para o teste de experiência de usuário os resultados foram os mostrados na figura 5.4.



**Figura 5.4** Resultados médios obtidos pelo User Experience Questionnaire.

**Fonte:** Autor

Em linhas gerais a plataforma obteve um desempenho na média mas a confiabilidade foi um ponto de avaliação que se mostrou como fraqueza da plataforma.

Após a reposta dos formulários para análise quantitativa, os usuários foram convidados a uma avaliação subjetiva da plataforma. Em geral, a ideia de uma ferramenta de assistência foi bem aceita pelos entrevistados, contudo um relato bastante pertinente foi o de E4, que citou: "*Se eu estou descrevendo mais ou menos o que eu quero ver, eu já tenho ideia do meu produto final.*". Isso nos faz refletir a abordagem utilizada e como podemos fazer para melhorá-la.

O tempo médio de execução do processo de desenvolvimento foi cerca de 45 minutos, para a criação das três imagens. Este processo engloba toda a execução, seja o download das bibliotecas necessárias para o funcionamento do código e a execução do modelo com os três *datasets*.

### 5.3 Discussão

A métrica final para avaliação do SUS é a média dos valores obtidos. Com um média de 70.625 a avaliação SUS da plataforma ficou um pouco acima da média geral de 68 pontos indicando que a plataforma tem um nível de usabilidade na média. Já o resultado negativo no quesito "confiabilidade" pode indicar que alguma pergunta pode não ter sido corretamente interpretada



ou que devido a abstração presente nas imagens geradas ainda é um fator imprevisibilidade do sistema.

Durante o processo de avaliação, ficou evidente que a falta de flexibilidade da interface foi um fator que incomodou os usuários, a necessidade de carregar uma imagem no Google Drive, a falta de uma forma de maior interação com o usuário e o processo consideravelmente lento, levando-se em conta que os usuários não tem conhecimento do quão complexo é a execução de de uma rede neural, foram fatores que desengajaram os usuários durante o processo. Isso se une ao fato de que para alguns propósitos, como propaganda para interno por exemplo, seria ideal gerar imagens realistas e devido a isso a ferramenta não cumpriu completamente seu propósito.

Ademais, o foco principal era avaliar a utilização de uma ferramenta de apoio nova, para profissionais que não possuem nenhum conhecimento técnico prévio. Neste contexto a ferramenta cumpriu com as expectativas, sendo de fácil utilização e permitindo aos usuários usar seus recursos sem a necessidade de auxílio de alguém especialmente capacitado para isso.

Por fim os relatos apresentados na análise qualitativa, indicam que para trabalhos futuros, seria interessante avaliar a ferramenta com pessoas que possuem pouco conhecimento no desenvolvimento de publicidade ou não tem recursos para por exemplo, contratar um profissional ou empresa do ramo publicitário.

## Conclusão

Neste trabalho foi proposto e desenvolvido uma plataforma que funciona como ferramenta de apoio no processo criativo de publicitários no Google Colaboratory para utilizar dos recursos de hardware necessários a fim de rodar uma arquitetura de VQ-GAN + CLIP abstraindo a complexidade e conhecimento necessário para tais profissionais.

Os dados foram obtidos através de uma análise quantitativa e qualitativa a partir de entrevistas e preenchimento de formulário e avaliam a ferramenta nos aspectos de usabilidade, experiência do usuário e uma análise subjetiva sobre expectativa x realidade, além de abrir um espaço para sugestões.

Algumas limitações enfrentadas durante a realização do trabalho que podem indicar que os dados não refletem o real cenário:

- A maioria dos entrevistados vêm do ambiente acadêmico com pouca experiência no mercado.
- Devido a limitação de tempo o número de publicitários que se dispuseram a participar da pesquisa não pôde ser maior.
- Pelo fato de não haver publicação dos pesos da rede ou código do DALL-e não foi possível avaliar a ferramenta gerando imagens hiper-realistas. Por se tratar de um trabalho multidisciplinar, houve um *gap* de conhecimento que demorou para ser solucionado.

Apesar das limitações, este trabalho contribui no entendimento do processo criativo de publicitários e na forma de como esse conhecimento pode ser utilizado no ramo de aprendizagem de máquina a fim de auxiliar publicitários durante o processo. Além disso, os resultados sugerem um potencial na ferramenta especialmente devido a melhorias contínuas que vêm havendo na síntese de imagens. Quanto a plataforma, pelo seu caráter inicial, obteve o resultado esperado quanto a simplicidade e indicou que há espaço para aprofundamento e melhorias.

### 6.1 Trabalhos futuros

A coleta de informações dos entrevistados permitiu identificar possíveis linhas de pesquisa e desenvolvimento que podem ser realizados, como:

- Eliminar a necessidade de se rodar o processo desde o início já que a sessão é reiniciada a cada nova conexão no Google Colaboratory. Isso pode ser feito por exemplo, utilizando um servidor próprio com os recursos de hardware necessário para execução da arquitetura.

- Utilizar mais redes além da VQ-GAN para avaliar os resultados de redes diferentes.
- Realizar um estudo com mais publicitários e com maior nível de experiência.
- adicionar *features* na plataforma a fim de torna-la mais completa.

## Referências Bibliográficas

- [ali] Alien dreams: An emerging art scene.
- [Ama96] Teresa M Amabile. Creativity in context: Update to. 1996.
- [B<sup>+</sup>96] John Brooke et al. Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7, 1996.
- [Bel92] Jack A Bell. Creativity, tv commercial popularity, and advertising expenditures. *International journal of advertising*, 11(2):165–172, 1992.
- [CL14] Erin Cherry and Celine Latulipe. Quantifying the creativity support of digital tools through the creativity support index. *ACM Trans. Comput.-Hum. Interact.*, 21(4), June 2014.
- [CWD<sup>+</sup>18] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1):53–65, 2018.
- [dO21] Claudio Carvalho de Oliveira. Utilizando redes adversárias generativas (gans) como agente de apoio à inspiração para artistas. *Universidade Federal de Pernambuco - Trabalho de Graduação*, 2021.
- [DPOK21] Sebastiaan De Peuter, Antti Oulasvirta, and Samuel Kaski. Toward ai assistants that let designers design. *arXiv preprint arXiv:2107.13074*, 2021.
- [DW06] Paul Dyson and Karl Weaver. Advertising’s greatest hits: profitability and brand value. *Admap*, 469:28, 2006.
- [ERO21] Patrick Esser, Robin Rombach, and Bjorn Ommer. Taming transformers for high-resolution image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12873–12883, 2021.
- [GJS<sup>+</sup>21] Shunan Guo, Zhuochen Jin, Fuling Sun, Jingwen Li, Zhaorui Li, Yang Shi, and Nan Cao. Vinci: An intelligent graphic design system for generating advertising posters. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2021.
- [GPAM<sup>+</sup>14] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

- [ima] Imagenet.
- [Kel19] John D Kelleher. *Deep learning*. MIT press, 2019.
- [KW13] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.
- [Lan21] Robin Landa. *Advertising by design: generating and designing creative ideas across media*. John Wiley & Sons, 2021.
- [LGZ18] Mu-Chu Lee, Bin Gao, and Ruofei Zhang. Rare query expansion through generative adversarial networks in search advertising. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 500–508, 2018.
- [LSL18] Changliang Li, Yixin Su, and Wenju Liu. Text-to-text generative adversarial networks. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–7. IEEE, 2018.
- [Pri14] I Priest. Client-agency relationship: Seven principles for better commercial creativity. *Market Leader, Quarter*, 3, 2014.
- [RKH<sup>+</sup>21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. *arXiv preprint arXiv:2103.00020*, 2021.
- [RPG<sup>+</sup>21] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. Zero-shot text-to-image generation. *arXiv preprint arXiv:2102.12092*, 2021.
- [SB11] Chris Solomon and Toby Breckon. *Fundamentals of Digital Image Processing: A practical approach with examples in Matlab*. John Wiley & Sons, 2011.
- [SC] Amy Smith and Simon Colton. Clip-guided gan image generation: An artistic exploration. *Evo\* 2021*, page 17.
- [SEB<sup>+</sup>18] Othman Sbai, Mohamed Elhoseiny, Antoine Bordes, Yann LeCun, and Camille Couprie. Design: Design inspiration from generative networks. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018.
- [SOU20] Camila Oliveira de SOUZA. Usando convolução separável em profundidade na otimização da arquitetura squeezeNet. Master’s thesis, Universidade Federal de Pernambuco, 2020.
- [TW17] Sarah Turnbull and Colin Wheeler. The advertising creative process: A study of uk agencies. *Journal of Marketing Communications*, 23(2):176–194, 2017.

- [VSP<sup>+</sup>17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Advances in neural information processing systems. *Proceedings of Machine Learning Research*, pages 5998–6008, 2017.
- [WCY<sup>+</sup>20] Lei Wang, Wei Chen, Wenjia Yang, Fangming Bi, and Fei Richard Yu. A state-of-the-art review on image synthesis with generative adversarial networks. *IEEE Access*, 8:63514–63537, 2020.