

DOSSIÊ

Narrativas de IA: tendências da
produção audiovisual



V. 14 – N. 3 - set./dez. 2023

ISSN: 2179-1465 / <https://www.revistageminis.ufscar.br>
DOI: <https://doi.org/10.14244/2179-1465.RG.2023v14i3p59-86>

ROTEIROS COMO PROMPTS: UMA EXPLORAÇÃO DE PRÉ-VISUALIZAÇÃO DE CENAS DE TITANIC COM FERRAMENTAS DE INTELIGÊNCIA ARTIFICIAL

SCRIPTS AS PROMPTS: A PREVIEW EXPLORATION OF SCENES FROM
TITANIC WITH ARTIFICIAL INTELLIGENCE TOOLS

GUIONES COMO PROMPTS: UNA EXPLORACIÓN PREVIA DE ESCENAS DE
TITANIC CON HERRAMIENTAS DE INTELIGENCIA ARTIFICIAL

Taís de Barros

Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
ORCID: <https://orcid.org/0000-0003-2570-0583>
Porto Alegre, RS, Brasil

Roberto Tietzmann

Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS)
ORCID: (<https://orcid.org/0000-0002-8270-0865>)
Porto Alegre, RS, Brasil

Recebido: 15/10/2023 / Aprovado: 25/02/2024

Como citar: BARROS, T. de; TIETZMANN, R. Roteiros como Prompts: uma exploração de pré-visualização de cenas de Titanic com ferramentas de Inteligência Artificial. Revista GEMInIS, v. 14, n. 3, p. 59–86, 2023.

Direito autoral: Sob a Licença Creative Commons-Atribuição 3.0 Internacional.



RESUMO

Este texto questiona como Plataformas de Inteligência Artificial Generativa (IAGs) podem colaborar com a pré-visualização de cenas, parte do processo de criação de uma obra audiovisual. Para isto, selecionamos uma cena do roteiro do filme *Titanic* (1997) como fonte dos comandos para a geração de imagens. Utilizamos uma metodologia exploratória, revisamos a literatura compilando conceitos básicos sobre Inteligência Artificial Generativa, testamos as plataformas *Stable Diffusion*, *Bing Image Creator* e *Genmo.AI* e comparamos as imagens geradas com os segmentos do filme original e discutimos suas aproximações e diferenças. Como resultados, destacamos que as ferramentas forneceram resultados consistentes com os segmentos de roteiro, embora nem sempre alinhadas com o que o filme apresentou, o que indica um processo de tomada de decisões pelas IAGs, o que as coloca como ferramentas relevantes para a pré-visualização de um roteiro, embora não abstraíam a presença humana na criação das cenas.

Palavras-chave: Cinema; Inteligência Artificial; Titanic.

ABSTRACT

This text questions how Generative Artificial Intelligence Platforms (GAIPs) can collaborate in the previsualization of scenes, a part of the audiovisual work creation process. To do this, we selected a scene from the script of the movie *Titanic* (1997) as the source of commands for generating images. We used an exploratory methodology, reviewed the literature by compiling basic concepts about Generative Artificial Intelligence, tested the *Stable Diffusion*, *Bing Image Creator*, and *Genmo.AI* platforms, and compared the generated images with segments from the original film, discussing their similarities and differences. As a result, we highlight that the tools provided consistent results with the script segments, although not always aligned with what the film presented, indicating a decision-making process by the GAIPs, making them relevant tools for script previsualization while not abstracting the human presence in scene creation.

Keywords: Cinema; Artificial Intelligence; Titanic.

RESUMEN

Este texto cuestiona cómo las Plataformas de Inteligencia Artificial Generativa (IAGs) pueden colaborar con la previsualización de escenas, parte del proceso de creación de una obra audiovisual. Para ello, seleccionamos una escena del guión de la película *Titanic* (1997) como fuente de los comandos para generar imágenes. Utilizamos una metodología exploratoria, revisamos la literatura recopilando conceptos básicos sobre Inteligencia Artificial Generativa, probamos las plataformas *Stable Diffusion*, *Bing Image Creator* y *Genmo.AI* y comparamos las imágenes generadas con los segmentos de la película original y discutimos sus aproximaciones y diferencias. Como resultado, destacamos que las herramientas brindaron resultados consistentes con los segmentos del guión, aunque no siempre alineados con lo que presentó la película, lo que indica un proceso de toma de decisiones por parte de los IAGs, que las ubica como herramientas relevantes para la previsualización de una película. guión, aunque no abstraen la presencia humana en la creación de las escenas.

Palabras Clave: Cine; Inteligencia artificial; Titanic.

1. INTRODUÇÃO

O ano de 2023 foi marcante por diversos motivos. Este foi o ano em que a OMS decretou o fim da emergência sanitária da pandemia de COVID-19 (OPAS, 2023) e que serviços e plataformas de inteligência artificial generativa (a partir daqui citadas como IAGs) ganharam a atenção do mundo ao se inserirem rapidamente em diversas áreas de atividade, cultura e economia, inclusive catalisando uma greve de atores e roteiristas estadunidenses em busca de condições de trabalho que contemplassem proteções frente a este novo contexto (PONIEWOZIK, 2023).

Quando exploramos as imagens estáticas e em movimento produzidas pelas plataformas e ferramentas de IAG como *Dall-E* (MEHDI, 2023), *Midjourney* (2023), *Genmo.AI* (GENMO, 2023), *Kaiber* (2023), *Stable Diffusion* (STABILITY AI, 2023), entre tantas outras, muitas vezes é difícil evitar um sorriso e a curiosidade ao interagir com estas cornucópias de imagens possibilitadas por frases simples. Por um lado, é como se boa parte da dificuldade de visualizar uma ideia tivesse sido removida. Imprecisas e abundantes, tais imagens sugerem não serem definitivas, mas rascunhos em um contínuo processo. Ao mesmo tempo, dão continuidade a um movimento de coautoria que incorpora de trabalhos de terceiros que começou séculos atrás com os clichês gráficos (MONTEIRO, 2003), dedicados a agregar com facilidade adereços gráficos pré-prontos para impressos, transitou pelas imagens em *clip art* e chegou aos bancos de imagens sons e vídeos licenciáveis na internet contemporânea. Com as IAGs, a autoria é diluída entre a plataforma que gera as imagens, o modelo que a alimenta, as imagens que formataram o modelo e o comando dado pelo usuário.

Mas que imagens estas plataformas de IAG criam? Após algumas explorações documentadas (TIETZMANN; PUHL, 2023; TIETZMANN; MONTARDO; PUHL, 2023) e outras em projetos de iniciação científica¹, podemos afirmar que elas são flexíveis em suas formas de representação, buscando atender aos comandos e podendo mimetizar com grande fidelidade imagens capturadas com câmeras com variados tipos de lentes, diversos estilos artísticos, incontáveis formas de iluminar as cenas, sugerir texturas e atmosferas, profundidade de campo e outros detalhes granulares que formam a experiência estética de uma imagem. Estas imagens de síntese podem ser de difícil identificação como tais, salvo pela presença de metadados, de uma autodeclaração de seus criadores ou da identificação de marcas contextuais e inconsistências em sua execução.

Esta versatilidade pode ser lida a partir de pelo menos duas abordagens. A primeira delas é, emprestando o termo cunhado por Umberto Eco (ECO, 1976), apocalíptica: seriam estas as tecnologias que encerrariam as profissões criativas em definitivo por serem capazes de resolver todas

¹ Agradecemos aqui aos bolsistas Mirelli Garcia, Guilherme Taborda e Maya Scuri da PUCRS pela colaboração em etapas iniciais do planejamento.

as necessidades visuais com o mínimo de investimento. Por outro lado, a mesma versatilidade visual despertou a curiosidade dos autores deste artigo e levou a pergunta da exploração empreendida neste texto: uma vez que o processo criativo de uma peça audiovisual é um intenso processo de elaboração, seleção, descarte e refinamento da obra que transita da ideia, à palavra, à imagem e à tela, seriam estas ferramentas auxiliares potencialmente relevantes para esta elaboração nos bastidores? Podem colaborar com a pré-visualização de cenas a partir de um roteiro?

Procuramos responder a estes questionamentos a partir de uma metodologia exploratória que alinha uma revisão de fontes bibliográficas, hemerográficas e do teste de ferramentas de IAGs. Organizaremos este artigo da seguinte maneira: uma explicação breve da tecnologia de base para estas inteligências artificiais generativas, a apresentação de nosso percurso metodológico que inclui as ferramentas *Stable Diffusion* versão 1.5 (ROMBACH *et al*, 2022), *Bing Image Creator* (MEHDI, 2023) e *Genmo.AI* (GENMO, 2023), concluindo o artigo com a aplicação da metodologia e a discussão dos resultados.

2. APROXIMANDO OS CONCEITOS DE RECUPERAÇÃO DA INFORMAÇÃO NA INTERNET COM O FUNCIONAMENTO DE MODELOS DE IAGS

Uma das primeiras coisas que impressiona ao utilizarmos um serviço amplamente acessível como o ChatGPT (OPENAI, 2023) ou o Google Bard (BARD, 2023) é a versatilidade em encontrar respostas para perguntas e proposições complexas, independente da precisão das réplicas. Ao invés de apenas encontrarmos o que já existe publicado em algum recanto da internet encontramos algo com algum grau de novidade a cada vez. A possibilidade de fazer perguntas em linguagem fluente e encontrar respostas traz uma ruptura com os paradigmas anteriores para a organização e recuperação de informações na internet, baseados em protocolos de comunicação entre as máquinas em rede estabelecidos há cerca de três décadas, migrando desde então para o senso comum a respeito do uso da tecnologia.

O mais simples deles, chamado Gopher² (toupeira, em português), foi criado em 1991 e é baseado em uma hierarquia de pastas e subpastas que continham documentos, semelhante aos sistemas de arquivos em computadores pessoais. Embora fosse fácil de entender e exigisse pouco processamento das máquinas e redes, era baseado em texto e tinha interação limitada. Em paralelo a

² Ele foi originalmente chamado de "Gopher" em homenagem ao mascote da Universidade de Minnesota, onde o protocolo foi desenvolvido. A sigla Gopher significa "Gopher Open Protocol for Distributed Information Retrieval", em tradução livre para o português, "Protocolo Gopher Aberto para Recuperação de Informações Distribuídas".

³ Usamos o tempo verbal no presente, uma vez que servidores e clientes deste protocolo ainda operam, mesmo que quase esquecidos após o surgimento de outras ferramentas.

isso, o HTTP (Protocolo de transferência de Hipertexto), criado em 1989, começou a ganhar espaço. Ele serviu de base para boa parte da experiência atual da Internet. A diferença fundamental é que o segundo foi projetado para transmitir hipertexto, o que facilitava a vinculação de conteúdos e a agregação de imagens, vídeos e outros conteúdos relacionados. Assim, enquanto o primeiro protocolo exigia rigor em guardar e recuperar a informação desejada, o segundo permitia localizar conteúdos de forma associativa.

O que nenhum dos dois possuía era um serviço de busca desde os primeiros momentos, deixando o trabalho pesado de encontrar as informações desejadas para os usuários, que precisavam elaborar e registrar caminhos de navegação. Empresas hoje amplamente conhecidas, como o Google, se destacaram a partir da metade da década de 1990 por definir o paradigma da busca por texto, em que uma simples frase recupera links para páginas que contenham esse texto, sendo as mais populares ou as patrocinadas mostradas em destaque. Nesses sistemas os conteúdos já existentes são indexados e podem ser buscados, reduzindo o esforço para acessá-los, mas sem a característica conversacional trazida pelas IAGs baseadas em modelos amplos de linguagem em seu fundamento.

Um *Large Language Model* (LLM), ou Modelo Amplo de Linguagem em tradução livre, é o que está na base da ruptura deste paradigma. Em essência é um modelo de inteligência artificial dedicado ao processamento de linguagem natural. Ele é projetado para manipular e gerar dados em larga escala. Esses modelos são baseados em redes neurais, que emulam o funcionamento do pensamento humano, e são treinados com a leitura de grandes quantidades de dados, principalmente textuais, para aprender padrões linguísticos e contextuais. Redes neurais têm sido objeto de pesquisa desde pelo menos a década de 1980, embora só tenham se tornado protagonistas recentemente, como coloca Crawford (2021). Em suma, eles são capazes de entender e gerar texto em uma variedade de assuntos, fornecer informações, realizar tarefas de linguagem natural e até mesmo participar de conversas como parceiros de diálogo virtual.

A arquitetura *Transformer* é predominante nas aplicações generativas, definida por Uszkoreit (2017) como “(...) uma arquitetura de rede neural inovadora baseada em um mecanismo de autoatenção que acreditamos ser particularmente bem adequada para o entendimento da linguagem”. O diferencial foi a introdução de uma abordagem de "atenção", que permite ao *Transformer* capturar informações contextuais em textos de maneira eficiente e atribuir diferentes pesos às palavras em uma sequência, dependendo de sua importância contextual e posição na frase para a demanda em questão (VASWANI, SHAZEER, PARMAR *et al*, 2017). Os GPTs (*Generative Pre-trained Transformer*) aplicam esse entendimento do repertório textual ao 'inverter o sentido' da leitura de materiais, criando uma resposta a partir de uma pergunta. Eles são capazes de usar o conhecimento

acumulado das relações e referências de maneira generativa, ou seja, são capazes de criar entregas originais e não baseadas em respostas previamente indexadas, como aquelas encontradas por um motor de busca na Internet.

Em um primeiro momento, é possível contemplar que esses modelos generativos seriam capazes de ocupar espaços na produção textual relacionada ao audiovisual, como no caso dos roteiros. Contudo, os modelos de linguagem não foram treinados apenas em linguagem textual, mas também em bancos de imagens, sequências de vídeo, sons e músicas, entre outros conteúdos, expandindo as possibilidades da proposta de atenção dos *Transformer*.

Um avanço que temos observado desde 2022 é a chegada ao grande público de plataformas generativas de imagem e a construção de conexões entre os modelos textuais e os modelos capazes de gerar todo o tipo de conteúdo. Estes modelos de linguagem capazes de traduzir texto em imagem recuperam de forma indireta um processo fundamental na criação audiovisual: a centralidade que o roteiro tem na relação com a obra que será realizada. O que está em palavras sugere que imagens surgirão ao final do processo de produção da obra, com diversas etapas de iteração criativa ao longo do processo, o que nos leva a de volta à pergunta inicial: podem essas novas tecnologias colaborar com este processo?

3. UM PERCURSO METODOLÓGICO EXPERIMENTAL: DA PALAVRA À IMAGEM COM APOIO DA IA

A proposta metodológica que apresentamos e aplicamos aqui se situa em um momento de tradução e transformação durante o processo de realização audiovisual: o momento em que o roteiro começa a ser interpretado em imagens. Rabiger (2008, p. 274) em seu livro sobre direção afirma que um filme “não encontra uma forma visual [sozinho]; você tem que impor ativamente uma” que o autor sugere ser alcançada através da coleção de referências e de esboços a serem posteriormente sistematizados. Esta costuma ser uma atividade que acontece durante o período de pré-produção da obra e envolve o núcleo de direção, o núcleo de arte e ilustradores ou artistas visuais para a produção de artes conceituais e *storyboards* que podem eventualmente ser desdobrados em rascunhos animados. Segundo Glebas (2009, p. 47) “Um roteiro é um plano verbal para uma história. Um *storyboard* é um plano para a visualização dessa história” e representa o momento em que a experimentação de soluções visuais envolve tanto a criação quanto o descarte destas soluções preliminares.

Este é um momento da realização de uma obra audiovisual em que entendemos haver um alinhamento direto entre as possibilidades das plataformas de geração de imagem e o processo

criativo que vai do texto até as telas. A ideia para esse percurso metodológico consiste em três etapas: selecionar uma cena de um roteiro, criar imagens que traduzam este roteiro e discutir os resultados.

Como critério de inclusão e seleção do roteiro, demos preferência a roteiros que já tenham sido filmados e sejam de amplo conhecimento das plateias. A justificativa é que consideraremos o filme final já lançado como uma referência de controle a partir do qual podem ser avaliados os resultados produzidos pelas IAGs.

Após o levantamento e a leitura de alguns roteiros, escolhemos o longa-metragem *Titanic* (1997), escrito e dirigido por James Cameron como ponto de partida. Justificamos a escolha desta obra por ser um filme com cenas de grande apelo visual, roteiro descritivo – complementado por alguns rebuscos de linguagem – e grande circulação. Baseado em fatos históricos, a obra recebeu diversos prêmios e foi acolhido pelas plateias globalmente nas diversas telas onde continua a ser exibido. Além disso, *Titanic* (1997) está situado em um ponto da evolução dos efeitos visuais em que a manipulação digital das imagens passou a permitir uma maior flexibilidade criativa aos realizadores, como coloca Rickitt (2000), representando um ponto de inflexão na história do meio.

O roteiro foi coletado em um repositório online e a cena selecionada foi extraída de uma cópia em DVD do filme, com as legendas desligadas. Em especial, escolhemos a cena nº 291 do roteiro de *Titanic* para fazer esta exploração. Esta é uma cena noturna ao final do filme, após o naufrágio, em que Jack (Leonardo DiCaprio) e Rose (Kate Winslet) flutuam nos destroços pouco antes da submersão de Jack.

Esta é uma das cenas icônicas do filme que, mesmo décadas após o lançamento, continua a levantar questionamentos para a equipe e os atores envolvidos (AMARO, 2019), especialmente sobre a possibilidade de sobrevivência de Jack. Muitos fãs acreditam que ele poderia ter compartilhado a tábua flutuante com Rose. Essa cena, que se tornou alvo frequente de memes⁴ na internet e até mesmo objeto de estudos científicos (HYPENESS, 2017), foi revisitada pelo diretor no especial de 25 anos do filme. James Cameron, juntamente com uma equipe de pesquisadores e dublês, tentou recriar o evento que levou à morte do protagonista e investigar se havia ou não espaço para Jack na tábua que salvou Rose (G1, 2023).

Escolhemos esta cena também por ela trazer algumas ambiguidades na representação dos acontecimentos. Ao contrário de momentos do início do filme como sua partida, ou mesmo da colisão com o iceberg, aqui temos a possibilidade de avaliar detalhes menos denotativos na interpretação visual feita pelas plataformas, como a representação de nuances emocionais dos personagens.

⁴ Imagens com apelo cômico com alta incidência de compartilhamentos nas redes sociais.

Selecionamos três momentos do roteiro referentes a cena já citada como comandos para a pré-visualização, uma vez que o texto do roteiro cinematográfico não corresponde exatamente à quantidade e sequência de segmentos de imagem – chamados planos – na obra concluída, uma ordenação que é feita posteriormente na edição e finalização. As escolhas de que segmentos selecionar do roteiro partiram da sugestão dada pelo texto de quanta ação e informação poderia se desenrolar em um plano, uma vez que o propósito era verificar a capacidade de tradução de texto em imagem, permanecendo aspectos de montagem a serem explorados em estudos futuros.

A metodologia de aplicação nas plataformas seguiu os passos: a) cada segmento de roteiro foi organizado em uma tabela com seu texto original e uma tradução para português (para conforto dos leitores) e um fotograma representativo; b) o bloco de roteiro foi inserido em sua versão original em inglês em cada uma das plataformas e foi solicitada a geração de oito imagens, após ter as falas removidas; c) identificamos categorias relevantes a observar em cada segmento do roteiro; d) os resultados foram coletados e ordenados; e) discutimos os resultados.

Os testes de pré-visualização envolveram inserir estes trechos do roteiro como comandos em plataformas geradoras de imagens estáticas, o que está alinhado ao tema de pré-visualização na forma de *storyboards*. Como recorte, escolhemos três delas: o modelo *Stable Diffusion* versão 1.5 (ROMBACH et al, 2022), o *Bing Image Creator* (MEHDI, 2023) e o *Genmo.AI* (GENMO, 2023). A geração das imagens foi realizada em um MacBook Pro modelo M1 com 8GB de RAM, com sistema operacional macOS Ventura em 27 de setembro de 2023.

Os textos foram inseridos primeiro no software gratuito *Diffusion Bee* (GUPTA, 2022)⁵, que roda sem precisar de conexão com a internet em computadores compatíveis com Mac OS, usando o modelo *Stable Diffusion 1.5* (STABILITY AI, 2023)⁶ para a geração das imagens. A escolha deste modelo é justificada por estar disponível de maneira livre na internet, podendo gerar imagens sem restrições de quantidade ou temática, ainda que seus autores (PATIL; CUENCA *et al*, 2022) recomendem usos positivos da tecnologia. O software rodou com configurações padrão e foi solicitado que gerasse oito imagens de 896x512 pixels para cada comando.

⁵ O software Diffusion Bee está disponível gratuitamente no endereço <https://diffusionbee.com> e é compatível com macOS. Após a instalação, ele indicou o download de um modelo para a geração de imagens, neste caso o Stable Diffusion versão 1.5.

⁶ Stable Diffusion é um modelo de geração de imagens da empresa Stability.AI, disponível no endereço <https://stability.ai>. Uma plataforma para geração de imagens sem necessidade de instalação pode ser consultada em <https://clipdrop.co/stable-diffusion-turbo>.

A segunda plataforma usada foi o *Bing Image Creator* (MEHDI, 2023)⁷, disponível online após um cadastro na Microsoft e gratuita, mas sujeita a uma limitação de uso diário baseada em créditos baseados em utilização de produtos da companhia. A interface e a utilização são simples: uma página mostra exemplos de comandos e resultados e em outra é possível inserir uma instrução, ainda que apenas em inglês. Em poucos segundos o sistema retorna quatro imagens de proporção 1:1 (quadradas), com resolução de 1024x1024 pixels, que podem ser vistas em detalhe e salvas localmente e na nuvem. O modelo por trás deste serviço é, conforme Mehdi (2023), uma versão aprimorada do DALL-E da OpenAI. Embora a própria OpenAI também ofereça uma plataforma online para usar o mesmo modelo em seu site⁸, um teste paralelo a este texto indicou que os resultados tendem a ser mais sofisticados na versão usada pela Microsoft. Aqui também foram geradas oito imagens por vez.

A terceira plataforma analisada foi o *Genmo.AI* (GENMO, 2023)⁹, disponível online com opções de acesso gratuito e planos pagos. A escolhemos por apresentar uma abordagem diferenciada, com a possibilidade de gerar imagens estáticas, vídeos animados e modelos tridimensionais a partir de uma interface conversacional baseada em linguagem natural. Isto significa que a interface é semelhante a uma conversa em um software de mensagens como o *WhatsApp* ou mesmo o *ChatGPT* (OPENAI, 2023b), ainda que as imagens produzidas estejam intercaladas entre os comandos e respostas. Com esta plataforma foram geradas oito imagens para cada prompt com a resolução de 1344x768 pixels, que corresponde aproximadamente à proporção de uma tela *widescreen* de uma televisão.

4. APLICAÇÃO DO MÉTODO E DISCUSSÃO DOS RESULTADOS

A cena nº 291 identifica sua locação como “Externa – Oceano” e usa uma formatação em que momentos-chave da trama, identificados pela palavra inglesa *beat* pontuam as mudanças de estados emocionais, sorte e revés dos personagens dentro da mesma cena. No filme, a cena se estende de 2h53min39s até aproximadamente 2h56min24s, durando 2min45s em nosso recorte. Após o naufrágio do Titanic, as cenas se passam em diversos pontos no mar aberto, justificando esta escolha

⁷ O Bing Image Creator foi consultado no endereço <https://www.bing.com/images/create>. Observamos que utiliza como base o Dall-E da OpenAI, e tem recebido constantes atualizações desde a redação do artigo, sendo incorporado em diversos serviços da Microsoft.

⁸ O Dall-E disponível quando da redação do texto continua disponível no endereço <https://openai.com/dall-e-2>.

⁹ O Genmo.AI está disponível no endereço <https://www.genmo.ai>. Depois da redação deste artigo, observamos empiricamente que o modelo de geração de imagens estáticas e em movimento continua a ser atualizado.

de redação de roteiro em *beats*. Uma síntese visual da cena pode ser vista na Imagem 1, com uma correção de cor aplicada para melhor visualização das ações.

Imagem 1: Síntese visual da cena nº 291 de Titanic (1997)



Fonte: produção dos autores a partir de cópia original do filme.

A Imagem 1 também ilustra os três planos que selecionamos para explorar a geração de imagens, identificados por uma moldura amarela e uma numeração sequencial no canto superior direito. Embora a cena nº 291 do roteiro possua 18 planos, contados em cada corte ou transição nos segmentos da imagem, ela apresenta uma maior variação de conteúdo visual principalmente na primeira metade. É nesse momento que são estabelecidos o local onde o casal de protagonistas se encontra, a situação em que estão, a distância de um possível barco de resgate e suas reações ao que é revelado.

Na visão geral da cena apresentada na Imagem 1, é possível observar que, uma vez que o contexto espacial, temporal e de status dos personagens é estabelecido, a cena se desenrola alternando entre eles com posições de câmera muito semelhantes, sendo planos abertos e *closeups* dos protagonistas; o que, percebemos, não traria tanta novidade na exploração metodológica baseada em criação visual.

A Tabela 1 apresenta a transcrição do início da cena nº 291, acompanhada de sua tradução para o português e de um fotograma ilustrativo desse trecho. Nesta cena, a câmera está posicionada a 90°, apontando para a balsa improvisada, construída com os destroços do naufrágio, onde Jack e Rose se encontram. Apesar de estarem em mar aberto, há uma sensação de intimidade entre os dois personagens, reforçando o tom romântico e trágico do filme.

Tabela 1 – Primeiro segmento selecionado, abertura da cena.

Texto no roteiro (inglês)	Tradução (português)
IN A HOVERING DOWNANGLE we see Jack and Rose floating in the black water. The stars reflect in the mill pond surface, and the two of them seem to be floating in interstellar space. They are absolutely still. Their hands are locked together. Rose is staring upwards at the canopy of stars wheeling above her. The music is transparent, floating... as the long sleep steals over Rose, and she feels peace.	Em um ângulo descendente flutuante, vemos Jack e Rose flutuando nas águas negras. As estrelas refletem na superfície serena, e ambos parecem estar flutuando no espaço interestelar. Eles estão absolutamente imóveis. Suas mãos estão entrelaçadas. Rose está olhando para cima, para o dossel de estrelas rodando acima dela. A música é transparente, flutuante... enquanto o longo sono toma conta de Rose, e ela sente paz.



Fonte: Tabela desenvolvida pelos autores.

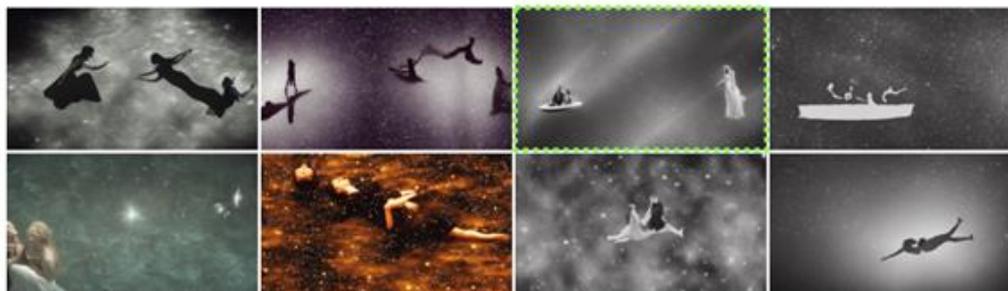
Ao analisarmos os conjuntos de imagens gerados pelas três plataformas de inteligência artificial generativa, a primeira percepção que se destaca é a diversidade de resultados que elas proporcionaram, de imagens semelhantes a ilustrações até quase fotográficas em sua aparência. No entanto, podemos melhor discutir os resultados estabelecendo a algumas categorias identificadas no roteiro e presentes em destaque também na versão final do filme, elas são: 1) Ângulo da câmera descendente; 2) Personagens Jack e Rose em cena; 3) Ambiente noturno, sob a luz das estrelas, em alto mar; 4) Flutuar olhando para o alto com as mãos entrelaçadas; 5) Sensação de paz. Uma vez que os modelos *Transformer* que embasam a interpretação de textos e a geração de imagens são baseados em um algoritmo que atribui pesos de atenção ao interpretar uma mensagem, podemos investigar

como esta atenção foi sendo atribuída e interpretada, e o que ficou de fora em cada plataforma. Os resultados estão na Imagem 2.

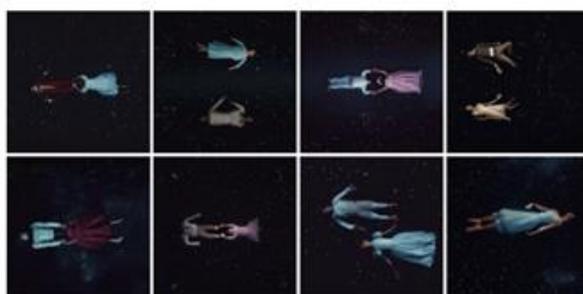
As imagens geradas pela plataforma *Stable Diffusion* (ROMBACH et al, 2022), em sua maioria, possuem características que evocam filmes antigos, com um apelo que remete aos filmes de truque do início do século XX e, posteriormente, no cinema surrealista. Neste modelo, há uma preferência por destacar representações de um céu estrelado e um estado híbrido do espaço que oscila entre a superfície serena do mar e o espaço entre as estrelas. A repetição das sugestões de flutuar e as repetidas menções às estrelas presentes no roteiro influenciaram a decisão de interpretação dada pelo modelo de privilegiar uma aparência extraterrena, inclusive por não ter conhecimento do contexto do restante da história, do filme e das convenções narrativas e estéticas da obra.

Imagem 2: Resultados de geração do primeiro bloco do roteiro.

Stable Diffusion



Bing Image Creator / Dall-E



Genmo.AI



Fonte: montagem realizada pelos autores.

As representações dos personagens são um tanto confusas; a repetição de corpos nas imagens sugere que várias frases do primeiro bloco do roteiro foram interpretadas e todos os resultados foram incorporados na mesma imagem. Isso ocorre porque os modelos não têm a certeza ou a consistência de um mundo seguro. Por exemplo, se o roteiro menciona que Jack e Rose estão flutuando numa noite serena e inicia uma nova frase dizendo que Rose observa as estrelas, nós, como leitores humanos, fazemos uma associação quase automática de que é a mesma Rose em ambas as frases. Portanto, se fôssemos representar essa cena, seria necessário apenas colocar dois personagens, um homem e uma mulher. No entanto, isso não ocorre em algumas das imagens aqui, possivelmente porque o sistema interpretou múltiplas vezes os personagens. Na terceira imagem da primeira fila, destacada com uma

moldura pontilhada verde, é possível observar esta interpretação visual de uma maneira bem clara, com Jack e Rose à esquerda da imagem e uma segunda Rose solitária de pé à direita, o que indica ser uma segunda representação da mesma personagem dentro da mesma imagem.

Temos aqui uma interpretação visual que diverge da maioria das cenas cinematográficas, exceto em *flashbacks*, momentos em que o personagem revisita uma cena passada de sua própria vida e revê a si como em um filme dentro do filme. No entanto, é crucial ressaltar que isso não é algo inédito na história da arte. Se observarmos a Coluna de Trajano (GOOGLE ARTS & CULTURE, 2023), construída para celebrar as vitórias militares do imperador de mesmo nome e atualmente em exibição no centro de Roma, notaremos que as representações dos personagens se repetem ao longo da coluna em um fluxo contínuo, estudada meticulosamente há séculos (POLLEN, 1874). Esta é uma maneira de condensar as elipses temporais e espaciais da história, sem recorrer à convenção dos quadros das histórias em quadrinhos, que, como coloca Scott McCloud (2005) só seria criada séculos mais tarde.

O modelo também tem dificuldades em interpretar a posição e angulação de câmera sugeridos: o ângulo de câmera descendente descrito não é explicitamente respeitado na maioria das imagens, estando claramente presente apenas na segunda imagem da segunda linha, com tom avermelhado.

As imagens geradas pela segunda plataforma, *Bing Image Creator* (MEHDI, 2023), acertaram, na maioria das vezes, o ângulo da câmera, representado em uma *plongée* de 90°. Além disso, em sete das oito imagens, os dois personagens principais estão representados através de estereótipos de gênero quanto ao código de vestimenta, como um vestido rodado para mulheres e calças para homens. O ambiente noturno, com uma luz suave, um salpicar de brilho das estrelas sobre a água¹⁰, e uma sensação de flutuação onde a gravidade está ausente, também foi respeitado.

No entanto, o fato de a plataforma gerar apenas imagens em formato quadrado compromete um pouco da leitura cinemática dos enquadramentos, normalmente associada a janelas retangulares. A relação entre os corpos representados nas imagens do *Bing Image Creator* (MEHDI, 2023), quando observada em detalhe, pode ser um pouco perturbadora; os corpos têm distorções e muitas vezes são mostrados de maneira incompleta em seus membros, além das fusões de membros e corpos entrelaçando Jack e Rose de forma literal. A representação do mar ou de uma espécie de balsa improvisada para sobreviver ao naufrágio não está presente; no entanto, este elemento não está descrito nesta cena do roteiro, o que justifica sua ausência.

¹⁰ A visualização das estrelas refletida sobre a água é praticamente impossível na redução das imagens, embora possa ser conferida nos arquivos originais.

A análise dos resultados da terceira plataforma revela uma peculiaridade técnica em sua execução: a plataforma *Genmo.Ai* (GENMO, 2023) gera imagens em lotes de quatro por vez. Isso é evidente, pois as quatro imagens superiores possuem uma característica mais fotográfica, apesar de os corpos apresentarem distorções significativas, enquanto as quatro imagens inferiores assemelham-se a histórias em quadrinhos, com personagens desenhados com aparência juvenil. O sistema tomou decisões autônomas a cada rodada de solicitações.

Dentre as três plataformas, a *Genmo.Ai* (GENMO, 2023) foi a que apresentou resultados mais semelhantes a estudos preparatórios para um filme. Esta é uma leitura pessoal dos autores, motivada pela interação dos personagens com o espaço do plano, da restrita quantidade de detalhes em cena em um enquadramento com um ponto focal, e da consistência nas ofertas a cada rodada. Claramente, os algoritmos decidiram oferecer variações sobre o mesmo tema, semelhantes entre si, ao invés de arriscar soluções mais radicais, como no *Stable Diffusion* (ROMBACH et al, 2022). Isto é efeito de uma proposta da companhia, que define esta ferramenta como um “copiloto criativo” (GENMO, 2023).

Há aspectos positivos e pontos a criticar na entrega da terceira plataforma. O ângulo de câmera descendente, destacado pelo do *Bing Image Creator* (MEHDI, 2023), não está presente em nenhuma das imagens, embora as quatro primeiras imagens estejam mais próximas disso. Por outro lado, os personagens Jack e Rose estão sempre em cena, e as quatro imagens com aparência de ilustração são charmosas em sua representação dos corpos, identificando a intimidade entre eles. Outros pontos positivos incluem a luz da noite e a locação em alto mar, acompanhadas da sensação de paz, mesmo que nem sempre as mãos estejam entrelaçadas.

Destacamos que parte da surpresa da cena é revelar que Jack, na verdade, já está morto, algo que o roteiro não deixa claro desde o primeiro momento. Como resultado, o algoritmo, ao interpretar este texto, manteve Jack vivo e aparentemente apaixonado nas quatro imagens estilo ilustração, algo que seria corrigido na planificação do filme com uma orientação do diretor ao ilustrador.

A Tabela 2 traz o segundo plano da mesma cena, que é apresentado após uma fusão e destaca o *closeup* do rosto de Rose. A escolha de uma fusão para a transição entre o primeiro plano e o segundo plano da cena acentua o aspecto de oscilação entre a realidade e o delírio. A alienação da personagem frente à situação dramática em que se encontra é reforçada por sua expressão neutra e a cantiga que entoava.

Tabela 2 – Segundo segmento selecionado, *closeup* no rosto de Rose.

Texto no roteiro (inglês)	Tradução (português)
---------------------------	----------------------

<p>CLOSE ON Rose's face. Pale, like the faces of the dead. She seems to be floating in a void. Rose is in a semi- hallucinatory state. She knows she is dying. Her lips barely move as she sings a scrap of Jack's song: ROSE^[L]_[SEP]"Come Josephine in my flying machine..."</p>	<p>CLOSE em no rosto de Rose. Pálido, como os rostos dos mortos. Ela parece estar flutuando em um vazio. Rose está em um estado semi- alucinatório. Ela sabe que está morrendo. Seus lábios mal se movem enquanto ela canta um pedaço da canção de Jack: ROSE "Vem Josephine na minha máquina voadora..."</p>
	

Fonte: Tabela desenvolvida pelos autores.

Da mesma maneira, estabelecemos algumas categorias para análise do segundo segmento do roteiro. Elas são: 1) Morte e Transição, representados pela palidez de Rose e sua flutuação em um vazio autorreflexivo; 2) Isolamento e Solidão, evidenciados pelo estado de Rose, destacando sua separação do mundo ao seu redor; 3) Desespero e Resignação, manifestados através do estado semi- alucinatório de Rose. Após submeter esta descrição às três plataformas e recolher os resultados, observamos que duas de três delas se aproximaram de maneira interessante de uma visualização adequada. Os resultados podem ser vistos na Imagem 3.

Imagem 3: Resultados de geração do segundo bloco do roteiro.

Stable Diffusion



Bing Image Creator / Dall-E



Genmo.AI



Fonte: montagem realizada pelos autores.

A análise dos resultados da primeira plataforma reforça a ideia de que a repetição de um nome pode gerar interpretações visuais inusitadas e peculiares, algo característico da geração de imagens com inteligência artificial quando da redação deste texto. Em cinco das oito imagens, o rosto da mesma mulher aparece duas vezes, com leves variações, refletindo a interpretação do segmento

de roteiro onde Rose é mencionada duas vezes: inicialmente como alguém de aparência pálida e, posteriormente, em um estado de delírio.

Destacamos a segunda imagem da segunda fileira, emoldurada por uma linha verde pontilhada, onde à esquerda vemos um rosto feminino distorcido com os olhos fechados e, à direita, um rosto pálido e visivelmente machucado, com uma expressão neutra, remetendo a um filme de terror. Variações dessas duas representações de Rose são evidentes em outras quatro imagens, sugerindo uma estrutura interpretativa baseada na constante atenção aos mesmos elementos do texto, mas com uma diversidade visual na geração das imagens.

Todos os resultados são tem um caráter perturbador com variados graus de clareza. Em particular, algumas delas revelam decisões significativas do modelo. Por exemplo, na primeira imagem do canto superior esquerdo, um rosto feminino com cabelos desgrelhados e olhos fechados, aparentando alienação da realidade, é talvez o resultado mais próximo do plano futuro do filme, na obra vemos a atriz de olhos abertos.

Na terceira imagem da segunda fileira, destacada por uma moldura pontilhada laranja, uma mulher aparece em meio a uma paisagem desfocada, prestes a abrir a boca, presumivelmente para cantar, algo referenciado no roteiro e presente em outras imagens. O ponto mais divergente é a última imagem do canto inferior direito da segunda fileira, onde rostos e corpos são representados de forma surrealista, desviando totalmente das demais interpretações. Curioso também é a escolha por imagens em branco e preto, tomada de forma autônoma pelo modelo.

As duas outras plataformas apresentaram resultados surpreendentemente convergentes, embora com interpretações distintas de rostos femininos baseadas no roteiro. O *Bing Image Creator* (MEHDI, 2023) gerou um conjunto de oito imagens que parecem retratar a mesma mulher, com *closeups* bem fechados no rosto, maquiagem intensa nos olhos e uma expressão de olhos quase cerrados, como se estivesse à beira do sono ou alcoolizada, sugestões visuais que se aproximam da perda de consciência sugerida pelo roteiro.

Neste conjunto, apenas em uma das oito imagens observamos a repetição dos rostos no canto inferior direito do quadro. Embora seja uma face vista apenas de relance, a forma do nariz indica o parentesco com os demais. Os lábios entreabertos, mencionados no roteiro, também são representados aqui. A ideia de flutuar no vácuo, alienada da realidade, é sugerida pela iluminação, que muitas vezes parece compatível com a personagem estando reclinada ou inclinada para o lado, diferenciando-se expressivamente em termos de fotografia das representações das outras plataformas.

Os resultados da terceira plataforma revelaram uma estratégia de interpretação do roteiro distinta das demais. O primeiro elemento que se destaca é que os oito rostos de mulheres, geralmente

representados de forma bastante verossímil, são caracterizados por uma iluminação marcada por luz e sombra. A noção de vácuo, presente no roteiro, é aqui traduzida como um fundo totalmente neutro, semelhante a um estúdio vazio e escuro. Embora isso também esteja presente nas criações do Bing, o formato quadrado utilizado por essa plataforma não oferece muito espaço para explorar este vazio, refletindo-se nas expressões.

Retornando à terceira plataforma, as expressões desses rostos transitam entre as diferentes emoções destacadas pelo roteiro. Uma espécie de surpresa, resignação, solidão e uma sensação de fim permeiam suas expressões, embora nenhum dos rostos represente todas as emoções simultaneamente. A palidez de Rose, citada no roteiro, é traduzida aqui na geração de imagens de mulheres com a pele clara, uma constante entre as três plataformas. De certa forma, é como se estivéssemos observando oito máscaras mortuárias diferentes, cada uma ainda retendo uma centelha de vida pouco antes de um final trágico.

Não há, assim como nas demais gerações realizadas, um conhecimento ou persistência do contexto do filme. É importante destacar que semelhanças com o rosto da atriz, referências à balsa improvisada ou ao período histórico do começo do século XX não são interpretadas, pois não estão presentes neste segmento do roteiro. De maneira intrigante, cada plataforma tomou as decisões de fotografia à sua maneira, privilegiando imagens em branco e preto na primeira, um tom pálido-esverdeado na segunda e uma luz de alto contraste na terceira, em suma, três convenções bem conhecidas na cultura para situações de alteridade.

O segmento da cena representada na Tabela 3 tem uma diferença fundamental dos demais por ser um plano que representa o ponto de vista da personagem Rose. Em termos da dramaturgia, ela está oscilando entre continuar junto de seu amado, que não sabe ter falecido ainda ou buscar um resgate no barco cuja tripulação tampouco a avistou. Ainda que a cena tenha uma resolução conhecida, uma vez que o filme é um longo flashback em que começamos com a personagem Rose idosa visitando o sítio onde o navio afundou, depois de três horas de romance e aventura, o momento de decisão da personagem cria tensão mesmo assim.

Tabela 3 – terceiro segmento selecionado, ponto de vista de Rose.

Texto no roteiro (inglês)	Tradução (português)
---------------------------	----------------------

ROSE'S POV... SLOW MOTION: The silhouetter of a boat crossing the stars. She sees men in it, rowing so slowly the oars lift out of the syrupy water, leaving weightless pearls floating in the air. The VOICES of the men sound slow and DISTORTED.	PONTO DE VISTA DE ROSE... CÂMERA LENTA: A silhueta de um barco cruzando as estrelas. Ela vê homens nele, remando tão lentamente que os remos se levantam da água viscosa, deixando pérolas sem peso flutuando no ar. As VOZES dos homens soam lentas e DISTORCIDAS.
	

Fonte: montagem realizada pelos autores

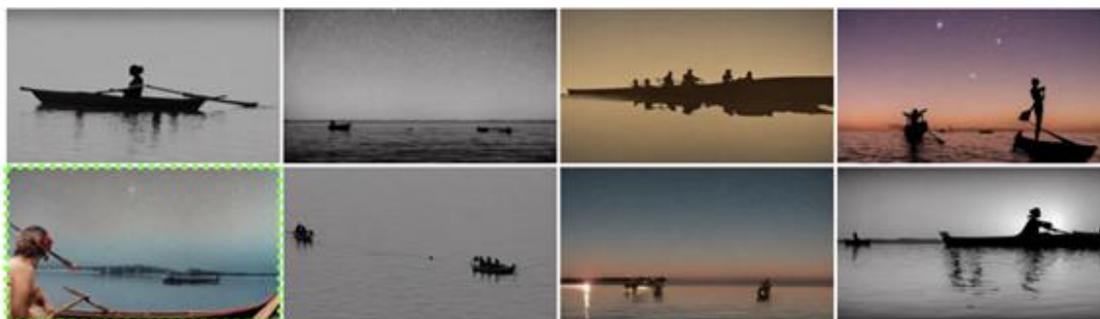
Como categorias identificadas no roteiro e que também estão em destaque na versão final do filme identificamos: 1) A perspectiva de Rose, indicando uma experiência e percepção pessoal da cena, além de uma posição de câmera compatível; 2) O uso de movimento lento, sugerindo um momento de intensidade emocional ou reflexão e possivelmente uma distorção da realidade percebida; 3) A presença da silhueta de um barco e homens remando lentamente, sustentando o suspense da cena; 4) A descrição da água como viscosa e a presença de vozes distorcidas, reforçando o distanciamento da realidade. Destacamos que este plano apresenta algumas informações visuais que não estão presentes no roteiro ainda neste ponto, como a presença de diversos corpos e fragmentos do naufrágio flutuando nos arredores e, portanto, dificultando a visão de Rose para o resgate. isto está presente na imagem, mas não no roteiro o que acontece alguns instantes mais tarde e fora de nosso escopo.

Este terceiro segmento do roteiro gerou resultados menos satisfatórios por parte de todas as plataformas, aparentemente devido a problemas na interpretação do texto, cada uma à sua maneira. Podemos afirmar que sempre faltou algo de essencial, o que é curioso, considerando os resultados mais positivos obtidos anteriormente. Ainda assim, é importante destacar que as ferramentas não

geraram imagens totalmente despropositadas; simplesmente, não conseguiram capturar nuances do roteiro de forma mais abrangente. Isto pode ser um resultado das cenas serem interpretadas separadamente sem que o contexto de uma colabore com a seguinte, algo que poderemos ver surgir em sistemas dedicados à pré-visualização de obras audiovisuais. Os resultados podem ser vistos na Imagem 4.

Imagem 4: Resultados de geração do terceiro bloco do roteiro.

Stable Diffusion



Bing Image Creator / Dall-E



Genmo.AI



Fonte: montagem realizada pelos autores.

A partir das imagens geradas pela primeira plataforma, seria difícil discernir que estamos visualizando uma cena noturna, uma vez que apenas um dos oito planos deixa claro que a cena ocorre

à noite, ao apresentar um conjunto de estrelas. Os outros sete parecem retratar momentos indefinidos do dia ou do amanhecer e, portanto, são incompatíveis com a proposta do filme. Como se pode observar na Tabela 3, o roteiro não especifica claramente que a cena ocorre à noite, embora menções à silhueta do barco cruzando diante das estrelas sejam suficientes para interpretar a cena como noturna.

A primeira plataforma enfatizou principalmente a temática da silhueta, com alguns remadores presentes em quase todas as imagens. Uma vez que se trata de uma representação exclusivamente visual e estática, não temos a possibilidade de ouvir as vozes distorcidas, mas a distância do ponto de vista é adequadamente representada. Um aspecto interessante das imagens geradas pela plataforma é que, em apenas uma delas, a primeira da segunda fileira, destacada com uma linha pontilhada verde, aparece uma personagem, aparentemente feminina, olhando para uma canoa ao longe. Não fica claro se a canoa distante tem remadores ou não, mas ali está, de certa forma a relação espacial de Rose com o objeto distante.

A interpretação realizada pela plataforma *Bing Image Creator* (MEHDI, 2023) do roteiro é, talvez, a mais fiel entre as três. Temos um conjunto de imagens onde barcos cruzam céus estrelados, provocando efeitos na água e frequentemente permitindo observar os reflexos do brilho das estrelas na superfície do mar. Mais uma vez, a noção de ponto de vista permanece implícita, precisando ser confirmada por algum outro plano da decupagem que mostraria Rose olhando em direção ao barco que se aproxima. Se os resultados são interessantes, o que há de errado, afinal?

O que principalmente chama a atenção é a noção de uma noite pautada por uma paleta de cores *instagramável* (ALVES, 2021), para emprestar uma expressão, onde um apelo a um código estético de algo colorido e harmonioso cromaticamente se sobrepõe a uma paleta de cores mais fechada, como aquela utilizada no filme. O roteiro não se pronuncia a respeito, limitando-se a dizer que a silhueta do barco passa em frente às estrelas. Isso sugere que até mesmo sistemas avançados de inteligência artificial possuem, junto de seus modelos e algoritmos, uma capacidade de agência e escolha autônoma quando permitido, e que esses modelos estão ajustados para agradar o que uma coletividade de usuários considera como bonito. Como contraponto, destacamos que seria possível solicitar, em uma iteração do processo, que esta cena fosse refeita com uma paleta de cores mais fechada ou uma noite menos luminosa, e possivelmente esta demanda seria atendida. Mas, neste caso, desejávamos investigar as respostas mais imediatas de cada uma das plataformas.

Os resultados da terceira plataforma mantêm a proposta de variações sobre o mesmo tema, já observada nas outras rodadas de investigação. Como ponto positivo, temos oito belas imagens de

um barco em meio a um lago, todas sob a luz das estrelas, com detalhes belamente ilustrados. No entanto, estão ausentes os remadores, a noção de movimento do barco, e a própria altura da "câmera" nem sempre é compatível com a perspectiva de alguém que estaria sobre a água, conforme inferido no texto do roteiro. Em termos da qualidade de iluminação é a que mais se aproxima de um resultado aproximado de uma obra audiovisual, mas toda atenção na interpretação do roteiro se prendeu nos aspectos do barco e sua silhueta e acabou-se por ignorar os humanos.

4. CONSIDERAÇÕES FINAIS

Ao chegarmos ao final desta primeira exploração de plataformas de imagens generativas, também é o momento de refletir sobre algumas limitações de nosso método de pesquisa. Ao escolhermos um filme já existente e bastante conhecido, é possível questionar o quanto o repertório de imagens amplamente circulado por Titanic influencia a nossa percepção do que que o roteiro deveria oferecer quando visualizado pelas ferramentas. Esta contaminação criaria uma percepção de que apenas a obra como a conhecemos seria desejável e outras interpretações teriam menos valor.

Entendemos que aplicações práticas desta metodologia seriam principalmente destinadas a visualizar filmes ainda por realizar, então a relação das equipes criativas com as ferramentas e seus resultados poderia acontecer de uma forma muito mais propositiva, com os resultados divergentes de uma escolha mais conservadora de enquadramentos servindo para provocar o processo criativo adiante.

Podemos argumentar que as plataformas de criação de imagens entregaram resultados que estão em conformidade com um entendimento comum dos comandos recebidos, sem cometer erros significativos. Destacamos também que utilizamos as plataformas com o mínimo de customização e apenas uma rodada de geração de resultados por vez. Não dedicamos um tempo sistemático a explorar refinamentos da escrita do roteiro para sugerir formas mais adequadas à visualização dos planos. A escolha de um roteiro de fins da década de 1990 também se alinha com esta escolha, provocando estes modelos e plataformas a lidarem com um conteúdo produzido décadas antes de sua elaboração. Potencialmente, a adaptação tanto das ferramentas quando das práticas de uso às ferramentas, é um campo aberto para dar continuidade a diversas investigações.

Ao final do processo, é possível observar como cada uma das plataformas desenvolve um perfil de resultados específico. Os modelos fazem escolhas diante de ambiguidades da linguagem, relacionando, de maneira consistente, uma ideia de prioridades na frase com destaques na imagem, alinhando gramáticas em ambos os campos. A primeira, *Stable Diffusion* (ROMBACH et al, 2022), consistentemente ofereceu os resultados mais variados, embora a repetição de rostos e a interpretação

inusitada de outros elementos sugere que pode ser mais adequada com comandos mais diretos, não operando tão bem com nuances da narrativa em roteiros.

A segunda, *Bing Image Creator* (MEHDI, 2023), proporcionou de forma consistente uma tradução bastante direta das imagens e dos comandos, resultando em algo de fácil apreciação. Isso pode levar a Microsoft e a OPEN AI, empresas responsáveis pela sua criação e funcionamento a ter aqui um serviço com possibilidade de apelo massivo e versátil. A terceira plataforma, *Genmo* (GENMO, 2023), demonstra seu posicionamento a atividades criativas ao apresentar os resultados com consistência estética a cada rodada de comandos (um comportamento semelhante ao que encontramos nas diferentes tomadas de um mesmo plano durante a realização de um filme) e incluir um duplo *Transformer*, que primeiro traduz as nuances do texto do roteiro para um prompt mais computável, depois acionando o gerador de imagens. Assim como utilizamos estas três plataformas para a criação deste artigo, é possível revisitar e aprimorar esta metodologia incluindo dezenas de outras. A possibilidade de treinar modelos próprios para superar vieses de interpretação e repertório das plataformas existentes é uma alternativa a ser considerada para a integração desses recursos a fins específicos, como a colaboração com a produção audiovisual.

Retomando a pergunta que nos levou à investigação deste texto, podemos afirmar que estas ferramentas podem colaborar com a pré-visualização de cenas a partir de um roteiro, dadas algumas limitações. A primeira delas é que colaborar com o processo criativo não significa substituir os indivíduos, mas impulsionar possibilidades e opções de criação visual que nem sempre seriam óbvias.

As inteligências artificiais generativas demonstram a capacidade de criar imagens com certo grau de originalidade, porém as desenvolvem a partir de um conjunto limitado de referências e podem tornar-se repetitivas rapidamente. Esta é consequência da observação de que nenhuma plataforma testada aqui se bastou em si para entregar a visualização em uma primeira rodada, ainda que tenham oferecido resultados instigantes. Observamos que todas as plataformas cometem erros e fantasiam ao representar o corpo humano, o que é uma das limitações técnico-estéticas atuais e não foi nosso foco por ser circunstancial. Portanto, o processo criativo de uma obra audiovisual como um filme continua sendo longo e minucioso, mesmo com estas ferramentas, mas pode ganhar com experimentações iterativas auxiliadas pela inteligência artificial.

Por fim, lembramos que esta é uma produção vinculada ao projeto Plataformização da Produção Cultural no Brasil. Edital Universal 2021 - Processo: 408110/2021-0.

REFERÊNCIAS

RAUEN, Fábio José. **Influência do sublinhado na produção de resumos informativos**. 1996. 200f. Tese (Doutorado em Letras/Linguística) – Curso de Pós-graduação em Letras/Linguística, Universidade Federal de Santa Catarina, Florianópolis, 1996.

ALVES, Angélica Luísa Santos da Fonseca Castro. **O admirável mundo instagramável: A estetização do comportamento de consumidor no Instagram**. 2021. Dissertação (Mestrado Comunicação, Cultura e Tecnologias da Informação) – Departamento de Sociologia da Iscte - Instituto Universitário de Lisboa, Lisboa, 2021. Repositório do Iscte. <http://hdl.handle.net/10071/23769>.

AMARO, Marina. Leonardo Dicaprio é questionado sobre cena icônica de ‘Titanic’ e tem reação mau humorada; Brad Pitt e Margot Robbie o provocam – vem ver!. Hugo Gloss: famosos, 2019. Disponível em: <https://hugogloss.uol.com.br/famosos/leonardo-dicaprio-e-questionado-sobre-cena-icônica-de-titanic-e-tem-reacao-mau-humorada-brad-pitt-e-margot-robbie-o-provocam-vem-ver/>. Acesso em: 25 set. 2023.

BARD. Disponível em: <https://bard.google.com>. Acesso em: 26 set. 2023.

CRAWFORD, K. **Atlas of AI**. New Haven: Yale University Press, 2021.

ECO, U. **Apocalípticos e integrados**. São Paulo: Perspectiva, 1976.

G1. Jack poderia ter subido na porta com Rose em 'Titanic', admite James Cameron. G1: pop & art, 2023. Disponível em: <https://g1.globo.com/pop-arte/noticia/2023/02/03/jack-poderia-ter-subido-na-porta-com-rose-em-titanic-admite-james-cameron.ghtml>. Acesso em: 23 de set. 2023.

GENMO. Meet Genmo, your new creative copilot. Disponível em: <https://www.genmo.ai>. Acesso em: 11 ago. 2023.

GLEBAS, Francis. **Directing the story: professional storytelling and storyboarding techniques for live action and animation**. Elsevier, 2009.

GOOGLE ARTS & CULTURE. **Can You Find Trajan's Column in the Streets of Rome?** How do you divide the Roman empire in two? With a pair of Ceasars. Google: Arts & culture. Disponível em: <https://artsandculture.google.com/story/UQUxiltOiTdlWw?hl=pt>. Acesso em 20 set. 2023.

GUPTA, D. **DiffusionBee - Stable Diffusion App for AI Art**. Disponível em: <https://diffusionbee.com>. Acesso em: 8 jul. 2023.

HYPENESS. Estudantes de matemática provam ‘por a + b’ que Jack não precisaria ter morrido em Titanic. HYPENESS: ciência, 2017. Disponível em: <https://www.hypeness.com.br/2017/11/estudantes-de-matematica-provam-por-a-b-que-jack-nao-precisaria-ter-morrido-em-titanic/>. Acesso em: 23 de set. 2023.

KAIBER. Bring your dreams to life Tell stories like never before with our advanced AI generation engine. Disponível em: <https://kaiber.ai/>. Acesso em: 11 ago. 2023.

MCCLOUD, Scott. **Desvendando os quadrinhos: história, criação, desenho, animação, roteiro**. São Paulo: M. Books, 2005.

MEHDI, Yusuf. Create images with your words - Bing Image Creator comes to the new Bing - The Official Microsoft Blog. Disponível em: <https://blogs.microsoft.com/blog/2023/03/21/create-images-with-your-words-bing-image-creator-comes-to-the-new-bing/>. Acesso em: 11 ago. 2023.

MIDJOURNEY. Disponível em: <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F>. Acesso em: 11 ago. 2023.

MONTEIRO, Salles. **Catálogo de clichês**. Cotia: Ateliê Editorial, 2003.

OPAS. OMS declara fim da Emergência de Saúde Pública de Importância Internacional referente à COVID-19. **Organização Pan-Americana da Saúde**. Disponível em: <https://www.paho.org/pt/noticias/5-5-2023-oms-declara-fim-da-emergencia-saude-publica-importancia-internacional-referente>. Acesso em: 03 de set. 2023.

OPENAI. **Introducing ChatGPT**. 30 nov. 2022. Disponível em: <https://openai.com/blog/chatgpt>. Acesso em: 8 jul. 2023.

OPENAI. **DALL·E 2**. Disponível em: <https://openai.com/dall-e-2>. Acesso em: 11 ago. 2023b.

PATIL, Suraj; CUENCA, Pedro; LAMBERT, Nathan; VON PLATEN, Patrick. Stable Diffusion with  Diffusers. Hugging Face Blog, 2022. Disponível em: https://huggingface.co/blog/stable_diffusion. Acesso em: 8 jul. 2023.

POLLEN, John Hungerford. **A Description of the Trajan Column**. Reino Unido: G. E. Eyre and W. Spittiswoode, and sold by Chapman & Hall, 1874.

PONIEWOZIK, James. TV's War With the Robots Is Already Here. **New York Times**. Disponível em: <https://www.nytimes.com/2023/05/10/arts/television/writers-strike-artificial-intelligence.html>. Acesso em: 25 set. 2023.

RABINGER, Michael. **Directing: film techniques and aesthetics**. 4ª ed. Elsevier, 2008.

RICKITT, Richard. **Special Effects, the history and the technique**. Nova Iorque: Billboard Books, 2000.

ROMBACH, R; BLATTMANN, A; LORENZ, D; ESSER, P; OMMER, B. **High-Resolution Image Synthesis With Latent Diffusion Models**. In: PROCEEDINGS OF THE IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR), June 2022, p. 10684-10695. Disponível em: <https://huggingface.co/runwayml/stable-diffusion-v1-5>. Acesso em: 8 jul. 2023.

STABILITY AI. AI by the people for the people. Disponível em: <https://stability.ai>. Acesso em: 11 ago. 2023.

Titanic. Direção: James Cameron. Produção: James Cameron, Jon Landau. Estados Unidos: Paramount Pictures, 1997. 195 min.

TIETZMANN, Roberto; PUHL, Paula Regina; MONTARDO, Sandra. **Artifícios Imagéticos: representações das identidades brasileiras a partir do modelo de inteligência artificial Stable**

Diffusion. *In:* Metamorfozes sociais: tecnologias, identidades e imaginários. Porto Alegre: Editora Sulina, 2023, no prelo (atualmente).

TIETZMANN, Roberto; PUHL, Paula Regina. Um modelo de país? Visões do Brasil pelas inteligências artificiais. In: 46º Congresso Brasileiro de Ciências da Comunicação, 2023, PUCMinas, Belo Horizonte. **Anais do 46º Congresso Brasileiro de Ciências da Comunicação.** Belo Horizonte: Intercom – Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, 2023.

USZKOREIT, Jakob. **Transformer: A Novel Neural Network Architecture for Language Understanding.** **Google Research Blog**, 2017. Disponível em: <https://blog.research.google/2017/08/transformer-novel-neural-network.html?m=1>. Acesso em 25 set 2023.

VASWANI, Ashish; SHAZEER, Noam; PARMAR, Niki. et al. **Attention is all you need.** **Advances in neural information processing systems**, v. 30, 2017. Disponível em: <https://research.google/pubs/pub46201/>. Acesso em: 25 set. 2023.

Informações sobre o Artigo

Resultado de projeto de pesquisa, de dissertação, tese: Fazendo filmes com Inteligência Artificial – um mapeamento de boas práticas e ferramentas para a produção audiovisual complementada por IA.

Fontes de financiamento: não se aplica

Apresentação anterior: não se aplica

Agradecimentos/Contribuições adicionais: aos bolsistas de iniciação científica Mirelli Garcia e Guilherme Taborda da PUCRS, integrantes do projeto.

Taís de Barros

Mestra em Comunicação Social (PPGCOM/PUCRS) na linha "Cultura e tecnologias das imagens e dos imaginários" (2022-2024). Possui graduação em Produção Audiovisual pela PUCRS (2021). Tem experiência na área de Comunicação, com ênfase em Audiovisual, atuando principalmente nos seguintes temas: cinema, produção audiovisual.

E-mail: tais.barros.rs@gmail.com

ORCID: <https://orcid.org/0000-0003-2570-0583>

Roberto Tietzmann

Doutor em Comunicação Social. Pesquisador e professor da Escola de Comunicação, Artes e Design – Famecos da PUCRS. Professor do Programa de Pós-Graduação em Comunicação da PUCRS. Tem como interesses de pesquisa zonas de contato entre audiovisual, design e tecnologia. Professor visitante na Faculté des Sciences du Sujet et de la Société na Universidade Paul Valéry, Montpellier III, na França. Coordenador do grupo de pesquisa ViDiCa – Cultura Digital Audiovisual e do LABiM – Laboratório de Imagem Digital. Pesquisador do programa Pesquisador Gaúcho da Fapergs.

E-mail: rtietz@pucrs.br

ORCID: <https://orcid.org/0000-0002-8270-0865>