
Filmando com inteligência artificial: relato de uma pesquisa exploratória¹

Guilherme TABORDA²

Mirelli GARCIA³

Roberto TIETZMANN⁴

Pontifícia Universidade Católica do Rio Grande do Sul, RS

RESUMO

Propomos uma pesquisa exploratória baseada na realização de um curta-metragem ficcional cujas imagens foram criadas total ou parcialmente com ferramentas de Inteligência Artificial Generativa (IAG). Nesta comunicação detalhamos os procedimentos de dois dos planos do curta. Como resultados de pesquisa, encontramos possibilidades e limitações implícitas das plataformas, a necessidade de construir um fluxo de trabalho iterativo para a geração das imagens e a conexão entre diversas IAGs para alcançar os resultados desejados.

PALAVRAS-CHAVE: comunicação, realização audiovisual, inteligência artificial generativa, curta-metragem.

INTRODUÇÃO

A chegada de ferramentas de inteligência artificial generativa (IAG) a uma multidão de consumidores e criadores a partir do segundo semestre de 2022 despertou curiosidade a respeito de como estas plataformas poderiam contribuir para a realização audiovisual. Desde 2023 surgiram uma série de ferramentas capazes de gerar segmentos de imagem em movimento a partir de prompts textuais, o que aproximou definitivamente os dois campos, prometendo substituir parte da etapa de gravação, e conduziu à criação do projeto de pesquisa que deu origem a esta comunicação. Nele, questionamos de que maneira ferramentas de inteligência artificial generativa podem contribuir com a criação de

¹ Trabalho apresentado no IJ04 - COMUNICAÇÃO AUDIOVISUAL, XIX Encontro dos Grupos de Pesquisas em Comunicação, evento componente do 47º Congresso Brasileiro de Ciências da Comunicação.

² Estudante de graduação do curso Superior Tecnológico em Produção Audiovisual da Escola de Comunicação, Artes e Design da PUCRS, E-mail: Guilherme.taborda@edu.pucrs.br.

³ Graduada do curso Superior Tecnológico em Produção Audiovisual da Escola de Comunicação, Artes e Design da PUCRS, E-mail: mirelligarcia1@gmail.com.

⁴ Doutor em comunicação, professor do programa de pós-graduação em comunicação social (PPGCOM) da PUCRS, orientador da pesquisa. E-mail: rtietz@pucrs.br.

conteúdos audiovisuais, incorporando estéticas e convenções narrativas visuais como debatido por Jullier e Marie (2007) e Aumont *et al* (2002).

Para responder a esta pergunta, escolhemos uma metodologia exploratória (Gil, 2010) que se estruturou em duas etapas. Nos primeiros meses da pesquisa, no segundo semestre de 2023, foram testadas diversas ferramentas para a criação de imagens e vídeos. No primeiro semestre de 2024, partimos para a criação de um curta-metragem cujos planos deveriam ser total ou parcialmente criados com IAGs. O roteiro, intitulado “A Realidade de Cada Um”, foi escrito pelo orientador da pesquisa, pelos bolsistas de iniciação científica (Mirelli Garcia e Guilherme Taborda) com a colaboração de um mestrando (Marcio Toson). Esta comunicação tem a forma de um relato de pesquisa em andamento, apresentando e discutindo estas duas etapas, e o processo de planos que fazem parte do filme, atualmente em edição.

O PRIMEIRO SEMESTRE DE EXPLORAÇÃO: SURPRESAS E FRUSTRAÇÕES

A partir de uma visitação no repositório de ferramentas de IAG chamado Future Tools (2023), foram experimentadas cerca de dez ferramentas capazes de gerar imagens estáticas e vídeos. A versatilidade da plataforma Vispunk (referenciar) em gerar tanto imagens em movimento quanto estáticas atraiu a atenção dos bolsistas. O fato de ser gratuita e não exigir cadastro também veio de encontro à experimentação.

Testes de decupagem no Vispunk, conduzidos pela bolsista Mirelli Garcia, tiveram como objetivo entender o quanto esta ferramenta poderia auxiliar a produção cinematográfica, incluindo a compreensão de linguagem específica, como enquadramentos e posições de câmera. Para tal, foi utilizado o roteiro de um curta-metragem previamente produzido, intitulado "Sozinha", em que segmentos do roteiro com descrições de cenas foram usados como prompts no Vispunk para fazer uma comparação entre a cena gerada pela IAG e a cena existente.

Na primeira tentativa, foi descrito um ambiente onde há uma poltrona em frente a um sofá, com uma menina sentada na poltrona de costas e outra menina sentada no sofá, em

uma sessão de psicanálise. O roteiro foi traduzido para o inglês como “*A room with an armchair in front of a sofa. We see the back of a girl sitting in the armchair and we see other girl sitting on the sofa*” e uma série de imagens foi gerada pelo Vispunk com ajustes nos prompts, buscando aproximar o resultado com a cena original do filme, sendo discutida nas reuniões de pesquisa. Uma amostra pode ser vista na figura nº1 abaixo.

Figura nº1 – Primeira experiência com o Vispunk: curta-metragem à esquerda, uma geração típica da IAG à direita



Fonte: Gerado por Mirelli Garcia

Mesmo com diferentes descrições sendo refinadas tentando especificar mais detalhes do que no roteiro original, os resultados não apresentaram uma qualidade "cinematográfica", uma vez que o Vispunk não reconheceu nomenclaturas de planos e posições de câmera específicos, bem como orientação espacial no set tanto em inglês quanto em português. Tampouco era capaz de gerar corpos humanos sem distorções significativas, o que representa um problema para a maioria das narrativas.

As limitações observadas no Vispunk destacam a complexidade e os desafios de replicar com precisão visual e emocional as cenas cinematográficas através de descrições textuais, evidenciando a necessidade de avanços na inteligência artificial para aplicações no campo da produção audiovisual. A ferramenta cumpriu seu propósito experimental, mas não era capaz de avançar.

Como resultado, os bolsistas passaram a acompanhar o amadurecimento de diversas ferramentas que, em questão de menos de um ano, passaram a se tornar mais aptas a atender a estas demandas.

CRIANDO OS PLANOS DO FILME: O PROCESSO

A realização audiovisual costuma ter etapas regulares em sua execução. Tradicionalmente consideramos três grandes etapas: a pré-produção, que envolve a preparação do roteiro e demais recursos necessários para a execução do projeto, a etapa de produção, identificado pela montagem do set de filmagem e a reunião de elenco e equipe e por fim a pós-produção, em que os segmentos registrados são editados e a obra ganha a forma final. As atividades com IAGs borram as barreiras entre a produção e a pós-produção, exigindo a investigação de fluxos de trabalho diversos tanto do audiovisual tradicional quanto da animação.

O ponto de partida continua a ser um roteiro, em que escolhemos não usar ferramentas generativas. O roteiro de “A realidade de cada um” propõe uma observação sobre como diversos personagens vistos, em um parque de uma grande cidade, estão imersos em seus pensamentos, e como podemos ouvi-los e enxergá-los transformando o que entendemos por realidade.

O curta abre com uma série de pessoas em atitude reflexiva em um parque, uma delas pode ser vista na figura nº1. O roteiro sugeria a imagem de um garoto de cerca de 14 anos.

Figura nº1 – Garoto no parque, cena do início do filme



Fonte: gerado por Guilherme Taborda Pereira da Cunha

A figura nº1 é o resultado de um primeiro experimento com o modelo de geração de imagens Realistic Vision v6.0 B1 (CIVITAI, 2024), utilizado através da plataforma online CivitAI. O Realistic Vision é um *fork* (derivado) do Stable Diffusion v1.5, modelo de geração de imagens de código aberto que é o principal produto da Stability AI (2024) - empresa essa que disponibiliza, para desenvolvedores, criadores, empresas e curiosos, uma variedade de outros modelos de inteligência artificial generativa que trabalham em outras modalidades (vídeo, áudio, texto), todos também *open source*.

Para atingir a imagem-resultado, foram necessárias sete gerações de quatro imagens cada, cada uma com leves alterações nos *prompts* (comandos) utilizados. Vinte-e-oito imagens foram geradas no total, e embora a escolhida como plano para o curta-metragem tenha sido apenas a décima segunda destas, mais quatro gerações (dezesseis imagens) foram necessárias para que o autor chegasse a tal conclusão. Todas as imagens foram geradas em *Draft Mode* (“Modo Rascunho”, em tradução direta), aplicando-se assim em sua geração a LoRA (*Low Rank Adaptation*) LCM-LoRA (SOFTWAREKEEP, 2024) a fim de simplificar e acelerar a mesma.

Durante a exploração dos recursos da plataforma CivitAI, o modelo Realistic Vision, chamou a atenção pela qualidade estética de suas imagens. O *prompt* fornecido inicialmente ao modelo, que compreende apenas a língua inglesa, foi “*teenager, male, school uniform, backpack, worried expression., wooded urban park, godrays*” (“adolescente, sexo masculino, uniforme escolar, mochila, expressão preocupada, parque urbano arborizado, raios de luz filtrados pelas copas das árvores”, em tradução direta), e ao longo das seis gerações seguintes, ele evoluiu até tornar-se “*child in school uniform with backpack, worried expression in face, visible face, urban park, godrays*” (“criança vestindo uniforme escolar com mochila, expressão preocupada no rosto, rosto visível, parque urbano, raios de luz filtrados pelas copas das árvores”), com o que gerou a imagem escolhida para ser utilizada no curta-metragem sendo “*boy, small backpack, worried expression, wooded urban park, godrays*” (“garoto, mochila pequena, expressão preocupada, parque urbano arborizado, raios de luz filtrados pelas copas das árvores”).

Como se pode ver na imagem-resultado, o modelo utilizado deu ao motivo gerado o corpo de uma criança do sexo masculino, mas ignorou o comando de aplicar uma expressão

preocupada a seu rosto, gerando-o inclusive de costas para a “câmera”, de forma que seu rosto não aparecesse na imagem. Isto se dá devido à impossibilidade de gerar imagens de crianças inerente a todos os modelos derivados do Stable Diffusion, um movimento consciente das empresas que criaram os modelos (ENGADGET, 2024). O bloqueio foi contornado, de certa maneira, no resultado apresentado acima; todas as imagens geradas em que o rosto do motivo é visível deram ao mesmo a aparência e o porte físico de um adulto. Além disso, duas imagens da 7ª geração e uma da 4ª apresentaram anomalias na forma de “objetos fantasmas”, e o motivo de uma da 6ª apresentou anatomia deformada – neste caso, um segundo braço à esquerda de seu corpo.

De uma maneira distinta, a geração de imagens de pessoas de outras faixas etárias não encontrou uma restrição explícita por parte dos modelos de IAG. Ao solicitar a geração de uma imagem de um idoso no parque, não foram encontrados impedimentos, como pode ser visto na figura nº2.

Figura nº2– Idoso no parque, cena do início do filme



Fonte: Gerado por Mirelli Garcia

A figura nº2 foi gerada pela plataforma de geração artificial *Stable Diffusion Online*, que utiliza um *prompt* para gerar uma imagem do zero. A versão mais recente desse modelo é o *Stable Diffusion XL*. Para chegar ao resultado foram feitas quatro diferentes gerações com diferentes prompts, sendo a figura final a segunda geração. De maneira semelhante à outra geração da figura nº1, a iteração dos prompts e a variedade de resultados que os modelos oferecem a cada entrega não permite encontrar a mesma certeza que se tem quando uma tomada chega a um resultado ideal em uma filmagem. Com IAGs sempre paira uma possibilidade de mais uma iteração ou ajuste entregar algo um pouco melhor,

com poucas consequências de tentar novamente, como desgastar a equipe, ficar sem orçamento etc.

A escolha pelo *Stable Diffusion* se deu ao pesquisar e descobrir que a plataforma havia ajustado o treinamento de seu modelo para gerar imagens de pessoas, deixando a imagem final o mais realista possível. O *prompt* que deu a imagem-resultado foi na língua inglesa: “A 65-year-old man, in a park, with a resentful, dejected and nostalgic expression.” traduzido a partir de “Um homem de 65 anos em um parque, tem uma expressão ressentida, abatida e nostálgica.” no roteiro. A intenção era que o personagem do idoso estivesse ressentido com a passagem do tempo e estivesse refletindo em um parque sobre como a vida era antigamente.

O resultado foi bastante convincente, da primeira até a última geração no *Stable Diffusion* com esse *prompt* foram eficazes e encaixaram no princípio conceitual acordado pela equipe; uma imagem realista e emocionante. Todas as quatro gerações trouxeram este mesmo homem idoso, mesma aparência e mesma roupa, apenas posições de câmera diferentes. No *prompt* não foi especificado qual seria o enquadramento da imagem, então a própria plataforma interpretou que seria um *close*, acredito que considerando captar todas as emoções que seriam demonstradas na imagem.

Outras cenas que exigiriam efeitos visuais e maior orçamento também foram geradas. O roteiro foi elaborado considerando as possibilidades de síntese de imagem e recursos de animação e redesenho de cena disponíveis nas plataformas utilizadas. A possibilidade de representar situações complexas foi assumida como normal, como em um passeio por um vulcão em erupção cujos fotogramas estão na figura nº2.

Figura nº2 – Frames de um vulcão em erupção



Fonte: gerado por Guilherme Taborda Pereira da Cunha

As imagens acima são *frames* de um vídeo de quatro segundos gerado através da plataforma Runway.ML (RUNWAYML, 2024), que utiliza um modelo multimodal de geração de imagens e vídeos próprio. O modelo está atualmente em sua segunda versão, e seu sucessor será um *General World Model* (Modelo de Mundo Geral, em tradução direta) que, na data da criação deste texto, está em fase de testes internos.

O *prompt* fornecido ao modelo que levou à geração dos resultados vistos acima foi “*Volcanic eruption at nighttime, flyby, drone shot, slow motion, clear sky*” (“Erupção vulcânica à noite, flyby, visão a partir de drone, câmera lenta, céu limpo”, em tradução direta), com variantes do mesmo incluindo também os termos *close-up* e *lava flow* (*fluxo de lava*, em tradução direta) gerando resultados insatisfatórios, que apresentaram deformação do terreno por onde a lava e a fumaça do vulcão fluía – demonstrando, assim, que o modelo atual da Runway.ML tem dificuldade em simular fluidos e seus movimentos de maneira convincente. Até mesmo nas imagens acima, pode-se ver a dificuldade do modelo em compreender completamente o comando fornecido a ele, pois o termo *clear sky* foi ignorado devido, possivelmente, a um conflito com o termo *nighttime*, o que resultou em um cenário com céus nublados ao invés de limpos.

Um passeio na lua também foi experimentado com o Runway.ML, na figura nº3 mais uma vez visto em fotogramas de um vídeo de quatro segundos.

Figura nº3 – Fotogramas de homem saltando na lua.



Fonte: gerado por Guilherme Taborda Pereira da Cunha

O *prompt* da figura nº3 foi “*Profile detail shot of astronaut's boots landing on the Moon's surface, slow-motion, dust cloud*” (“Plano-detalle do pé em perfil da bota de um astronauta aterrissando na superfície da lua, câmera lenta, nuvem de poeira”, em tradução direta). Ao bolsista Guilherme foi dado o objetivo de gerar um plano de ser mais

ambicioso artisticamente e subsequentemente não obedecer completamente apenas ao que estava no roteiro, tanto buscando uma escolha criativa por parte do autor, que a fez com a intenção de testar os limites do modelo de geração multimodal da plataforma Runway.ML.

Antes que o resultado apresentado acima fosse atingido, uma série de cliques insatisfatórios foram gerados, todos apresentando deformações e anomalias anatômicas e de movimento – novamente demonstrando, assim, a dificuldade do modelo de movimento disponível Gen-2 de simular o movimento de fluidos e, também, de tecidos (como o utilizado na roupa do motivo dos cliques). Foram necessárias, ao autor do texto, duas gerações a partir do mesmo *prompt* para que um resultado satisfatório fosse atingido, e mesmo assim, certos elementos do clipe demonstrado acima não são integralmente perfeitos, como o texto indecifrável no calçado em primeiro plano (o uso de modelos de *inpainting* de vídeo sobre a geração final pode corrigir isso, permitindo a edição de detalhes internos à imagem) e o movimento levemente anômalo, artificial da perna e do pé em segundo plano.

Estes planos seguiram um fluxo de produção em que estas imagens estáticas foram criadas a partir de trechos do roteiro usados como prompts, sendo refinados neste processo. Depois, foram alimentadas em outras plataformas como a Runway.ML para que fossem animadas, tanto em movimentos de câmera como em ações dos personagens, provocando diferentes ciclos de criação a ser discutida na apresentação do trabalho.

CONSIDERAÇÕES DE UMA PESQUISA EM ANDAMENTO

O processo de realização audiovisual amadureceu desde os primeiros dias do cinema, sendo adaptado e transformado para a televisão e as produções de vídeo pela internet, tanto em propostas sofisticadas e profissionais das empresas de streaming quanto em canais de vídeo pessoais e de influenciadores em redes sociais. Em comum a todos, é preciso preparar, registrar e organizar que foi registrado para poder exhibir algo coerente. Com as IAGs, este processo pode ser reposicionado mais uma vez, transformando-se em um movimento de rodadas de interação entre humano e máquina, potencializando a criação e levantando novas questões sobre autoria que continuaremos a investigar.

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

AUMONT, J.; BERGALA, A.; MARIE, M.; VERNET, M. **A estética do filme (2a edição)**. Campinas: Papirus Editora, 2002.

CIVITAI. Disponível em: <https://civitai.com/>. Acesso em: 27 jun. 2024.

_____. **Realistic Vision V60-B1**. Disponível em: <https://civitai.com/models/4201/realistic-vision-v60-b1>. Acesso em: 27 jun. 2024.

ENGADGET. **The world's leading AI companies pledge to protect the safety of children online**. Disponível em: <https://www.engadget.com/the-worlds-leading-ai-companies-pledge-to-protect-the-safety-of-children-online-213558797.html>. Acesso em: 27 jun. 2024.

GIL, A.C. **Como elaborar projetos de pesquisa**. 5º ed. São Paulo: Atlas, 2010.

JULLIER, L.; MARIE, M. **Lire Les Images de Cinéma**. Paris: Larousse, 2007.

RUNWAYML. Disponível em: <https://runwayml.com/>. Acesso em: 27 jun. 2024.

_____. **Introducing GEN-3 Alpha**. Disponível em: <https://runwayml.com/blog/introducing-gen-3-alpha/>. Acesso em: 27 jun. 2024.

_____. **Introducing general world models**. Disponível em: <https://research.runwayml.com/introducing-general-world-models>. Acesso em: 27 jun. 2024.

SOFTWAREKEEP. **How to use Stable Diffusion LoRA models**. Disponível em: <https://softwarekeep.com/blogs/how-to/how-to-use-stable-diffusion-lora-models>. Acesso em: 27 jun. 2024.

STABILITY AI. Disponível em: <https://stability.ai/>. Acesso em: 27 jun. 2024.

_____. **Stable Image**. Disponível em: <https://stability.ai/stable-image>. Acesso em: 27 jun. 2024.

THE AI NAVIGATOR. **What is inpainting**. Disponível em:
<https://www.theainavigator.com/blog/what-is-inpainting>. Acesso em: 27 jun. 2024.

WIKIPÉDIA. **Código aberto**. Disponível em: https://pt.wikipedia.org/wiki/Código_aberto.
Acesso em: 27 jun. 2024.