



Instituto Federal de Educação, Ciência e Tecnologia da Paraíba
Campus Campina Grande
Coordenação do Curso Superior de Tecnologia em Telemática

O Uso de LLMs na Geração de Vídeos: Um Estudo sobre a Construção de *prompts* Eficientes

Washington Cipriano Silva Albuquerque

Orientador: Dr. Éwerton Rômulo Silva Castro

Campina Grande, Dezembro de 2025

© Washington Cipriano Silva Albuquerque



Instituto Federal de Educação, Ciência e Tecnologia da Paraíba
Campus Campina Grande
Coordenação do Cursos Superior de Tecnologia em Telemática

O Uso de LLMs na Geração de Vídeos: Um Estudo sobre a Construção de *prompts* Eficientes

Washington Cipriano Silva Albuquerque

Monografia apresentada à Coordenação do
Curso Superior de Tecnologia em Telemática
do IFPB - Campus Campina Grande, como
requisito parcial para conclusão do curso de
Tecnologia em Telemática.

Orientador: Dr. Éwerton Rômulo Silva Castro

Campina Grande, Dezembro de 2025

Catálogo na fonte:
Ficha catalográfica

A345u Albuquerque, Washington Cipriano Silva.
O uso de LLMs na geração de vídeos: Um Estudo sobre
a Construção de prompts Eficientes / Washington Cipriano
Silva Albuquerque. - Campina Grande, 2025.
45f.: il.

Trabalho de Conclusão de Curso (Graduação em
Tecnologia em Telemática) - Instituto Federal da Paraíba,
2025.

Orientador: Prof. Dr. Éwerton Rômulo Silva Castro.

1. Inteligência artificial. 2. Modelos de Linguagem de Grande Escala (LLM). 3. Engenharia de Prompt. 4. Vídeos digitais - Produção. 5. Telemática. I. Castro, Éwerton Rômulo Silva. II Título.

CDU 004.85

O Uso de LLMs na Geração de Vídeos: Um Estudo sobre a Construção de *prompts* Eficientes

Washington Cipriano Silva Albuquerque

Dr. Éwerton Rômulo Silva Castro
Orientador

Msc. Iana Daya Cavalcante Facundo Passos
Membro da Banca

Dr. José Antônio Cândido Borges da Silva
Membro da Banca

Campina Grande, Paraíba, Brasil
Dezembro/2025

Dedicatória (Dedico este trabalho à minha família, amigos e professores, pelo apoio,
incentivo e inspiração ao longo desta jornada.)

"O sucesso é a soma de pequenos esforços repetidos dia após dia."
Robert Collier

Agradecimentos

Agradeço ao curso de Tecnologia em Telemática do Instituto Federal da Paraíba, Campus Campina Grande, que ao longo do tempo se tornou não apenas um lugar de aprendizado, mas também de crescimento pessoal e profissional.

A cada disciplina, projeto e desafio enfrentado, foi possível desenvolver não apenas competências técnicas, mas também valores essenciais, como disciplina, trabalho em equipe, persistência, comprometimento, entre outros.

A Telemática, enquanto área de conhecimento, abriu caminhos para compreendermos a importância da integração entre redes, comunicação e tecnologia da informação, mostrando como essas áreas, juntas, têm impacto significativo na sociedade.

O IFPB de Campina Grande, mais do que um espaço acadêmico, foi um ambiente de descobertas e transformações, onde construímos amizades e transformamos cada dificuldade em um degrau para o sucesso, aprendendo e evoluindo a cada dia.

Agradeço também aos professores, colegas e a todos que fizeram parte dessa jornada, pois cada contribuição, seja em sala de aula, em um laboratório ou em uma simples conversa de corredor deixou marcas que levarei para sempre.

Resumo

Os LLMs (*Large Language Models* - Modelos de Linguagem de Grande Escala) representam um avanço importante na área da inteligência artificial, principalmente no campo do PLN (*Natural Language Processing* - Processamento de Linguagem Natural). Sua origem remonta aos primeiros estudos em linguística computacional, com o objetivo de compreender e gerar linguagem natural de maneira eficiente. Isso possibilita a automação de tarefas antes exclusivas dos humanos. No entanto, como se tratam de ferramentas recentes, ainda é possível identificar dificuldades na sua utilização pelo público geral, como a falta de clareza, ambiguidade, especificidade insuficiente e pouco domínio técnico. Esta monografia explora como os LLMs vêm sendo utilizados para gerar vídeos a partir de comandos textuais (*prompts*). Apresenta-se uma análise das melhores práticas na construção de prompts eficientes, detalhando elementos como clareza, especificidade, estilo visual, emoção e contexto, bem como suas aplicações em áreas como *marketing*, educação e entretenimento. Conclui-se que *prompts* com determinadas características produzem resultados mais eficientes. O estudo destaca que dominar a escrita de prompts é uma competência essencial para explorar o potencial da IA (*Artificial - Intelligence*, Inteligência artificial) na produção audiovisual.

Palavras-chave: modelos de linguagem, geração de vídeo, inteligência artificial, *prompts*.

Abstract

Large Language Models (LLMs) represent a major advancement in the field of artificial intelligence, particularly within Natural Language Processing (NLP). Their development stems from early studies in computational linguistics, aiming to understand and generate natural language efficiently, enabling the automation of tasks previously performed exclusively by humans. However, because these tools are relatively new, users still face challenges in employing them effectively, including lack of clarity, ambiguity, insufficient specificity, and limited technical knowledge. This article explores how LLMs are being used to generate videos from textual commands (*prompts*). It presents an analysis of best practices for constructing effective prompts, highlighting elements such as clarity, specificity, visual style, emotion, and narrative context, as well as their applications in areas such as marketing, education, and entertainment. The study concludes that prompts containing specific characteristics yield more efficient results. It also emphasizes that mastering prompt writing has become an essential skill for harnessing the potential of Artificial Intelligence (AI) in audiovisual production.

Keywords: language models, video generation, artificial intelligence, *prompts*.

Sumário

Lista de Siglas	xii
Lista de Figuras	xiii
Lista de Quadros	xiv
1 Introdução	1
1.1 Justificativa e Relevância do Trabalho.	2
1.2 Objetivos	3
1.2.1 Objetivo Geral.	3
1.2.2 Objetivos Específicos	3
1.3 Metodologia.	3
1.3.1 Etapas do Estudo	3
1.3.2 Abordagem Analítica.	4
1.3.3 Resultados Esperados.	4
1.4 Palavras-Chave.	4
2 Fundamentação Teórica	6
2.1 História dos Modelos de Linguagem de Grande Escala (LLMs).	6
2.2 Ferramentas Criativas e Artísticas.	7
2.3 Engenharia de Prompts.	8
2.4 Geração de Vídeos por IA	9
2.5 Sora.	10
2.5.1 Justificativa da Escolha do Modelo Sora.	11
2.6 Desafios e Limitações da Geração de Vídeo por IA.	12
2.6.1 Estrutura Recomendada para um <i>prompts</i> de Vídeo.	12
3 Resultados	14
3.1 Interface Sora 2.	14
3.2 Exemplo de <i>prompt</i> Bem Estruturado.	17
3.2.1 Análise do <i>prompt</i> de Vídeo.	18
3.2.2 Análise.	23
3.3 Exemplo de <i>prompt</i> ruim e ineficiente	25

3.4	Análise das Deficiências do <i>prompt</i>	26
3.5	Considerações Finais.	30
3.6	Sugestões Para Trabalhos Futuros.	30
Referências Bibliográficas		32

Lista de Siglas

AI	<i>Artificial Intelligence</i> (<i>Inteligência Artificial</i>)
GPT	<i>Generative Pre-trained Transformer</i> (Transformador Pré-Treinado Generativo)
HMM	<i>Hidden Markov Model</i> (Modelo Oculto de Markov)
LLM	<i>Large Language Model</i> (Modelo de Linguagem de Grande Porte)
PNL	<i>Natural Language Processing</i> (Processamento de Linguagem Natural)

Lista de Figuras

1	Evolução dos LLMs.	7
2	Ferramentas criativas.	8
3	Engenharia de <i>prompts</i>	9
4	Tela inicial.	15
5	Sora 2.	16
6	Resultado primeiro exemplo.	17
7	Exemplo	22
8	Resultado do segundo exemplo.	22
9	Exemplo 3.	26
10	Resultado do terceiro exemplo.	26

Lista de Quadros

1.1 Termos de busca utilizados na pesquisa 5

Capítulo 1

Introdução

A criação de conteúdo digital tem passado por uma transformação significativa com os avanços da IA (*Artificial Intelligence* - Inteligência Artificial), especialmente após o surgimento dos LLMs (*Large Language Models* - Modelos de Linguagem de Grande Escala). Esses modelos são baseados em arquiteturas neurais são formas de organizar uma rede neural, ou seja, o desenho da estrutura que define como os neurônios artificiais estão conectados e como a informação passa por eles, como *Transformer* (Transformador Pré-Treinado Generativo), [Hadi *et al.* 2025]. Essa inovação revolucionou o processamento de linguagem natural permitindo compreender melhor o contexto das palavras, treinar em larga escala e servir de base para grande parte das IAS modernas de linguagem.

Esse progresso mudou a forma como sistemas Computacionais interagem com a linguagem natural, possibilitando aplicações como assistentes virtuais, sistemas de diálogo, ferramentas de tradução automática, produção de textos e análise semântica [Liu *et al.* 2023]. Mais recentemente, a capacidade desses modelos foi direcionada para aplicações diferentes, conectando a linguagem escrita à geração de imagens, áudios e vídeos, criando novas opções de expressão e produção criativa [Wang *et al.* 2024].

Dentro desse contexto, um dos usos ascendentes mais promissores é a geração automatizada de vídeos a partir de comandos textuais, conhecidos como *prompts*. Essa abordagem consiste em traduzir instruções escritas em produções audiovisuais, permitindo que o usuário crie conteúdo de alta qualidade com base em apenas descrições textuais.

A motivação deste trabalho nasce de constatação de que, apesar de crescente popularização dessas ferramentas, muitos usuários ainda enfrentam dificuldades na formulação de *prompts* eficazes. Instruções mal elaboradas podem resultar em vídeos inconsistentes, pouco realistas ou desalinhados com os objetivos do usuário, o que limita o potencial criativo dessas tecnologias [Mienye *et al.* 2025].

Assim compreender como estruturar *prompts* mais eficientes torna-se um passo fundamental para democratizar o uso das IAS generativas na criação audiovisual. Esse tema se insere em um contexto de relevância social, cultural e econômica, dado que a produção de vídeo tem impacto em diferentes áreas como educação *marketing*, entretenimento e comuni-

cação digital.

1.1 Justificativa e Relevância do Trabalho.

A produção de vídeo vem aumentando muito e de forma acelerada e prometendo mudar diferentes áreas. Na **educação**, professores e instituições de ensino passaram a utilizar vídeos interativos, animações e tutoriais gerados por IA para tornar o aprendizado mais dinâmico e acessível. Estudos como o de [Xu *et al.* 2024] demonstram que vídeos instrucionais criados por modelos generativos apresentam desempenho de aprendizagem comparável e, em alguns casos, superior aos vídeos tradicionais, favorecendo o engajamento e a retenção do conhecimento. De forma complementar, [Netland 2025] compararam vídeos produzidos por humanos e por IA, observando que os conteúdos automatizados podem reduzir custos e tempo de produção sem comprometer a qualidade educacional. Na **indústria audiovisual e no entretenimento**, o uso de IA tem transformado profundamente o processo de criação. Ferramentas de geração de vídeo automatizadas estão sendo adotadas por estúdios, agências e criadores independentes, permitindo a elaboração de roteiros, *storyboards* e cenas completas de maneira rápida e eficiente. Segundo [Yu *et al.* 2024], A adoção de ferramentas de geração de vídeo por IA na indústria ainda enfrenta desafios como a necessidade de qualificação técnica e a adaptação dos fluxos de trabalho, mas apresenta um enorme potencial para reduzir custos e acelerar o desenvolvimento de conteúdo multimídia.

Esses avanços evidenciam que a IA não apenas acelera a produção de vídeos, como também redefine a forma como o conteúdo é planejado, produzido e utilizado, impactando áreas como educação, entretenimento, marketing e comunicação. No entanto muitos usuários não conseguem obter bons resultados devido à falta de conhecimento sobre como elaborar *prompts* claros detalhados etc. Assim, o trabalho justifica-se pela necessidade de buscar boas práticas para gerar *prompts* eficientes permitindo que a tecnologia seja usada de forma produtiva e mais confiável [Yu *et al.* 2024], a adoção dessas ferramentas enfrenta barreiras técnicas e sociais, e uma das formas de superá-las é fornecer maior clareza aos usuários sobre como elaborar instruções mais precisas e adequadas.

A relevância deste estudo está em promover a democratização do uso da IA na produção audiovisual, visto que a elaboração de *prompts* constitui uma competência essencial para explorar o potencial criativo dos Modelos. Ao reunir diretrizes e exemplos práticos, o estudo oferece orientações claras e demonstrações para guiar a elaboração de *prompts*, este trabalho pretende auxiliar tanto pesquisadores quanto profissionais que desejam utilizar a IA na criação de vídeos, ampliando o alcance e a qualidade do conteúdo gerado. Além disso, ao discutir aspectos éticos e técnicos da geração automática de vídeos, o estudo contribui para reflexões importantes sobre os limites e as possibilidades dessa tecnologia no cenário contemporâneo.

1.2 Objetivos

A definição dos objetivos deste trabalho orienta o desenvolvimento da pesquisa e estabelece com clareza o que se pretende alcançar. Nesta seção, apresentam-se o objetivo geral e os objetivos específicos que estruturam a investigação sobre as boas práticas na elaboração de *prompts* para geração de vídeos por Inteligência Artificial. Esses objetivos organizam e orientam a aplicação prática e fundamentam a análise dos resultados obtidos, garantindo coerência metodológica ao estudo.

1.2.1 Objetivo Geral.

Analisar, por meio de um estudo de caso, as boas práticas na elaboração de *prompts* aplicados à geração de vídeos por Inteligência Artificial, destacando como cada elemento textual contribui para a clareza, coerência e fidelidade visual do resultado produzido.

1.2.2 Objetivos Específicos

- Investigar os fundamentos teóricos dos Modelos de Linguagem de Grande Escala (LLMs) e da engenharia de *prompts*
- Descrever as principais características estruturais que compõem um *prompt* eficiente, como clareza, detalhamento, estilo visual e coerência narrativa.
- Aplicar um *prompt* detalhado na ferramenta **Sora**, modelo multimodal da OpenAI, a fim de gerar um vídeo representativo para [Brooks *et al.* 2025].
- Explicar, de forma minuciosa, cada característica textual presente no *prompt*, demonstrando sua função e relevância na construção do comando completo.

1.3 Metodologia.

Este trabalho foi desenvolvido como um **estudo de caso prático e exploratório**, com o objetivo de analisar como a formulação dos *prompts* influencia o resultado dos vídeos gerados pela Inteligência Artificial *Sora 2*, modelo OpenAI.

A pesquisa buscou compreender, a importância dos elementos textuais dentro de um *prompts* o nível de detalhe, coerência e fidelidade visual dos vídeos produzidos.

1.3.1 Etapas do Estudo

O estudo foi dividido em três etapas principais:

Etapa 1 — Elaboração dos *prompts* Iniciais. Foram criados dois *prompts* complexos, contendo descrições completas de ambiente, personagens, ações, iluminação e estilo visual.

Etapa 2 — Geração e Observação dos Resultados. Cada versão de *prompts* foi inserida na ferramenta *Sora 2*, que gerou automaticamente os vídeos correspondentes. Os resultados foram analisados visualmente, utilizando imagens extraídas dos vídeos para demonstrar o comportamento da IA frente a cada tipo de comando.

Etapa 3 — Simplificação Progressiva dos *prompts*.

A partir dos *prompts* iniciais, foram realizadas modificações graduais, retirando partes das descrições e simplificando as instruções. O objetivo foi descrever a importância de cada elemento na construção de um *prompts*, demonstrando como cada parte como cenário, ação, iluminação e emoção contribui para a fidelidade e a consistência do resultado final.

1.3.2 Abordagem Analítica.

A análise foi de caráter **qualitativo e descritivo**, baseada na observação direta dos resultados visuais. Não foram aplicados critérios de pontuação ou métricas numéricas; o foco esteve na interpretação do comportamento da IA, considerando aspectos como:

- Fidelidade ao texto do *prompt*.
- Clareza e consistência das cenas geradas.

Essas observações embasaram a linha de raciocínio sobre a construção de *prompts* eficientes para geração de vídeos.

1.3.3 Resultados Esperados.

Com essa abordagem, espera-se demonstrar que ***prompts* mais detalhados produzem resultados mais coerentes e fiéis**. As imagens extraídas dos vídeos servirão como comprovação prática, sustentando a discussão apresentada nos capítulos seguintes.

1.4 Palavras-Chave.

A definição das palavras-chave foi realizada com o objetivo de representar de forma clara os principais conceitos abordados neste estudo. Cada termo foi selecionado para refletir os elementos centrais da pesquisa, incluindo a geração de vídeos por Inteligência Artificial, a construção de *prompts*. O quadro 1.1 apresenta os principais termos de busca utilizados na pesquisa, formulados para identificar estudos sobre engenharia de prompts e geração de vídeos por Inteligência Artificial. As combinações de palavras e operadores booleanos foram escolhidas para abranger diferentes abordagens, incluindo modelos multimodais e

ferramentas como a Sora 2. Essa estrutura permitiu localizar publicações relevantes em bases científicas, assegurando uma busca ampla e coerente com o tema investigado.

Quadro 1.1: *Termos de busca utilizados na pesquisa*

Conjunto de Termos	Combinação Utilizada na Busca
Engenharia de Prompts	("Prompt Engineering"OR "Engenharia de Prompts") AND ("AI Video Generation"OR "Geração de Vídeos por IA")
Prompt Design	("Prompt Engineering"OR "Prompt Design") AND ("Sora 2"OR "OpenAI Video Model")
IA Generativa	("Text-to-Video"OR "IA Generativa") AND ("Artificial Intelligence"OR "Modelos Multimodais")
Otimização de Prompts	("Prompt Optimization") AND ("Video Creation"OR "Criação Audiovisual com IA")

A definição das palavras-chave constitui uma etapa central na estruturação de qualquer pesquisa científica, pois influencia diretamente a precisão, a abrangência e a qualidade das referências encontradas. No âmbito deste estudo, que aborda a engenharia de *prompts* aplicados à geração de vídeos por Inteligência Artificial, a escolha criteriosa dos termos de busca foi fundamental para garantir que as publicações identificadas refletissem, de maneira fiel, os conceitos e tecnologias discutidos ao longo do trabalho.

Os termos selecionados representam os pilares conceituais deste tema, abrangendo desde fundamentos teóricos relacionados aos Modelos de Linguagem de Grande Escala (LLMs) até técnicas específicas de construção de comandos textuais capazes de orientar sistemas multimodais, como o modelo Sora, no processo de criação audiovisual. Ao empregar operadores booleanos e combinações semânticas cuidadosamente formuladas, buscou-se atingir um equilíbrio entre especificidade e amplitude, permitindo localizar estudos que tratam tanto da estruturação de *prompts* quanto do comportamento de modelos generativos em tarefas complexas de vídeo.

Essa abordagem ampliada possibilitou identificar pesquisas que descrevem práticas recomendadas, estratégias de detalhamento textual, desafios de consistência visual e limitações dos sistemas de geração de vídeo. Além disso, as palavras-chave utilizadas favoreceram a recuperação de conteúdos que tratam de , aplicações práticas e análises Dessa forma, a construção criteriosa desse conjunto de termos não apenas orientou a pesquisa, mas também assegurou que o embasamento teórico fosse sólido e coerente com os objetivos propostos, mesmo considerando a aplicação prática centrada em um único modelo de Inteligência Artificial.

Capítulo 2

Fundamentação Teórica

A presente fundamentação teórica tem como objetivo contextualizar os principais conceitos, tecnologias e abordagens relacionadas à geração de vídeos por Inteligência Artificial, com ênfase na importância da engenharia de *prompts* como elemento central nesse processo. Para isso, são abordados aspectos históricos da evolução dos Modelos de Linguagem de Grande Escala (LLMs), seu papel como ferramentas criativas e artísticas, e as transformações que essas tecnologias vêm promovendo na produção audiovisual. Em seguida, discute-se o funcionamento e as capacidades do modelo Sora, como exemplo de aplicação avançada na geração de vídeos, além de suas limitações e desafios atuais. Por fim, são apresentadas diretrizes recomendadas para a construção de *prompts* eficazes.

2.1 História dos Modelos de Linguagem de Grande Escala (LLMs).

Os modelos de linguagem progrediram desde os anos 1950, quando os sistemas baseados em regras linguísticas buscavam compreender padrões de linguagem natural. Embora pioneiros, esses métodos apresentavam limitações significativas, já que não conseguiam capturar a complexidade da linguagem humana [Hadi *et al.* 2025]

Nos anos de 1980 e 1990, surgiram modelos probabilísticos, como n-grams e os(HMMs) modelos ocultos de Markov , que representaram um avanço, mas ainda não eram capazes de lidar com dependências de longo prazo [Hadi *et al.* 2025].

Um momento decisivo ocorreu em 2017, com a publicação do artigo *Attention is All You Need*, de Vaswani et al, que apresentou a arquitetura *Transformer* (Transformador Pré-Treinado Generativo).

Esse progresso permitiu treinar modelos em larga escala e compreender melhor o contexto das palavras transformando o campo do PNL (Processamento Natural de Linguagem) [Hadi *et al.* 2025] A partir daí modelos como o GPT1, GPT3 e GPT4 (OpenAI) [OpenAI 2025], LLaMA (Meta) [Meta 2024] e Bard (Google) [Google 2023] consolidaram os LLMS como base de aplicações modernas em tradução automática, assistentes virtuais, análise semântica

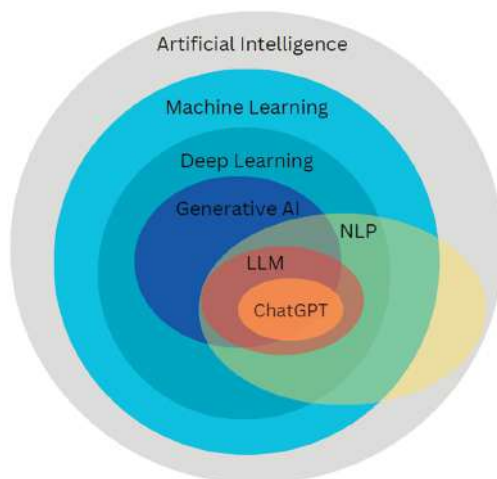
e geração diversificada de conteúdo. Entre essas possibilidades, destaca-se o uso criativo e artístico, em que a capacidade dos modelos de compreender e gerar linguagem natural vem sendo explorada para expandir as fronteiras da produção cultural e multimodal.

A Figura 1 pode ser interpretada como uma linha histórica em camadas que reflete a evolução da Inteligência Artificial. A trajetória inicia nos anos 1950, com o surgimento da *Artificial Intelligence* (IA), que buscava criar sistemas capazes de executar tarefas inteligentes. Nos anos 1980, o avanço dos métodos estatísticos impulsiona o *Machine Learning*, permitindo que máquinas aprendam padrões a partir de dados. A partir de 2010, o aumento do poder computacional consolida o *Deep Learning*, baseado em redes neurais profundas, que transforma áreas como visão e linguagem.

Por volta de 2020, crescem os modelos de *IA Generativa*, capazes de criar textos, imagens e vídeos, ao mesmo tempo em que o *NLP* evolui rapidamente após a introdução dos *Transformers* em 2017. Esse avanço leva ao surgimento, entre 2018 e 2023, dos *Large Language Models* (LLMs), grandes modelos de linguagem treinados em bilhões de palavras. No centro dessa evolução está o ChatGPT, lançado em 2022, representando um marco prático e acessível dessa nova geração de modelos.

Assim, a Figura sintetiza décadas de progresso tecnológico, mostrando como cada camada da IA clássica aos LLMs modernos se desenvolveu ao longo do tempo até culminar em sistemas como o ChatGPT.

Figura 1: *Evolução dos LLMs.*



Fonte: <https://www.researchgate.net/figure/Evolution-of-LLM-Illustrating-the-interplay-and-overlap_fig2_381943056>

2.2 Ferramentas Criativas e Artísticas.

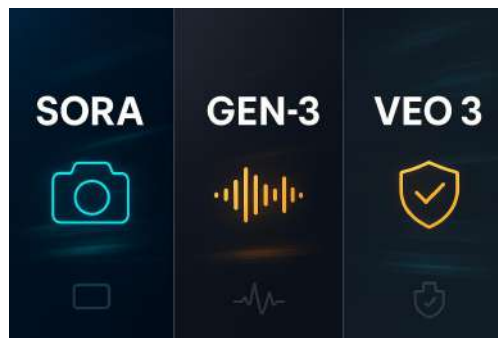
Os LLMS vêm se destacando como ferramentas criativas, aplicadas na geração de conteúdos artísticos como poesia, roteiros, descrições visuais e até no *design* (projeto) de jogos digitais. Essas aplicações demonstram o potencial de colaboração entre humanos e IA em diferentes

contextos como literatura, artes visuais e entretenimento [Mienye *et al.* 2025].

No campo áudio visual, as IAS auxiliam desde a escrita de roteiros até a descrição de cenas para *softwares*(programas) de geração de imagens e vídeos auxiliando e ampliando o processo criativo, ao oferecer novas perspectivas narrativas e apoiar artistas na experimentação de estilos inovadores

A Figura 2 são apresentados três modelos avançados de inteligência artificial utilizados na geração multimodal: **Sora**, **Gen-3** e **Veo 3**. Cada modelo é representado por um ícone que simboliza sua principal funcionalidade Sora está associado à geração de vídeos e conteúdo visual, Gen-3 à síntese de áudio e dinâmicas sonoras, enquanto Veo 3 está relacionado a mecanismos de segurança e controle. A imagem destaca como esses modelos atuam de forma complementar dentro do ecossistema atual de IA generativa.

Figura 2: *Ferramentas criativas.*



Fonte: <<https://skywork.ai/blog/sora-2-vs-gen-3-vs-veo-comparison/>>

No entanto, para que esses usos criativos e artísticos sejam realmente eficazes, torna-se essencial definir de forma clara como o usuário interage com os modelos. É nesse contexto que surge a engenharia de *prompts*, responsável por estruturar instruções precisas que orientam a geração de conteúdos consistentes e alinhados aos objetivos criativos.

2.3 Engenharia de Prompts.

A engenharia de *prompts* consiste na formulação de instruções textuais claras e detalhadas para guiar os resultados produzidos por LLMs. Segundo [Singh 2023], a qualidade da Saída depende diretamente de precisão das instruções Um *prompt* bem construído deve conter elementos estilo, contexto narrativa e características visuais desejadas. Essa prática permite maior controle sobre a geração de conteúdo e amplia a consistência dos resultados, especialmente em aplicações como geração de vídeo [Ekin 2023].

Assim, ao mesmo tempo em que a engenharia de *prompts* fornece os meios para orientar e controlar a saída dos modelos, ela se torna ainda mais relevante quando aplicada à geração de vídeos por IA, onde a complexidade narrativa e visual exige instruções detalhadas para garantir resultados consistentes e alinhados ao objetivo do usuário.

A Figura 3 apresenta um esquema em formato de funil que representa as etapas essenciais do processo de elaboração de *prompts* para modelos de linguagem. O diagrama organiza quatro fases principais: a construção do prompt, a interpretação realizada pelo LLM, a validação da saída gerada e, por fim, a otimização contínua. Cada etapa ilustra como a criação e o refinamento cuidadoso do prompt influenciam diretamente a qualidade e a precisão das respostas produzidas pelo modelo.

Figura 3: *Engenharia de prompts.*



Fonte: <https://thebuilderscompass.substack.com/p/engenharia-de-prompts-a-diferenca?utm_campaign=post&utm_medium=web>

2.4 Geração de Vídeos por IA .

A geração de vídeos por meio de IA tem como principal motivação a necessidade de acelerar a criação de conteúdo, tornando o processo mais rápido, diversificado, eficiente e econômico. Esses sistemas vêm sendo aplicados em diversas áreas, como *marketing*, educação e entretenimento, com o objetivo de aumentar o engajamento e melhorar a experiência do usuário [Liu e Özsu 2009]

Apesar de seu potencial, a geração automática de vídeos ainda apresenta desafios. Entre eles, a necessidade de grandes conjuntos de dados para treinamento, possibilidade de desvios entre a mensagem pretendida e a saída produzida pelo modelo. Ou seja, os sistemas de IA nem sempre conseguem compreender com precisão o significado ou a intenção do *prompt*, o que pode resultar em vídeos com erros de contexto, narrativa ou estilo [Vargas, Rocha e Freire 2007]

O vídeo pode ser definido como a representação de uma sequência de imagens organizadas para simular movimento, acompanhada ou não de som. Tradicionalmente, a produção audiovisual envolve três etapas principais: pré-produção, produção e pós-produção [Liu e Özsü 2009]. Na pré-produção, ocorre o planejamento, a definição do público-alvo, a elaboração do roteiro e do *storyboard*, na produção, realiza-se a gravação e a captação de imagens e áudios; e na pós-produção, são feitas a edição, a sonorização e a finalização do material.

Desse modo, modelos avançados como o Sora surgem como exemplos práticos do potencial e das capacidades atuais da Inteligência Artificial na geração de vídeos. Ao integrar processamento de linguagem natural, áudio nativo e alto realismo visual, esses modelos ilustram como os desafios conceituais mencionados como a consistência narrativa e a precisão da saída podem ser parcialmente superados, oferecendo soluções concretas para a criação de conteúdo rápido, diversificado e de qualidade

2.5 Sora.

O Sora é um modelo avançado de geração de vídeo desenvolvido para transformar instruções textuais e visuais em conteúdo audiovisual de alta qualidade.

Sora, é capaz de gerar vídeo com alta fidelidade. Os resultados sugerem que a ampliação de modelos de geração de vídeo é um caminho promissor para a criação de simuladores de propósito geral do mundo físico. Isso significa que o modelo poderia ser usado para simular cenários complexos. [Brooks *et al.* 2025]

O Sora pode gerar vídeos em *widescreen* 1920x1080p, vídeos verticais 1080x1920 e todas as variações intermediárias. Isso permite que o Sora crie conteúdo para diferentes dispositivos diretamente em suas proporções nativas. também pode ser acionado com outros tipos de entrada, como imagens ou vídeos pré-existent. Essa capacidade permite que o Sora realize uma ampla gama de tarefas de edição de imagem e vídeo, E é capaz de estender vídeos, seja para frente ou para trás no tempo. É possível usar o Sora para interpolar gradualmente entre dois vídeos de entrada, criando transições suaves entre vídeos com sujeitos e composições de cena totalmente diferentes. também é capaz de gerar imagens de tamanhos variados com resolução de até 2048x2048. [Brooks *et al.* 2025]

É possível simular alguns aspectos de pessoas, animais e ambientes do mundo físico. Essas propriedades surgem sem qualquer viés indutivo explícito para 3D, objetos, etc, são fenômenos puramente relacionados à escala.

O Sora pode gerar vídeos com movimento dinâmico de câmera. Por exemplo a medida que a câmera se desloca e gira, pessoas e elementos da cena se movem de forma consistente através do espaço tridimensional. O Sora demonstra, a capacidade de modelar de forma eficaz tanto dependências de curto quanto de longo prazo. Por exemplo, o modelo é capaz de manter a consistência de pessoas, animais e objetos mesmo quando estes estão ocultos ou saem do quadro. Além disso, é capaz de gerar múltiplas tomadas de um mesmo personagem em uma

única amostra, preservando sua aparência ao longo de todo o vídeo.

O modelo também é capaz, em certas situações, de simular ações que alteram o estado do ambiente de forma simples. Por exemplo, um pintor pode aplicar novos traços em uma tela que permanecem ao longo do tempo, ou um indivíduo pode comer um hambúrguer, deixando marcas de mordida.

O Sora também consegue simular processos artificiais, como em *videogames*. Por exemplo, o modelo é capaz de controlar simultaneamente o jogador no *Minecraft*, utilizando uma política básica, enquanto renderiza o mundo e suas dinâmicas com alta fidelidade. [Brooks *et al.* 2025].

Mesmo com o progresso alcançado pelo Sora, a geração de vídeos por IA ainda não estão livre de problemas. É necessário reconhecer as limitações atuais e os desafios que precisam ser superados para que essa tecnologia seja consolidada de forma confiável e responsável.

2.5.1 Justificativa da Escolha do Modelo Sora.

A escolha do modelo Sora como foco deste estudo se justifica por sua relevância tecnológica e por representar um dos avanços mais significativos na geração de vídeos por Inteligência Artificial. Entre os diversos modelos de texto para vídeo existentes, o Sora destaca-se por sua capacidade multimodal refinada, interpretando instruções textuais complexas com elevado nível de fidelidade visual e mantendo consistência de personagens, coerência espacial e continuidade temporal ao longo das cenas.

Além disso, o Sora utiliza uma arquitetura baseada em LLMs adaptada para raciocínio visual, permitindo compreender nuances semânticas presentes nos *prompts* e traduzi-las em elementos cinematográficos, como movimentos de câmera, ambientação, expressões faciais, iluminação e interações dinâmicas com o cenário. Essa característica é fundamental para esta pesquisa, que tem como objetivo analisar como diferentes componentes textuais influenciam a qualidade do vídeo produzido. A utilização do Sora funciona como um mecanismo eficaz para comprovar empiricamente a eficiência dos *prompts* avaliados, uma vez que o modelo responde de maneira sensível e detalhada às instruções fornecidas.

O modelo apresenta resultados superiores em aspectos como coerência física, comportamento natural de objetos, profundidade de campo, iluminação realista e integração entre narrativa e estética fatores essenciais para embasar as análises aqui propostas e ajudar a validar a eficácia das características textuais presentes nos *prompts*.

Dessa forma, utilizar o Sora como estudo de caso possibilita investigar uma tecnologia de ponta que está transformando processos de criação em áreas como cinema, *marketing* digital, educação e entretenimento. Sua adoção neste trabalho reforça a importância da engenharia de *prompts* como competência essencial para a produção audiovisual assistida por IA e destaca o caráter inovador da pesquisa ao explorar uma das soluções mais avançadas no campo da Ia ao mesmo tempo em que fortalece a comprovação prática da eficiência dos prompts analisados.

2.6 Desafios e Limitações da Geração de Vídeo por IA.

Apesar dos avanços recentes, a geração de vídeos por IA ainda enfrenta uma série de desafios técnicos, sociais e de mercado que limitam sua adoção em larga escala [Yu *et al.* 2024]. Um dos principais obstáculos está relacionado à qualidade do conteúdo gerado, uma vez que os vídeos muitas vezes apresentam falta de solidez temporal, entre outros aspectos, o que pode comprometer a credibilidade do material produzido.

Outro desafio relevante está relacionado ao alto custo computacional dessas ferramentas. A geração de vídeos requer grande capacidade de processamento de dados e o uso de recursos de *hardware* (equipamentos físicos) especializados, o que torna o processo caro e de difícil acesso para muitos usuários. Além disso, há barreiras na aceitação do mercado, uma vez que diversos profissionais do setor audiovisual demonstram resistência ao uso de IA por receio de substituição da mão de obra humana. A ausência de normas regulatórias claras e as preocupações éticas como o risco de *deepfakes* (...) e a manipulação indevida de imagens também reforçam o cenário de cautela.

Superar esses obstáculos é fundamental para que a tecnologia alcance maior confiabilidade e seja incorporada de forma sustentável nos diversos setores da sociedade.

A geração de vídeos por IA tem potencial para transformar diferentes setores em escala global, ampliando a produção audiovisual e possibilitando que indivíduos e pequenas empresas produzam conteúdo de alta qualidade sem depender de grandes estúdios ou recursos elevados [Yu *et al.* 2024]. Esse impacto se reflete em diferentes segmentos da sociedade, nos quais a automação de vídeos pode reduzir custos e acelerar processos criativos.

Em relação à perspectiva futura, é possível imaginar um cenário no qual as IAS se tornarão cada vez mais multitarefas e interativas, permitindo maior controle criativo dos usuários sobre os resultados. A expectativa é que a evolução da engenharia de *prompts* e o desenvolvimento de algoritmos mais eficientes tornem a geração de vídeos por ia mais acessível, realista e confiável, consolidando seu papel como ferramenta essencial na produção de conteúdo.

2.6.1 Estrutura Recomendada para um *prompts* de Vídeo.

A criação de vídeos por meio de Inteligência Artificial depende diretamente da elaboração cuidadosa do *prompt* utilizado. Um *prompt* bem estruturado não apenas transmite a intenção do usuário, mas também orienta a IA na geração de material coerente, visualmente consistente e alinhado ao resultado desejado. Estudos sobre modelos texto-para-vídeo destacam a importância da clareza nas descrições e da qualidade das instruções para guiar os sistemas generativos [Singh 2023].

Esses aspectos envolvem desde a clareza e objetividade na descrição, até elementos específicos como tipo de tomada, descrição de personagens e objetos, ações, cenário, estética, adjetivos descritivos, contexto narrativo, movimentos de câmera e elementos temporais. Cada

um desses componentes desempenha um papel essencial na comunicação com a IA, garantindo que a interpretação do *prompt* seja eficiente em relação à visão criativa do usuário.

A seguir, são apresentadas diretrizes detalhadas para a construção de *prompts* eficazes, capazes de orientar o processo de geração de vídeos de forma criativa, objetiva e próxima ao resultado esperado.

1. **Seja objetivo e claro:** Ao escrever sua solicitação, utilize o máximo de detalhes possível sobre iluminação, estilo cinematográfico, cor, clima e estética. Uma linguagem simples e precisa evita ambiguidades e ajuda a IA a compreender com proximidade o tipo de vídeo que se deseja gerar [Inc. 2025]..
2. **Tipo de tomada:** Refere-se à perspectiva da câmera e como ela deve se mover. Por exemplo: *“close-up com zoom lento de uma personagem olhando pela janela”*.
3. **Descrição do personagem e objetos:** Explique quem é o personagem, o que está vestindo e como são os objetos em cena. Exemplo: *“um grande urso polar alegre usando cachecol azul”*.
4. **Ação:** Especifique o que o personagem está fazendo. Exemplo: *“o urso está correndo pela neve”* [Inc. 2025].
5. **Local e ambientação:** Descreva o cenário, incluindo clima e condições visuais. Exemplo: *“ambiente gelado, com neve caindo e céu nublado”* [Inc. 2025]..
6. **Estética:** Defina o estilo visual do vídeo: realista, animado, cinematográfico, etc. Detalhe a aparência desejada, como profundidade de campo e tipo de lente. Exemplo: *“vídeo cinematográfico em 35mm com tons frios”* [Inc. 2025]..
7. **Adjetivos descritivos:** Use termos que evoquem sensações e atmosfera desejada, como calmo, sombrio, misterioso ou vibrante. Isso ajuda a IA a captar a emoção do vídeo [Inc. 2025]..
8. **Contexto ou história de fundo:** Para vídeos com enredo, forneça um contexto narrativo. Isso contribui para a coerência entre cenas e personagens. Também é possível usar imagens como base, garantindo continuidade visual [Inc. 2025]..
9. **Movimentos de câmera:** Indique se deseja movimentos como panorâmica, zoom ou travellings. Por exemplo: *“panorâmica lateral da cidade ao entardecer”* [Inc. 2025]..
10. **Elementos temporais:** Seja específico quanto ao horário e clima, pois a luz e o ambiente mudam com o tempo. Exemplos: *“amanhecer com névoa densa”, “fim de tarde com sol alaranjado”, “noite com trovões ao fundo”* [Inc. 2025].

Ao definir claramente elementos como estética, narrativa e movimentos de câmera, os resultados tornam-se mais precisos, criativos e fiéis à visão desejada.

Capítulo 3

Resultados

Este capítulo apresenta uma análise sobre a estrutura e a construção de *prompts* eficazes para a geração de vídeos por Inteligência Artificial (IA), com foco no modelo Sora. A elaboração de *prompts* bem estruturados é essencial para orientar corretamente o comportamento do modelo e garantir que o resultado final seja coerente, esteticamente agradável e narrativamente consistente.

Inicialmente, são explorados os principais elementos que compõem um *prompt* de qualidade, como clareza, objetividade e detalhamento técnico. Em seguida, são apresentados exemplos práticos que demonstram como diferentes componentes — como enquadramento, ambientação, iluminação e emoções — influenciam diretamente o processo de geração visual.

A seção inclui a análise de dois *prompts* cinematográficos, ambos elaborados com o objetivo de ilustrar boas práticas na criação de instruções textuais para modelos de geração de vídeo. Cada *prompt* é decomposto em partes, evidenciando o papel de aspectos como a descrição de personagens e objetos, a definição da ação, o ambiente, a estética visual e os movimentos de câmera.

Por fim, é apresentado o *Checklist Único*, uma proposta de estrutura padronizada para criação de *prompts* eficientes, dividida em cinco categorias principais: Personagem, Cenário e Ambiente, Emoção e Clima Narrativo, Estilo Visual e Técnica Cinematográfica. Esse modelo serve como guia para pesquisadores e criadores, permitindo desenvolver *prompts* mais precisos, expressivos e tecnicamente consistentes.

O objetivo deste capítulo é demonstrar, de forma analítica, como a qualidade da descrição textual impacta diretamente o desempenho dos modelos de IA na geração de vídeos, destacando a importância da engenharia de *prompts* como ferramenta criativa e técnica no campo do audiovisual inteligente.

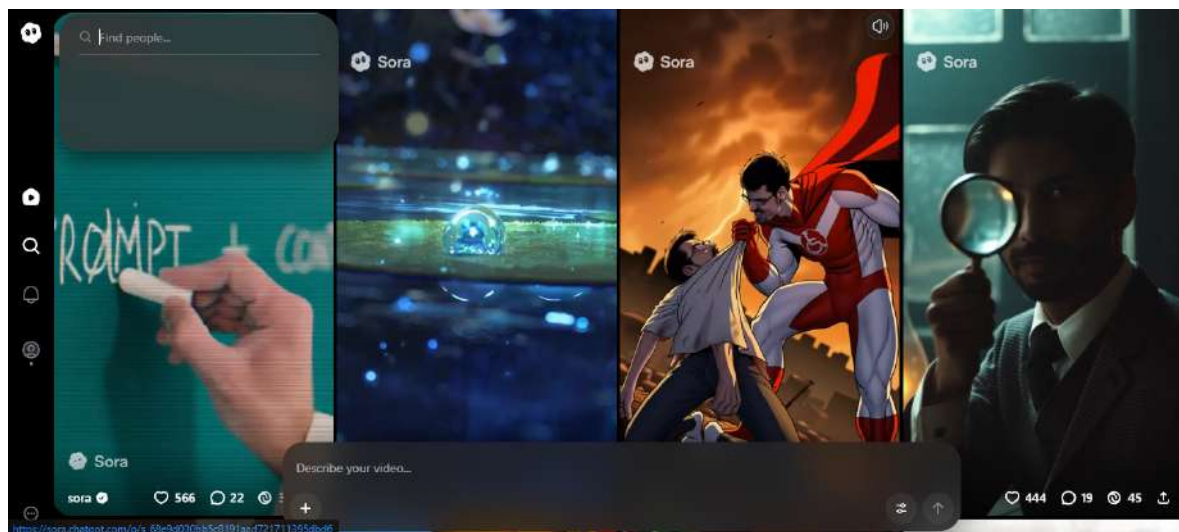
3.1 Interface Sora 2.

A Figura 4 a seguir mostrada a tela inicial da plataforma Sora 2, um ambiente moderno e intuitivo voltado para a geração de vídeos por Inteligência Artificial. **A interface é dividida**

em colunas verticais que exibem exemplos de vídeos criados, cada um com diferentes estilos e temáticas desde cenas realistas até animações. **No canto esquerdo** , há uma barra lateral preta com ícones que permitem navegar entre as principais seções, como o Feed (rede social) principal, pesquisa, notificações e perfil. Acima dela, uma caixa de busca convida o usuário a “Find people...”,(Encontre pessoas) reforçando o aspecto social da plataforma. **Ao centro da tela**, destacam-se miniaturas de vídeos gerados pela Sora, cada uma identificada com o logotipo da ferramenta. Entre os exemplos, é possível ver: uma mão escrevendo a palavra *prompts* em um quadro, uma gota d’água animada em um cenário realista, uma ilustração de super-herói em estilo de quadrinhos e uma cena cinematográfica com um homem segurando uma lupa. **Na parte inferior central** , há uma barra escura com o campo “Describe your video...”(Descreva seu vídeo), onde o usuário pode digitar seu *prompts* a descrição que orientará a ia na criação do vídeo. Ao lado, um ícone de “+” sugere a adição de novos projetos ou prompts.

A estética geral é interativa, combinando tons escuros com elementos visuais . A tela transmite a ideia de uma galeria criativa, onde o usuário pode explorar inspirações e criar vídeos personalizados a partir de suas próprias ideias.

Figura 4: *Tela inicial.*



Fonte: <<https://openai.com/pt-BR/index/sora/>>

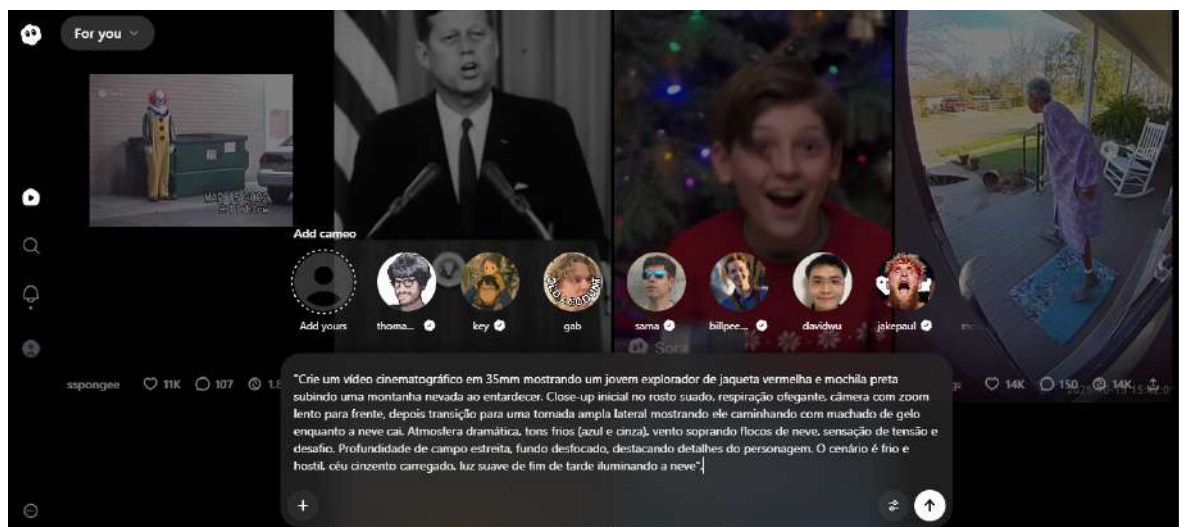
Para que a Inteligência Artificial gere vídeos de alta qualidade, é fundamental que o *prompt* seja construído de forma clara, detalhada e estruturada. Os elementos descritos anteriormente desde a objetividade na solicitação até a definição de tipo de tomada, descrição de personagens e objetos, ações, cenário, estética, adjetivos descritivos, contexto, movimentos de câmera e elementos temporais funcionam como diretrizes que orientam a ia a interpretar corretamente a intenção do usuário [Inc. 2025].

Quando combinados de maneira precisa, esses componentes permitem que o modelo compreenda não apenas o que deve aparecer na cena, mas também como cada elemento

deve ser apresentado, garantindo coerência visual, narrativa e emocional. Para ilustrar como essas diretrizes se aplicam na prática, é apresentado um exemplo de *prompt* bem estruturado, que integra todos os aspectos discutidos

A Figura 5 é apresentada a seguir, na qual é mostrada a interface do Sora. No centro da tela, há um campo de texto onde o usuário insere um *prompt* descritivo para gerar um vídeo neste caso, uma cena cinematográfica em 35 mm de um jovem explorador subindo uma montanha nevada ao entardecer. A interface também exibe exemplos de vídeos e criadores que já utilizaram o modelo, permitindo ao usuário explorar resultados e inspirações.

Figura 5: Sora 2.



Fonte: <<https://openai.com/pt-BR/index/sora/>>

A Figura 6 a seguir mostra o resultado visual obtido após a inserção de um *prompt* na inteligência artificial Sora, voltada para geração de vídeos. O comando descrevia uma cena cinematográfica de um jovem explorador em ambiente frio e nevado, e o modelo produziu uma imagem fiel à descrição retratando um homem com expressão intensa, vestindo gorro escuro e jaqueta vermelha, sob iluminação suave e atmosfera gelada. O resultado evidencia a capacidade da Sora de interpretar instruções textuais detalhadas e convertê-las em representações visuais realistas e emocionalmente expressivas.

Figura 6: *Resultado primeiro exemplo.*



Fonte: <<https://openai.com/pt-BR/index/sora/>>

3.2 Exemplo de *prompt* Bem Estruturado.

A seguir, apresenta-se um exemplo prático de *prompt* elaborado que foi inserido anteriormente Figura 6 na IA para a geração de vídeo na inteligência artificial Sora. O objetivo deste exemplo é ilustrar como a descrição textual detalhada e coerente pode orientar o modelo na criação de resultados visuais mais consistentes e cinematográficos. O comando a seguir foi desenvolvido com foco na ambientação, nos elementos visuais e na emoção transmitida pela cena, buscando explorar o potencial do modelo para representar ambientes e personagens de forma realista.

É importante destacar que, neste trabalho, a ênfase não recai sobre a avaliação do desempenho técnico da inteligência artificial Sora, mas sim sobre a análise da estrutura e das características do *prompt* utilizado. O objetivo é compreender como a formulação textual influencia o resultado visual gerado, observando elementos como clareza, detalhamento, coerência narrativa e direcionamento estético. Assim, a discussão a seguir concentra-se na qualidade e na construção do comando inserido, evidenciando as boas práticas na elaboração de *prompts* para geração de vídeos por IA.

prompt:

"Crie um vídeo cinematográfico em 35mm mostrando um jovem explorador de jaqueta vermelha e mochila preta subindo uma montanha nevada ao entardecer. *Close-up* inicial no rosto, respiração ofegante, câmera com *zoom* lento para frente, depois transição para uma tomada ampla lateral mostrando ele caminhando com machado de gelo enquanto a neve cai. Atmosfera dramática, tons frios (azul e cinza), vento soprando flocos de neve, sensação de tensão e desafio. Profundidade de campo estreita, fundo desfocado, destacando detalhes do personagem. O cenário é frio e hostil, céu cinzento carregado, luz suave de fim de tarde iluminando a neve".

As IAs de geração de vídeo são capazes de interpretar prompts em diferentes idiomas. No entanto, o desempenho pode variar conforme a língua utilizada, sendo o inglês geralmente o idioma que apresenta melhores resultados, devido à maior presença nos dados de treinamento dos modelos.

3.2.1 Análise do *prompt* de Vídeo.

A análise tem como objetivo examinar a estrutura de um *prompt* elaborado para geração de vídeo por meio de inteligência artificial. O intuito é compreender como a formulação textual influencia diretamente o resultado visual produzido, evidenciando a importância de uma descrição clara, coerente e rica em detalhes. Nesse contexto, são observados aspectos como a definição de cenário, personagens, ações e enquadramentos de câmera, além de elementos estéticos e emocionais que contribuem para a construção de uma narrativa cinematográfica.

A identificação desses componentes permite reconhecer boas práticas na criação de *prompts* eficazes, capazes de orientar a IA de forma mais precisa e expressiva. Assim, o foco desta análise está na estrutura e na qualidade do comando, e não na capacidade técnica do modelo em si, destacando o papel do texto como principal guia criativo no processo de geração audiovisual.

A seguir, são destacados os principais elementos presentes no *prompt* citado anteriormente Figura 6 , com ênfase em suas características descritivas e narrativas.

1. Clareza e objetividade:

“Crie um vídeo cinematográfico em 35mm mostrando um jovem explorador de jaqueta vermelha e mochila preta subindo uma montanha nevada ao entardecer.” A descrição aqui refere-se à forma como o *prompt* comunica sua ideia ao modelo de geração de vídeo. A clareza está presente porque a descrição é precisa e compreensível, permitindo que o modelo visualize com facilidade o que deve ser criado: há um personagem definido (“jovem explorador”), elementos visuais específicos (“jaqueta vermelha”, “mochila preta”), um ambiente claramente descrito (“montanha nevada”) e um momento temporal (“ao entardecer”). A objetividade se manifesta porque o *prompt* vai diretamente ao ponto, sem incluir informações desnecessárias ou subjetivas. Ele comunica o que deve ser mostrado, como deve ser mostrado (formato cinematográfico em 35mm) e em que contexto, de forma concisa

2. Tipo de tomada:

“*Close-up* inicial no rosto suado, respiração ofegante, câmera com *zoom* lento para frente, depois transição para uma tomada ampla lateral...” Neste caso especifica perspectiva da câmera e movimento. Trata da forma como o *prompt* orienta a posição, o enquadramento e o movimento da câmera durante o vídeo. A descrição demonstra clareza na definição da perspectiva e da dinâmica visual. O *close-up* indica

um enquadramento fechado, focado no rosto do personagem, destacando emoção e esforço. O *zoom* lento para frente acrescenta movimento cinematográfico e intensidade, aproximando o espectador da cena. Por fim, a transição para uma tomada ampla lateral amplia o campo de visão, mostrando o ambiente e o deslocamento do personagem.

3. Descrição do personagem e objetos:

“...um jovem explorador de jaqueta vermelha e mochila preta...” Neste trecho Define - se quem é o personagem e seus objetos principais. Destaca como o *prompt* apresenta os elementos centrais da cena por meio da caracterização do protagonista e dos itens que o acompanham. A descrição é direta e visualmente específica, permitindo que o modelo identifique quem é o personagem e quais são seus atributos visuais principais. A menção à “jaqueta vermelha” e à “mochila preta” fornece contrastes de cor marcantes que ajudam a destacar o explorador no cenário nevado, além de transmitir aspectos de sua função (explorador preparado, em ambiente hostil). Assim, o *prompt* demonstra clareza na caracterização e objetividade na escolha dos detalhes, definindo o personagem e seus objetos de forma suficiente para orientar a construção visual sem adicionar informações irrelevantes.

4. Ação:

“...subindo uma montanha nevada ao entardecer... caminhando com machado de gelo...” Mostra - se claramente o que o personagem está fazendo. Descreve o movimento e o comportamento do personagem dentro da cena. Ação é explicitamente definida, mostrando o que o personagem faz e em que contexto. A expressão “subindo uma montanha nevada” transmite esforço físico e desafio, enquanto “caminhando com machado de gelo” reforça a ideia de exploração e sobrevivência em um ambiente hostil. Dessa forma, o *prompt* apresenta a ação de forma clara, dinâmica e coerente com o cenário, permitindo ao modelo compreender não apenas a pose do personagem, mas também o ritmo e a intenção da cena, o que contribui para uma representação mais realista e envolvente.

5. Local e ambientação:

“...montanha nevada ao entardecer... neve cai... cenário é frio e hostil, céu cinzento carregado...” Define o ambiente, clima e condições visuais. Estabelece o espaço físico e a atmosfera emocional da cena. Há uma descrição rica e precisa do ambiente, que situa o personagem em um local específico (“montanha nevada”) e em um momento do dia (“ao entardecer”). Além disso, elementos como “neve cai” e “céu cinzento carregado” criam uma sensação visual e climática de frio, isolamento e desafio. Demonstra clareza na construção do cenário e objetividade na escolha dos detalhes, fornecendo informações suficientes para que o modelo gere um ambiente coerente, imersivo e alinhado ao tom dramático da narrativa.

6. Estética:

“Vídeo cinematográfico em 35mm... atmosfera dramática, tons frios (azul e cinza)... profundidade de campo estreita, fundo desfocado...”

estilo visual detalhado, referência técnica de lente, estética dramática. o tom artístico do vídeo, orientando o modelo quanto à aparência final da cena. Há uma indicação clara da linguagem cinematográfica desejada, com a menção à filmagem em 35mm, que remete a uma textura clássica e realista. A atmosfera dramática e os tons frios (azul e cinza) reforçam a sensação de tensão e solidão no ambiente gelado. Já a profundidade de campo estreita e o fundo desfocado direcionam o foco para o personagem, criando um efeito estético profissional e emocionalmente intenso. Apresenta clareza nas escolhas visuais e técnicas, e objetividade ao definir a estética e o clima da cena, guiando a geração de um resultado coeso, cinematográfico e visualmente expressivo.

7. Adjetivos descritivos:

“Atmosfera dramática... frio e hostil... sensação de tensão e desafio...” OS adjetivos reforçam emoção e atmosfera desejada. Utiliza palavras qualificadoras para intensificar a emoção e caracterizar o clima da cena. Os adjetivos escolhidos são precisos e expressivos, ajudando a construir a identidade emocional e visual do vídeo. Como “dramática” e “hostil” evocam um ambiente de dificuldade e superação, enquanto “tensão” e “desafio” comunicam o estado psicológico do personagem diante das condições adversas. Demonstra clareza na escolha dos adjetivos e objetividade ao empregar palavras que direcionam o tom emocional, garantindo que o modelo compreenda a intensidade, o clima e o sentimento que devem ser transmitidos na cena.

8. Contexto ou história de fundo:

“...um jovem explorador... subindo uma montanha nevada...” Há uma narrativa implícita (um explorador enfrentando desafios na natureza hostil). Indica quem é o personagem e qual é sua jornada, permitindo ao modelo inferir um enredo de superação e aventura. A figura do “explorador” e a ação de “subir uma montanha nevada” remetem a um conflito entre o homem e a natureza, sugerindo esforço, coragem e resistência diante de condições extremas. Sugere uma situação narrativa reconhecível e objetividade ao transmitir o contexto sem precisar de longas explicações, permitindo que a cena tenha profundidade simbólica e propósito dramático.

9. Movimentos de câmera:

“...close-up inicial... zoom lento para frente... transição para uma tomada ampla lateral...” É discutido aqui o dinamismo e movimentação de câmera. Orienta o dinamismo visual da cena por meio da descrição dos deslocamentos e transições da câmera. Há uma sequência clara de movimentos cinematográficos, que conduz o olhar do espectador e reforça a intensidade emocional do momento. O *close-up* inicial cria intimidade e foco no personagem; o *zoom* lento para frente aumenta a tensão e aproxima o público

da ação; E a transição para a tomada ampla lateral amplia o campo de visão, contextualizando o ambiente e o desafio enfrentado. Movimentos de câmera e objetividade ao descrever a sequência visual, o que contribui para uma filmagem fluida, envolvente e com forte impacto narrativo.

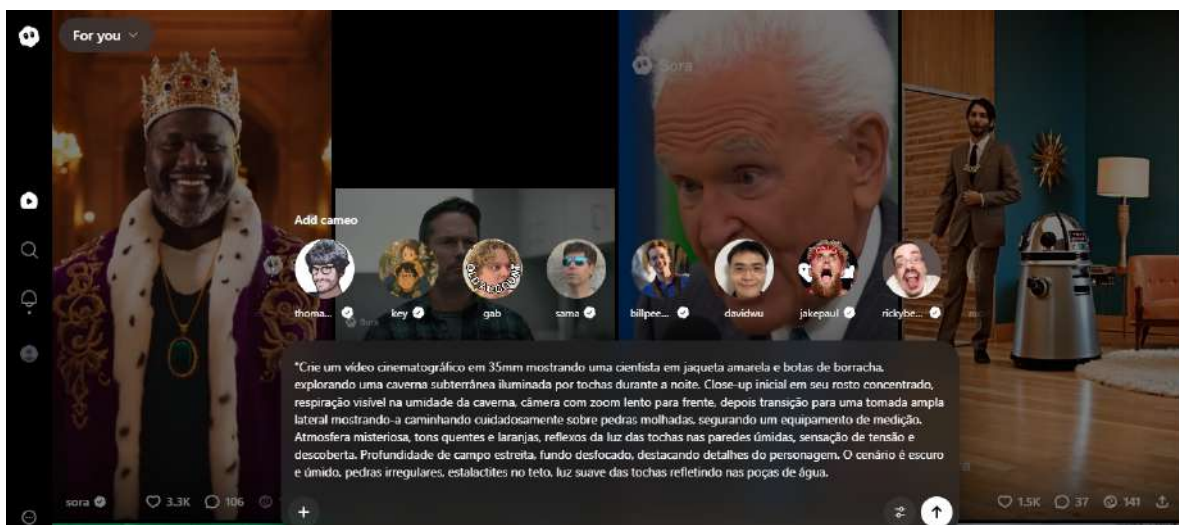
10. Elementos temporais:

“...ao entardecer... luz suave de fim de tarde iluminando a neve...” Elementos que define claramente o horário e a iluminação correspondente. Especifica o momento do dia e as condições de iluminação que compõem a atmosfera da cena. Há uma definição precisa do tempo e da luz, situando a ação em um período de transição entre dia e noite. O uso do entardecer cria uma iluminação natural e difusa, que reforça o tom melancólico e dramático da narrativa. A “luz suave” reflete calma e realismo, enquanto o brilho sobre a neve adiciona contraste visual e profundidade à composição da imagem.

A análise detalhada do *prompt* evidencia como cada elemento contribui para a construção de um vídeo cinematográfico coerente, visualmente rico e emocionalmente envolvente. Aspectos como clareza e objetividade garantem que o modelo compreenda de forma precisa o que deve ser gerado, enquanto a descrição do personagem, de suas ações e dos objetos fornece identidade e função ao protagonista. A definição do ambiente, a ambientação climática, os elementos temporais e os adjetivos descritivos trabalham juntos para criar atmosfera, tensão e profundidade narrativa. Além disso, a especificação de tipos de tomada, movimentos de câmera e estética detalham o estilo visual e a dinâmica da cena, assegurando um resultado técnico e esteticamente consistente.

Na sequência, um novo *prompt* é inserido na IA Sora com o objetivo de gerar outro vídeo. Assim como no exemplo anterior, o resultado é analisado detalhadamente, observando as principais características do *prompt*, como clareza, estrutura, estética e elementos narrativos, a fim de compreender como cada componente influencia a criação visual e o comportamento do modelo durante a geração do vídeo.

A Figura 7 a seguir apresenta a interface do Sora, onde foi inserido um *prompt* descritivo no campo central para gerar vídeos com base em instruções textuais. Neste caso, o *prompt* descreve uma cena cinematográfica em 35mm mostrando uma cientista vestindo jaqueta amarela e botas de borracha, explorando uma caverna subterrânea iluminada por tochas durante a noite. A interface também exibe exemplos de vídeos e criadores que já utilizaram o modelo, permitindo ao usuário explorar inspirações e compreender diferentes estilos visuais

Figura 7: *Exemplo .*

Fonte: <<https://openai.com/pt-BR/index/sora/>>

A Figura 8 a seguir apresenta o resultado visual obtido após a inserção de um *prompt* na inteligência artificial Sora, voltada para a geração de vídeos. O comando descrevia uma cena cinematográfica em 35mm de uma cientista explorando uma caverna subterrânea iluminada por tochas durante a noite, e o modelo produziu uma imagem coerente e fiel à descrição, retratando uma mulher com expressão concentrada, vestindo jaqueta amarela e botas de borracha, caminhando com cautela entre pedras molhadas e reflexos de luz nas paredes úmidas. A iluminação em tons quentes e alaranjados reforça a atmosfera misteriosa e de descoberta, enquanto o enquadramento e a profundidade de campo estreita destacam os detalhes do cenário e da personagem. O resultado demonstra a precisão e a sensibilidade visual da Sora em interpretar instruções textuais complexas.

Figura 8: *Resultado do segundo exemplo.*

Fonte: <<https://openai.com/pt-BR/index/sora/>>

3.2.2 Análise.

O *prompt* a seguir foi inserido anteriormente na IA Sora, servindo como base para a geração de um vídeo cinematográfico que resultou na Figura 8. Suas características e elementos descritivos serão analisados posteriormente, com o objetivo de compreender como cada componente contribui para a construção visual, narrativa e estética da cena.

prompt:

"Crie um vídeo cinematográfico em 35mm mostrando uma cientista em jaqueta amarela e botas de borracha, explorando uma caverna subterrânea iluminada por tochas durante a noite. *Close-up* inicial em seu rosto concentrado, respiração visível na umidade da caverna, câmera com *zoom* lento para frente, depois transição para uma tomada ampla lateral mostrando-a caminhando cuidadosamente sobre pedras molhadas, segurando um equipamento de medição. Atmosfera misteriosa, tons quentes e laranjas, reflexos da luz das tochas nas paredes úmidas, sensação de tensão e descoberta. Profundidade de campo estreita, fundo desfocado, destacando detalhes do personagem. O cenário é escuro e úmido, pedras irregulares, estalactites no teto, luz suave das tochas refletindo nas poças de água."

1. Clareza e objetividade:

"Crie um vídeo cinematográfico em 35mm mostrando uma cientista em jaqueta amarela e botas de borracha, explorando uma caverna subterrânea iluminada por tochas." Descrição clara, detalhada e direta, sem ambiguidades. Instruções diretas e bem estruturadas permitem que a inteligência artificial compreenda facilmente o que deve ser gerado. A descrição é precisa e sem excessos, delimitando claramente o tipo de vídeo (cinematográfico), o enquadramento (em 35mm), o personagem principal (uma cientista), e o ambiente (caverna subterrânea iluminada por tochas).

Essa objetividade garante que o modelo interprete corretamente os elementos essenciais da cena, enquanto a clareza evita ambiguidades, resultando em uma geração visual coerente com a intenção do criador.

2. Tipo de tomada:

"*Close-up* inicial em seu rosto concentrado, respiração visível na umidade da caverna, câmera com *zoom* lento para frente, depois transição para uma tomada ampla lateral..." Esse trecho especifica perspectiva da câmera e movimentos." indica como o *prompt* define o enquadramento e a perspectiva da câmera para orientar a narrativa visual. O *close-up* evidencia a expressão e a concentração da cientista, aproximando o espectador da emoção do personagem. O *zoom* lento para frente cria dinamismo e intensidade, enquanto a transição para uma tomada ampla lateral amplia a visão do ambiente, mostrando detalhes da caverna e o cuidado da personagem ao caminhar sobre pedras

molhadas. Dessa forma, o *prompt* fornece orientações claras sobre os movimentos e enquadramentos da câmera, garantindo que a cena seja visualmente envolvente.

3. Descrição do personagem e objetos:

“...uma cientista em jaqueta amarela e botas de borracha... segurando um equipamento de medição...” Aqui define - se quem é o personagem e seus objetos principais, evidencia como o *prompt* apresenta o protagonista e os elementos que o acompanham, fornecendo identidade e função na cena, A descrição identifica claramente quem é o personagem (uma cientista) e seus atributos visuais (jaqueta amarela, botas de borracha), além de especificar os objetos que ela utiliza (equipamento de medição), que reforçam seu papel e sua ação na narrativa.

4. Ação:

“...explorando a caverna, caminhando cuidadosamente sobre pedras molhadas...” . Mostra - se claramente o que o personagem está fazendo, descreve o movimento e o comportamento do personagem dentro da cena. A ação é apresentada de forma clara e direta, mostrando o que a cientista está fazendo e como ela interage com o ambiente. O verbo “explorando” indica a atividade principal, enquanto “caminhando cuidadosamente sobre pedras molhadas” reforça o cuidado e a tensão da situação, transmitindo realismo e envolvimento na narrativa.

5. Local e ambientação:

“...caverna subterrânea iluminada por tochas, pedras irregulares, estalactites no teto, luz suave das tochas refletindo nas poças de água...” Define - se o ambiente, clima e condições visuais, Estabelece o cenário físico e a atmosfera da cena. A descrição define claramente (caverna subterrânea), os detalhes visuais (pedras irregulares, estalactites) e a iluminação (luz suave das tochas refletindo nas poças de água), criando uma atmosfera de mistério e tensão.

6. Estética:

“Vídeo cinematográfico em 35mm, atmosfera misteriosa, tons quentes e laranjas, profundidade de campo estreita, fundo desfocado...” . É apresentado aqui o estilo visual detalhado, Referência técnica de lente, estética cinematográfica, o tom artístico da cena, orientando a IA sobre a aparência final do vídeo.

7. Adjetivos descritivos:

“Atmosfera misteriosa, sensação de tensão e descoberta...” . OS adjetivos reforçam emoção e atmosfera desejada, utiliza termos qualificadores para intensificar a emoção e caracterizar o clima da cena. Os adjetivos e expressões escolhidos são precisos e expressivos, ajudando a transmitir mistério, suspense e curiosidade, enquanto comunicam o estado emocional da personagem durante a exploração da caverna.

8. Contexto ou história de fundo:

“...uma cientista explorando ruínas subterrâneas em busca de medições e descobertas...”. Aqui gera - se narrativa implícita para dar coerência à cena, Evidencia como o *prompt* sugere uma narrativa implícita, conferindo sentido e propósito à cena. O texto indica quem é o personagem e qual é sua missão, permitindo à IA inferir um enredo de investigação e exploração. Essa descrição cria coerência narrativa, mostrando que as ações da cientista caminhar cuidadosamente, usar equipamentos e observar o ambiente fazem parte de um objetivo maior.

9. Movimentos de câmera:

“...*close-up* inicial no rosto concentrado, câmera com *zoom* lento para frente, depois transição para uma tomada ampla lateral...”. Aqui define - se o dinamismo e movimentação de câmera. evidencia como o *prompt* orienta a trajetória da câmera para enriquecer a narrativa visual. A descrição indica claramente como a câmera deve se comportar: o *close-up* aproxima o espectador da expressão e concentração da cientista, o *zoom* lento para frente aumenta a intensidade e o envolvimento, e a tomada ampla lateral amplia o campo de visão, mostrando detalhes do ambiente e da interação da personagem com ele.

10. Elementos temporais:

“...caverna subterrânea iluminada por tochas durante a noite, com umidade visível na caverna, luz suave das tochas refletindo nas paredes e poças de água...” . Isto define horário e iluminação com precisão, Especifica o momento do dia e as condições de iluminação, contribuindo para a atmosfera da cena. A descrição define claramente o horário (noite) e a iluminação ambiente (luz suave das tochas refletindo nas superfícies úmidas), criando um clima de mistério, tensão e descoberta.

Essas instruções técnicas aprimoram significativamente a qualidade da produção, permitindo que a IA reproduza uma estética cinematográfica profissional.

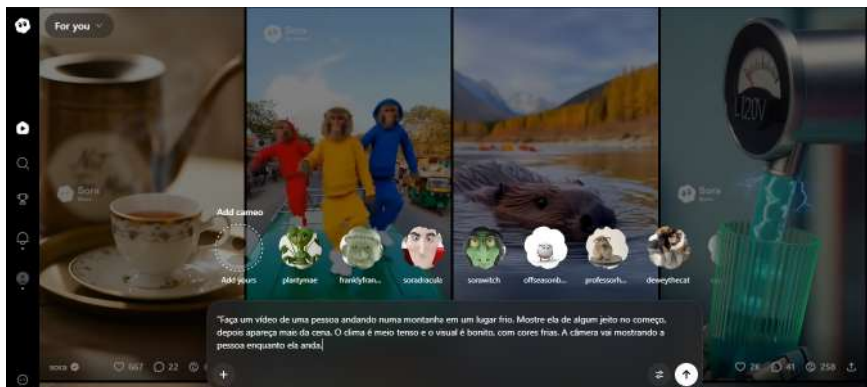
Em síntese, os exemplos de *prompts* analisados ilustram boas práticas na elaboração de comandos para geração de vídeos por IA, ao combinar clareza, detalhamento técnico, ambientação precisa e elementos emocionais, garantindo resultados visuais mais consistentes, envolventes e de alto impacto.

3.3 Exemplo de *prompt* ruim e ineficiente

Um *prompt* foi inserido na plataforma Sora para demonstrar o funcionamento do modelo e ilustrar como a descrição textual influencia o resultado visual gerado. O comando utilizado, apresentado na Figura 9 a seguir, descreve uma cena simples, solicitando que a inteligência artificial produza um vídeo de uma pessoa caminhando em uma montanha fria, com progressão de enquadramento, atmosfera tensa e estética baseada em cores frias. O *prompt*

empregado foi o seguinte: “Faça um vídeo de uma pessoa andando numa montanha em um lugar frio. Mostre ela de algum jeito no começo, depois apareça mais da cena. O clima é meio tenso e o visual é bonito, com cores frias. A câmera vai mostrando a pessoa enquanto ela anda.”

Figura 9: *Exemplo 3.*



Fonte: <<https://openai.com/pt-BR/index/sora/>>

A Figura 10 a seguir apresenta o resultado gerado pela Sora a partir do *prompt* inserido previamente. Nela, é possível observar a interpretação visual produzida pela inteligência artificial, que traduz a descrição textual em uma cena.

Figura 10: *Resultado do terceiro exemplo.*



Fonte: <<https://openai.com/pt-BR/index/sora/>>

3.4 Análise das Deficiências do *prompt*.

A seguir apresenta-se a análise completa do *prompt* inadequado, relacionando cada trecho às dez características para a construção de comandos eficazes para geração de vídeo, destacando por que cada elemento é considerado insuficiente ou problemático.

1. Clareza e objetividade

Trecho problemático: “uma pessoa”, “de algum jeito”, “um lugar frio”. Faltam informações sobre iluminação, estilo cinematográfico, cor, clima e estética, além de descrições vagas e imprecisas.

Justificativa: A IA não sabe quem aparece, como é o ambiente, qual o estilo visual, nem qual é a intenção criativa, comprometendo a fidelidade da geração.

2. Tipo de tomada

Trecho problemático: “mostre ela de algum jeito no começo”, “a câmera vai mostrando a pessoa”. Não há definição de enquadramento, como *close-up*, plano aberto, *zoom* ou panorâmica.

Justificativa: A ausência de orientação clara pode levar a ângulos incoerentes, pouco cinematográficos ou desconexos.

3. Descrição do personagem e objetos

Trecho problemático: “uma pessoa”. Não há detalhes sobre idade, vestimentas, acessórios ou objetos significativos.

Justificativa: A IA gera personagens genéricos, sem identidade visual, prejudicando consistência e narrativa.

4. Ação

Trecho problemático: “andando”. A ação é vaga, sem intensidade, emoção, objetivo ou contexto.

Justificativa: A IA não sabe se o personagem está subindo, explorando, correndo, escalando ou realizando algo relevante para a cena.

5. Local e ambientação

Trecho problemático: “montanha em um lugar frio”. Não há descrição de neve, vento, condições climáticas, visibilidade ou textura do ambiente.

Justificativa: O cenário torna-se genérico, sem atmosfera cinematográfica ou coerência com a intenção de tensão e desafio.

6. Estética

Trecho problemático: “visual é bonito”. Não há definição de estilo (realista, animado, cinematográfico), nem indicações de lente, cor ou profundidade de campo.

Justificativa: A IA pode misturar estilos ou gerar estética incompatível com o objetivo criativo.

7. Adjetivos descritivos

Trecho problemático: “clima é meio tenso”. O adjetivo “meio” enfraquece a emoção desejada e torna a atmosfera imprecisa.

Justificativa: A IA não capta adequadamente a emoção pretendida, prejudicando a ambientação dramática.

8. Contexto ou história de fundo

Trecho problemático: *nenhum contexto fornecido*. Não há explicação sobre quem é o personagem, por que está ali ou qual o objetivo de sua ação.

Justificativa: Sem contexto narrativo, o vídeo pode parecer desconexo e sem propósito.

9. Movimentos de câmera

Trecho problemático: “a câmera vai mostrando a pessoa”. Não são especificados movimentos como panorâmica, *zoom*, *tilt* ou *travelling*.

Justificativa: A IA pode aplicar movimentos inadequados, bruscos ou nada cinematográficos.

10. Elementos temporais

Trecho problemático: *nenhuma indicação do horário do dia*. O *prompt* apenas menciona “lugar frio”, sem referências a amanhecer, entardecer, noite ou condições luminosas.

Justificativa: A iluminação pode ser aleatória e incoerente, prejudicando a construção da atmosfera.

A criação de vídeos por meio de Inteligência Artificial depende fortemente da qualidade do *prompt* utilizado. Quando o *prompt* não especifica características como personagem, ambiente, estilo visual, iluminação, emoção, movimentos de câmera ou elementos temporais, a IA tende a preencher essas lacunas com base em padrões estatísticos aprendidos durante o treinamento. Isso significa que tudo o que não é explicitamente descrito acaba sendo decidido automaticamente pela IA, e não pelo usuário. De fato, de acordo com a OpenAI, quanto mais descritiva for a legenda ou o comando, melhor será a fidelidade do vídeo gerado, já que a IA usa esse detalhe para guiar sua síntese visual [Brooks *et al.* 2025].

Checklist Único.

O **Checklist Único** oferece um padrão estruturado para a criação de *prompts* eficientes na geração de vídeos por IA, funcionando como um guia para garantir clareza, detalhamento e qualidade visual. Embora outras abordagens sejam possíveis, este modelo organiza os elementos essenciais para alcançar resultados cinematográficos consistentes.

O checklist é dividido em **cinco categorias principais**:

1. Personagem

- Aparência: idade, etnia, vestuário, acessórios, expressões faciais.

- Estado emocional: ex. cansado, confiante, assustado.
- Movimento ou ação: ex. correndo, observando, sentado, lutando.

2. Cenário e Ambiente

- Localização: floresta, cidade futurista, montanha, deserto, etc.
- Hora do dia: manhã, entardecer, noite, madrugada.
- Clima e condições: chuva, sol, neve, vento, tempestade.
- Detalhes sensoriais: vento soprando, folhas caindo, luz refletindo.

2. Cenário e Ambiente

- Localização: floresta, cidade futurista, montanha, deserto, etc.
- Hora do dia: manhã, entardecer, noite, madrugada.
- Clima e condições: chuva, sol, neve, vento, tempestade.
- Detalhes sensoriais: vento soprando, folhas caindo, luz refletindo.

4. Estilo Visual

- Estilo cinematográfico: 35mm, animação, sci-fi realista, noir.
- Paleta de cores: tons frios, quentes, dessaturados, vibrantes.
- Iluminação: luz difusa, luz lateral, contra-luz, luz natural.
- Profundidade de campo: estreita (fundo desfocado) ou ampla.

5. Técnica Cinematográfica

- Movimentos de câmera: *close-up*, panorâmica, *travelling*, *zoom* lento.
- Composição: regra dos terços, enquadramento central, plano médio.
- Ângulo da câmera: ângulo normal, *plongée*, *contra-plongée*, *over-the-shoulder*.
- Transições: corte seco, fade-in/out, câmera lenta, plano sequência.

Em conjunto, o checklist fornece uma estrutura lógica e objetiva para criar *prompts* que geram vídeos com **narrativa coerente, estética profissional e impacto emocional**, facilitando o trabalho da IA e elevando a qualidade do resultado final.

A construção de *prompts* bem elaborados é fundamental para maximizar o potencial das inteligências artificiais na geração de vídeos. Instruções precisas e detalhadas permitem que o modelo compreenda a intenção do criador, resultando em cenas mais realistas, coerentes e visualmente impactantes. Um bom *prompt* não apenas orienta a IA, mas também garante que o produto final transmita de forma eficaz narrativa, emoção e estilo desejados.

3.5 Considerações Finais.

O presente trabalho teve como objetivo analisar, por meio de um estudo de caso, as boas práticas na construção de *prompts* para geração de vídeos utilizando o modelo multimodal Sora. A partir da fundamentação teórica sobre LLMs, engenharia de *prompts* e geração audiovisual por Inteligência Artificial, foi possível compreender a relevância dos elementos textuais que influenciam diretamente a qualidade dos resultados produzidos.

A análise demonstrou que a elaboração de *prompts* bem estruturados, contendo descrições detalhadas de cenário, personagem, estética visual, movimentos de câmera e elementos temporais, é fundamental para garantir vídeos mais coerentes, realistas e alinhados com a intenção criativa do usuário. Ao mesmo tempo, verificou-se que instruções vagas, genéricas ou incompletas levam a resultados inconsistentes, reforçando a importância da clareza, objetividade e detalhamento.

Por meio do estudo de caso realizado, foi possível observar, na prática, como a remoção ou simplificação de partes essenciais do *prompt* afeta negativamente a fidelidade visual do vídeo gerado. Dessa forma, conclui-se que a engenharia de *prompts* constitui uma competência central para a utilização eficiente de modelos generativos de vídeo, tornando-se uma habilidade indispensável para pessoas, profissionais e pesquisadores que pretendem explorar o potencial da IA na produção audiovisual.

Este estudo contribui para a compreensão de como a formulação textual impacta a performance de ferramentas como a Sora, oferecendo diretrizes claras e exemplos práticos que auxiliam no desenvolvimento de *prompts* mais eficientes e tecnicamente consistentes.

3.6 Sugestões Para Trabalhos Futuros.

A partir da análise realizada neste trabalho, é possível identificar diversas oportunidades para aprofundamento e expansão do estudo relacionado à geração de vídeos por Inteligência Artificial e à elaboração de *prompts* eficientes. A seguir, são apresentadas sugestões que podem orientar pesquisas futuras:

- **Avaliação comparativa entre diferentes modelos de geração de vídeo.** Investigar o desempenho de outras ferramentas multimodais, como Gen-3, Veo ou modelos open-source, comparando aspectos como fidelidade visual, consistência temporal, interpretação de *prompts* e limitações técnicas.

- **Criação de um dataset anotado de prompts e vídeos gerados.** Elaborar um conjunto de dados contendo *prompts* classificados segundo suas características (clareza, estética, contexto, nível de detalhamento, etc.), junto aos vídeos produzidos, possibilitando pesquisas quantitativas e treinamento de sistemas avaliativos.
- **Análise da percepção do usuário em relação à qualidade dos vídeos.** Realizar estudos com pessoas de todos os segmentos, avaliando como diferentes públicos interpretam a qualidade do vídeo gerado, o realismo da cena e a precisão narrativa.
- **Investigação sobre limitações éticas e mecanismos de detecção de inconsistências.** Analisar riscos relacionados a deepfakes, manipulação visual e desinformação, além de propor diretrizes ou ferramentas que auxiliem na identificação de conteúdos sintéticos.
- **Automação da criação de prompts por meio de IA.** Explorar modelos capazes de gerar *prompts* otimizados automaticamente, a partir de intenções do usuário, esboços visuais, áudio ou outros meios multimodais.
- **Estudos sobre adaptação de prompts para diferentes áreas.** Investigar como a engenharia de prompts pode ser aplicada em cenários específicos, como educação, publicidade, cinema independente, redes sociais ou acessibilidade digital.
- **Análise aprofundada do impacto social da popularização da geração de vídeos por IA.** Examinar como pessoas de diferentes segmentos utilizam essas ferramentas, identificando barreiras de acesso, desafios culturais, oportunidades de inclusão digital e impactos no mercado de trabalho.

Essas direções futuras não apenas ampliam a compreensão sobre o funcionamento dos modelos multimodais, mas também contribuem para o desenvolvimento de práticas mais seguras, eficientes e acessíveis no uso de Inteligência Artificial para criação audiovisual. Estudos posteriores podem complementar e aperfeiçoar os resultados aqui discutidos, fortalecendo o campo da engenharia de *prompts* e suas aplicações práticas.

Referências Bibliográficas

- [Brooks *et al.* 2025] BROOKS, T. *et al.* Video generation models as world simulators. 2025. Acesso em: 3 out. 2025. Disponível em: <<https://openai.com/research/video-generation-models-as-world-simulators>>.
- [Ekin 2023] EKIN, S. *Prompt Engineering For ChatGPT: A Quick Guide To Techniques, Tips, And Best Practices*. 2023. TechRxiv.
- [Google 2023] GOOGLE. *O Bard está disponível em português do Brasil*. 2023. <<https://blog.google/intl/pt-br/produtos/o-bard-esta-disponivel-em-portugues-do-brasil/>>. Acesso em: 30 nov. 2025.
- [Hadi *et al.* 2025] HADI, M. U. *et al.* Large language models: A comprehensive survey of its applications, challenges, limitations, and future prospects. feb 2025.
- [Inc. 2025] INC., A. *Escreva solicitações de texto eficazes para geração de vídeo*. 2025. <<https://helpx.adobe.com/br/firefly/work-with-audio-and-video/work-with-video/writing-effective-text-prompts-for-video-generation.html#clear-and-descriptive>>. Acesso em: 03 out. 2025.
- [Liu e Özsu 2009] LIU, L.; ÖZSU, M. T. Video. In: *Encyclopedia of Database Systems*. [S.l.]: Springer, 2009.
- [Liu *et al.* 2023] LIU, P. *et al.* Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, v. 55, n. 9, p. 1–35, 2023.
- [Meta 2024] META. *Meta Llama 3.1*. 2024. <<https://ai.meta.com/blog/meta-llama-3-1/>>. Acesso em: 30 nov. 2025.
- [Mienye *et al.* 2025] MIENYE, I. D. *et al.* Large language models: an overview of foundational architectures, recent trends, and a new taxonomy. *Discover Applied Sciences*, v. 7, p. 1027, 2025.
- [Netland 2025] NETLAND, T. Comparing human-made and ai-generated teaching videos: An experimental study on learning effects. *Computers & Education*, Elsevier, 2025. Indexed in Scopus and Web of Science. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360131524001787>>.
- [OpenAI 2025] OPENAI. *ChatGPT*. 2025. <<https://openai.com/pt-BR/index/chatgpt/>>. Acesso em: 30 nov. 2025.
- [Singh 2023] SINGH, A. A survey of ai text-to-image and ai text-to-video generators. In: *2023 4th International Conference on Artificial Intelligence, Robotics and Control (AIRC)*. [S.l.]: IEEE, 2023. p. 32–36.
- [Vargas, Rocha e Freire 2007] VARGAS, A.; ROCHA, H. V. da; FREIRE, F. M. P. Promídia: produção de vídeos digitais no contexto educacional. *Renote*, v. 5, n. 2, 2007.

[Wang *et al.* 2024] WANG, Z. *et al.* History, development, and principles of large language models—an introductory survey. *arXiv preprint arXiv:2402.06853*, 2024.

[Xu *et al.* 2024] XU, T. *et al.* From recorded to ai-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology*, Wiley, v. 56, n. 4, p. 1463–1487, 2024. Indexed in Scopus and Web of Science. Disponível em: <<https://bera-journals.onlinelibrary.wiley.com/doi/10.1111/bjet.13530>>.

[Yu *et al.* 2024] YU, T. *et al.* Barriers to industry adoption of ai video generation tools: A study based on the perspectives of video production professionals in china. *Applied Sciences*, v. 14, n. 13, p. 5770, 2024.

	INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DA PARAÍBA
	Campus Campina Grande - Código INEP: 25137409
	R. Tranquílino Coelho Lemos, 671, Dinamérica, CEP 58432-300, Campina Grande (PB)
	CNPJ: 10.783.898/0003-37 - Telefone: (83) 2102.6200

Documento Digitalizado Ostensivo (Público)

trabalho de conclusão de curso

Assunto:	trabalho de conclusão de curso
Assinado por:	Washington Albuquerque
Tipo do Documento:	Anexo
Situação:	Finalizado
Nível de Acesso:	Ostensivo (Público)
Tipo do Conferência:	Cópia Simples

Documento assinado eletronicamente por:
■ **Washington Cipriano Silva Albuquerque, DISCENTE (202321210013) DE TECNOLOGIA EM TELEMÁTICA - CAMPINA GRANDE**, em 29/12/2025 15:31:38.

Este documento foi armazenado no SUAP em 29/12/2025. Para comprovar sua integridade, faça a leitura do QRCode ao lado ou acesse <https://suap.ifpb.edu.br/verificar-documento-externo/> e forneça os dados abaixo:

Código Verificador: 1718051
Código de Autenticação: d9d048b493

