Effective Strategies for Large Language Model Video Prompt Engineering: A Comprehensive Technical Report

Executive Summary

1.1 Research Objective and Core Findings

This report presents a comprehensive investigation into effective strategies for creating Large Language Model (LLM) video prompts, with the primary objective of developing a systematic, evidence-based framework for enhancing interaction with video generation APIs. The analysis synthesizes findings from academic literature, industry best practices, and expert-level technical guides. The central finding of this research is that optimal video generation necessitates a paradigm shift away from simple, conversational requests towards a structured, cinematic briefing model. This approach treats the LLM not as a conversationalist but as a virtual production crew, requiring precise, technical, and visually grounded instructions to achieve high-fidelity, controllable, and engaging video outputs.

1.2 The Universal Prompt Framework

The investigation reveals a consensus around a modular and hierarchical prompt structure as the primary driver of output quality and creative control. This framework prioritizes elements in a logical sequence, typically beginning with the overall framing and moving to granular details: Shot Type > Subject > Action > Setting > Camera Behavior > Lighting > Style. The order of these components is not arbitrary; it directly influences the model's processing and adherence, with elements placed earlier in the prompt receiving greater weight. The

JavaScript template provided in the initial research scope serves as a powerful, production-ready implementation of this universal framework, codifying best practices into an actionable tool.

1.3 Key Recommendations

Based on the extensive analysis, this report puts forth several key recommendations for practitioners. First is the adoption of a "director's mindset," which involves leveraging a precise lexicon of cinematic terminology to guide the AI model. Second is the implementation of an iterative, feedback-driven workflow. Prompt engineering is not a single-shot process but a cycle of design, generation, evaluation, and refinement. This can range from simple A/B testing to more sophisticated Human-in-the-Loop (HITL) evaluation and, ultimately, contributing to model alignment through Reinforcement Learning from Human Feedback (RLHF). Finally, the report highlights the strategic necessity of audience-centric prompt design, where audience personas are systematically translated into specific visual and stylistic cues within the prompt to ensure the final video resonates with its intended viewers.

1.4 Report Structure Overview

The report is structured to guide the reader from foundational principles to advanced applications. It begins by deconstructing the anatomy of a high-performance prompt, establishing the core tenets of clarity and specificity. It then provides a detailed lexicon of cinematic language required for effective "AI direction." A comparative analysis of various prompt structures follows, culminating in a detailed case study of a best-in-class template. The report then explores systematic workflows for prompt refinement, model-specific nuances, and the critical process of translating audience personas into visual styles. Finally, it presents a multi-faceted framework for evaluating prompt-generated video, concluding with a synthesized, universal prompt template and actionable recommendations for implementation.

The Anatomy of a High-Performance Video Prompt: Core Principles of Clarity and Specificity

2.1 From Problem Formulation to Prompt Engineering

The genesis of an effective video prompt precedes the act of writing. It begins with a clear and rigorous definition of the creative objective, a process known as problem formulation. This involves delineating the focus, scope, and boundaries of the intended video clip.[1] Prompt engineering, defined as the art of crafting the optimal textual input by selecting the appropriate words, phrases, and formats, is the subsequent step that translates this well-defined problem into a set of instructions the AI can execute.[1] A meticulously engineered prompt cannot salvage a poorly formulated creative goal; therefore, the initial and most critical step is to define the intent and constraints of the desired output in a single, clear sentence before any prompt is drafted.[2] This principle, "Focus More on Problems, Less on Prompts," underscores that the strategic and creative work of defining the video's purpose is a non-negotiable prerequisite for technical success.[1]

2.2 The Primacy of Specificity and Visual Grounding

Generative video models are not conceptual interpreters; they are visual synthesizers that thrive on concrete, descriptive detail. Vague, abstract, or purely emotional language is a primary source of failure, leading to generic or irrelevant outputs.[5] The models require prompts that are direct, easily understood, and descriptive of a visual scene.[7] This principle of specificity operates at every level of the prompt. For instance, a description like "a weathered oak table" provides the model with specific textural and material cues that are absent in a generic phrase like "a nice table".

Crucially, prompts must describe what the camera *sees*, not what a character *feels*. Emotional intent must be translated into visually grounded cues. A prompt for a "sad man" is weak because sadness is an internal state. A strong prompt would describe the physical manifestation of that emotion: "a man with slumped shoulders, his face partially obscured by the collar of his coat, walks slowly down a wet, reflective city street at night".[5] Similarly, conceptual actions must be replaced with physical ones. "A man hacking into the mainframe" is an abstract narrative idea. "A close-up of a man's hands vigorously typing on a keyboard with glowing green characters in a dark server room" is a set of visual instructions the model can render.[7] This commitment to visual grounding is the most fundamental characteristic of an effective video prompt.

2.3 The "One Clip, One Action" Principle

A recurring and critical best practice across all leading video generation platforms is the principle of limiting each prompt to a single, clear, and specific action.[3] Prompts that attempt to describe a complex narrative sequence or multiple simultaneous actions within a single generation request consistently result in a degradation of output quality, including visual artifacts, loss of character consistency, and a failure to adhere to the prompt's instructions.[5] Overloading a prompt with too many conflicting ideas—such as multiple actions, scene changes, or incompatible styles—overwhelms the model's capacity to maintain coherence.[5]

This is not merely an empirical best practice but a direct consequence of the underlying architecture of video diffusion models. These models function by extending image generation capabilities into the temporal dimension, a process that involves denoising a sequence of frames while maintaining object identity, style, and physical plausibility over time.[13] The core challenge is managing spatio-temporal consistency.[12] Each subject and action introduced in a prompt adds a new vector of motion and a new set of required consistencies that the model must track across every frame of the generated clip. This exponentially increases the computational complexity of the denoising process. When faced with multiple independent trajectories of motion and attribute binding within a short time window (typically 4-8 seconds), the model's attention mechanisms can fail to allocate resources effectively. This can lead to flickering, where an object's attributes change mid-clip, or a complete breakdown in motion coherence.[12] By constraining the prompt to a single subject performing a single, well-defined action, the prompt engineer effectively simplifies the optimization problem the model needs to solve. This constraint allows the model to dedicate its full computational capacity to rendering that one action with the highest possible fidelity, resulting in a more stable, coherent, and prompt-adherent video. Therefore, complex narratives should always be deconstructed into a sequence of single-action clips, generated individually and edited together in post-production.

Mastering the Virtual Camera: A Lexicon of Cinematic and Technical Prompting

3.1 The "AI as Film Crew" Paradigm

The most effective mental model for interacting with advanced text-to-video LLMs is to abandon conversational language and adopt the role of a director briefing a film crew.[3] The prompt should function as a "concise director's note" or a shot description for a "cinematographer who has never seen your storyboard".[3] This paradigm shift necessitates fluency in the professional language of cinematography, as these models are trained on vast datasets of film and video content, making them highly responsive to industry-standard terminology.[10] Using this technical lexicon provides a level of precision and control that is unattainable with plain-language descriptions.

3.2 Essential Cinematic Vocabulary

A robust video prompt leverages a specific vocabulary to control the visual output. The following terms are fundamental for directing the virtual camera and production team.

Camera Shots

The shot type defines the framing of the scene and should be one of the first elements specified in the prompt to establish the foundational composition.[10]

- **Wide Shot (WS) / Establishing Shot:** Captures a large area to provide context and set the scene. Example: "Wide establishing shot of a bustling medieval marketplace, showcasing vibrant stalls and busy shoppers." [16]
- **Medium Shot (MS):** Frames the subject from approximately the waist up, balancing detail with environmental context. Example: "Medium shot of a chef meticulously plating a dish in a brightly lit kitchen." [10]
- **Close-Up (CU):** Tightly frames a person or object, highlighting detail and emotion. Example: "Close-up on the protagonist's face, capturing a single tear rolling down her cheek." [16]
- **Extreme Close-Up (ECU):** Focuses on a very small detail, such as a person's eyes or the texture of an object. Example: "Extreme close-up of a blinking reptilian eye, with city lights reflected in the pupil." [10]
- **Over-the-Shoulder (OTS):** Frames the shot from behind a character, looking at another character, often used in conversations. Example: "Over-the-shoulder shot of a mentor

observing an apprentice's work on a complex machine." [16]

- **High-Angle Shot:** The camera is positioned above the subject, looking down, which can make the subject appear vulnerable or small. Example: "High-angle shot revealing a lone hiker standing in a vast, empty desert." [16]
- **Low-Angle Shot:** The camera is positioned below the subject, looking up, which can make the subject appear powerful or imposing. Example: "From a low-angle shot, show the towering figure of a CEO addressing a crowd." [10]
- **Dutch Angle:** The camera is tilted on its axis, creating a sense of unease, tension, or disorientation. Example: "Dutch angle shot of a dark alleyway, conveying a sense of danger." [16]
- **Bird's-Eye View:** Looks directly down on the scene from above, offering an omniscient perspective. Example: "Bird's-eye view of city streets, bustling with midday traffic." [16]

Camera Movements

Describing camera movement brings dynamism to a scene. For optimal results, prompts should be limited to a single, clear camera move per clip.[11]

- **Static / Tripod Shot:** The camera does not move. Example: "Static shot of a serene lake at sunrise, with mist rising from the water." [10]
- **Pan:** The camera pivots horizontally from a fixed point. Example: "Pan across the city skyline at sunset, slowly revealing a hidden rooftop garden." [16]
- **Tilt:** The camera pivots vertically from a fixed point. Example: "Tilt downward from the canopy of a giant redwood tree to the forest floor, unveiling a hidden cabin." [16]
- **Dolly (In/Out):** The entire camera moves forward or backward, often on a track, changing the perspective and depth. Example: "Slow dolly in toward an antique music box on a table, intensifying curiosity as it begins to open." [16]
- **Truck (Left/Right):** The entire camera moves horizontally, parallel to the subject. Example: "Truck right, following a dancer as she moves across a stage." [16]
- **Crane / Jib Shot:** The camera is mounted on a crane, allowing for sweeping vertical movements. Example: "Crane shot rising up from a character's face to reveal the massive army assembled behind them." [16]
- **Handheld:** The camera is held by the operator, creating a sense of immediacy or documentary-style realism. Example: "Handheld shot following a journalist as she navigates through a chaotic protest." [16]

Lighting

Lighting is a powerful tool for establishing mood and tone. Prompts should specify the source, direction, and quality of the light.

- **Source & Quality:** Examples include "soft, golden-hour sunlight streaming through windows" [17], "dramatic side lighting creating deep shadows" [17], or "soft, diffused lighting from an overcast sky".[8]
- **Named Techniques:** Referencing professional lighting setups like "three-point lighting," "Rembrandt lighting," or "rim lighting" can provide the model with precise instructions.[17]
- **Color & Temperature:** Specify the color and emotional feel of the light, such as "neon city lights reflecting on wet streets" [17] or "a warm, amber glow from a candlelit room". A highly specific prompt might detail the entire setup: "soft window light with a warm lamp fill and a cool edge from the hallway".[11]

Lens and Depth of Field (DOF)

Specifying the virtual lens and its properties adds a final layer of professional control, influencing perspective and focus.

- **Lens Type:** Mentioning a focal length can guide the model's composition. Examples: "shot with an 85mm lens" for portrait-like compression, or "wide-angle 24mm lens" for a more expansive, distorted perspective.[3]
- **Depth of Field (DOF):** This controls how much of the scene is in focus. "Shallow depth of field" (or "shallow DOF") blurs the background, isolating the subject and creating a cinematic look. "Deep focus" keeps both foreground and background sharp. "Rack focus" refers to the action of shifting focus from one subject to another within the shot.[11]

3.3 Style and Aesthetic Modifiers

Beyond technical specifications, style modifiers guide the overall look and feel of the video. However, generic terms like "cinematic" are weak unless substantiated with specific details.[3] More effective prompts use:

- **Film Stock & Format References:** Terms like "shot on 35mm film," "16mm black-and-white," "IMAX-scale," or "VHS home video aesthetic" provide strong visual anchors.[8] Adding "film grain texture" can further enhance this effect.[17]
- **Artistic & Genre References:** Citing specific aesthetic movements ("Art Nouveau,"

"Surrealism," "Cyberpunk") or film genres ("Film Noir") gives the model a rich visual library to draw from.[8] Direct film references, such as "neon reflections on wet pavement, inspired by Blade Runner," are also highly effective.[10]

- **Technical Output Specifications:** For precise control over the final product, prompts or API parameters should include the desired aspect ratio (e.g., "16:9 aspect ratio" for YouTube, "9:16 vertical format" for TikTok) and frame rate (e.g., "24fps" for a filmic look).[15]

A Comparative Analysis of Prompt Structures: From Simple Statements to Modular Frameworks

4.1 The Evolution of Prompt Formats

The format and structure of a video prompt have a significant impact on its efficacy. The evolution of prompting techniques reflects a growing understanding of how video LLMs process textual information. Initial or novice approaches often rely on simple, conversational statements, such as "a car driving through a city".[17] While easy to formulate, these prompts cede creative control to the model and typically yield generic, amateurish results.[7] Models like RunwayML explicitly advise against such conversational or command-based phrasing (e.g., "Can you please make me a video of...") in favor of direct, descriptive statements.[7]

From an academic perspective, the prompt formats used in text-to-video generation fall under the category of "hard prompts," which are human-readable instructions designed to elicit a specific task-oriented response. This contrasts with "soft prompts," which are non-human-readable learned embeddings or vectors.[23] The most effective hard prompts for video generation have evolved from simple declarative sentences into more complex, structured frameworks that organize information hierarchically.

4.2 Analysis of Structured Prompt Templates

As prompt engineering has matured, several structured templates have emerged from industry best practices and expert analysis. While they vary slightly in their terminology, they share a common principle: deconstructing a creative vision into discrete, logical components that align with the model's interpretive capabilities.

- **Structure 1: The Layered Framework:** This approach organizes prompt elements in a descending order of importance, starting with the broadest framing and moving to the finest details. A common implementation is: Shot Type > Subject & Action > Camera Movement > Lighting & Mood > Technical Details.[17] This structure ensures that the most critical compositional decisions are established first.
- **Structure 2: The Modular Slots Approach:** This popular and highly effective structure treats the prompt as a collection of modular "slots" that are filled in to build a complete scene description. The typical order is: Subject + Action + Setting + Style + Camera + Lighting + Motion Physics + Audio.[3] This method is systematic and ensures all key cinematic variables are considered.
- **Structure 3: The RunwayML Recommended Format:** RunwayML's official documentation suggests a concise, three-part structure: [camera movement]: [establishing scene]. [additional details]..[7] This format is particularly effective for its clarity and directness, prioritizing motion and scene context.
- **Structure 4: The V.I.D.E.O. Framework:** This conceptual framework organizes the prompt around five key pillars: **V**isuals (scene, camera), **I**ntent (goal, focus), **D**etails (lighting, props), **E**motion (mood), and **O**utput (duration, aspect ratio).[24] While other frameworks embed intent and emotion within visual descriptions, this model makes them explicit components, which can be a useful creative aid for prompters.

Across these structures, a clear pattern emerges: the most successful prompts are not freeform prose but rather a systematic assembly of technical and descriptive components. The consensus points towards a hierarchical structure where framing (shot type), subject, and action are defined first, followed by modifiers for camera, lighting, and style.

4.3 Case Study: Deconstruction of the User-Provided JavaScript Template

The generateVideoPrompt function provided in the research objective represents a state-of-the-art synthesis of these structured prompting principles. It is more than a simple text formatter; it is an engineered system for producing high-quality, reliable video prompts by guiding the user and mitigating common errors.

Component Breakdown and Analysis

- **Core Structure:** The template's primary instruction, in/at,,,, is a direct and robust implementation of the modular, layered framework identified as a best practice across multiple expert sources.[3] It is concise, logical, and covers all essential cinematic elements.
- **Hierarchical Ordering:** The list of REQUIRED ELEMENTS (in order of importance) is a critical feature. It codifies the principle that the order of elements in a prompt dictates their priority in the AI's processing. By placing Shot type first, the template ensures the fundamental framing of the scene is established before more granular details are considered. This is a sophisticated, actionable instruction that directly improves prompt adherence.
- **Rule-Based Error Prevention:** The WRITING RULES and AVOID sections function as a built-in linter for prompt quality. The rule to limit word count to 100-150 words addresses the tendency for models to perform better with shorter, more focused prompts. The emphasis on "film language" and "one main action" directly reinforces the "AI as Film Crew" paradigm and the "One Clip, One Action" principle, preventing the common pitfalls of vagueness and prompt overloading.[5] The AVOID list explicitly warns against known failure modes like negative phrasing, which models like RunwayML handle poorly [7], and complex narratives, which are better handled with multiple clips.
- **Advanced Features for Iteration:** The inclusion of TECHNICAL SPECS and ALTERNATIVE APPROACHES elevates the template from a single-prompt generator to a tool for systematic experimentation. The ALTERNATIVE APPROACHES section encourages the user to think iteratively, exploring variations in camera angle or lighting. This aligns with advanced prompt engineering workflows that rely on A/B testing and iterative refinement to optimize results.[25]

Cognitive Scaffolding for the Prompt Engineer

The true value of this template lies in its function as a form of "cognitive scaffolding." It guides the user through the complex process of translating an abstract creative idea into the precise, machine-readable format that video LLMs require. A user may begin with a vague concept like "a video about Abraham Lincoln." The template's rigid structure immediately forces a series of critical, disambiguating decisions. The `` requirement compels the user to choose between a wide shot of a speech or a close-up of writing. The need for "2-3 distinctive visual details" prevents a generic, stock-image Lincoln and grounds the character in specifics. The "one clear, specific action" rule is a hard guardrail against the natural human tendency to describe a story rather than a single visual moment, thereby preventing the primary cause of generation failure.[5] By systematically leading the user through each component—Camera, Lighting, Style—the template ensures that no critical cinematic element is overlooked. In this way, the template is not merely a passive formatter; it is an active, interactive guide that operationalizes the "director's mindset" and preemptively solves the most common user errors.

The Iterative Workflow: A Framework for Systematic Prompt Refinement and Feedback Integration

5.1 The Principle of Iterative Prompt Development

Crafting a perfect prompt in a single attempt is a rare exception, not the rule. The consensus among experts is that prompt engineering is an inherently iterative process that requires extensive experimentation to achieve optimal results.[27] The standard workflow mirrors the scientific method: formulate a hypothesis (the prompt), conduct an experiment (generate the video), analyze the results (evaluate the output), and refine the hypothesis (modify the prompt).[25] This cycle begins with a clear, specific base prompt that outlines the core idea. Based on the initial output, the prompter makes small, targeted adjustments to refine the result and move it closer to the desired outcome.[28] For instance, an effective iteration strategy involves changing only one or two elements at a time—such as clarifying the lens type or adjusting a line of dialogue—to precisely isolate the impact of each modification on the final video.[3] This methodical approach is fundamental to mastering prompt control.

5.2 Human-in-the-Loop (HITL) for Prompt Science

To move beyond ad-hoc trial and error, a more structured approach known as Human-in-the-Loop (HITL) can be employed. This methodology aims to transform prompt engineering from a subjective art into a systematic science by introducing rigor, objectivity, and replicability into the evaluation process.[29] A formal HITL workflow for prompt validation involves using multiple human assessors to review AI-generated outputs against a predefined set of criteria, often called a "codebook".[29]

The process unfolds in several steps:

1. **Generation:** A set of videos is generated using the prompt being tested.

2. **Independent Assessment:** Multiple human assessors independently evaluate each video according to the criteria in the codebook. For video generation, these criteria might include "Prompt Adherence" (does the video accurately reflect the prompt?), "Visual Quality" (are there artifacts or glitches?), "Motion Coherence" (is the movement natural and consistent?), and "Aesthetic Alignment" (does the style match the prompt's intent?).
3. **Reliability Measurement:** The level of agreement between assessors is calculated using a metric like Inter-Coder Reliability (ICR). A high level of agreement indicates that the criteria are objective and consistently understood.[29]
4. **Deliberation and Refinement:** If agreement is low, the assessors discuss their disagreements. This crucial step helps to identify ambiguities in the evaluation criteria or flaws in the prompt itself, leading to revisions of either the codebook or the prompt. This process is repeated until a sufficient level of agreement is achieved.[29]

This systematic approach fosters objectivity by mitigating individual biases and creates a transparent, documented process that allows others to validate and replicate the findings.[29]

5.3 Reinforcement Learning from Human Feedback (RLHF) for Model Alignment

While HITL refines the prompt, Reinforcement Learning from Human Feedback (RLHF) is a more advanced technique that uses human feedback to refine the model itself.[32] RLHF aims to align the model's behavior more closely with human preferences, reducing the need for perfectly engineered prompts to achieve desired outcomes.[33]

The RLHF process for video generation involves two key stages:

1. **Training a Reward Model:** Humans are presented with pairs of videos generated from the same prompt and asked to choose which one they prefer. This creates a large dataset of human preference data (e.g., for prompt X, Video A is better than Video B).[32] A separate "reward model" is then trained on this dataset to predict which video a human would likely prefer. This model learns to assign a higher score to videos with desirable characteristics (e.g., smoother motion, better prompt adherence).[32]
2. **Fine-Tuning the Generative Model:** The main text-to-video model is then fine-tuned using reinforcement learning. The reward model provides the "reward" signal. The generative model is encouraged to produce videos that receive a high score from the reward model, effectively steering its output towards what humans have indicated they prefer.[32]

Recent research is actively applying this technique to the text-to-video domain. Initiatives are underway to build large-scale video preference datasets (like VIDEOPREFER) and train specialized video reward models (like VideoReward) to specifically improve aspects like motion quality, visual fidelity, and alignment with textual prompts.[13]

5.4 A/B Testing for Empirical Prompt Optimization

For practitioners in a production environment, A/B testing provides a robust and empirical method for optimizing prompts.[26] This strategy involves systematically comparing the performance of two or more prompt variants against each other to determine which is more effective at achieving a specific goal.[26]

The methodology for A/B testing prompts includes the following steps:

1. **Define Objectives and Metrics:** Establish a clear goal for the test (e.g., increase visual quality, improve prompt adherence) and define the key metrics to measure success. These can include automated scores, user satisfaction ratings, or human evaluations of relevance and correctness.[26]
2. **Design Variants:** Create two or more versions of a prompt. Each variant should test a specific hypothesis (e.g., Prompt A uses a "35mm lens" description, while Prompt B uses an "85mm lens" description to see which produces a more desirable aesthetic).
3. **Randomize and Deploy:** Randomly assign the prompt variants to different generation requests to ensure an unbiased comparison.
4. **Collect and Analyze Data:** Collect the performance data for each variant and analyze the results to identify any statistically significant differences.
5. **Iterate and Roll Out:** Select the winning prompt variant and deploy it to production. The learnings from the test should be documented to inform future prompt design.[26]

The progression from simple manual iteration to structured HITL evaluation, large-scale RLHF, and empirical A/B testing represents a clear trajectory towards increasing automation, scale, and rigor in prompt optimization. This evolution suggests a future in which the role of the prompt engineer shifts from being a hands-on "prompt crafter" to a "prompt system designer." In this new role, the focus will be less on tuning individual prompts and more on designing and managing the automated feedback loops, evaluation frameworks, and reward models that continuously improve the AI system's performance at scale. The intelligence is gradually being transferred from the prompt itself to the underlying model and the systems that align it.

Navigating the Generative Video Ecosystem:

Model-Specific Nuances and Pitfalls

6.1 Overview of Leading Models

The field of text-to-video generation is dominated by a handful of powerful models, each with its own architectural strengths, training data biases, and prompting idiosyncrasies. The leading models as of this report include OpenAI's Sora, Google's Veo, RunwayML's series of models, Kuaishou's Kling, and Luma Labs' Dream Machine. While the core principles of cinematic and structured prompting are broadly applicable, achieving optimal results requires an understanding of each model's specific "dialect" and preferred interaction style.

6.2 Model-Specific Prompting Guides

OpenAI Sora

Sora's prompting philosophy is firmly rooted in the "cinematographer briefing" model. It responds well to detailed, descriptive prose that sets a scene.[11] A key characteristic is the trade-off between prompt length and creative freedom: shorter, simpler prompts grant the model more latitude for interpretation, often leading to surprising results, while longer, more detailed prompts offer greater control and consistency.[11] Best practices for Sora include limiting each clip to a single, clear camera move and one primary subject action. For timing control, describing actions in discrete "beats" or "counts" (e.g., "Actor takes four steps to the window, pauses...") is highly effective. For visual consistency across multiple shots, specifying the quality of light sources and defining a stable color palette with three to five "color anchors" is recommended.[11]

Google Veo

Veo operates on a similar "director's note" paradigm, favoring prompts structured in modular slots (Subject, Action, Setting, Style, etc.).[3] A distinguishing feature of Veo is its native audio generation capability. This requires prompts to include explicit cues for dialogue, sound effects, and ambient soundscapes. The recommended format for dialogue is clear and direct, such as: "Character says: 'Let's move' (no subtitles)".[3] The (no subtitles) directive is crucial for preventing the model from rendering unwanted text on the video.

RunwayML

RunwayML's documentation provides explicit guidance on what to avoid. Prompts should be descriptive, not conversational, command-based, or negative.[7] The model does not respond well to instructive language like "don't show" or "avoid." Instead, users should positively describe the desired scene.[7] The recommended structure is a concise, three-part format: [camera movement]: [establishing scene]. [additional details].[7] When using an image as an input alongside text, the text prompt should focus exclusively on describing the desired *movement*, as the model will derive the visual content from the image itself.[7]

Luma Labs Dream Machine

Dream Machine distinguishes itself by favoring natural, conversational language over rigid, technical structures.[39] It is designed for an iterative workflow within a "board," where the model retains context from previous generations. This allows users to start with a simple idea and progressively refine it through conversational follow-up prompts (e.g., "Add a golden hour glow," "Now include a small cottage in the distance").[39] It also offers advanced, tool-like prompt commands, such as using @character with an uploaded image for character consistency or @style to use an image as a visual reference.[39]

Kling

Kling provides users with a choice between different model versions within its interface, such as a "Turbo" model tuned for speed and creativity and a "Master" model optimized for superior dynamics and strict prompt adherence.[40] This allows users to select the appropriate tool for their specific task. The interface also includes dedicated fields for adding sound effects and background music, separating these elements from the main visual prompt.[40]

6.3 Common Pitfalls and Mitigation Strategies

Despite their differences, all video generation models are susceptible to a common set of prompting errors. Understanding these pitfalls is essential for efficient and effective video creation. The following table synthesizes the most frequent errors identified in the research and provides concrete examples of how to mitigate them. This table serves as a practical, diagnostic guide for prompt engineers. Its value lies in its ability to help users quickly identify failure patterns in their own prompts by comparing them against the "Ineffective Prompt" examples and immediately see the corrected, effective structure. This accelerates the learning process and provides an actionable framework for troubleshooting.

Pitfall	Example of Ineffective Prompt	Example of Improved Prompt & Rationale	Sources
Vagueness / Conceptual Language	"A man hacking into the mainframe."	"Close-up on a man's hands vigorously typing on a glowing green keyboard in a dark server room." (Describes visible action rather than an abstract concept.)	5
Overloading / Multiple Actions	"A woman walks into a cafe, orders a coffee, and sits down by the window as it starts to rain."	*Clip 1:* "Medium shot of a woman in a trench coat pushing open the glass door of a cozy cafe." *Clip 2:* "Close-up of a barista handing a steaming latte to	3

		the woman." (Breaks the narrative into single-action clips to maintain quality and coherence.)	
Negative Phrasing	"A sunny beach with no people."	"A pristine, empty tropical beach with golden sand and clear blue water. The camera is static." (Uses positive phrasing to describe the desired elements of the scene, which models interpret more reliably.)	7
Ungrounded Emotional Cues	"A sad man walking in the city."	"A man with slumped shoulders walks slowly down a wet, reflective city street at night, his face partially obscured by the collar of his coat." (Translates the internal emotion of 'sadness' into observable, physical details and atmospheric context.)	5
Ambiguous Camera/Motion	"The camera moves around the room."	"Slow 360-degree orbital shot around a central oak table in a sunlit library." (Specifies the exact	12

		type, path, and speed of the camera movement, removing all ambiguity.)	

Audience-Centric Prompt Design: Translating Personas into Compelling Visual Styles

7.1 The Importance of Audience Analysis

The technical proficiency of a prompt is only one component of its success; the other is its strategic alignment with the target audience. Creating video content without a deep understanding of the intended viewer is a critical strategic error that can lead to poor engagement and a failure to achieve communication goals, even if the video is technically well-executed.[41] A thorough audience analysis—which involves gathering and interpreting data about audience demographics, interests, preferences, and online behaviors—is an essential prerequisite for crafting content that truly resonates.[43] This analysis informs crucial decisions about content format, tone of voice, and, most importantly for video generation, the specific visual style that will be most appealing and effective for that audience segment.[42]

7.2 Creating and Applying Audience Personas

The primary output of audience analysis is the creation of detailed audience personas (sometimes called "buyer personas"). A persona is a semi-fictional representation of an ideal audience segment, built from real-world data and insights. A comprehensive persona includes demographic information (age, gender, location, income), psychographic information (values, interests, lifestyle), and behavioral data (media consumption habits, online activity).[42]

Generative AI can be a powerful tool in this process. By providing an LLM with research data—such as survey results, interview transcripts, or market reports—practitioners can use a structured prompt to generate detailed, well-articulated personas efficiently.[49] An effective prompt for persona generation should specify the role (e.g., "You are a market researcher"), the task (e.g., "Create a detailed persona"), the context (e.g., "for a B2B SaaS company"), and the desired format (e.g., "list their goals, pain points, and media habits").[49] Once a persona is established, it can be used as a foundational context for all subsequent content-related prompts. By instructing the AI to generate content *for that specific persona*, all outputs become more targeted and relevant.[1]

7.3 Translating Personas into Prompt Elements

The most sophisticated application of audience analysis in prompt engineering is the ability to translate the abstract characteristics of a persona into the concrete, technical elements of a video prompt. This involves mapping demographic and psychographic traits to specific choices in lighting, color, camera movement, and overall aesthetic style. For example, content aimed at a younger, tech-savvy Gen Z audience might favor dynamic, fast-paced visuals, whereas content for a B2B professional audience would likely benefit from a more stable, clean, and polished aesthetic.[46]

The following table provides a practical framework for this translation process. It bridges the gap between high-level marketing strategy and the technical execution of prompt engineering. By demonstrating how to convert audience traits into specific prompt keywords and cinematic techniques, it makes the goal of audience-centric content creation directly actionable at the API level. This represents a significant step beyond generic prompt advice, offering a strategic layer of control for creating more impactful and resonant video content.

Persona Segment (Demographics/Psychographics)	**Resonant Visual Themes**	**Corresponding Prompt Elements (Style, Lighting, Camera, Color)**	**Sources**
Gen Z (18-24), Tech-Savvy, Prefers Short-Form Content	Dynamic, authentic, visually stimulating, fast-paced	9:16 aspect ratio, handheld camera, fast motion, glitchcore aesthetic, neon color palette, dynamic motion,	[42]

		quick cuts	
Millennials (25-40), Professional, Values Authenticity & Story	Cinematic, clean, nostalgic, high-quality, aspirational	cinematic style shot on 35mm, soft golden-hour lighting, shallow depth of field, slow dolly forward, warm, natural color palette	[45]
Gen X / Boomers (40+), Corporate/B2B, Values Clarity & Trust	Professional, clear, informative, stable, polished	static tripod shot, clean studio lighting, minimalist aesthetic, motion graphics, high-key lighting, brand-aligned color palette	[47]
Luxury Consumers, High-Income, Aspirational	Elegant, sophisticated, moody, dramatic, exclusive	slow motion, macro cinematography, low-key lighting, dramatic contrast, monochromatic or desaturated color palette, smooth, deliberate camera moves	[8]

Measuring Success: A Multi-faceted Evaluation Framework for Prompt-Generated Video

8.1 The Need for Objective Evaluation

Evaluating the output of generative video models is an inherently complex task. A "successful" generation must be assessed along multiple, often subjective, dimensions, including its overall visual quality, aesthetic appeal, alignment with the input prompt, narrative coherence, originality, and lack of harmful biases.[58] To move beyond subjective impressions and enable systematic improvement, a multi-faceted evaluation framework that combines both automated metrics and structured human judgment is required.

8.2 Automated Evaluation Metrics (for Visual Quality & Diversity)

For assessing the fundamental visual fidelity and variety of generated videos, several metrics adapted from the image generation domain are widely used. These metrics compare the statistical properties of generated content to those of a real-world dataset.

- **Fréchet Inception Distance (FID):** FID has become the de facto industry standard for evaluating the quality of generative models. It measures the similarity between the feature distributions of a set of generated images (or video frames) and a set of real images. The features are extracted from a pre-trained Inception neural network. A lower FID score indicates that the generated distribution is closer to the real distribution, signifying higher quality and diversity. It is valued for its strong correlation with human perception of visual quality.[59]
- **Inception Score (IS):** An earlier metric, IS, evaluates two properties: the quality of individual images (are they confidently classified as a recognizable object?) and the diversity of the entire set (do they cover a wide range of classes?). A higher IS is better. While largely superseded by FID, it can still provide complementary information.[59]
- **Precision and Recall for Distributions:** These metrics offer a more nuanced evaluation by separating quality from diversity. **Precision** measures the fraction of generated samples that are realistic (high quality), while **Recall** measures the fraction of the real data distribution that the model is able to generate (high diversity). This pair of metrics is particularly useful for diagnosing model failure modes, such as "mode collapse," where a model generates high-quality but low-variety outputs (high Precision, low Recall).[59]

8.3 Human-Centric Evaluation Metrics (for Prompt Alignment & Engagement)

While automated metrics are useful for assessing pixel-level quality, they cannot evaluate how well a video aligns with the semantic and stylistic intent of a prompt. For this, human evaluation is indispensable.

- **Qualitative Expert Reviews:** This method involves subject matter experts (such as filmmakers, animators, or prompt engineers) assessing the generated videos against a rubric of criteria. These criteria typically include prompt adherence, relevance, accuracy, motion quality, and overall aesthetic appeal. This is a core component of the HITL workflow discussed previously.
- **Human eYe Perceptual Evaluation (HYPE):** HYPE is a more formalized framework for human evaluation. It uses structured experimental protocols, such as time-limited "real vs. fake" identification tasks, to quantitatively measure how convincingly AI-generated content can pass as authentic to human observers.[59]
- **User Engagement Metrics:** Once a video is deployed, its performance can be measured using standard video analytics. Key metrics include **view duration**, **audience retention/drop-off points**, and **interaction rates** (likes, shares, comments).[43] These metrics provide direct, real-world feedback on how engaging and effective the generated content is for its target audience. A high drop-off rate at a specific point in a video, for example, can signal a failure in narrative coherence or visual quality that can be traced back to the prompt.

8.4 T2V-Specific Benchmarks

As the field matures, specialized benchmarks are being developed to test the unique challenges of text-to-video generation. These benchmarks go beyond general quality to probe specific compositional and dynamic capabilities.

- **Compositionality Benchmarks (e.g., T2V-CompBench):** These benchmarks are designed with suites of prompts that specifically test a model's ability to handle complex compositional requests. This includes correctly binding attributes to objects (e.g., "a red cube and a blue sphere"), maintaining spatial relationships ("the cube is on top of the sphere"), and accurately depicting object interactions and motion.[63]
- **Dynamics Evaluation Protocols (e.g., DEVIL):** These protocols focus on evaluating whether the "dynamics" of the generated video match the dynamics described in the prompt. For example, a prompt containing high-action words like "explodes" or "races" should result in a video with high motion and visual change, while a prompt describing a "serene landscape" should yield a low-dynamics video. This measures a model's ability to translate the energy and pacing of language into appropriate visual motion.[64]

A comprehensive evaluation strategy should leverage a combination of these approaches: automated metrics like FID to track baseline visual quality, and structured human evaluations and T2V-specific benchmarks to assess the more nuanced aspects of prompt alignment and compositional accuracy.

The Comprehensive Video Prompt Template: Synthesis, Implementation, and Recommendations

9.1 Synthesis of Findings

The preceding analysis has established a clear and consistent set of principles for effective text-to-video prompt engineering. The core takeaway is the transition from ambiguous, conversational language to a structured, technical, and visually grounded methodology. This "director's mindset" requires specificity in describing subjects and actions, fluency in the language of cinematography to control the virtual camera, and a systematic, iterative workflow to refine and optimize outputs. Furthermore, advanced prompting incorporates a strategic layer of audience awareness, translating persona characteristics into specific aesthetic choices. The most effective prompts are not merely written; they are engineered according to these principles to maximize clarity, control, and impact.

9.2 The Finalized Universal Prompt Template

Synthesizing the best practices from across the research, the user-provided JavaScript template stands as an exemplary foundation. It is here refined and annotated to create a universal, gold-standard template for practitioners. This version incorporates an explicit `` component to accommodate models with native sound generation capabilities like Google Veo, and its internal comments are enhanced with the rationale behind each rule, transforming it from a simple formatter into an educational tool.

JavaScript

```
/**
```

```
 * Generate an optimized, production-ready video prompt for AI video generation models.
 * This template is designed for models like Sora, Veo, RunwayML, Kling, and Luma.
 * It enforces a structured, cinematic approach to maximize prompt adherence and output quality.
 *
 * @param {string} userConcept - The user's core creative idea for the video clip.
 * @returns {string} A formatted and optimized prompt ready for an AI video generation API.
 */
function generateUniversalVideoPrompt(userConcept) {
  return `
Transform the following user concept into a production-ready AI video prompt (100-150 words).

User's concept: "${userConcept}"

---

**PROMPT:**

Write ONE descriptive paragraph following this precise structure:

 of performing in/at,, under, in a.

---

**GUIDING PRINCIPLES (in order of importance):**
1.  **Shot Type:** Start with the framing (e.g., wide shot, medium shot, close-up, extreme close-up).
This establishes the entire scene's composition first.
2.  **Subject:** Clearly define the main subject. Include 2-3 specific, visible details (e.g., "a woman in a
red trench coat with blonde hair") to ensure consistency and avoid generic outputs.
3.  **Action:** Describe ONE clear, specific, and physically plausible action. Multiple actions in one
prompt severely degrade quality due to the architectural constraints of video models.
4.  **Setting:** Ground the action in a specific location and time (e.g., "a neon-lit Tokyo alley at
midnight," "a serene beach at golden hour").
5.  **Camera:** Detail the camera's behavior. Specify both movement (e.g., static, slow dolly in,
handheld tracking) and angle (e.g., low angle, high angle, eye-level).
6.  **Lighting:** Describe the light to control the mood. Specify its source (e.g., "soft window light"),
direction (e.g., "from the left"), and quality (e.g., "creating deep shadows").
7.  **Style:** Provide a specific aesthetic reference. Avoid generic terms like "cinematic." Instead, use
film stock ("shot on 35mm film"), genre ("film noir aesthetic"), or an artist/director reference ("in the
style of Wes Anderson").

---

**WRITING RULES:**
✓ Total prompt length should be 100-150 words. Models follow concise, dense prompts more reliably.
✓ Use professional cinematic and photographic language (e.g., dolly, crane, rack focus, shallow DOF,
f/1.8, Rembrandt lighting).
✓ The order of elements in the prompt paragraph dictates their priority in the AI's processing. Place
```

the most important element FIRST.
✓ Describe only what the camera can SEE. Translate emotions and concepts into visible actions and environmental details.
✓ Be specific over generic: "a weathered oak desk" is superior to "a nice desk."

AVOID:
✗ Prompts longer than 150 words.
✗ Describing multiple simultaneous actions or a sequence of events. (Use multiple clips for a narrative.)
✗ Using ungrounded emotional adjectives (e.g., "a sad man").
✗ Using negative or instructive language (e.g., "don't show," "avoid," "no people"). Describe what you *want* to see instead.

EXAMPLE STRUCTURED PROMPT:
"Close-up of weathered, wrinkled hands slowly turning the pages of a large, leather-bound book on a dark oak desk. The camera slowly dollies back to reveal an elderly historian in a candlelit study. The scene is lit by a single, warm candle on the left, creating dramatic, low-key lighting with a high contrast ratio. The style is moody and atmospheric, shot on 35mm film with a shallow depth of field, f/1.8, creating significant bokeh in the background."

After the main prompt paragraph, add the following technical and creative specifications:

TECHNICAL SPECS
- **Duration:** 4-8s (Optimal for instruction-following and quality)
- **Aspect Ratio:** 16:9 (Landscape), 9:16 (Vertical), or 2.39:1 (Cinematic)
- **Frame Rate:** 24fps (Filmic) or 30fps (Standard Video)
- **Audio:**

ALTERNATIVE APPROACHES (2 variations, 40-50 words each)
Provide two brief, alternative prompts that explore different creative choices for the same core concept. This facilitates A/B testing and rapid iteration.
- **Variation 1 (Different Camera):** Explore a different camera angle or movement (e.g., "High-angle shot looking down...").
- **Variation 2 (Different Lighting/Mood):** Explore a different lighting setup or mood (e.g., "Bright, high-key lighting from a large window...").

OUTPUT FORMAT:
Begin directly with the main prompt paragraph. Do not include any preamble, explanations, or conversational text.
`;
}

9.3 Implementation Guide

To effectively integrate this template into a development or creative workflow, the following step-by-step process is recommended:

1. **Define Objective & Audience Persona:** Before writing any code or prompts, clearly articulate the goal of the video clip and the characteristics of the target audience. Use the framework in Section 7 to translate audience traits into potential visual styles.
2. **Draft Initial Prompt using Template:** Use the generateUniversalVideoPrompt function or a manual version of the template to structure the initial creative concept. Methodically fill in each component, from shot type to style, adhering to the writing rules.
3. **Generate and Evaluate:** Send the generated prompt to the video generation API. Assess the resulting video using the multi-faceted evaluation framework from Section 8. At a minimum, perform a qualitative review for prompt adherence, visual quality, and motion coherence.
4. **Iterate and Refine:** Based on the evaluation, make small, targeted changes. Use the "Alternative Approaches" section of the template to guide this iteration. For example, if the initial shot feels too static, try the variation with more camera movement. This process aligns with the HITL and A/B testing principles from Section 5.
5. **Finalize and Deploy:** Once a prompt consistently produces the desired output, it can be "locked" and saved into a prompt library for reuse in production.

9.4 Future Outlook and Recommendations

The field of generative video is evolving at an accelerated pace. While the principles outlined in this report represent the current state-of-the-art in prompt engineering, practitioners must remain adaptable.

Recommendation 1: Establish Internal Prompt Libraries. Organizations should create and maintain internal libraries of validated, high-performing prompts. These libraries, categorized by task, style, or target audience, will accelerate development, ensure brand consistency, and serve as a valuable internal knowledge base.

Recommendation 2: Adopt Systematic Evaluation Processes. To ensure consistent quality and objectivity, creative and technical teams should adopt a formal, HITL-based evaluation process for all critical video generation tasks. This moves the workflow from subjective preference to data-driven decision-making.

Future Outlook: The continued advancement of RLHF for video models will likely have a profound impact on prompt engineering.[34] As models become better aligned with general human preferences, the need for hyper-detailed, manually tuned prompts for common requests may diminish. The role of the prompt engineer will evolve. The focus will shift from crafting individual prompts to designing the systems that facilitate model improvement: curating high-quality preference datasets, designing effective reward models, and building the robust evaluation pipelines that power the next generation of more intuitive and capable video LLMs. The future prompt engineer is a systems thinker, an AI trainer, and a creative strategist, operating at the intersection of art and data science.

References

Acar, O. A. (2023). Cited in [1].

AWA Digital. (n.d.). Cited in [36].

Blattmann, A., et al. (2023). Cited in [14].

Brooks, T., et al. (2024). Cited in [14].

Chen, Y., et al. (2024). Cited in [14].

Cook, L. (2023). Cited in [1].

Google Research. (n.d.). Cited in [36].

Ho, J., et al. (2022). Cited in [14].

Johnmaeda. (2023). Cited in [1].

Krippendorff, K. (2011). Cited in [31].

Su, Y., et al. (2024). Cited in [14].

Sun, B., et al. (2024). Cited in [14].

Warrens, M. J. (2015). Cited in [31].

Wang, T., et al. (2022). Cited in [65].

Wyzowl. (2023). Cited in [42].

Yin, W., et al. (2023). Cited in [14].

Yu, J., et al. (2024). Cited in [14].

Works cited

1. Effective Prompts for AI: The Essentials - MIT Sloan Teaching & Learning Technologies, accessed October 16, 2025, https://mitsloanedtech.mit.edu/ai/basics/effective-prompts/
2. Prompt Engineering Importance and Applicability with Generative AI - ResearchGate, accessed October 16, 2025, https://www.researchgate.net/publication/384803603_Prompt_Engineering_Importance_and_Applicability_with_Generative_AI
3. Veo 3 Prompt Writing Best Practices: Actionable Guide for Reliable ..., accessed October 16, 2025, https://skywork.ai/blog/veo-3-prompt-writing-best-practices-actionable-ai-video-guide/
4. Prompt Engineering for AI Guide | Google Cloud, accessed October 16, 2025, https://cloud.google.com/discover/what-is-prompt-engineering
5. Best Practices for Prompting AI Video Generators - Leylinepro, accessed October 16, 2025, https://www.leylinepro.ai/blog/best-practices-for-prompting-ai-video-generators
6. Common AI Prompt Mistakes and How to Fix Them - AI Tools, accessed October 16, 2025, https://www.godofprompt.ai/blog/common-ai-prompt-mistakes-and-how-to-fix-them
7. Gen-3 Alpha Prompting Guide – Runway, accessed October 16, 2025, https://help.runwayml.com/hc/en-us/articles/30586818553107-Gen-3-Alpha-Prompting-Guide
8. Runway Prompt Guide: Advanced Settings for Cinematic AI Video Creation - AI Automation, accessed October 16, 2025, https://createxflow.com/runway-prompt-engineering/
9. Complete Guide to Writing Prompts for AI: Text-to-Image, Text-to-Video, Image-to-Video, and Visual Direction | by Shweta Kaushal | Bootcamp | Medium, accessed October 16, 2025, https://medium.com/design-bootcamp/complete-guide-to-writing-prompts-for-ai-text-to-image-text-to-video-image-to-video-and-visual-f59ccd2929cb
10. AI Video Prompt Writing: Structured Templates and Description for Better AI Clips - Reddit, accessed October 16, 2025, https://www.reddit.com/r/StableDiffusion/comments/1j4yhlb/ai_video_prompt_writing_structured_templates_and/

11. Sora 2 Prompting Guide - OpenAI Cookbook, accessed October 16, 2025, https://cookbook.openai.com/examples/sora/sora2_prompting_guide
12. Troubleshooting AI Video Generation in Scenario: Fix Common Issues, accessed October 16, 2025, https://help.scenario.com/en/articles/troubleshooting-video-generations/
13. InstructVideo: Instructing Video Diffusion Models with Human Feedback - CVF Open Access, accessed October 16, 2025, https://openaccess.thecvf.com/content/CVPR2024/papers/Yuan_InstructVideo_Instructing_Video_Diffusion_Models_with_Human_Feedback_CVPR_2024_paper.pdf
14. CVPR Poster OpenHumanVid: A Large-Scale High-Quality Dataset ..., accessed October 16, 2025, https://cvpr.thecvf.com/virtual/2025/poster/35209
15. Ultimate Sora 2 Prompt Guide: Craft Perfect AI Video Instructions for Cinematic Results, accessed October 16, 2025, https://www.glbgpt.com/hub/ultimate-sora-2-prompt-guide/
16. Crafting Cinematic Sora Video Prompts: A complete guide · GitHub, accessed October 16, 2025, https://gist.github.com/ruvnet/e20537eb50866b2d837d4d13b066bd88
17. Create Hollywood-Quality AI Videos In 30 Seconds [2025] - Sora Prompt Generator, accessed October 16, 2025, https://promptsgenerators.com/sora-prompt-generator/
18. Best Prompts for Runway Gen-2 (AI Video Generation Techniques) - Aituts, accessed October 16, 2025, https://aituts.com/runway-prompts/
19. Veo on Vertex AI video generation prompt guide - Google Cloud, accessed October 16, 2025, https://cloud.google.com/vertex-ai/generative-ai/docs/video/video-gen-prompt-guide
20. Mastering Veo 3: An Expert Guide to Optimal Prompt Structure and Cinematic Camera Control | by miguel ivanov | Medium, accessed October 16, 2025, https://medium.com/@miguelivanov/mastering-veo-3-an-expert-guide-to-optimal-prompt-structure-and-cinematic-camera-control-693d01ae9f8b
21. How to write effective text prompts to generate AI videos? - FlexClip Help Center, accessed October 16, 2025, https://help.flexclip.com/en/articles/10326783-how-to-write-effective-text-prompts-to-generate-ai-videos
22. RunwayML Prompt Guide: How to Prompt in Runway (Examples Included) - Pollo AI, accessed October 16, 2025, https://pollo.ai/hub/runway-prompt-guide
23. JindongGu/Awesome-Prompting-on-Vision-Language-Model - GitHub, accessed October 16, 2025, https://github.com/JindongGu/Awesome-Prompting-on-Vision-Language-Model
24. AI Video Generation Prompt Template System You Can Use | by Shushant Lakhyani, accessed October 16, 2025, https://medium.com/@slakhyani20/ai-video-generation-prompt-template-system-you-can-use-696b4672dcc1
25. Prompt iteration strategies | Generative AI on Vertex AI - Google Cloud, accessed

October 16, 2025,
https://cloud.google.com/vertex-ai/generative-ai/docs/learn/prompts/prompt-iteration
26. How to Perform A/B Testing with Prompts: A Comprehensive Guide ..., accessed October 16, 2025, https://www.getmaxim.ai/articles/how-to-perform-a-b-testing-with-prompts-a-comprehensive-guide-for-ai-teams/
27. General Tips for Designing Prompts - Prompt Engineering Guide, accessed October 16, 2025, https://www.promptingguide.ai/introduction/tips
28. Iterative Prompt Development - Medium, accessed October 16, 2025, https://medium.com/@aanshsavla2453/iterative-prompt-development-14252e0597f5
29. From Prompt Engineering to Prompt Science with Humans in the Loop, accessed October 16, 2025, https://cacm.acm.org/research/from-prompt-engineering-to-prompt-science-with-humans-in-the-loop/
30. [2401.04122] From Prompt Engineering to Prompt Science With Human in the Loop - arXiv, accessed October 16, 2025, https://arxiv.org/abs/2401.04122
31. From Prompt Engineering to Prompt Science With Human in the Loop - arXiv, accessed October 16, 2025, https://arxiv.org/html/2401.04122v3
32. Reinforcement learning from human feedback - Wikipedia, accessed October 16, 2025, https://en.wikipedia.org/wiki/Reinforcement_learning_from_human_feedback
33. What Is Reinforcement Learning From Human Feedback (RLHF)? - IBM, accessed October 16, 2025, https://www.ibm.com/think/topics/rlhf
34. Boosting Text-to-Video Generative Model with MLLMs Feedback - NIPS papers, accessed October 16, 2025, https://proceedings.neurips.cc/paper_files/paper/2024/file/fbe2b2f74a2ece8070d8fb073717bda6-Paper-Conference.pdf
35. [2501.13918] Improving Video Generation with Human Feedback - arXiv, accessed October 16, 2025, https://arxiv.org/abs/2501.13918
36. How to write AI prompts for A/B testing teams | Kameleoon, accessed October 16, 2025, https://www.kameleoon.com/blog/how-write-ai-prompts-ab-testing-teams
37. A/B Test Prompts and Models - Portkey Docs, accessed October 16, 2025, https://portkey.ai/docs/guides/getting-started/a-b-test-prompts-and-models
38. Runway Gen-2 Prompts: Everything You Need to Know | Pollo AI, accessed October 16, 2025, https://pollo.ai/hub/runway-gen-2-prompt-guide
39. Dream Machine Guide: Best practices - Luma AI, accessed October 16, 2025, https://lumalabs.ai/learning-hub/best-practices
40. Kling AI Text to Video | AI Video Generation from Prompts, accessed October 16, 2025, https://app.klingai.com/global/text-to-video/new
41. Top 10 Mistakes to Avoid When Using an AI Video Generator | by Ram Bharat - Medium, accessed October 16, 2025, https://medium.com/@ram-bharat/top-10-mistakes-to-avoid-when-using-an-ai-

video-generator-6e37a250e62d
42. How To Create Video Content Tailored To A Specific Target Audience - Pickle Pictures, accessed October 16, 2025, https://picklepictures.com/how-to-create-video-content-tailored-to-a-specific-target-audience/
43. Social Media Audience Analysis: A Guide to Understanding Your Followers | Sprinklr, accessed October 16, 2025, https://www.sprinklr.com/blog/social-media-audience-analysis/
44. Mastering Audience Connection: Unveiling Your Video Viewership - CineSalon, accessed October 16, 2025, https://www.cine.salon/resources/how-to-find-your-target-audience-for-your-videos
45. How to Tailor Content Creation for Your Audience's Needs - Homerun Entertainment, accessed October 16, 2025, https://www.homerunent.com/blog/2025/6/30/how-to-tailor-content-creation-for-your-audiences-needs
46. Mastering Social Media Audience Analysis: A How-To Guide - Socialinsider, accessed October 16, 2025, https://www.socialinsider.io/blog/social-media-audience-analysis/
47. Mastering Target Audience Analysis: Proven Strategies Unveiled - Contentoo, accessed October 16, 2025, https://www.contentoo.com/blog/target-audience-analysis-proven-strategies
48. Personalization Strategies: Tailoring Content to Individual User Needs - Common Ninja, accessed October 16, 2025, https://www.commoninja.com/blog/personalization-strategies-tailoring-content-to-individual-user-needs
49. How to write effective AI prompts | Help center - Formaloo, accessed October 16, 2025, https://help.formaloo.com/en/articles/9797669-how-to-write-effective-ai-prompts
50. How to Create an AI Marketing Persona: 8 Prompts For Deep ..., accessed October 16, 2025, https://www.orbitmedia.com/blog/ai-marketing-personas/
51. How I'm using AI to streamline persona and journey map creation ..., accessed October 16, 2025, https://uxdesign.cc/using-ai-to-streamline-persona-and-journey-map-creation-37fa859dafb0
52. 10 Best Customer Persona Generators In 2025 [Reviewed] | Team-GPT, accessed October 16, 2025, https://team-gpt.com/blog/customer-persona-generators
53. 20 Best Animated Explainer Video Examples - Content Beta, accessed October 16, 2025, https://www.contentbeta.com/blog/animated-explainer-video/
54. How To Pick The Right Animated Video Style For Your Project?, accessed October 16, 2025, https://videoigniter.com/animated-video-style/
55. How To Tailor Content To Different Demographics: Detailed ..., accessed October 16, 2025, https://penfriend.ai/blog/tailor-content-to-different-demographics
56. Animation Techniques 101: Best Style for Advert Success - Educational Voice,

accessed October 16, 2025, https://educationalvoice.co.uk/animation-techniques-for-your-commercial/
57. Best Types of Animation Styles (and Their Marketing Uses), accessed October 16, 2025, https://www.yumyumvideos.com/blog/types-of-animation-styles/
58. Challenges in generated video evaluation - At Scale Conferences, accessed October 16, 2025, https://atscaleconference.com/videos/challenges-in-generated-video-evaluation/
59. An Essential Guide for Generative Models Evaluation Metrics | by ..., accessed October 16, 2025, https://pub.towardsai.net/an-essential-guide-for-generative-models-evaluation-metrics-255b42007bdd
60. Evaluating Generative AI: A Comprehensive Guide with Metrics, Methods & Visual Examples, accessed October 16, 2025, https://medium.com/genusoftechnology/evaluating-generative-ai-a-comprehensive-guide-with-metrics-methods-visual-examples-2824347bfac3
61. Video Analytics 101: A Beginner's Guide to Data-Driven Content Creation - Wowza, accessed October 16, 2025, https://www.wowza.com/blog/video-analytics-101
62. Video Content Creation Strategy, Tips & Tools - YouTube Creators, accessed October 16, 2025, https://www.youtube.com/creators/how-things-work/content-creation-strategy/
63. T2V-CompBench: A Comprehensive ... - CVF Open Access, accessed October 16, 2025, https://openaccess.thecvf.com/content/CVPR2025/papers/Sun_T2V-CompBench_A_Comprehensive_Benchmark_for_Compositional_Text-to-video_Generation_CVPR_2025_paper.pdf
64. NeurIPS Poster Evaluation of Text-to-Video Generation Models: A Dynamics Perspective, accessed October 16, 2025, https://neurips.cc/virtual/2024/poster/93313
65. Iterative Prompting Strategies - Emergent Mind, accessed October 16, 2025, https://www.emergentmind.com/topics/iterative-prompting-strategies