ENDLIVITY
CREATIVE LABS

Anleitung: Eigenen GPT in ChatGPT für Videoprompting (Veo 3 & Veo 3.1) erstellen

Schritt 1: Neues Projekt erstellen

Öffne ChatGPT u nd erstelle ein neues Projekt. Gib ihm einen passenden Namen – zum Beispiel: „VEO 3 Prompt Engineer“. Dieses Projekt dient als dein persönlicher Arbeitsbereich für den Veo 3 Video Prompt Engineer.

Schritt 2: Innerhalb des Projekts einen neuen Chat starten

Starte innerhalb deines Projekts einen neuen Chat. Ab jetzt führst du alle zukünftigen Unterhaltungen, Trainings und Prompt-Erstellungen ausschließlich in diesem Chat durch. So bleibt der Kontext erhalten und dein GPT kann gezielt weiterlernen.

Schritt 3: GPT-Rolle initialisieren

Gib im Chat folgende Nachricht ein:

"Bitte agiere in Zukunft als Veo 3 Prompt Engineer. Ich werde dir Trainingsdaten, Prompt-Strukturen und Beispiele zur Verfügung stellen, damit du mir professionelle Prompts auf Englisch mit bis zu 2000 Zeichen erstellen kannst. Sollten Informationen für einen professionellen Prompt fehlen, bitte frage nach. Hast du dies verstanden und bist du bereit für die Informationen?"

Wenn gewünscht kannst du dies hinzufügen: “Bitte erstelle mir die Prompts für Veo3 im JSON-Stil.”

Schritt 4: Trainingsinformationen eingeben

Füge nun den gesamten Trainings- und Strukturtext für Veo 3 ein (siehe weiter unten). Dieser enthält alle wichtigen Punkte wie:
- Szenenbeschreibung
- Visueller Stil
- Kamerabewegung
- Hauptmotiv
- Hintergrund / Umgebung
- Licht / Stimmung
- Audio (optional)
- Farbpalette
- Dialog / Hintergrundgeräusche
- Untertitel und Spracheinstellungen

Dieser Abschnitt beinhaltet auch Beispiele für sehr gute Prompts und die detaillierten Erklärungen zum Thema Dialog, Untertitel vermeiden, Kamerabewegungen und Szenenstruktur aufbauen.<

Diese Trainingsinformationen kannst du auch von Zeit zu Zeit erweitern, ändern, anpassen, etc.

Schritt 5: Verständnis bestätigen

Frage den GPT, ob er alle Informationen verarbeitet und verstanden hat. Sobald er dies bestätigt, ist dein Veo 3 Prompt Engineer bereit für die Erstellung professioneller Prompts.

Schritt 6: Ausgangsbild hochladen & Szene beschreiben

Lade dein Ausgangsbild (Startframe und eventuell Endframe) hoch und formuliere deine Szene, z. B.:

„Bitte erstelle mir einen Prompt für das angehängte Bild. Die Gemüseverkäuferin im Jahr 80 nach Christus steht mitten auf dem Marktplatz vor dem Kolosseum und sagt in die Kamera: 'Ich bin die Händlerin des Marktes, bringe Gemüse für das Volk. Wir alle vergehen mit der Zeit. Doch das Kolosseum bleibt bestehen.'"

Schritt 7: Video erstellen

Kopiere den erstellten Prompt in Freepik oder ein anderes Tool mit Veo 3 Anbindung. Erstelle das Video und sieh dir das Ergebnis an.

Schritt 8: Anpassungen und Feinschliff

Wenn du Änderungen wünschst, schreibe sie einfach im selben Chat. Gib an, worauf sich das Video stärker konzentrieren soll – etwa die Kamerafahrt, die Stimmung, die Beleuchtung oder die Handlung. Der GPT wird den Prompt entsprechend anpassen und optimieren.

Trainingsinformationen:

Before we get into advanced techniques, we need to get the basics right. Every strong Veo 3 prompt includes a few core elements that compose the building blocks of your shot. According to Google's own guidance, the most important parts of any prompt are:

Subject: Who or what is the focus of the scene? A person, animal, object, or abstract form.

Context: Where is this taking place? Think background, setting, environment, time of day.

Action: What is the subject doing? This can be simple (walking, sitting, turning) or layered (pausing, reacting, adjusting posture).

Style: The overall look and feel. You can reference film genres (*film noir*, *animated short*), visual tone (*gritty realism*, *warm and soft*), or artistic aesthetics (*oil painting*, *Pixar-like*).

These four are essential, but you can also add optional modifiers for more control:

Camera motion: Is the camera static, tracking, pulling in, or panning?

Camera framing: What type of shot is this? Wide, medium, close-up?

Camera angle: Do we see the subject at eye level, from below (low angle), from above (high angle), or directly overhead?

Ambiance: Describe the lighting and color tone (e.g., "pale morning light," "cool blue shadows," "sunset orange haze").

Audio: If you want sound, specify it. Dialogue, ambient noise, or background music can all be included—check out our separate guide on prompting for audio.

Example Prompt

A man in worn clothing walks slowly across an open desert, one hand raised to shield his face from the sun. The camera begins at shoulder height behind him, then rises in a smooth, drone-style lift into an overhead wide shot, revealing the vast, empty landscape stretching endlessly in all directions. The horizon shimmers with heat beneath a pale blue sky. Style: Cinematic, tense, minimalist. Audio: A slow-building thriller film score, layered with low strings and subtle pulses beneath the silence.

Observations

The result shows an almost perfect prompt adherence! Let's unpack this prompt:

Subject: A man in worn clothing.

Context: An open desert under a pale blue sky, with the horizon shimmering from heat.

Action: He walks slowly, one hand raised to shield his face from the sun.

Style: Cinematic, tense, minimalist.

Camera: The camera begins behind him at shoulder height, then lifts smoothly into an overhead shot.

Audio: A slow-building thriller film score.

What else could you add to make this shot even stronger? Maybe a dust storm on the horizon? You could even reverse the camera movement—have it approach instead of pulling away. Try it in Leonardo.Ai and see if the new version pushes the mood in a direction you like.

Prompt Length

What It Is

Prompt length is how much you tell Veo about the shot. Too short, and the output may feel generic. Too long, and the model might get confused or try to do too much.

Best Practice

Aim for a balanced range: roughly 3–6 sentences, or 100–150 words. This gives you room to describe the subject, context, action, and style, with optional space for other elements like camera, ambiance, or sound. Long-winded paragraphs or overly compressed one-liners tend to perform worse. Think of each prompt as a single, self-contained shot.

Example Prompt

A wide, eye-level cinematic shot captures a man walking slowly across a frost-covered bridge at dawn, his hands tucked into the pockets of a heavy coat. Pale morning light glows faintly through soft, curling fog that clings to the bridge railings. In the distance, bare trees fade into the mist, their skeletal branches barely visible. The pace is unhurried and reflective, evoking a naturalistic and quiet mood. The scene is filled with subtle, atmospheric sounds—faint footsteps crunching on frost, steady breaths in the cold air, and the distant caw of a crow echoing across the stillness.

Observations

Notice that while the output keeps the camera below eye level (instead of at eye level), it gets so many other elements impressively right:

Subject: A man with his hands tucked into the pockets of a heavy coat.

Action: He's walking at an unhurried, reflective pace.

Context: A frost-covered bridge on a foggy morning; in the distance, bare trees fade into the mist, their skeletal branches barely visible.

Style: Cinematic.

Composition: Wide shot.

Ambiance: Pale morning light glowing faintly through soft fog.

Sound: Faint footsteps crunching on frost, with the distant caw of a crow echoing across the stillness.

Prompt Structure

What It Is

Prompt structure refers to the overall way your shot description is organized. It's not just about word order, but about how the information is arranged and presented. For example, you might write your prompt as a flowing cinematic paragraph or break it into labeled sections like "Subject," "Action," and "Camera." Veo may interpret the same scene differently depending on the structure you use or which element it encounters first.

Best Practice

There's no one-size-fits-all format, but Veo tends to interpret structure literally. The way you format your prompt can affect how the model handles pacing, focus, and motion. One of the most important things is to clarify your goal and visualize the shot beforehand so you know what you want to see. If you want the background to stand out or the lighting to carry more weight, it often makes sense to mention it first. For example, let's rewrite the prompt above in a modular format with labeled sections, this time placing the context before the subject.

Example Prompt (Modular Format With Labeled Sections)

Context: A frost-covered bridge at dawn, with bare trees fading into the mist in the distance.

Subject: A man with his hands tucked into the pockets of a heavy coat.

Action: He walks slowly across the bridge at an unhurried, reflective pace.

Style: Cinematic.

Composition: Wide shot, eye level

Lighting and Ambiance: Pale morning light glowing faintly through soft, curling fog that clings to the bridge railings.

Audio: Faint footsteps crunching on frost, steady breaths in the cold air, and the distant caw of a crow echoing across the stillness.

Observations

One major difference we're seeing compared to the previous prompt is that we now see much more of the bridge in the first few seconds. While a video generation model is a bit of a black box and we can never know for sure, we can hypothesize that specifying the bridge first forced the model to give more attention to that element.If you want to try more variations, you can go on Leonardo.Ai and experiment with different prompt structures. Remember that finding the right structure takes a lot of experimentation, which is why having a clear goal is key: it helps you experiment purposefully, not aimlessly.

Camera Shots

What It Is

Camera shot refers to how much of the subject and environment is visible in the frame. A wide shot might show a landscape and a distant figure. A close-up might focus on a single gesture, like a hand clenching or a tear falling. Veo understands classic film language (wide shot, medium shot, close-up, etc.) and generally follows these instructions accurately.

Best Practice

Always specify the shot type if you want control over framing. If you leave it out, Veo will often default to the most common framing patterns it encountered during training for videos similar to

what you're trying to create. You can combine it with a camera angle or motion, but avoid describing multiple conflicting shot types in a single prompt.

Because of how video generation models are trained, they sometimes have a strong tendency to default to a certain camera framing, even if you repeatedly and clearly ask for a different one. This happens when the subject, action, and context you're trying to depict are strongly tied in the training data to a specific framing. The best approach here isn't to repeat the same prompt over and over, hoping the model will eventually get it right—instead, try a workaround, like changing some aspect of the subject or action, even if it means making a small compromise.

Example

Close-up. Style: TV commercial. A woman in her 30s takes her first sip of coffee while sitting on a small balcony overlooking a quiet city street. She's wrapped in a soft sweater, morning light grazing her face. It's chilly—steam rises gently from the mug and curls past her cheek. Her shoulders drop slightly as the warmth hits. Her eyes close for just a moment—not dramatic, just real. Audio: background music.

Observations

Prompt adherence is nearly perfect here. Building on our advice about prompt structure, notice how frontloading the camera framing and style helped ensure we got a close-up in a TV commercial tone.

Camera Movements

What It Is

Camera movement refers to how the camera behaves within the shot. Is it staying locked in place, pulling back, panning across a space, or tracking a subject from behind? These choices affect pacing, emotional tone, and how much of the scene unfolds over time.

Best Practice

If you want the camera to move, you need to say so clearly in your prompt. If you don't mention any movement, Veo will default to something that might not match what you had in mind. Starting with a visual goal is key because it directs your experimentation.

Movement instructions typically work better when separated from subject actions. For example, write "The camera pulls back" as a standalone sentence, rather than embedding the motion within a longer description. This helps the model parse your intent and match the framing more reliably.

Again, if you don't get the output you want, it's not a good idea to fight the model by repeating the same prompt over and over again. Think of Veo as a highly complex statistical machine—instead of forcing it to behave the way you want (which, statistically, it won't), it's better to experiment with different approaches to achieve your goal.

Example Prompt

A sleek smartwatch sits on a rugged rock near the edge of a mountain cliff. The camera begins close, then pulls back in a smooth, continuous drone-style shot. As it rises, a vast alpine landscape unfolds—jagged peaks, mist rolling through the valley, and golden sunrise light washing over everything. The tone is cinematic and epic, emphasizing the contrast between modern technology and untamed nature.

Observations

Prompt adherence was high for this shot, and we got nearly everything right, especially the camera movement. Notice how the sentence describing the camera movement is clearly separated from those describing other elements.

Camera Angles

What It Is

Camera angle refers to where the camera is positioned in relation to the subject. Are we looking up at the subject, looking down on them, or seeing them straight on? This choice changes how the subject is perceived:

Low angle (camera below eye level): Makes the subject look large, powerful, or imposing.

High angle (camera above eye level): Can make the subject feel small, isolated, or vulnerable.

Eye-level: Neutral, grounded, balanced.

Best Practice

If the angle matters, name it early, because Veo won't always choose the most expressive or cinematic angle on its own. Like with other aspects of prompting, if the angle isn't specified, the model often defaults to an angle that might not be what you want.

Specifying angle alone isn't always enough. For high angles, describe what the camera sees beneath the subject. For low angles, describe what's behind or above them. This helps Veo lock onto the perspective you want.

Since they are all elements related to the camera, it's always worth trying to include angle, framing, and movement together.

Example Prompt

A low-angle medium shot frames a boxer from below as he bounces in place before a match, lit by harsh overhead fluorescents. Sweat glistens on his jawline and neck, catching the cold light. His breath is slow and controlled, his shoulders rolling with each inhale. Huge crowd. Style: Gritty, cinematic realism.

Observations

The output shows strong prompt adherence, especially for camera angle and framing. Notice that because we didn't specify any camera movement, Veo defaulted to a handheld shot. To understand why, consider its training data: most boxing scenes—whether from live broadcasts or films—are often shot handheld at the start of a match.If you want to experiment further with this prompt, go to Leonardo.Ai and try adding a camera movement that feels counterintuitive (for example, a dolly-out, which in this case might release the tension instead of building it).

Repetition vs. Variation

What It Is

Repetition is using the same words many times in the same prompt to force the model to emphasize an element more. Variation involves using synonyms, rephrasings, or different visual references to express the same idea—sometimes with added nuance—without repeating the exact same words.

Best practice

Veo doesn't understand language the way we do. It's a statistical model, and under the hood, your prompt is converted into a sequence of numbers. So repeating a word—like you're trying to clarify something to a toddler—usually doesn't help. Our tests show that repeating the same word multiple times in one prompt doesn't create stronger emphasis. It often results in a noisier, less focused output.

Variation can be helpful when you're describing layered details like weather, light, or ambiance. For example, instead of writing:

Rain falls. Rain drips. Rain hits the pavement. Rain reflects the light. Rain gathers in puddles.

You could try:

Cold drizzle falls under dim streetlights. Droplets tap against rusted metal and scatter into puddles. A sheen of water coats the sidewalk, reflecting pink signage.

However, even varied phrases can still map to similar patterns in the model's embedding space, which means Veo might interpret them as simple repetition. As with other aspects of prompting, the most effective approach is goal-driven experimentation.

Example Prompt

Wide shot. A narrow alley glows under pulsating signage as cold drizzle falls from the sky. Droplets tap against rusted pipes and ripple across the soaked pavement. A sheen of water coats the sidewalk, reflecting pink signage. A hooded figure walks slowly past corroded vending machines, their reflection bending across the wet stone. Style: Cinematic, urban night.

Audio: A distant mechanical alarm blares once, then fades. Neon buzzes softly. Static crackles from unseen speakers. A low electrical hum pulses beneath the rain.

Observations

Even though we didn't specifically mention "rain" in our prompt, Veo generated a drizzle (a light rain). On the left side of the shot, a few droplets tap against rusting pipes. There's also a sheen of water on the sidewalk, reflecting pink signage, with raindrops clearly visible. The stone looks wet. Overall, we got solid results and achieved the variation we aimed for.

Object Count and Complexity

What It Is

This refers to how many distinct elements you include in a single shot—both different types of objects (like characters, props, background details) and how many instances of the same object you ask for.

Best Practice

Based on our tests, Veo 3 can handle low-to-moderate object counts with good fidelity—typically up to around 15 of the same item. Beyond that, there's a risk of vague shapes, inconsistent spacing, or objects merging together. If your shot depends on visual precision with high numbers, you'll usually get better results by focusing the scene around one or two key elements and keeping the background simple or implied.

If your scene is complex and includes many different objects—where count also matters—keep in mind that asking Veo to render a dozen lanterns, twenty birds, or a crowd of people means asking it to manage spatial logic, scale, and interaction all at once. You can always try and see what happens, but if it doesn't work the first time, don't keep repeating it. Instead, think like a screenwriter: which objects actually serve the story? Once you have an answer, cut the rest.

Example Prompt

Only six lanterns float slowly across the surface of a misty lake, forming a wide ring. Their warm glow flickers across the glassy water, each reflection trembling softly in the haze. The lake is silent, still, encircled by tall, dark trees fading into fog. Style: Cinematic, eerie stillness.

Audio: Low, tension-building horror score; faint water movement beneath the music.

Observations

Specifying "only" helped the model understand that we needed exactly six lanterns. Since our goal was to get the correct number, we front-loaded this element to make sure it was prioritized. The shot is quite cinematic and could easily belong in a horror film.

Tone and Writing Style

What It Is

This refers to how you phrase your prompt. Are you writing in clear, present-tense directions? Or are you leaning into poetic language, technical phrasing, or something closer to screenplay format? While Veo doesn't understand tone and style exactly like a human, it may react differently based on the style and rhythm of your words.

Best Practice

Consider this short poem: *Time knelt quietly in the corner, counting breaths it never took*. Can you visualize it? Neither can Veo. While you don't have to flatten your writing, make sure you're always describing something that can be seen or heard in the frame.

Veo tends to respond well to prompts that read like visual direction—simple, image-focused, and written in the present tense. Think like a filmmaker: you need to create emotion through what is seen and heard, not through text.

Example Prompt

Wide shot. Style: cinematic. A curved corner diner glows brightly on a dark, empty street at night. Inside, three customers sit at the long counter—two men in suits and fedoras, one woman in a red dress, all quietly facing forward. A server sits quietly behind the counter, avoiding eye contact. The interior is stark and clean, lit with warm overhead light that spills out onto the sidewalk. Outside, the storefront windows reflect empty green-tinted buildings and a quiet, empty road. Audio: strong wind outside.

Observations

Our goal with this shot was to evoke loneliness, but we never explicitly used the word in the prompt, nor did we rely on poetic or literary language. Instead, we leaned on visual cues to suggest the emotion indirectly: the lack of interaction between the customers, the distance between them, the server avoiding eye contact, the empty streets, the sound of the wind.

And if this shot feels vaguely familiar, you're not wrong—it was inspired by Hopper's famous Nighthawks painting. Try choosing a painting you love (one that conveys a strong emotion) and recreate it in Leonardo.Ai without explicitly naming the feeling you're trying to evoke. Focus on the visual (and, if possible, audio) cues that build that emotion naturally.

Prompting for Image-to-Video

What It Is

Image-to-Video (I2V) is a powerful workflow that fundamentally changes the prompting process. Unlike Text-to-Video (T2V), where you describe a scene from scratch, I2V starts with an existing image. Think of your uploaded image as the foundation—it has already defined the subject, context, style, and composition.

Your text prompt's job is no longer to create a world, but to bring that world to life. The prompt's focus narrows significantly to defining three key elements: the action, the camera movement, and the audio.On the Leonardo.Ai app, you can use an I2V workflow by selecting a start frame:

Best Practice

When using I2V, your prompting strategy should shift from description to direction. Your prompt should describe what changes or moves within the static frame you've provided. Cut out any descriptions already present in the image. If your image shows a knight in a forest, you don't need to mention the knight or the forest—just the action he takes.

I2V is the best modality for ensuring character or asset consistency across multiple clips. If you need the same character in different scenes, start each generation with the same reference image. A pro tip is to also use a standardized text description for that character in each prompt to further solidify their visual details.

I2V often produces more authentic and higher-quality motion than T2V, especially for action-heavy prompts. By providing a static image, you allow the model to reallocate its resources from inventing a scene to perfecting the physics of the movement within it.

Example Prompt

A great use case of an I2V workflow is animating a logo, so let's try doing that. First, we'll create an image using Lucid Origin:

A sleek, modern tote bag with a clean, minimalist mountain logo, rendered in a simple, sans-serif font, sitting on a reclaimed wooden table, with a subtle grain texture, against a neutral background, where the tote bag's smooth, matte fabric has a slight sheen, and the mountain logo is depicted in a combination of thin lines and basic geometric shapes, with the bag's handle made of a sturdy, woven material, and the overall color palette is dominated by earthy tones, with the logo's color being a deep, rich blue, and the bag itself is a light, creamy beige, with the table's wood grain visible in warm, golden hues.

Next, we'll animate this logo using the image above as a start frame:

The mountain logo on the tote bag subtly animates, with clean lines tracing the peaks. The camera slowly zooms in, focusing on the movement. Audio: A gentle whooshing sound as the lines animate, followed by a soft, satisfying click.

Observations

Notice that the video inherited the visual information provided by the image almost perfectly—even the bag folds are the same! Our prompt focused on the action (the logo subtly animates), the camera movement (zooming in), and the audio.

Prompting for Start and End Frames

What It Is

The Start and End (S/E) Frame workflow offers the highest level of narrative and cinematic control available in Veo 3. In this mode, you provide two images: one for the exact starting composition and one for the exact ending composition. Veo's task is to generate a smooth, controlled video that connects these two specific visual states.

Your prompt doesn't need to describe what's in the frames themselves; instead, it must function as a detailed roadmap for the camera path, the subject's transformation, and the audio cues that bridge the gap between point A and point B.On the Leonardo.Ai app, you can use an S/E workflow by adding both a start and an end frame:

Best Practice

To get the most out of the S/E workflow, your prompt must be precise and descriptive, focusing entirely on the transition. Your primary goal is to guide the transformation. The prompt should explicitly describe the camera's movement, the action happening in between the frames, and the soundscape that accompanies the transition.

Defining the camera path is the most critical element for a successful S/E generation. You must use clear, descriptive cinematic terms to link the start and end compositions. A prompt like "Slow dolly out combined with a crane shot rising smoothly to reveal the protagonist" dictates a continuous, deliberate movement that physically connects two different shots.

Clearly state the physical or stylistic change that occurs between the frames. Whether a character performs an action or the entire scene undergoes a visual metamorphosis—for instance, "shifting the visual style from photorealistic to animation"—it must be articulated in the prompt.

The S/E workflow is perfect for creating seamless video loops. To do this, use a start frame and an end frame that are nearly identical, and include the instruction "a seamless loop" in your prompt to constrain the action.

Example Prompt

Let's start by creating the start frame using Lucid Origin:

An empty, unfurnished living room with plain, light gray walls, a polished hardwood floor, and a large window allowing natural light to flood the space, seen from a wide angle, with the room's

corners and edges sharply defined, a neutral colored ceiling, and no visible fixtures or furniture, creating a sense of minimalism and openness.

Next, we use Nano Banana to edit the image into we're going to use as our end frame:

The same room, now fully furnished and decorated in a modern style, with vibrant wall color and art.

Finally, we generate the videos using an S/E workflow:

A fast, shimmering wave of energy washes across the room, leaving a trail of sparkling particles in its wake. Over the next seconds, these particles coalesce and elegantly construct the furniture and decorations, which settle into place one by one until the room is fully transformed. Audio: A quick whoosh, followed by a subtle, sparkling sound effect that builds and resolves into a soft, magical chime as the final item appears.

Observations

With the exception of a few minor artifacts, the result feels a bit like magic! Notice how the prompt doesn't waste words describing the empty room (start frame) or the decorated room (end frame); its entire focus is on detailing the transformation process that connects them. The audio instructions are detailed and synced to the visual action.

Conclusion

Veo 3 and Veo 3.1 are powerful creative tools, but they are both highly sensitive to how you write. As you've seen throughout this guide, even small changes in prompt length, structure, or phrasing can shift the tone, pacing, or focus of a shot. That's why clarity and intention are essential.

To get strong, consistent results, make sure your prompt includes the core elements: subject, context, action, and style. For more control, you can also add optional modifiers: camera language, ambiance, or audio.

Be mindful of how you manage prompt length and structure. Keep prompts focused, with one clear idea per shot. Use straightforward sentence structure and avoid overloading the frame with too many actions or visual elements.

Above all, have a goal. If you know what you want to see, you'll be able to shape your prompt around that outcome and experiment with purpose. Whether you're aiming for a specific mood, pacing, or visual payoff, clear intent makes a difference.If you want to keep exploring, try experimenting in Leonardo.Ai and take a look at our separate guide on audio prompting.