

KI in der Film- und Videoproduktion

Der praktische Einstieg

» Hier geht's
direkt
zum Buch

DIE LESEPROBE

Kapitel 1

Von Worten zu bewegten Bildern

Videos mit KI-Textprompts erzeugen

Schreib ein paar Worte – und die KI baut dir daraus bewegte Bilder: Kamera fährt, Licht sitzt, Figuren agieren. Hier lernst du, wie du aus einer Idee präzise Prompts formst, die wie Regieanweisungen funktionieren, und welche Tools dafür aktuell die besten Ergebnisse liefern. So wird aus deiner Vorstellung ein stimmiger Clip – schnell, kontrollierbar und überraschend nah an »echtem« Film.

Die wohl einfachste – und aktuell auch beliebteste – Methode, mit künstlicher Intelligenz Videos zu erzeugen, funktioniert allein über Sprache oder Text: Du gibst der KI eine Beschreibung deiner Szene – also einen sogenannten Prompt – und bekommst im besten Fall ein vollständiges Video zurück. Diese Technologie hat in den letzten Monaten einen riesigen Sprung gemacht. Das Internet ist inzwischen voll mit Clips, bei denen man zweimal hinschauen muss, um zu erkennen: Moment mal, das ist ja gar kein echtes Video!

Oft fällt einem erst durch den Kontext oder das Motiv auf, dass etwas nicht stimmen kann – etwa eine surreale Szene oder eine Kombination von Elementen, die so nicht existieren kann. Und selbst dann ist es manchmal schwierig, Fehler zu finden. Viele dieser Clips sehen mittlerweile so realistisch aus, dass man sie auf den ersten Blick für Hollywood-Material halten könnte. Ob dein Video am Ende auch so beeindruckend wird, hängt im Wesentlichen von zwei Faktoren ab: Erstens natürlich von der KI-Plattform, die du verwendest – denn die Ergebnisse unterscheiden sich teils deutlich in Qualität, Stil und technischer Umsetzung. Und zweitens – und das ist fast noch wichtiger – von der Qualität deines Prompts, also der Beschreibung, mit der du das Video generierst.

In diesem Kapitel zeige ich dir beides: Welche Tools aktuell besonders gut sind – und wie du deine Prompts so formulierst, dass du das Maximum aus ihnen herausholst.

Ein kleiner persönlicher Tipp gleich vorweg: Ich lasse alle meine Prompts von ChatGPT schreiben. Nicht, weil ich es nicht selbst könnte, sondern weil es schneller, präziser und einfach effizienter ist. Ich spreche meine Szene mit dem Mikrofon ein, beschreibe, was ich vor Augen habe, und ChatGPT wandelt meine Gedanken in einen klar formulierten, englischen Prompt um. Natürlich funktionieren viele Tools auch mit deutschen Prompts, aber mein Eindruck ist: Auf Englisch funktionieren sie besser.

Außerdem hat die Nutzung eines Large Language Models wie ChatGPT noch einen riesigen Vorteil: Ich kann die komplette Geschichte zuerst grob planen, also ein Storyboard mit ChatGPT erstellen – unterteilt in Szenen, die wiederum in einzelne Shots zerlegt werden. Und weil ChatGPT diese Gesamtgeschichte kennt, muss ich später nicht jedes Mal das komplette Setting neu beschreiben. Ich gebe einfach nur noch an, was diesen einen Shot besonders macht – und ChatGPT füllt automatisch alle stilistischen und erzählerischen Details ein, die zur Szene passen. Dadurch entsteht ein stimmiger Look mit durchgehender Erzählweise – und das mit erstaunlich wenig Aufwand.

In den nächsten Abschnitten erfährst du also, wie ein guter Prompt aufgebaut ist (mit konkreten Beispielen), welche Elemente wie Handlung, Licht oder Kamerabewegung wichtig sind und welche Plattformen sich gerade besonders lohnen – von Seedance bis Google Veo3. Los geht's – von Worten zu Bildern. Und von deiner Idee zum fertigen Film.

1.1 Regie statt Beschreibung – Die Denkweise hinter einem guten Prompt

Wer KI-Videos erstellen möchte, sollte nicht einfach nur beschreiben, was zu sehen ist – sondern Regie führen. Ein Prompt ist keine Inhaltsangabe, sondern eine präzise Anweisung wie am Filmset. Stell dir vor, du gibst einem Kamerateam, einem Schauspieler und einem Lichttechniker zeitgleich Regieanweisungen. Genau das ist ein guter Prompt: Er setzt eine kreative Vision in präzise Anweisungen um. Frage dich bei jedem Prompt: Warum ist dieser Moment wichtig? Was soll

er auslösen? Das verändert sofort den Fokus – von reinem »Was passiert?« hin zu »Wie wirkt es?«.

Die acht Elemente für perfekte Videoprompts

Ein gutes KI-generiertes Video beginnt immer mit einem klaren, durchdachten Prompt. Je präziser du deine Szene beschreibst, desto überzeugender wird das Ergebnis. Dafür genügt es nicht, nur ein Motiv zu nennen – entscheidend sind die Details. In den folgenden acht Elementen erfährst du, wie du Schritt für Schritt die richtigen Angaben machst, um aus einer bloßen Idee eine lebendige und filmreife Szene entstehen zu lassen.

1. **Subjektbeschreibung:** Sei präzise: Wer oder was ist in deiner Szene? Beschreibe nicht nur »ein Mann«, sondern »ein Mann in elegantem Anzug, mit nassem Haar, der nervös wirkt«. Je klarer die Details, desto besser kann die KI dein Bild erzeugen.
2. **Handlung:** Erkläre nicht nur, was passiert, sondern auch wie. Statt »Der Mann rennt über die Straße« beschreibe: »Der Mann sprintet panisch über eine regennasse Straße, schaut sich ständig um und wirkt gehetzt.« So entsteht Dynamik.
3. **Umgebung:** Setze die Szene mit einer klaren Kulisse. Wo spielt das Ganze? Drinnen oder draußen? Tag oder Nacht? Beispiel: »Eine futuristische Großstadt bei Nacht, Neonlichter spiegeln sich auf dem Asphalt, Drohnen schweben zwischen den Wolkenkratzern.«
4. **Licht und Atmosphäre:** Beschreibe Stimmung und Beleuchtung so genau wie möglich. Warmes Morgenlicht, kaltes Neonlicht oder dramatische Schatten machen einen riesigen Unterschied. Beispiel: »Gedämpftes, goldenes Licht fällt durch große Fenster, Staubpartikel schweben in der Luft, die Szene wirkt ruhig und nostalgisch.«
5. **Kamerabewegung:** Die KI kann nicht nur Bilder, sondern ganze Shots umsetzen, wenn du sie anleitest. Gib klare Anweisungen wie bei einem Drehbuch: »Kamerafahrt von links nach rechts in Zeitlupe« oder »Nahaufnahme auf das Gesicht, dann Zoom-Out, um die gesamte Landschaft zu zeigen.«

Wenn du mit einer KI arbeitest, die nicht nur Bilder, sondern auch Ton erzeugen kann, lassen sich diese fünf Elemente sinnvoll um drei erweitern:

6. **Sprache:** Was sagen die Figuren? Gib den exakten Text an, idealerweise mit Stimmcharakter (z. B. »tiefe, ruhige Männerstimme« oder »aufgeregte, helle Kinderstimme«).
7. **Musik:** Beschreibe Genre, Stimmung oder Instrumentierung der Hintergrundmusik (»sanfte Klavierballade«, »treibender Techno-Beat«).
8. **Soundeffekte:** Notiere Geräusche, die zur Szene passen (»Donnergrollen in der Ferne«, »leises Knistern eines Lagerfeuers«). Diese akustischen Elemente machen den Unterschied zwischen einem stummen Bild und einer echten Filmszene – und sie lassen sich in vielen modernen Text-to-Video-Tools inzwischen erstaunlich präzise umsetzen.

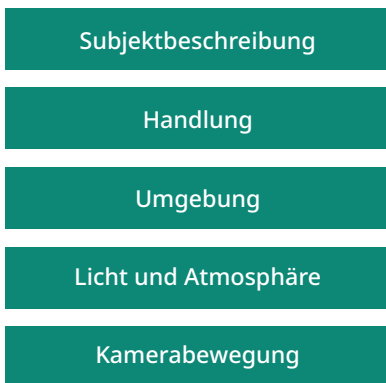


Abbildung 1.1 Fünf Videoelemente (ohne Audio) für perfekte Videoprompts

Tipp

Kombiniere alle fünf Elemente in einem Satz oder Absatz. Je vollständiger dein Prompt, desto realistischer und filmischer wird das Ergebnis.

Ein Beispiel für einen kompletten Prompt: »Eine junge Frau in einem roten Kleid läuft barfuß über einen nassen Asphalt in einer futuristischen Stadt bei Nacht. Sie hält einen transparenten Regenschirm und schaut nervös nach hinten. Neonlichter spiegeln sich auf der Straße, Drohnen fliegen über die Dächer. Die Szene ist in kühles blau-violettes Licht getaucht, leichter Regen fällt, Nebel zieht über den Boden. Kamera: Nahaufnahme auf ihr Gesicht, dann Kamerafahrt nach rechts in Zeitlupe, um die Stadt im Hintergrund zu zeigen.«

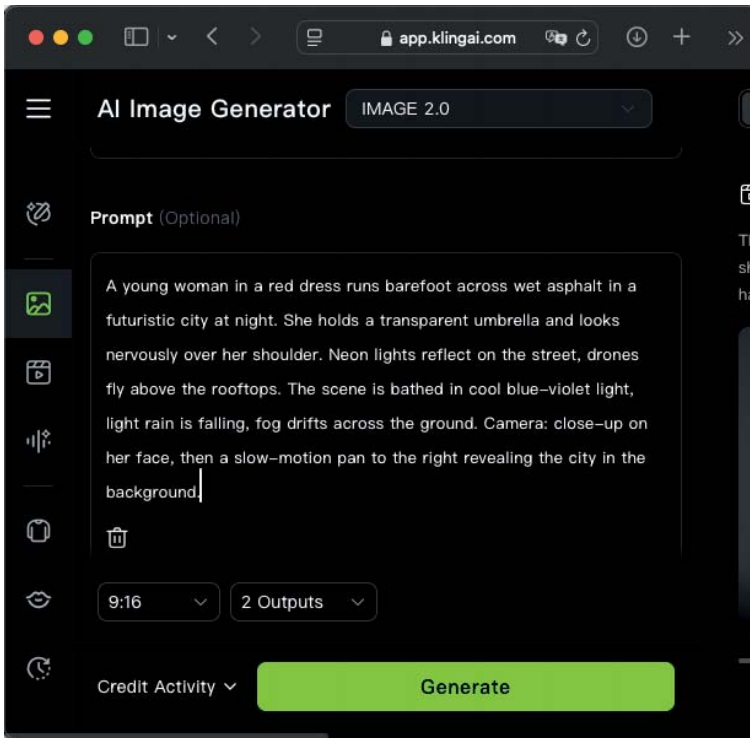


Abbildung 1.2 Ein Prompt in Kling AI

Mein Praxis-Tipp

Ein Prompt darf gerne umfangreich sein – die KI hat kein Problem mit vielen Details. Doch Vorsicht: Wenn die Handlung zu komplex wird, steigen die meisten Tools aus. Besonders schwierig wird es, wenn du mehreren Figuren unterschiedliche Aktionen geben möchtest. Beispiel: »Die Frau läuft nach links und nimmt ein Buch, während der Mann sich setzt und etwas anderes macht.« Solche Kombinationen führen oft zu fehlerhaften Ergebnissen. Gleiches gilt für zu viele Handlungsabfolgen: »Er macht zuerst das, dann geht er dorthin, danach setzt er sich.« Besser ist es, komplexe Abläufe in einzelne Szenen aufzuteilen. Arbeite mit mehreren Prompts: Zeige zuerst den einen Charakter und seine Aktion, dann den anderen in einer eigenen Szene. So erhältst du saubere und kontrollierbare Ergebnisse.

Nun möchte ich dir noch ein oft unterschätztes Werkzeug vorstellen: Viele Tools bieten eine Funktion, die kaum jemand nutzt, aber extrem hilfreich ist: **Negative Prompts**. Hier beschreibst du, was nicht im Bild erscheinen soll. Das ist besonders nützlich, wenn das KI-Ergebnis Fehler enthält oder ungewollte Details auftauchen. Beim nächsten Versuch schreibst du genau das in den negativen Prompt. Mit der Zeit wirst du lernen, typische Fehler im Voraus zu vermeiden, indem du sie direkt ausschließt.

Ein positiver Prompt würde so lauten: »Eine junge Frau in einem roten Kleid läuft barfuß über einen nassen Asphalt in einer futuristischen Stadt bei Nacht, Neonlichter spiegeln sich auf dem Boden, leichter Regen fällt.« Der negative Prompt wäre dazu: »keine verzerrten Gesichter, keine zusätzlichen Personen, keine verschwommenen Hände, keine unnatürlichen Körperhaltungen.«

Tipp

Wenn du häufig feststellst, dass z. B. Hände oder Gesichter unsauber generiert werden, füge diese Begriffe immer in den negativen Prompt ein. Das erhöht die Qualität enorm.

Licht gezielt im Prompt beschreiben

In der Filmproduktion ist Licht mehr als nur Helligkeit – es ist ein kreatives Werkzeug. Es entscheidet, ob eine Szene gemütlich wirkt oder bedrohlich, ob wir uns einer Figur nahe fühlen oder distanziert. Genau das gilt auch für KI-generierte Szenen. Wer Licht richtig beschreibt, bekommt nicht nur schönere Ergebnisse – sondern erzeugt gezielt Stimmung. Dabei gibt es verschiedene Aspekte zu beachten, von den Lichtarten über Lichtfarbe bis zur Entfernung der Lichtquelle. Gehen wir sie nacheinander durch.

Es gibt in der klassischen Filmproduktion drei fundamentale **Lichtarten** (3-Point-Lighting), die du im Prompt bedenken musst:

1. Das Hauptlicht (Key Light) ist die stärkste Lichtquelle in einer Szene. Sie bestimmt, wo das Licht herkommt, wie klar die Konturen sind und wie stark der Kontrast wirkt. In Prompts kann man das z. B. beschreiben mit: »Key Light von schräg oben rechts, warmes Licht, starke Schatten«.

2. Das Aufhelllicht (Fill Light) mildert die Schatten, die das Key Light erzeugt. Es ist weicher und weniger intensiv – oft spricht man von diffuser Beleuchtung. Im Prompt kannst du etwa schreiben: »Fill Light links vorne, weich, kaum Schatten«. In der Praxis bestimmt das Fill Light maßgeblich den Kontrast einer Szene: Je stärker es eingesetzt wird, desto gleichmäßiger und ruhiger wirkt das Bild; je schwächer, desto dramatischer fällt der Schatten.
3. Von hinten erzeugt das Gegenlicht (Back Light) eine Kontur, die das Motiv vom Hintergrund abhebt – ideal für Silhouetten oder räumliche Tiefe. Ein typischer Prompt wäre in diesem Fall: »Back Light erzeugt helle Umrisse, Hintergrund bleibt dunkel«.

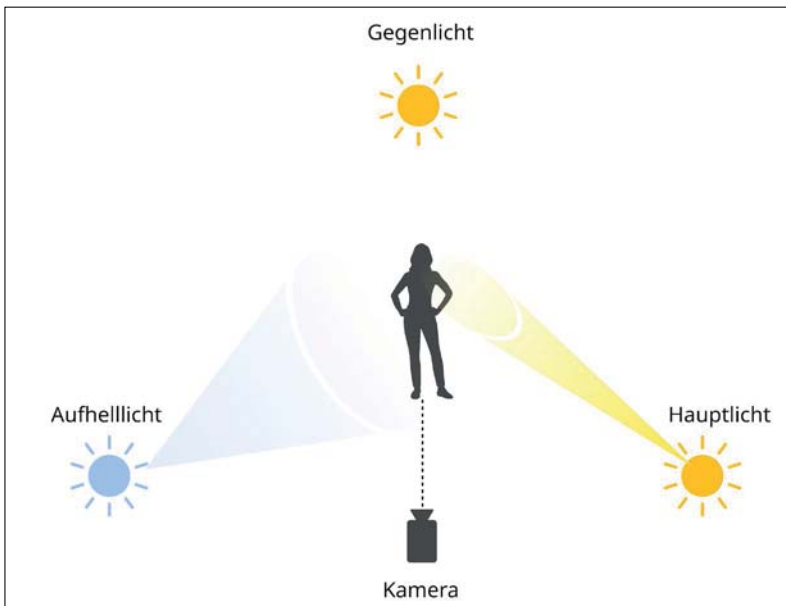


Abbildung 1.3 Drei-Punkt-Beleuchtung (Top-Down)

Auch die **Lichtqualität** sollte aufgenommen werden. Hartes Licht erzeugt scharfe Schatten und wirkt oft dramatisch, direkt, mit starkem Kontrast. Weiches Licht hingegen verteilt sich sanfter, erzeugt kaum Schatten und wirkt natürlich, schmeichelnd oder nostalgisch. Im Prompt beschreibst du das zum Beispiel so: »Weiches Licht, wie durch ein dünnes Tuch gefiltert« oder »Hartes Sonnenlicht, starke Schatten auf dem Boden«.

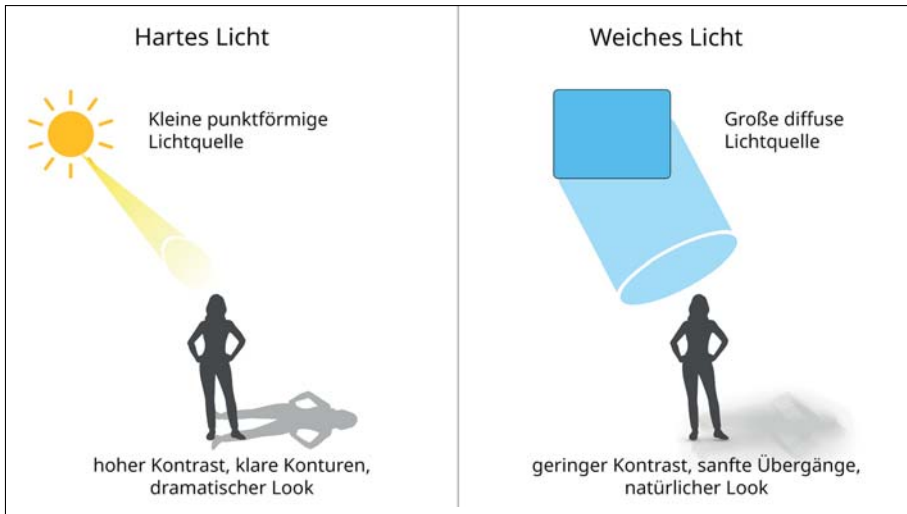


Abbildung 1.4 Lichtqualität: Hartes oder weiches Licht?

Jede Lichtquelle hat eine Temperatur – gemessen in Kelvin, die die **Lichtfarbe** beeinflusst. Warmes Licht (z. B. Kerze oder Abendsonne) liegt bei 2.000–3.500K, neutrales Tageslicht bei etwa 5.500K und Kaltes Licht (z. B. Himmel oder Neonröhren) bei 6.000–7.000K. Diese Werte musst du nicht exakt kennen, aber es hilft, wenn du Begriffe wie »warmes Licht«, »blaues Gegenlicht« oder »kaltes Neonlicht« verwendest. Auch »goldenes Abendlicht« oder »kühles Morgenlicht« sind für viele KI-Tools bereits klar interpretierbar.

In der Film- und Videoproduktion spielt die Glaubwürdigkeit von **Lichtquellen** eine zentrale Rolle, da sie Atmosphäre, Stimmung und Realismus entscheidend beeinflussen. Practical Lights sind in diesem Zusammenhang Lichtquellen, die im Bild sichtbar sind – wie eine Lampe, ein Fernseher oder Kerzen. Motivated Lighting bedeutet: Das Licht hat eine logische Quelle – es passt zur Szene. Eine Lichtquelle, die aussieht wie Sonne durch ein Fenster, sollte z. B. auch Schatten in diese Richtung werfen. Ein Beispiel für einen Prompt: »Die Szene spielt nachts, das einzige Licht kommt von einer Straßenlaterne. Gelbliches Licht fällt schräg auf das Gesicht des Mannes, der Rest bleibt im Dunkeln. Eine schwache Spiegelung im Fenster. Atmosphäre: still, melancholisch.«

Als letztes musst du den **Lichtabfall** (Falloff) und die Entfernung bedenken. Was heißt das? Licht wird schwächer, je weiter es sich vom Objekt entfernt – dieser Effekt nennt sich Falloff. Ein Licht ganz nah an der Figur erzeugt harte Übergänge

von hell zu dunkel. Wenn du das Licht weiter entfernst, wird der Übergang sanfter. Das solltest du bedenken, wenn du im Prompt beschreibst, wie stark der Kontrast sein soll. Typische Lichtstile sind:

- **High Key Lighting:** Sehr hell, kaum Schatten, freundliche Stimmung – z. B. in Werbespots oder Komödien.
- **Low Key Lighting:** Dunkle Szenerie, starke Kontraste, geheimnisvoll oder bedrohlich – oft in Thriller oder Film Noir.
- **Chiaroscuro (Rembrandt-Licht):** Kontrastreiche Ausleuchtung mit gezielten Lichtakzenten im Gesicht. Oft künstlerisch, emotional, dramatisch.

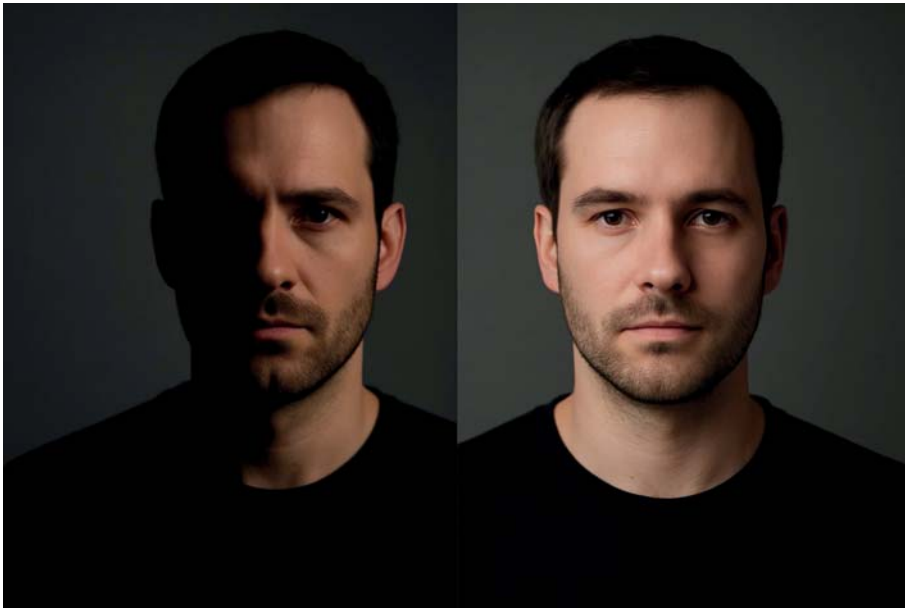


Abbildung 1.5 Hartes vs. weiches Licht (Split-Screen) – KI-generiert mit ChatGPT (DALL-E 3). Links: kleines, hohes Key bei 45°, starker Kontrast, Rembrandt-Dreieck. Rechts: große, diffuse Quelle, weiche Schatten, sanfte Übergänge

Fazit: Wer Licht beschreiben kann, steuert die Wirkung seiner Szene. Du musst kein Kameramann sein, um gute Lichtprompts zu schreiben – aber du solltest wissen, wie Licht aussieht, sich anfühlt und was es ausdrückt. Je klarer dein Prompt ist, desto besser versteht die KI deine Szene – und desto filmischer wird das Ergebnis. Wenn du möchtest, kannst du mit Begriffen wie »diffuses Fens-

terlicht«, »weiches Morgenlicht von links« oder »silber-blauer Lichtschein vom Kühlschrank« deiner Szene nicht nur Atmosphäre geben – sondern sie führen wie ein echter Regisseur.

Soviel erst einmal zur Lichtsituation. Kommen wir nun zum zweiten elementaren Aspekt: der Kamera. Eine Szene lebt nicht nur von dem, was sie zeigt, sondern auch davon, wie sie gezeigt wird. Schon kleinste Unterschiede in Kamerabewegung und -perspektive können die Wirkung komplett verändern – Spannung erzeugen, Nähe schaffen oder das Publikum bewusst irritieren. Daher schauen wir uns nun an, welche etablierten Kameratechniken es gibt, wie du sie in Prompts sinnvoll einsetzt und mit welchen Tricks sich sogar bestimmte Looks und Stimmungen nachbilden lassen – ganz ohne echtes Filmset.

Kamerabewegungen überzeugend prompten

Wenn du einen guten Prompt schreiben willst, solltest du nicht nur wissen, was in der Szene passiert, sondern auch, wie die Kamera sich dabei bewegt. Profis in der Filmwelt sprechen dabei ganz selbstverständlich von Pan, Dolly oder Crane Shot für diese Kamerabewegungen – Begriffe, die oft aus der klassischen Kameraarbeit stammen. Sie sind über Jahrzehnte gewachsen und haben sich bewährt, weil sie gezielt Emotionen erzeugen oder die Aufmerksamkeit der Zuschauer lenken. Deshalb lohnt es sich, diese Techniken zu kennen – nicht nur für das echte Filmset, sondern auch für die Arbeit mit KI-generierten Videos.

- **Pan (horizontaler Schwenk):** Beim Pan bewegt sich die Kamera horizontal von links nach rechts oder umgekehrt, ohne ihre Position zu verändern. Man nutzt diesen Schwenk häufig, um einen Raum zu zeigen, eine neue Information im Bild zu enthüllen oder einer Bewegung zu folgen. Besonders wirkungsvoll ist der Pan, wenn er die Perspektive eines Charakters übernimmt oder die Dramatik eines Moments unterstützt.
- **Tilt (vertikaler Schwenk):** Der Tilt funktioniert ähnlich wie der Pan, nur in der Vertikalen: Die Kamera neigt sich nach oben oder unten. Mit einem langsamen Tilt nach oben kannst du zum Beispiel ein beeindruckendes Bauwerk oder eine bedrohliche Erscheinung ins Bild rücken. Ein Tilt nach unten wiederum erzeugt oft das Gefühl von Kontrolle oder Entdeckung.

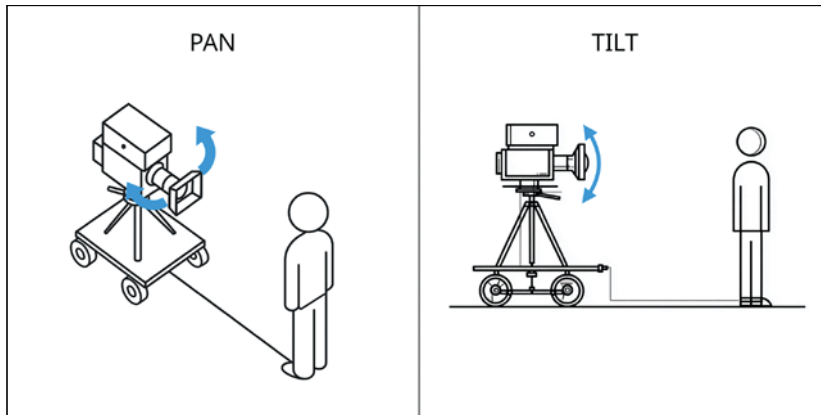


Abbildung 1.6 Pan und Tilt

- **Dolly Shot (Kamerafahrt vor oder zurück):** Ein Dolly Shot entsteht, wenn sich die Kamera physisch auf den Darsteller zu oder von ihm wegbewegt. Diese Bewegung wirkt viel natürlicher und emotionaler als ein Zoom. Sie kann Nähe und Intensität schaffen – oder umgekehrt eine emotionale Distanz aufbauen, etwa wenn sich die Kamera langsam von einer Figur entfernt.
- **Tracking Shot (seitliche Kamerafahrt oder Mitverfolgung):** Der Tracking Shot bewegt sich mit einer Figur mit – meist seitlich oder hinterher. Dadurch entsteht das Gefühl, Teil der Bewegung zu sein. Besonders in Verfolgungsszenen oder Momenten mit viel physischer Aktion erzeugt diese Technik Spannung und Immersion.

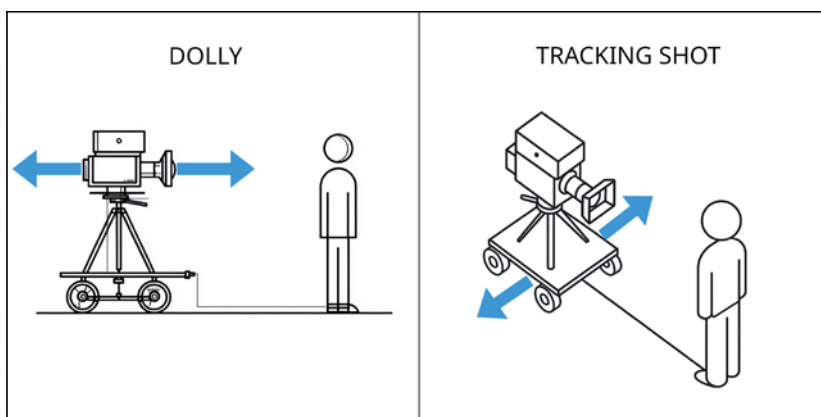


Abbildung 1.7 Dolly Shot und Tracking Shot

- **Crane Shot (Kamerabewegung von oben nach unten oder umgekehrt):** Mit einem Kran oder Kameraarm lassen sich große vertikale Bewegungen realisieren. Eine Kamera, die sich von einem hohen Punkt senkt, kann eine Szene feierlich eröffnen oder die Dramatik eines Moments unterstreichen. Umgekehrt kann sie sich auch nach oben bewegen, um einen Überblick zu geben oder das Ende einer Szene einzuleiten.
- **Zoom (optische Vergrößerung ohne Positionsveränderung):** Der Zoom verändert die Brennweite der Kamera, ohne dass sie sich selbst bewegt. Damit wird ein Objekt größer oder kleiner im Bild – doch weil sich die Perspektive nicht verändert, wirkt der Zoom oft künstlich oder »filmisch distanziert«. Genau deshalb wird er gerne in Thrillern oder surrealen Szenen eingesetzt. Da der Zoom schlecht bildlich darzustellen ist, gibt es hierzu keine Abbildung.
- **Dolly Zoom (Vertigo-Effekt/kombinierte Fahrt und Zoom):** Der Dolly Zoom kombiniert eine Vor- oder Rückwärtsbewegung der Kamera mit einem gleichzeitigen Zoom in die entgegengesetzte Richtung. Das Ergebnis ist ein optisch verstörender Effekt, bei dem das Motiv scheinbar gleich groß bleibt, während sich der Hintergrund verzerrt. Bekannt wurde diese Technik durch Alfred Hitchcock – sie eignet sich besonders für Momente emotionaler Desorientierung.

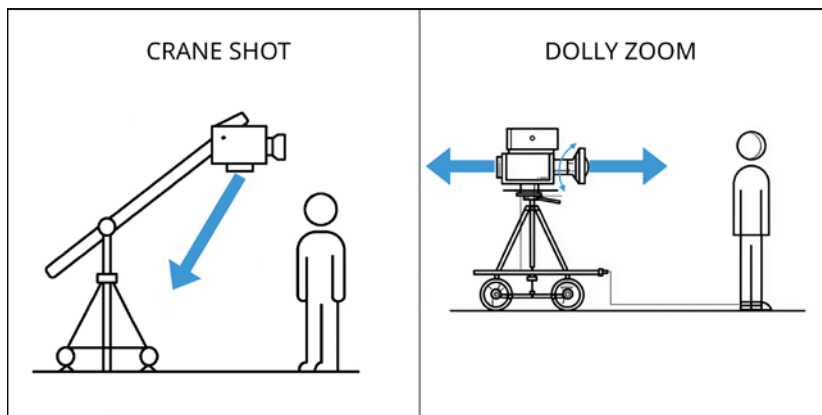


Abbildung 1.8 Crane Shot und Dolly Zoom

- **Handheld (wackelige Handkamera):** Bei dieser Technik wird die Kamera von der Hand getragen – entweder bewusst unruhig oder subtil vibrierend. Sie vermittelt Nähe, Authentizität oder auch Chaos. Oft fühlt man sich dabei

wie mitten im Geschehen, was sich besonders für dokumentarische, subjektive oder intensive Szenen eignet. Auch in KI-generierten Szenen kann ein »Handheld« im Prompt die Wahrnehmung natürlicher und spontaner wirken lassen – etwa als würde die Szene aus der Perspektive eines Reporters oder Protagonisten gefilmt.

- **Steadicam (stabilisierte Kamerabewegung):** Die Steadicam sorgt für ruhige, fließende Kamerabewegungen, selbst wenn der Kameramann sich bewegt. Sie ist das Gegenteil zur Handkamera – gleich stabil, aber mit mehr Mobilität als ein Dolly. Mit ihr können Figuren über längere Strecken verfolgt werden, ohne dass die Bewegung unruhig wirkt.

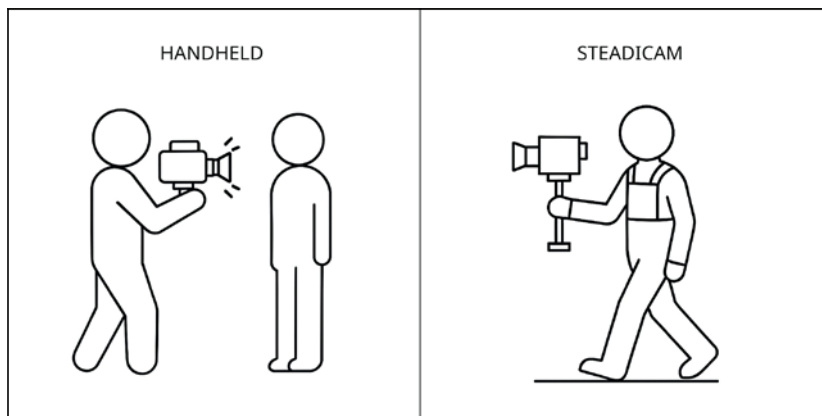


Abbildung 1.9 Handheld und Steadicam

- **Whip Pan (schneller Schwenk mit Bewegungsunschärfe):** Der Whip Pan ist ein rasanter horizontaler Schwenk, bei dem das Bild oft verschwimmt. Diese Technik wird eingesetzt, um Dynamik zu erzeugen oder als Übergang zwischen zwei Szenen. Sie kann Spannung aufbauen oder visuell ausdrücken, dass sich etwas blitzschnell verändert hat.
- **Arc Shot (kreisförmige Bewegung um das Motiv):** Beim Arc Shot bewegt sich die Kamera in einem Halb- oder Vollkreis um eine Figur oder ein Objekt. Dadurch entstehen visuelle Spannung und ein Gefühl von Umrundung, Kontrolle oder innerer Bewegung. In Dialogszenen wirkt er dramatischer als eine statische Einstellung.

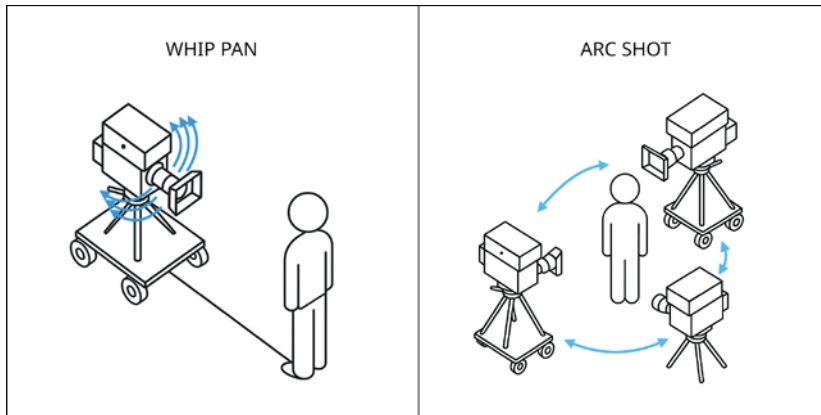


Abbildung 1.10 Whip Pan und Arc Shot

- **Boom Shot (sanfte Vertikalbewegung mit Kran oder Jib):** Der Boom Shot ähnelt dem Crane Shot, ist aber oft kleiner und feiner in der Ausführung. Die Kamera wird sanft nach oben oder unten bewegt, etwa um ein neues Element ins Bild zu bringen oder einen Ort langsam zu enthüllen.
- **Static Shot (statische Kamera):** Manchmal ist es am wirkungsvollsten, wenn sich die Kamera gar nicht bewegt. Ein statischer Shot erzeugt Ruhe, Fokus und lässt dem Spiel der Schauspieler*innen Raum. Gerade in intensiven Dialogszenen wirkt eine unbewegte Kamera oft stärker als jede Bewegung.

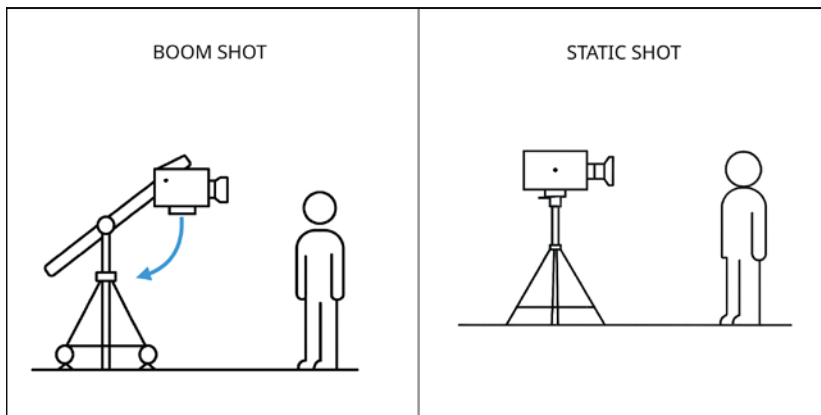


Abbildung 1.11 Boom Shot und Static Shot

- **Roll (Kameradrehung um die eigene Achse):** Bei der Roll-Bewegung dreht sich die Kamera seitlich – also nicht im Raum, sondern um ihre eigene Längsachse. Das erzeugt einen verzerrten, schwindelerregenden Effekt, der Desorientierung oder den Bruch der Realität darstellen kann. Man kennt diesen Effekt aus Bird's-Eye View.

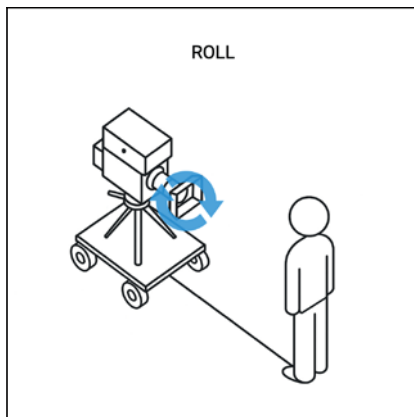


Abbildung 1.12 Roll

Die Kameraperspektive im Prompt wiedergeben

Das waren also die klassischen Kamerabewegungen. Wenn du eine Szene planst oder einen Prompt formulierst, geht es nicht nur darum, *wie* sich die Kamera bewegt, sondern auch, *von wo aus* sie das Geschehen zeigt. In der Filmsprache spricht man dabei von Perspektiven oder Blickwinkeln. Sie haben über Jahrzehnte hinweg ikonische Bedeutungen entwickelt und steuern ganz gezielt, wie ein Bild auf das Publikum wirkt. Deshalb lohnt es sich, diese Techniken genauso gut zu kennen wie die klassischen Kamerabewegungen – egal, ob am echten Filmset oder in KI-generierten Videos. Die Wahl der Perspektive beeinflusst unmittelbar die Wahrnehmung und Emotion einer Szene. Natürlich entscheidet auch in KI-generierten Szenarien die präzise Beschreibung dieser Blickwinkel darüber, ob eine Szene emotional funktioniert oder nur wie eine statische Darstellung wirkt. Diese Blickwinkel sollte man kennen:

- **Bird's-Eye View (Vogelperspektive):** Hier blickt die Kamera senkrecht von oben auf die Szene. Diese Draufsicht schafft Übersicht, betont Strukturen oder

lässt eine Situation fast wie ein Spielbrett wirken. Oft vermittelt die Vogelperspektive Distanz und kann Machtverhältnisse deutlich machen – etwa, wenn Figuren klein und verletzlich erscheinen.

- **Low Angle Shot (Froschperspektive):** Bei der Froschperspektive schaut die Kamera von unten nach oben auf eine Figur oder ein Objekt. Das lässt Personen mächtig, bedrohlich oder heroisch wirken. Ein klassisches Beispiel sind Superheldenfilme, in denen die Kamera bewusst aus der Froschperspektive filmt, um die übermenschliche Größe der Figuren zu unterstreichen.

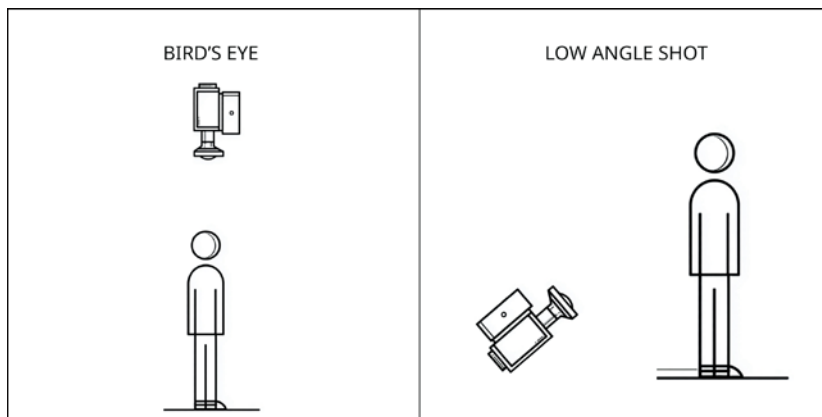


Abbildung 1.13: Bird's Eye und Low Angle Shot

- **Eye-Level Shot (Augenhöhe):** Die Kamera befindet sich auf der Höhe der Augen einer Figur. Das wirkt neutral, natürlich und unauffällig. Diese Einstellung eignet sich besonders für Dialogszenen oder Momente, in denen das Publikum auf Augenhöhe mit den Figuren sein soll – ohne Überhöhung oder Verfremdung.
- **Dutch Angle (schräger Kamerawinkel):** Beim Dutch Angle oder »verkanteten Winkel« ist die Kamera seitlich gekippt. Dadurch wirkt das Bild instabil, schräg oder unruhig. Filmemacher nutzen diese Perspektive gerne, um innere Verwirrung, Bedrohung oder Spannung zu zeigen. In Thrillern oder Horrorfilmen ist der Dutch Angle ein beliebtes Stilmittel, um den Zuschauer emotional aus dem Gleichgewicht zu bringen.

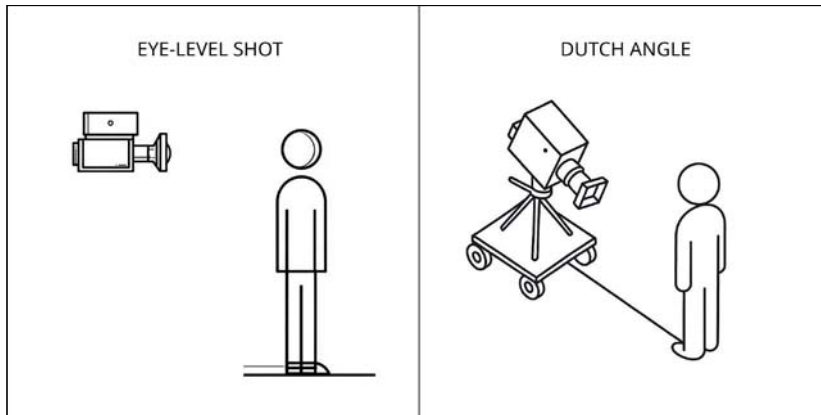


Abbildung 1.14 Eye-Level Shot und Dutch Angle

Der Kameral-Look – Stil und Charakter durch Technik

Neben den grundlegenden Elementen eines Prompts – Subjekt, Handlung, Umgebung, Licht und Kamerabewegung – gibt es zusätzliche Stellschrauben, mit denen du den Look deiner Szene noch präziser formen kannst. Eine davon ist der Kamera-Look. Er gehört nicht direkt zu den zentralen Bausteinen des Prompts, weil er nicht beschreibt, was in der Szene passiert, sondern wie diese Szene wahrgenommen wird. Trotzdem ist er ein mächtiges Stilmittel, um der KI einen visuellen Charakter vorzugeben. Der Kamera-Look bestimmt die gesamte Bildästhetik – von der Farbtiefe über Kontraste bis hin zu Schärfe, Rauschen oder Objektivverzerrungen.

Selbst wenn keine echte Kamera im Spiel ist, lassen sich diese Stile im Prompt simulieren – etwa durch Hinweise wie »gefilmt mit 35mm«, »Handycam-Ästhetik« oder »im Stil einer IMAX-Aufnahme«. Damit steuerst du also weniger den Inhalt als vielmehr die ästhetische Wirkung.

- **35mm, Super 8 & Co – der analoge Look:** Analoge Formate wie 35mm-Film oder Super 8 erzeugen ein warmes, leicht körniges Bild mit natürlicher Farbwiedergabe und weichem Kontrast. Besonders beliebt bei nostalgischen oder künstlerischen Szenen. Ein Prompt-Beispiel: »Gefilmt auf 35mm Film, warmes Farbspektrum, leichte Körnung, nostalgischer Look wie in alten Kinofilmen.«



Abbildung 1.15 Porträt einer Frau im 35mm-Filmlook – kreiert mit Midjourney

- **Digitale Kino-Kameras:** Formulierungen wie »ARRI Alexa«, »RED Komodo« oder »Blackmagic URSA« stehen für digitale High-End-Kameras mit extrem hoher Dynamik und cineastischem Look. Sie liefern saubere, gestochen scharfe Bilder – perfekt für moderne, professionelle Szenen. Ein Prompt-Beispiel: »Aufgenommen mit ARRI Alexa, weicher Bokeh-Hintergrund, sanfte Farben, cinematische Bildtiefe.«



Abbildung 1.16 Porträt einer Frau im digitalen Kinolook (ARRI Alexa) – kreiert mit Midjourney

- **Handylook & Found Footage:** Ein Smartphone-Video wirkt oft spontaner, roher, unmittelbarer. Wird gerne eingesetzt bei Doku-Stil, Vlogs oder Found-

Footage-Szenen. Ideal, wenn eine Szene »echt« oder »ungestellt« wirken soll. Ein Prompt-Beispiel: »Aufgenommen mit Smartphone, leicht überbelichtet, harte Schatten, Bild leicht verwackelt.«



Abbildung 1.17 Porträt einer Frau im Smartphone-/Found-Footage-Look – kreiert mit Midjourney

- **Vintage, VHS & Retro-Filter:** Manche Szenen sollen wirken wie aus einem alten Musikvideo oder Horrorfilm der 80er. Hier helfen Begriffe wie »VHS-Ästhetik«, »Retro Camcorder«, »Scanlines«, »verblasste Farben« oder »interlaced footage«. Ein Prompt-Beispiel: »Gefilmt mit VHS-Camcorder, Scanlines im Bild, starkes Rauschen, blasse Farben, Retro-Ästhetik.«



Abbildung 1.18 Porträt einer Frau im VHS-/Retro-Look der 80er – kreiert mit Midjourney

- **Schwarzweiß & Stilmittel:** Auch Farbe (oder deren Fehlen) ist ein Kamera-Look: Schwarzweiß erzeugt einen zeitlosen, dramatischen Stil, Sepia wirkt historisch oder warm-nostalgisch und Bleach Bypass erzeugt einen entsättigten, kontrastreichen Look, bekannt aus Filmen wie Saving Private Ryan. Ein Prompt-Beispiel: »Schwarzweißfilm, harter Kontrast, stilisiert wie ein klassischer Film Noir der 40er.«



Abbildung 1.19 Porträt einer Frau in Schwarzweiß- und Sepia-Looks – kreiert mit Midjourney

Fazit: Der Kamera-Look ist ein mächtiges Stilmittel. Ob analog oder digital, retro oder hypermodern – mit der richtigen Beschreibung kannst du den visuellen Charakter deiner Szene stark beeinflussen. Viele KI-Tools reagieren gut auf Begriffe wie »35mm«, »VHS«, »IMAX« oder »Bokeh«, weil sie gelernt haben, welche Ästhetik damit gemeint ist. Wenn du dich darauf einlässt, kannst du visuelle Handschriften großer Regisseure imitieren – oder aber deinen ganz eigenen Stil finden.

Was die Theorie nicht verrät: KIs interpretieren Prompts manchmal überraschend

All diese Hinweise und Techniken helfen dir, einen professionellen Prompt zu schreiben – aber wie die jeweilige KI diesen Prompt tatsächlich umsetzt, ist eine andere Geschichte. Nicht jede Plattform versteht die gleichen Begriffe gleich gut – und manchmal kommt es zu verblüffenden Missverständnissen. Ich habe

zum Beispiel schon erlebt, dass ich eine Kameradrohnenfahrt beschreiben wollte – und am Ende flog im fertigen Clip eine Drohne durchs Bild, statt dass die Kamera wie eine Drohne bewegt wurde. Ein Freund von mir hatte sogar eine Szene, in der plötzlich eine klassische Filmkamera durchs Bild rollte – offenbar, weil die KI die Kamerabewegung zu wörtlich genommen hat. Viele Plattformen arbeiten deshalb inzwischen mit vorgegebenen Kamerafahrten, die man aus Presets auswählen kann. Das hilft, Fehler zu vermeiden – schränkt aber auch die kreative Freiheit ein. Kurz gesagt: Die Theorie ist wichtig, aber in der Praxis musst du testen, was bei welchem Tool wirklich funktioniert. Manches klappt erstaunlich gut – anderes leider (noch) gar nicht.

1.2 Welche KI ist die beste für Text-to-Video?

Wenn du ein komplettes Video nur mit einem Textprompt erzeugen willst, stehen dir mittlerweile eine ganze Reihe an Plattformen zur Verfügung – und die Ergebnisse können beeindruckend sein. Die Qualität hat sich rasant verbessert, und manche Szenen wirken heute schon fast so, als wären sie mit einer echten Kamera gedreht worden.

Das aktuell wohl bekannteste Tool in diesem Bereich ist **Google Veo 3**. Das liegt nicht nur an der exzellenten Bildqualität, sondern vor allem an einem revolutionären Feature: Veo kann Videos inklusive Ton generieren – und dieser Ton kann sogar Sprache enthalten, die perfekt lip-synchron mit den Lippenbewegungen der dargestellten Figuren ist. Besonders bemerkenswert: Die Spracheingabe funktioniert nicht nur auf Englisch, sondern auch auf Deutsch. Du kannst also direkt in den Prompt schreiben, was deine Figur sagen soll – und das erzeugte Video liefert eine überzeugende Vertonung gleich mit. Diese Qualität hat allerdings ihren Preis. Veo 3 ist mit Abstand das teuerste Tool in dieser Kategorie.

Wenn du dagegen vor allem auf die visuelle Qualität Wert legst, kam in vielen Tests ein anderes Tool besser weg: **Seedance AI** von der chinesischen Firma ByteDance. Vor allem bei klaren Prompts und realistischen Szenen zeigt Seedance oft die besten Resultate. Die Pro-Version liefert deutlich bessere Details und flüssigere Schnitte als die Light-Version, die manchmal Artefakte zeigt.

Sehr kreativ, manchmal aber auch etwas unberechenbar, arbeitet **Kling AI**. Kling bietet einige Alleinstellungsmerkmale wie die Funktion Multi-Elements, bei der

man einzelne Elemente im Video gezielt hinzufügen, austauschen oder löschen kann. Auch der Bild-zu-Video-Modus ist spannend, da Kling hier interessante Kamerafahrten erzeugt – etwa einen langsamen Zoom aus einem AI-generierten Bild heraus.

Einen weiteren großen Namen in diesem Feld stellt **Sora** von OpenAI dar. Sora hat vor allem durch seine extrem hochwertige Text-to-Video-Qualität viel Aufmerksamkeit erregt. Die Clips wirken oft erstaunlich realistisch, mit feinen Details bei Licht, Schatten, Texturen und Bewegungen – besonders überzeugend bei Szenen, die in komplexen, realen Umgebungen spielen. Sora punktet außerdem mit längeren Videoausgaben von bis zu einer Minute, was für Storytelling und zusammenhängende Szenen ein echter Vorteil ist. Allerdings ist das Tool aktuell noch nicht frei verfügbar und wird in erster Linie professionellen Partnern sowie Testern bereitgestellt. Wer Zugang hat, berichtet jedoch von einer beeindruckenden filmischen Wirkung, die viele andere Tools derzeit nicht erreichen.

Hailuo AI

Ebenfalls empfehlenswert ist Hailuo AI, das in vielen Szenen durch eine besonders authentische Kameraführung auffällt. Hier wirken die Bewegungen oft filmischer als bei der Konkurrenz. Besonders positiv fiel Hailuo in Szenen auf, bei denen Reflexionen, Lichtstimmungen oder technische Details (wie ein Schutzhelm beim Schweißen) realistisch wiedergegeben wurden.

Was die verschiedenen Tools kosten, ist leider nicht ganz einfach zu vergleichen. Fast alle Plattformen arbeiten mit einem Credit-System, das man über ein Monatsabo auflädt. Wie viele Credits man für welches Abo bekommt, und wie viele davon ein Clip verbraucht, ist von Plattform zu Plattform unterschiedlich – genauso wie die Länge, Qualität oder der Tonumfang des fertigen Videos. Manche bieten Full HD oder 4K, andere nur einfache Auflösung. Einige generieren Ton, andere nicht.

Ich habe mir dennoch die Mühe gemacht, die durchschnittlichen Preise pro Clip (zwischen 5 und 10 Sekunden) zu analysieren. Die genauen Preise findest du in der folgenden Tabelle. Bitte beachte aber: Diese Daten sind eine Momentaufnahme und können sich jederzeit ändern. Und natürlich ist es ein bisschen so, als würde man Äpfel mit Birnen vergleichen – aber zumindest bekommst du eine grobe Orientierung, welche Plattform was kostet.

Plattform	Plan	Preis pro Monat	Credits pro Monat	Credits per Video	Preis pro Video (Euro)	Besonderheiten
Google Veo 3	Pay-per Use	kein Abo (optional \$75 Pro-Plan)	keine Credits	\$ 0,25–0,50 pro Sekunde (ohne Audio)	1,84–3,68 €	Fast vs. Standard Pricing
Hailuo AI	Ultra	\$124,99	12.000	25 (768 p), 80 (1080 p)	0,23–0,76 €	Premium Features
Kling AI	Standard/ Pro/ Premium	variiert	variiert	20 (Kling 1.6) 100 (Kling 2.0)	0,14–0,98 €	unterschiedliche Modelle
Pika Labs	Standard	\$10	700	Turbo: 5 Credits	0,01–0,02 €	viele Kreativ-Features
Seedance	Pro (jährlich)	\$62,90	156.000 (pro Jahr)	320	1,42 €	große Mengen, günstiger pro Credit
Seedance	Starter	\$29,90	3.000	320	0,87 €	geeignet für Social Media
Runway ML	Standard	ca. \$15	625	Gen-4: 12 Credits/Sek (5s=60)	0,55 €	Zugang zu Gen-3/4
OpenAI Sora	Plus	\$20	ca. 1.000	10s Video ca. 20 Credits	0,37 €	in ChatGPT Plus integriert
OpenAI Sora	Pro	\$200	ca. 10.000	20s Video ca. 40 Credits	0,37 €	1080 p, 20 s, ohne Wasserzeichen
Runway ML	Pro	ca. \$28	2.250	Gen-4 Turbo: 5 Credits/Sek	0,36 €	schnellere Modelle
Pika Labs	Fancy	\$76	6.000	Turbo: 5 Credits	0,01 €	Roll-over teilweise

Tabelle 1.1 Kostenvergleich der wichtigsten KI-Video-Plattformen

Fazit: Keines dieser Tools ist perfekt. Manche Szenen muss man mehrmals generieren (»re-rollen«), um gute Ergebnisse zu bekommen. Je besser dein Prompt ist, desto besser wird meist auch das Resultat. Wenn du einen realistischen, filmischen Look suchst, wirst du mit Hailuo oder Seedance Pro wahrscheinlich am glücklichsten. Willst du dagegen kreative Freiheit und verrückte Effekte – etwa einen Tiger, der sich plötzlich in einen Schmetterling verwandelt – dann ist Kling AI vielleicht die spannendste Option. Und wenn du Wert auf High-End-Qualität inklusive Ton legst und bereit bist, entsprechend zu investieren, ist Google Veo 3 das derzeitige Nonplusultra. Aber wie so oft gilt: Am besten ist es, du probierst selbst aus, was für dein Projekt am besten passt – denn auch die beste KI kann nur so gut sein wie der Mensch, der sie füttert.

1.3 Google Veo 3 – Kurzvideo mit integriertem Audio

Als Google im Frühjahr 2025 sein neues Videogenerierungsmodell Veo 3 vorstellte, staunten viele nicht schlecht. Zum ersten Mal war es möglich, mit Googles KI nicht nur extrem realistische Kurzvideos zu erzeugen, sondern diese auch gleich mit einer passenden Audiospur auszustatten – von Umgebungsgeräuschen über Musik bis hin zu gesprochenem Dialog mit synchronisierten Lippenbewegungen. Die Clips sind derzeit auf acht Sekunden begrenzt, können aber eine erstaunliche erzählerische Wirkung entfalten. In meinem eigenen Flow-Account ist die deutsche Sprache bei Sprachausgabe bislang noch nicht auswählbar; über bestimmte Drittanbieter-Interfaces lässt sie sich zwar nutzen, doch offiziell wird sie in Flow selbst noch nicht angeboten.

Veo 3 ist Teil der Plattform Flow, die in enger Zusammenarbeit mit Filmschaffenden entwickelt wurde. Sie kombiniert Googles modernste KI-Modelle – darunter DeepMinds Veo und Imagine sowie das Sprachmodell Gemini – zu einem Arbeitsbereich, in dem man Projekte verwalten, Szenen erstellen und Clips zu ganzen Sequenzen verbinden kann. Wer ein Google AI Pro-Abo abschließt (aktuell 21,99 Euro pro Monat), bekommt Zugriff auf Veo 3. Die Zahl der möglichen Generationen ist limitiert, in Gemini sind es momentan drei Veo-3-Videos pro Tag; in Flow hängt das Kontingent zusätzlich vom gewählten Modell ab. Aber auch Einsteiger können Veo nun testen: 100 Credits sind monatlich umsonst!