# The image prompt

Realism is decided here, before anything moves. Paste this whole doc into a chatbot and ask it to write you a prompt and it has everything it needs.

## The skeleton

Every prompt has the same six parts in the same order. Missing any one of them is where the fake look comes from.

IMAGE PROMPT SKELETON

```
[SUBJECT] {age range} {ethnicity} {gender}, {one clothing detail, plain},
{expression cue}.

[DISTANCE] Waist or mid-torso up, whole head with clear headroom, room visible
behind them.

[CAMERA] Phone propped on {surface} at {height}, looking {angle} at the person,
front camera wide-angle, {chin and ceiling visible / desk edge in frame}.

[LIGHT] {one real light source with direction and flaw}.

[SET] {room type}. Not colour-sorted, some things leaning, worn edges, small
honest clutter, not symmetrical.

[CAPTURE] Shot on iPhone 15 front camera. Kodak Portra 400 color. Natural film
grain visible in shadow areas, ISO 1600. Slight chromatic aberration at edges,
gentle vignette. RAW photo, slight camera noise, natural tonal variation.

[BANS] Only one person in frame, including posters, prints, photographs and
reflections. Plain clothing, no graphics, no text, no faces on any garment. At
most one held object. No string lights, no potted plants, no macrame, no
colour-sorted shelves.
```

Generate at 2K, not 1K. Same price on most models for four times the texture, and 1K is a direct cause of the painted look.

## Fill these in per format

Every value below was read off a real clip frame by frame. Open them and look before you write the prompt, because the table is a summary and the clip is the actual spec.

**GRWM.** leilannyjolie (https://www.tiktok.com/@leilannyjolie/video/7548934829945982263) and itsariannetorres (https://www.tiktok.com/@itsariannetorres/video/7649468140853693709). Watch where the phone is sitting and how rarely she looks at the lens.

**Podcast.** Camp Chaos (https://www.tiktok.com/@campchaospodcast/video/7641697751721397518) and this clip (https://www.tiktok.com/@podcastviralclips0/video/7631721814577663254). Watch the body angle to the unseen host and how boring the camera is.

**Street.** chris.stocks (https://www.tiktok.com/@chris.stocks/video/7564450806158707982) and geeonthestreet (https://www.tiktok.com/@geeonthestreet/video/7341745604432842027). Watch the handheld sway and how little of the interviewer you ever see.

| | GRWM / STORYTIME | PODCAST | STREET INTERVIEW | STREET POV-MIC |
|---|---|---|---|---|
| **Camera on** | vanity, propped | desk, static | handheld | handheld |
| **Height** | chin, just below eye level | chest, level straight across | chest | eye level |
| **Framing** | head-and-shoulders to mid-chest, air above hair | waist-up, modest headroom | **tight two-shot, waist up, filling most of the frame** | medium-wide, mid-thigh up, generous headroom |
| **Eyeline** | off lens, on mirror or down at products | on an unseen host, body angled 30–45° | on each other, not the lens | at the lens, flicking just past it |
| **Motion** | none, it's propped | none | real sway and micro-jitter | slight micro-motion |
| **Foreground** | products kissing bottom edge | desk surface, mic-boom base, water glass | handheld mic in frame near the subject's chin | a hand and a small wireless mic from the bottom-left |
| **Light** | bright window daylight, or genuinely dim bedroom | bright lived-in room | bright uneven daylight, faint shadows | overcast flat daylight, washed out |
| **Set** | plain wall or door | bookshelf with genuinely mixed books | storefronts, cars, signs, **only a little street at the edges** | storefronts softly blurred, plenty of street |
| **Hands** | ONE brush in ONE hand, motion blur | still, or holding a mug | interviewer holds the mic | subject's free, mic held steady |

Those last two columns are different formats and people mix them up constantly. A street interview is a tight two-shot where the mic is in frame and there is barely any street. A POV-mic clip is one person, further back, plenty of street, talking at the lens. Using POV framing for an interview gives you a distant two-shot where nobody can read the faces.

A locked-off street shot reads staged immediately, so both get handheld sway. And podcast without the 30–45° body angle reads as a monologue instead of one side of a conversation.

## Two of ours, and what the prompt asked for



Native testimonial. Phone propped low on the table below chin height, tilted up, she leans over into the lens. Face large in the upper third, one everyday item cut off at the bottom edge. (video at socialrollout.com/craft/testimonial.mp4)




Street interview. Tight two-shot, waist up, filling the frame, mic near the chin, only a little street at the edges. (video at socialrollout.com-/craft/street-interview.mp4)

The left one is the arm's-length crop doing the work. Face large, real skin, already mid-word at 0.2 seconds. That is the thumb-stop, and it holds the open even though this particular render has a text problem (see below). The right one is a street interview framed the way a street interview is supposed to be framed: close enough that both faces are readable and the mic sits in shot near the chin.

## What happens when the render ignores it




Same format as the clip above. Shot from several steps back, faces too small to read, bottom third empty pavement, four people in the background. (video at socialrollout.com/craft/street-wide.mp4)

The prompt that produced this contains the words "NOT a wide shot from a couple of steps back, NOT knees-up with empty sidewalk around them". It did it anyway.

Worth separating those two cases, because they need different fixes. If your prompt never specified a distance, add one. If it did and you got this, the answer is to regenerate or say it harder in the camera clause, not to rewrite the spec. And never fix it with a post-hoc crop, because a crop on a wide generation still reads staged.

## One hook layer, not two

The testimonial above has a real defect and it is worth seeing.

The engine baked its own caption into the footage, and the post overlay landed on top of it. Two hook layers collide on the opening frame, so the text is illegible from 0 to about 1.6 seconds, and the baked line dissolves mid-air like a glitch. That is exactly the thumb-stop window.

Add an explicit negative to the scene prompt:

BAN BAKED-IN TEXT

```
No on-screen text, captions, subtitles or watermarks rendered into the footage.
No lettering on props, walls, screens or clothing.
```

The overlay you add afterwards should be the only text layer in the clip.

## Light lines that work

Never write lighting like a DP. "Warm key from above-left, soft fill" gives you light that's too even, which is what a studio looks like and what a real room never does. Pick one of these and adapt:

LIGHT LINES

```
Single overhead bulb, shadows under the chin and eye sockets.
Cold window light from camera-left clashing with a warm lamp behind.
Green fluorescent cast from a ceiling panel, slightly overexposed forehead.
Dim bedroom, one lamp out of frame, most of the room falling into shadow.
Bright overcast daylight through a window, flat and slightly blue.
```

Bad lighting is what real footage looks like.

## The never list

| NEVER WRITE | WHAT YOU GET |
|---|---|
| **"photorealistic", "high quality"** | painted CGI, the opposite of the ask |
| **"phone", "selfie", "holding phone"** | an actual phone rendered as a prop |
| **"35mm lens"** | cinema lens, so bokeh. Name a phone camera |

| NEVER WRITE | WHAT YOU GET |
|---|---|
| **"warm natural lighting"** | golden-hour studio look |
| **"medium shot, slightly low angle"** | level, centred, professional. Describe placement instead |
| **"lived-in", "cozy", "curated"** | it art-directs the adjective and escalates it |
| **"generic, unbranded" scoped to the frame** | rows of identical blank spines, sterile |
| **stacked realism directives** | they compete. A few chosen cues beat a pile |

On some models the negative prompt is silently ignored. No error, the exclusions just do nothing. Put exclusions in as positive framing instead: "free of airbrushed smoothness, without digital art quality."

## Why the bans block exists

Widen a prompt without bans and you get invented content in the space you just opened up. A plain band tee came back with four printed faces across the chest, which at a glance reads as four extra people in the shot, and a single hairbrush became two mirrored brushes.

Every pixel you add is a pixel the model gets to invent on. Widening and banning happen in the same edit or not at all.

Scope the bans to what the person holds and wears. Scoping "generic, unbranded" to the whole frame instead gives you rows of identical blank book spines, which is fake in a different direction.

## Use a real photo as the plate

Biggest single realism gain available. Backgrounds are the number one tell because the room gets invented twice, once by the image model and again by the video model, and two rounds of invention compound.

So don't generate the room. Composite the person into an actual photograph of an actual room and hand the model a change list:

PLATE CHANGE LIST

```
Use the attached photo as the room. Keep its structure, lighting direction and
clutter. Change: {swap the sofa colour}, {different mug on the table},
{remove the wall art}. Do not restyle or tidy the room. Do not make it
symmetrical or better lit.
```

Counterintuitively, the modified plate reads as more real than a faithful one, because you keep the photo's structure, lighting physics and clutter statistics while losing the one thing you don't want, which is it being recognisable.

Your own phone photos beat stock every time. Stock is lit and staged and that's the exact quality you're escaping. A badly lit photo of a real kitchen is a better input than a professional one.

One that prompting can't fix: plates whose implied camera is at arm's length, bedrooms especially, make the model render an extended arm holding the camera. Negative phrasing fails and so does positive pose language, because the plate's geometry beats the prompt. Fix it downstream with a cheap edit pass instead:

ARM REMOVAL EDIT

```
Remove the bare arm reaching into the foreground. Keep everything else exactly
the same.
```

## Worked examples

GRWM, skincare product, mid-twenties woman.

GRWM EXAMPLE PROMPT

```
Mid-twenties woman, plain grey crewneck, slight eyebrow raise with mouth
slightly open as if about to speak.

Waist or mid-torso up, whole head with clear headroom, bathroom visible behind
her.

Phone propped on the counter at chin height, slightly below eye level, looking
faintly up at her face, front camera wide-angle, chin and ceiling visible. No
camera motion.

Single overhead bulb, shadows under the chin, slightly overexposed forehead.

Plain painted wall behind, a few skincare bottles kissing the bottom edge of
frame, one towel not folded straight.

Shot on iPhone 15 front camera. Kodak Portra 400 color. Natural film grain
visible in shadow areas, ISO 1600. Slight chromatic aberration at edges, gentle
vignette. RAW photo, slight camera noise, natural tonal variation.

Only one person in frame, including posters, prints, photographs and
reflections. Plain clothing, no graphics, no text, no faces on any garment. At
most one held object. No string lights, no potted plants, no macrame, no
colour-sorted shelves.
```

PODCAST EXAMPLE PROMPT

```
Late-twenties man, plain dark tee, mid-sentence with a slight frown as if
disagreeing.

Waist-up, modest headroom, desk surface and mic-boom base in the lower
foreground.

Camera on the desk at chest height, level straight across, front camera
wide-angle. Body angled 40 degrees toward someone off-frame to the right,
eyeline on them, never at the lens. No camera motion.

Bright room, one window out of frame camera-left, warm lamp behind causing a
colour clash.

Bookshelf behind with genuinely mixed books, different heights, some leaning,
worn spines, not colour-sorted, not symmetrical. Compact black mic on a black
boom entering the lower left corner.

Shot on iPhone 15 front camera. Kodak Portra 400 color. Natural film grain
visible in shadow areas, ISO 1600. Slight chromatic aberration at edges, gentle
vignette. RAW photo, slight camera noise, natural tonal variation.

Only one person in frame, including posters, prints, photographs and
reflections. Plain clothing, no graphics, no text, no faces on any garment. At
most one held object. No string lights, no potted plants, no macrame, no
colour-sorted shelves.
```

STREET EXAMPLE PROMPT

```
Two people, mid-twenties, plain everyday clothing, mid-conversation.

Tight two-shot from up close, both framed from roughly the waist up, filling
most of the frame. Only a little street visible at the edges. Interviewer on the
left holding a black stick mic clearly in frame near the subject's chin, subject
on the right.

NOT a wide shot from a couple of steps back. NOT knees-up with empty sidewalk
around them.

Handheld at chest height with real sway and micro-jitter. Not gimbal-locked, not
tripod-steady.

Bright overcast daylight, flat and slightly blue, no hard sun.

Real storefront depth behind them, awnings and parked cars, pavement with
visible joins. Nobody else on the street.

Shot on iPhone 15 front camera. Kodak Portra 400 color. Natural film grain
visible in shadow areas, ISO 1600. Slight chromatic aberration at edges, gentle
vignette. RAW photo, slight camera noise, natural tonal variation.

Exactly two people in frame and no others, including in the background, in
reflections, or on posters. Plain clothing, no graphics, no text, no faces on
any garment. One held object each at most. No string lights, no potted plants.
```

## Before you render

Is a distance stated. Is the camera described as a physical object on a surface at a named height. Is there exactly one light source with a direction and a flaw. Are the bans present. Is "photorealistic" anywhere in there. Is it set to 2K. Is there a real plate with a change list.