

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
Camera on	vanity, propped	desk, static	handheld
Height	chin, just below eye level	chest, level straight across	chest
Framing	head-and-shoulders to mid-chest, air above hair	waist-up, modest headroom	two-shot, mid-thigh up, medium-wide
Eyeline	off lens, on mirror or down at products	on an unseen host, body angled 30–45°	on the interviewer, off lens
Motion	none, it's propped	none	real sway and micro-jitter
Foreground	products kissing bottom edge	desk surface, mic-boom base, water glass	interviewer's hand and a tiny mic in the corner
Light	bright window daylight, or genuinely dim bedroom	bright lived-in room	real daylight, storefront depth
Set	plain wall or door	bookshelf with genuinely mixed books	real street with depth
Hands	ONE brush in ONE hand, motion blur	still, or holding a mug	free

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

What the distance actually does

Two of our own street clips, same format, same engine, posted a week apart. These are the frames at 0.2 seconds, which is all most people ever see.



Full body, both subjects small, bottom third is empty pavement, four background extras. About a third of viewers were still watching two seconds in.



Mid-thigh up, faces readable, mic in frame, storefront depth behind, nobody in the background. Held about twice as many.

Nothing about the second one is a better idea. Same script quality, same engine. The camera is closer and the frame is emptier of things nobody asked for.

Wide is the default failure because it feels safer and more cinematic. It puts the faces too far away to read at a glance, fills the bottom third with ground, and hands the model room to add people. Fix the distance before you touch anything else.

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
"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 example

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.

Two-shot, medium-wide, both from mid-thigh up, faces clearly readable.
Interviewer on the left holding a black stick mic, subject on the right.

Handheld at chest height with real sway and micro-jitter, front camera
wide-angle. Not locked off.

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.