The clip and the edit

The video prompt

Two clauses belong on every image-to-video prompt. Without them you lose the framing you just paid for and the speech pads itself.

APPEND TO EVERY I2V PROMPT

```
Movement: {one specific action, small}. Speed: natural, unhurried.
Framing: hold the starting framing throughout.
End: finish with the same framing and camera position as the first frame.
```

The end-state pin is the important one. Image-to-video creeps tighter across the clip, so you start waist-up and end on a face, which quietly undoes the framing work. Banning motion ("no zoom") half-works at best. Saying where it has to end up works.

That generalises: if something drifts, say where it has to arrive rather than forbidding how it moves. Models follow destinations much better than prohibitions.

Match the duration to the line

Hand a model a five-second line and a nine-second clip and it won't slow down or pause. It pads, usually by repeating a chunk of a sentence, and the result is uncanny in a way that's hard to place. The instinct is to blame speech quality and go looking in the wrong place entirely.

LINE LENGTH	CLIP LENGTH
Count words, divide by 2.7	snap to the nearest tier at or below that

Never round up. Add a timing cue to the prompt so it knows how words map onto seconds. Keep a repeat-detection pass downstream if you like, but that's a net, not the fix.

More generally, when a model does something strange, check whether you handed it a shape it was obliged to fill.

The beat map

BEAT	TIME	JOB
Hook	0–3s	the problem or the promise, landing by 1.5s
Value	3s to ~70%	escalate. One idea per line

BEAT	TIME	JOB
Payoff	~70–90%	the turn, or the reveal
CTA	last 2–3s	logo, one line, button

The hook gets a second and a half, which is about four words at speaking rate. First cut before three seconds, always. Being slow in that window is the single most common reason an otherwise fine ad dies.

Cut anything that doesn't serve the hook or the message. Not "cut the bad bits". Most overlong ads are full of individually fine footage, which is exactly why they stay overlong.

Pacing

Don't cut on a timer. Every 1.5 seconds forever feels energetic while you're making it and reads as mechanical on playback.

PACING SPEC

```
0-3s        micro-cuts, 1-2s each
key line    hold 3-4s so it lands
final 25%   tighten, accelerate the cuts
never       hold a static shot past 5s
always      something changes every ~4s: cut, b-roll, graphic, or reframe
```

That four-second interrupt is the cheapest retention lever in the edit. Watch this street clip (https://www.tiktok.com/@chris.stocks/video/7564450806158707982) and count how rarely two seconds pass without something changing.

Cut on the emphasis word, not between words. Cutting in the gap between sentences is technically clean and feels dead.

Transitions

TRANSITION	WHAT IT IS	WHEN
Blur-whip	zoom plus heavy blur across the seam, ~0.14s	cutting to different b-roll or a new location
Zoom-punch	quick scale-in and settle on the incoming shot	emphasis on a claim
Flash cut	one or two white frames	fast snappy jump
Hard cut	nothing	between talking-head shots

A transition has to land on something different. A whoosh bridging a shot to itself, a zoom of the same footage, is a squish. Nothing was revealed so it reads as a gimmick. If there's nothing to cut to, don't add the transition.

And don't use one transition everywhere. The same one ten times reads as a template.

Sound

SOUND	WHERE	HOW OFTEN
whoosh, alternating two	ordinary hard cuts	most cuts, light
swipe	directional transition	occasional
punch	zoom-in or emphasis	1–3 times
click	flash cut	occasional
shutter	a reveal: product, UI, app	reveals only
boom	the one big moment	once
riser into impact	build into the CTA, land on the card	once
pop	caption keyword	subtle, low

The whoosh's peak has to land on the cut, not its start. Sound files have a lead-in, so triggering at the cut point puts the peak late and everything feels loose. Lead the sample by its peak offset.

Boom once, riser once. If either fires twice they both stop meaning anything. No meme sounds in a produced ad.

The ten tells

1. The same sound on every cut, or one dramatic hit on repeat.
2. The same transition on every cut.
3. A transition landing on the same footage. The squish.
4. Static captions. One position, one font, one colour, word-by-word pop.
5. All b-roll in one register.
6. Static b-roll held as a frozen frame.
7. The same talking-head frame across multiple cuts.
8. A flat CTA slide, or a trailing black frame.
9. Fixed-timer pacing.
10. Meme sounds.

Eight of those ten are one mistake: repetition where there should be variety. Amateur edits pick one solution and apply it everywhere.

Fix captions first. Static single-position word-by-word captions are the fastest read on a scroll and the most common tell in AI-generated ads.

Before you export

Does the hook land by 1.5s with the first cut before 3s. List the cut timecodes and check the gaps aren't all within half a second of each other, because that's a timer. Does something change every four seconds. Does every transition land on something new. Count the most-used sound, and over four firings it's punctuation rather than design. Do the captions move.