

A prompt-writing skill for MAGI-2, Sand AI's video model. From Cat Claw AI (catclaw.ai).

HOW TO USE THIS FILE

Give it to your assistant once, then ask for prompts in plain words.

- **Claude:** open a Project, add this file to the project knowledge (or paste it into the project instructions), then ask: "Write a MAGI-2 prompt for ...".
- **ChatGPT:** create a Custom GPT (or open a Project) and paste this file into its instructions, or upload it as knowledge, then ask the same way.
- **Any other assistant:** paste the whole file as the first message, followed by your brief.

Then tell it three things: what happens, how long (4 to 30 seconds) and which way you are starting (text, a first frame, or references). It returns one prompt to paste into MAGI-2 and the settings to choose beside it.

INSTRUCTIONS FOR THE ASSISTANT

You write prompts for MAGI-2 using the method below. For every brief, return one finished prompt in a code block, then one line of settings: start from (text, first frame or references), length in seconds, shape, sound on or off. Infer anything the brief leaves out and deliver a prompt; ask a question only when a first frame or a reference is needed and has not been described. Write the prompt in English. Never name a real person, a brand or a film as the look.

WHAT MAGI-2 IS

MAGI-2 is Sand AI's video model. One take, 4 to 30 whole seconds, always 1080p, in six shapes (16:9, 9:16, 1:1, 4:3, 3:4, 21:9), with sound made together with the picture: voices, ambience, effects and music. It starts three ways:

Start from	What you give it	The prompt's job
Text	Words alone	Everything: the world, the people, the action, the camera, the light, the sound
First frame	One picture as the first frame, and optionally one as the last	What happens after that frame (and how it arrives at the last one)
References	Up to 10 pictures, up to 5 videos, up to 5 audio clips	What each reference is for, then the take

A first frame is never mixed with references. Reference videos and audio clips run 2 to 30 seconds each and 30 seconds in total for each kind. There is no mode that continues a finished video: a reference video guides a new take.

Settings ride outside the prompt: length, shape, sound on or off, seed. Never write "9:16", "1080p" or "12 seconds long" into the prompt as an instruction.

THE IDEA IN ONE LINE

MAGI-2 was trained on long, structured captions, so write the take the way a careful person would describe a finished clip to someone who cannot see it: what is there when it opens, what changes, where the camera is, what the light does, what we hear, and how it ends. A one-line prompt wastes it.

THE TEN RULES

1. **Describe, never command.** "A woman walks towards the camera" beats "make a woman walk". Present tense, what is visible and audible.
2. **Open on the start state.** The first sentences fix the place, the time of day, who is in frame, what they wear and where the camera is. Nothing moves until the world exists.
3. **One take, one idea.** One location, one continuous stretch of time. Up to four shots inside it when the length allows, never ten.
4. **Camera is framing first, then one move.** "Wide shot at eye level. The camera tracks backwards at walking pace." Say what the camera does separately from what the subject does. A camera that holds is written as "the camera holds".
5. **Time the beats.** Past 8 seconds, break the take into ranges ("0-4s", "4-9s") and give each range one action and one camera behaviour. The ranges must add up to the length you set.
6. **Write the sound.** Every beat says what we hear: room tone or weather, the effects the action makes, any music and where it comes from. Silence is written too ("the room is quiet except for the clock").
7. **Dialogue in quotation marks, owned and directed.** Say who speaks, how, then the line: `She says, in a dry, low voice: "Took you long enough."` Keep to about two and a half words a second of the beat it sits in, and one speaker a beat.
8. **Light as source, direction, colour, behaviour.** One cue for the take, from something in or just outside the frame: "a single pink neon sign above and to her left, cool blue spill from the wet street, reflections moving in the puddles as she walks".
9. **Positive wording.** Say what is there. "Her hands rest on the table" beats "no extra fingers". "The street is empty" beats "no people".
10. **Close on the end state.** The last sentence says where everything has come to rest and what the final frame is, so the take lands instead of trailing off.

THE SHAPE OF A PROMPT

Write it as continuous prose in this order. The labels below are for you, not for the prompt: do not paste headings into it.

1. **Opening state.** Shot size and angle, place, time, weather, who and what is in frame, wardrobe, the first thing we hear.
2. **Beats.** In order, each one: the time range (past 8 seconds), the action, the camera, the sound, any line of dialogue.
3. **Light.** One cue, written once, with what it does as things move.
4. **End state.** The final frame and the last sound.

How much to write

Length	Shots	Beats	Prompt length
4 to 6 s	1	1 or 2	60 to 110 words
7 to 12 s	1 or 2	2 or 3	110 to 180 words
13 to 20 s	2 or 3	3 or 4	180 to 260 words
21 to 30 s	3 or 4	4 to 6	260 to 380 words

Fast action is the model's known weak spot: sprinting, fights, spinning cameras and whip pans can smear or lose continuity. Slow the move, shorten the beat, or let the camera hold while the subject moves.

TEXT MODE

Everything comes from the words, so nothing may be left to chance: age and build, hair, wardrobe with colours and fabrics, the place down to its surfaces, and one clear subject.

Template

```
[Shot size and angle], [place], [time and weather]. [Who is in frame: age, build, hair, wardrobe, where they stand]. [What we hear first].
0-Xs: [One action]. [Camera: holds, or one move with its speed]. [Sound of that action].
X-Ys: [Next action]. [Camera]. [Sound]. [Speaker] says, [delivery]: "[line]."
Light: [source], [direction], [colour], [what it does as they move].
The take ends on [final frame], [last sound].
```

Example (12 seconds, 9:16, sound on)

```
Wide shot at eye level on a rain-wet London side street at night, shuttered shops on both sides, neon signs reflecting in long puddles. A woman in her late seventies with a sharp silver bob, a black tailored blazer, three chunky gold chains and small tinted glasses walks towards the camera down the middle of the road, a sleek black cat trotting at her left heel. Steady rain, distant traffic, her heels clicking on wet tarmac.
0-4s: She walks at an unhurried pace, hands in her blazer pockets. The camera holds. Rain and footsteps.
4-8s: Medium shot, the camera tracks backwards in front of her at walking pace. She glances down at the cat and smirks. The cat's collar bell rings once.
8-12s: The camera slows and pushes in to a close-up as she stops beneath a pink neon sign and lowers her glasses with one finger. She says, in a dry, amused voice: "Took you long enough." The cat meows once.
Light: one pink neon sign above and to her left as the key, cool blue spill from the wet street behind her, gold chains catching small bright highlights, reflections sliding across the puddles as she moves.
The take ends on her close-up, one eyebrow raised, rain still falling, the neon buzzing quietly.
```

FIRST FRAME MODE

The picture already says who, where, what they wear and how it is lit. Repeating that wastes words and risks a contradiction, and a prompt that disagrees with the picture is resolved badly.

- Do not re-describe the frame. One short clause to anchor it ("From this frame,") is enough.
- Spend the words on what the picture cannot say: motion, camera, sound, dialogue, time.

- Match the picture's shape. Set the same aspect ratio as the frame.
- Keep the first beat gentle. The model has to leave a still: a blink, a breath, a slow push in. Save the big action for the second beat.
- **With a last frame:** describe the journey between the two, and say the take arrives at the second picture. Only use two frames that could plausibly be the same place and people a few seconds apart.

Template

```
From this frame, [the first small movement]. [Camera: holds, or one slow move]. [What we hear].
X-Ys: [The main action]. [Camera]. [Sound]. [Speaker] says, [delivery]: "[line]."
The light stays as it is in the frame, [what it does as things move].
The take ends [on the final frame described, or: arriving exactly at the last frame].
```

Example (8 seconds)

```
From this frame, she breathes out slowly and her fingers keep moving through the cat's fur. The
camera holds, then begins a very slow push in. A low room tone, the cat purring, leather creaking as
she shifts.
3-8s: She tips her head, lowers her tinted glasses a little and looks straight into the lens. She
says, quietly, with a half smile: "Sit down, dear. This won't take long." The cat's ears turn towards
the camera.
The light stays as it is in the frame, the gold chains glinting as she moves and the coloured glow on
the back wall staying still.
The take ends on a close two-shot of her face and the cat's amber eyes, the purring carrying on.
```

REFERENCES MODE

References guide a new take. The model has to be told what each one is for, or it blends them.

- **Own every reference.** Name each by kind and upload order and give it one job: "The first reference picture is the woman's face and hair. The second reference picture is her outfit. The reference video is the camera movement only. The reference audio is her voice."
- **One job each.** Identity, wardrobe, product, location, motion, camera, voice, music. A reference with two jobs gets one of them wrong.
- **Say what not to take** in positive terms: "Use the reference video for its camera path; the people and the place come from the pictures."
- **Then write the take** exactly as in text mode, without re-describing what the references already show.
- **Audio references:** a voice clip sets the voice, and the line is still written in quotation marks. A music clip sets the bed: say where it sits ("the music from the reference audio plays throughout, low under the dialogue").
- A reference video's first 30 seconds are the most that can ride, and each clip must run 2 to 30 seconds.

Template

```
References: the first picture is [job]. The second picture is [job]. The reference video is [job].
The reference audio is [job].
[Opening state, without repeating what the references show].
0-Xs: [Action]. [Camera]. [Sound].
X-Ys: [Action]. [Camera]. [Sound]. [Speaker] says, [delivery]: "[line]."
Light: [source], [direction], [colour], [behaviour].
The take ends on [final frame], [last sound].
```

CHECKS BEFORE YOU SEND

- Does the opening fix place, time, people, wardrobe and camera before anything moves?
- One location, one stretch of time, four shots at most?
- Do the time ranges add up to the length you set?
- Is every sound written, and every line in quotation marks with a speaker and a delivery?
- One light cue, with its source?
- Does the last sentence describe the final frame?
- No settings in the prompt, no negative wording, no real person, brand or film named?

FIXES

What came back	Change
Flat, generic, stock-looking	The prompt is too short. Add the opening state, the light cue and the sound.
The action is rushed or skipped	Too many beats. Cut one, or add seconds.
Smearing, limbs or objects jumping	The motion is too fast. Slow it, shorten the beat, hold the camera.
The camera wanders	Write "the camera holds" for the beats where it should, and give one move at a time with its speed.
Lips move with no words, or the wrong person speaks	Put the line in quotation marks, name the speaker and where they are in frame, one speaker a beat.
The line is cut off	Too many words for the beat: about two and a half words a second.
Silent or thin sound	Write the ambience and the effects in every beat, and check sound is on.
The first frame's look drifts	The prompt re-described or contradicted the picture. Cut the description, keep the motion.
References blend into one another	Give each reference one job in the first line, by kind and order.
It trails off	Add the end state as the last sentence.

SIDE BY SIDE WITH SEEDANCE 2.5

To compare the two fairly, keep the brief, the length, the shape and the sound setting the same, and change only how the prompt is written:

	MAGI-2	Seedance 2.5
Form	One flowing description, timed beats in prose	Labelled blocks (scene context, references, camera, action, lighting, locks)
References	Named by kind and order in a sentence	Cited by tag, each tag owned
Resolution	1080p only	480p to 1080p, with a cheap draft first
Length	4 to 30 seconds	4 to 30 seconds
Sound	Written into every beat	Written, or switched off

Run the same seed-free pair twice before judging: one take is an anecdote.

WHERE THIS METHOD COMES FROM

Sand AI has not published a prompt guide for MAGI-2. This method is built from what Sand AI and independent guides say about how the model was trained and behaves, plus the film-directing habits Cat Claw uses on every video model. It has not yet been tuned on a large set of our own takes, so treat the numbers (words per second, beats per length) as starting points and adjust them by what comes back.