Writing a prompt

Full course PDF

Learn framing, camera, light, prompts and delivery through short lessons and practice.

6. Writing a prompt

What does step mean in Sampling and guidance: the two dials that shape a generation?

One pass of the model plus one scheduler update, moving the latent partway from noise toward a picture.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

What does scheduler mean in Sampling and guidance: the two dials that shape a generation?

The rule that decides step sizes and how each noise prediction is applied; also called the sampler.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

What does classifier-free guidance mean in Sampling and guidance: the two dials that shape a generation?

Running the model with and without conditioning and pushing away from the unconditional prediction to strengthen prompt adherence.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

What does guidance scale mean in Sampling and guidance: the two dials that shape a generation?

How strongly that push is applied, trading prompt fidelity against variety.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

What does distillation mean in Sampling and guidance: the two dials that shape a generation?

Training a student model to reproduce a many-step sampler's output in far fewer steps.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

What does text encoder mean in Text encoders and conditioning: how a prompt steers the picture?

A separate network that converts prompt tokens into vectors the diffusion model can attend to.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

What does contrastive training mean in Text encoders and conditioning: how a prompt steers the picture?

Learning by matching images to their captions among many wrong options, as CLIP does.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

What does cross-attention mean in Text encoders and conditioning: how a prompt steers the picture?

The mechanism by which each image patch reads a weighted mixture of the prompt token vectors at every denoising step.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

What does conditioning mean in Text encoders and conditioning: how a prompt steers the picture?

Any extra input that biases generation, including text, a start image or structural maps.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

What does frozen encoder mean in Text encoders and conditioning: how a prompt steers the picture?

A text encoder whose weights stay fixed while the diffusion model trains around it.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

In the prompt reference’s suggested order, which comes first, second and third? Shot size / Aspect ratio / Duration

Duration → Aspect ratio → Shot size

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

In the prompt reference’s suggested order, which comes first, second and third? Lens and focal length / Camera movement / Camera angle

Camera angle → Camera movement → Lens and focal length

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

In the prompt reference’s suggested order, which comes first, second and third? Colour and grade / Lighting / Depth of field

Depth of field → Lighting → Colour and grade

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

In the prompt reference’s suggested order, which comes first, second and third? Mood / Subject motion / Medium and style

Medium and style → Subject motion → Mood

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

In the prompt reference’s suggested order, which comes first, second and third? Resolution / quality tier / Pacing and cuts / Audio bed

Audio bed → Pacing and cuts → Resolution / quality tier

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

In the prompt reference’s suggested order, which comes first, second and third? Edit instruction (video editing workflows) / Ad format / mode (branded video) / First/last frame anchoring

First/last frame anchoring → Ad format / mode (branded video) → Edit instruction (video editing workflows)

Work down the aspects in order. The order is the recommended order in the prompt too.

https://shotcraft.video/prompts/

Study this original prompt example: A luxury fragrance film. A plain rectangular amber glass perfume bottle with a brushed brass cap stands on dark stone. Begin in a close product shot and slowly move the camera toward the bottle, keeping its whole silhouette visible. A broad soft light makes a clean reflection along the glass; a narrow warm backlight defines the edges. Deep charcoal surroundings, restrained amber colour, crisp cap detail. The bottle remains still as the camera settles on the final composition. Which choice describes its Framing?

Close product shot

Original starting points, not tested generation results. Adapt them to your model and inspect the output.

https://shotcraft.video/prompts/luxury-video-prompts/

Study this original prompt example: A luxury fragrance film. A plain rectangular amber glass perfume bottle with a brushed brass cap stands on dark stone. Begin in a close product shot and slowly move the camera toward the bottle, keeping its whole silhouette visible. A broad soft light makes a clean reflection along the glass; a narrow warm backlight defines the edges. Deep charcoal surroundings, restrained amber colour, crisp cap detail. The bottle remains still as the camera settles on the final composition. Which choice describes its Movement?

Slow camera push-in

Original starting points, not tested generation results. Adapt them to your model and inspect the output.

https://shotcraft.video/prompts/luxury-video-prompts/

Study this original prompt example: A luxury fragrance film. A plain rectangular amber glass perfume bottle with a brushed brass cap stands on dark stone. Begin in a close product shot and slowly move the camera toward the bottle, keeping its whole silhouette visible. A broad soft light makes a clean reflection along the glass; a narrow warm backlight defines the edges. Deep charcoal surroundings, restrained amber colour, crisp cap detail. The bottle remains still as the camera settles on the final composition. Which choice describes its Light?

Soft reflection and warm edge light

Original starting points, not tested generation results. Adapt them to your model and inspect the output.

https://shotcraft.video/prompts/luxury-video-prompts/

Study this original prompt example: A luxury fragrance film. A plain rectangular amber glass perfume bottle with a brushed brass cap stands on dark stone. Begin in a close product shot and slowly move the camera toward the bottle, keeping its whole silhouette visible. A broad soft light makes a clean reflection along the glass; a narrow warm backlight defines the edges. Deep charcoal surroundings, restrained amber colour, crisp cap detail. The bottle remains still as the camera settles on the final composition. Which choice describes its Palette?

Amber and charcoal

Original starting points, not tested generation results. Adapt them to your model and inspect the output.

https://shotcraft.video/prompts/luxury-video-prompts/

Explain why this claim is wrong: more steps always improves the image.

Sampling converges, and past the plateau the frame stops changing while the render time keeps climbing.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

Explain why this claim is wrong: the scheduler is a quality setting where one option is simply better.

It is a different route through the same trained weights. Different schedulers give genuinely different pictures from the same seed, and none is a strict upgrade.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

Explain why this claim is wrong: higher guidance means better prompt adherence.

Only up to a point. Beyond it, contrast and saturation blow out, composition collapses to one reading, and variety across seeds disappears, which is exactly the fidelity-versus-diversity trade the guidance paper describes.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

Explain why this claim is wrong: few-step distilled models are just fast versions of the same model.

They are separately distilled models with their own look, and they often fix guidance internally, so a control you relied on may not be exposed.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

Explain why this claim is wrong: the seed is the only thing that decides the output.

Scheduler, step count, guidance value and model version all change the result, so all of them belong in the record if a shot has to be reproduced.

https://shotcraft.video/learn/concepts/sampling-and-guidance/

Explain why this claim is wrong: the model reads the prompt as a sentence and follows it as an instruction.

It receives a set of token vectors, and each image patch weights those vectors for itself. There is no stage where the instruction is parsed and obeyed as a whole.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

Explain why this claim is wrong: longer, more detailed prompts always help.

Adding tokens spreads attention across more of them and can weaken the ones already working. Specific concrete nouns usually beat length.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

Explain why this claim is wrong: negative words remove things from the image.

Attention has no natural mechanism for not. Where systems do support negatives, they implement them as a separate conditioning pass, not as a word inside the prompt.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

Explain why this claim is wrong: any competent text encoder is interchangeable with any other.

Encoders capture different things. Imagen's result that scaling the text encoder beat scaling the diffusion model shows how much rides on this, and models built on different encoders answer the same phrasing differently.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

Explain why this claim is wrong: the text encoder is trained alongside the image model.

It is often frozen, trained separately on its own objective, and simply queried by the diffusion model.

https://shotcraft.video/learn/concepts/text-encoders-conditioning/

Progress can be exported to a file and imported again. With no server, that file is the only real guarantee — browsers do clear storage.