Overview

Text → Diffusion model → Image → (Style model) → Upscale model → final Image

Text

Prompt

Tokenizer

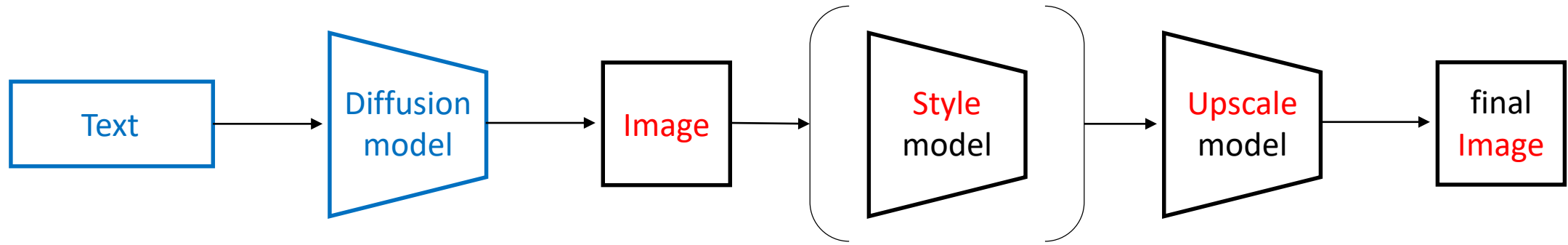Token

Embedding model

Embeddings

or

LLM

Token Embeddings

???CLIP???
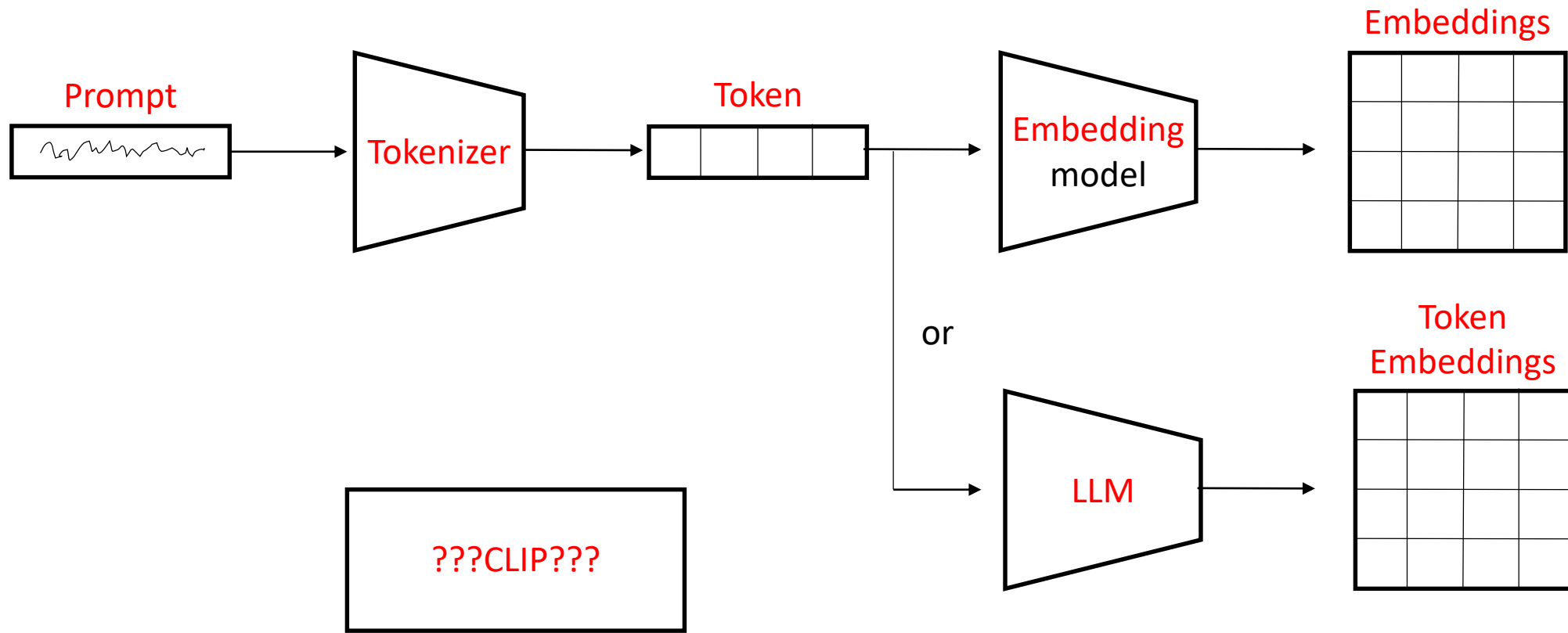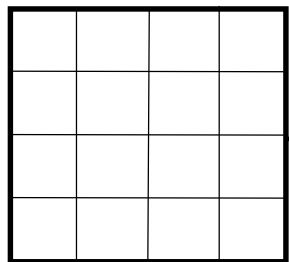
Diffusion model

Train

Embeddings

Diffusion model

output Noise

optimization

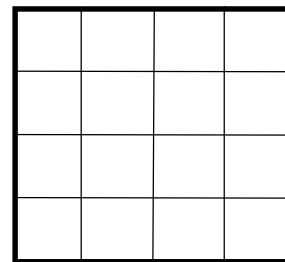random Noise

Loss

Inference

Embeddings

Diffusion model

output Noise

random Noise

subtraction

Scheduler

result

output Image

output Image → (Style model) → Upscale model → final Image