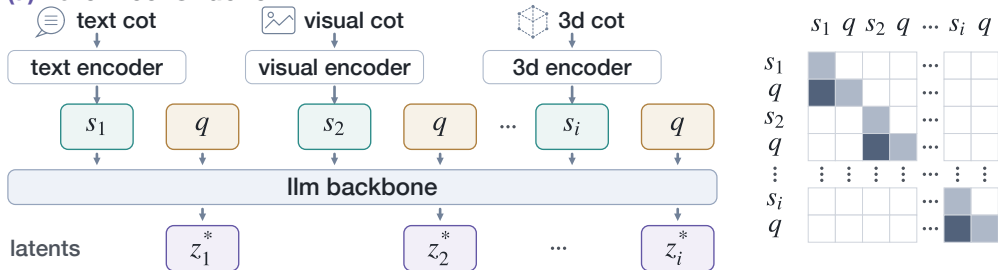
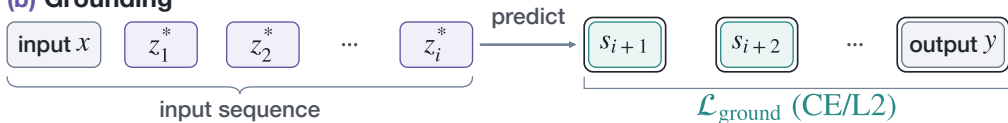


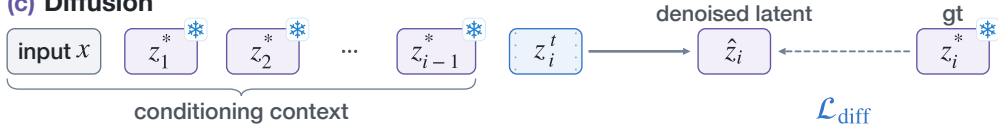
### (a) Latent construction



### (b) Grounding



### (c) Diffusion



### Inference



Legend:  $s_i$  step,  $q$  learnable embedding,  $z_i^*$  clean latent,  $z_i^t$  noisy latent,  $\text{stop grad.}$  (snowflake icon),  $\square$  loss