Table 1: CIFAR-10 posterior sampling results for CNF prior. We report expected classifier log probability, FID scores, and ELBO for the class posteriors, averaged over all 10 classes.

Model	Sampler	$\mathbb{E}[\log p(\mathbf{y} \mid \mathbf{x})]$ (↑)	FID (↓)	ELBO (↑)
I-CFM	Prior	-5.88	84.79	-24.04
	DPS	-2.22	84.96	-
	RTB	-4.20	90.77	-147.69
	Latent HMC	-2.80	46.69	-
	Adj. Matching	-3.09	19.45	-17.23
	Outsourced Diff.	-3.35	34.28	-20.36

Table 2: SD 1.5 fine-tuning results, averaged across three prompts used in [Venkatraman et al.][2]. DPOK, DDPO and RTB results taken from the same paper.

Sampler	$\mathbb{E}[\log r(\mathbf{x}, \mathbf{y})]$(↑)	CLIP diversity (↑)
Prior	-0.17	0.18
DDPO	1.37	0.09
DPOK	1.23	0.13
RTB	1.4	0.11
Outsourced Diff.	1.26	0.14

Table 3: SD3 prior and posterior results with ELBO for each prompt

Prompt	**Prior**			**Outsourced Diff.**		
	Reward	**Diversity**	ELBO	**Reward**	**Diversity**	ELBO
A cat and a dog.	0.5	0.14	3.55	1.23	0.09	27.25
A cat riding a llama.	0.79	0.18	1.01	1.53	0.14	10.83
A quiet village is disrupted by a meteor strike.	0.65	0.24	1.29	0.94	0.21	23.2
A human with a horse face and a human with a wolf face.	1.22	0.2	19.32	1.36	0.18	26.1
AVG	**0.79**	**0.19**	**6.29**	**1.27**	**0.16**	**21.85**

[2] 'A green rabbit.', 'A cat and a dog.', and 'Four roses.'

Algorithm 1 Training loop for Outsourced Diffusion Sampler

Initialize: deterministic prior function f (e.g. CNF integrators, GAN generator), randomly initialized noise posterior model p_F^ϕ, randomly initialized $Z^\phi(\mathbf{y})$ (scalar for fixed $\mathbf{y}$), VP-SDE backward policy p_B, log reward function $\log r(\mathbf{x}, \mathbf{y})$, on-policy update fraction p.
for each step $n = 1, 2, \ldots, N$ **do**
  Sample a batch of trajectories: $\{\tau^{(i)}\}_{i=1}^B \sim p_F^\phi(\tau \mid \mathbf{y})$
  for $i = 1, \ldots, B$ **do**
    Compute log density: $\log R^{(i)} \leftarrow \log \mathcal{N}(\mathbf{z}^{(i)}; \mathbf{0}, \mathbf{I}) + r\Big(f(\mathbf{z}^{(i)}), \mathbf{y}\Big)$
    Store experience $(\tau^{(i)}, \log R^{(i)})$ in replay buffer $\mathcal{D}$
  end for
  Draw $u \sim \text{Uniform}(0, 1)$
  if $u \leq p$ **then**
    Keep on-policy batch $\{(\tau^{(i)}, \log R^{(i)})\}_{i=1}^B$
  else
    Sample off-policy batch $\{(\tau^{(i)}, \log R^{(i)})\}_{i=1}^B \sim \mathcal{D}$
  end if
  Compute $\mathcal{L}_{\text{TB}}(\tau; \mathbf{y}, \phi)$ for batch using TB loss eq(4).
  Update p_F^ϕ, $Z^\phi(\mathbf{y})$ using $\nabla_\phi \mathcal{L}_{\text{TB}}(\tau; \mathbf{y}, \phi)$.
end for