

# Representation Entanglement for Generation: Training Diffusion Transformers Is Much Easier Than You Think

## Supplementary Materials

### A Discriminative semantics in inference

We posit that REG’s semantic reconstruction capability during inference stems from two key design elements: (1) the architectural entanglement of class token with image latents during training, and (2) the consistent application of SiT’s [2] velocity prediction loss to both them. Comparative analysis reveals: (1) The One learnable Token (OLT) method (see Fig. 4(a)) concatenates noised latents with a learnable token and only calculates velocity prediction loss  $\mathcal{L}_v$  on dense features. In contrast, REG (see Fig. 4(c)) entangles one high-level noised class token with low-level noised latent features while computing velocity prediction losses  $\mathcal{L}_{pred}$  for both components.

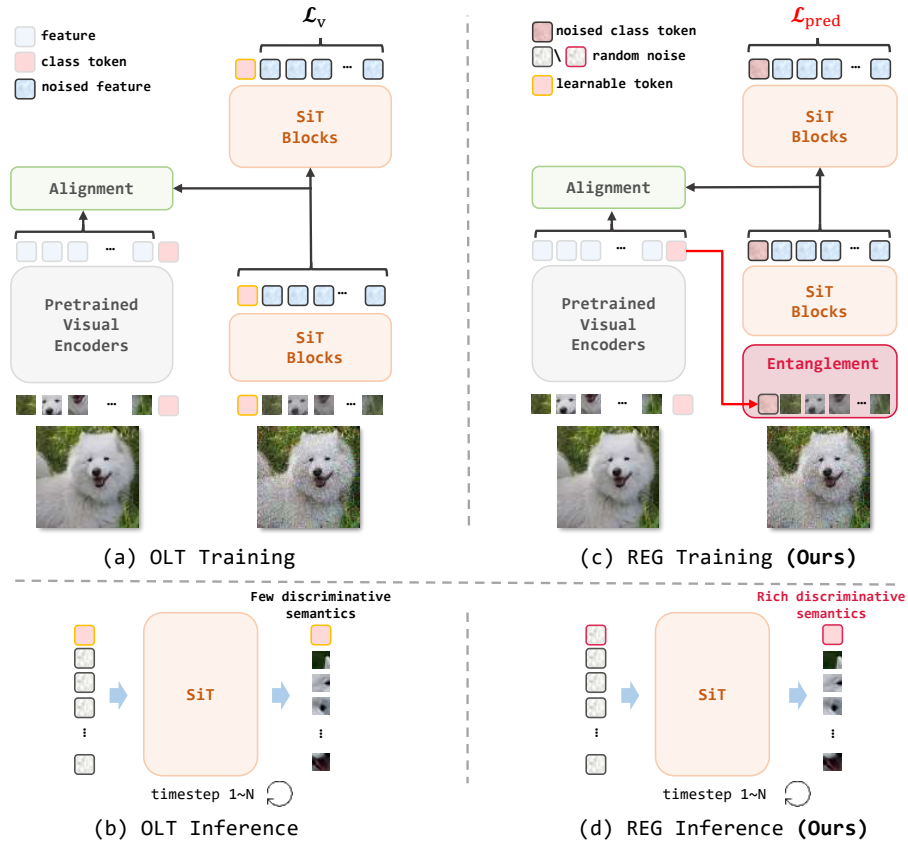


Figure 4: Comparison between the One Learnable Token (OLT) method and REG for generation. (a) During training, OLT simply concatenates noised latents with a learnable token while computing velocity prediction loss  $\mathcal{L}_v$  on dense features. (b) OLT’s output learnable token demonstrates minimal discriminative semantics after multi-step denoising. (c) REG entangles one high-level noised class token with low-level noised latent features while computing velocity prediction losses  $\mathcal{L}_{pred}$  for both components. (d) REG can reconstruct the corresponding global semantics of image latents with rich discriminative semantics.

To quantitatively validate the two methods’ discriminative semantics in inference, we use 10,000 ImageNet validation images as input processed through identical noise injection via the VAE encoder [15]. REG’s inference integrates noised latents with the noise-initialized class token through concatenation before multi-step denoising (see Fig. 4(d)), while OLT similarly processes noised latents with its learnable token (see Fig. 4(b)). Then, we process these ImageNet validation images through DINOv2 [23] to obtain the reference class token. Fig. 5 computes both CKNNA and cosine

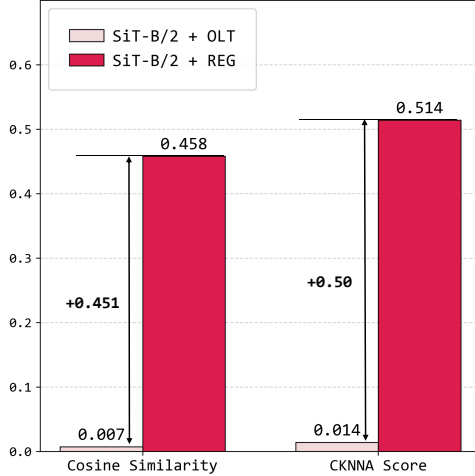


Figure 5: Quantitative evaluation of cosine similarity and CKNNA score between OLT’s learnable token, REG’s class token, and DINOv2-g’s reference token. Results demonstrate REG’s superior semantic retention during inference compared to OLT.

similarity between the reference class token and the output from REG’s class token/OLT’s learnable token using SiT-B/2 as backbone. The results demonstrate a significant disparity: OLT’s learnable token achieves only 0.007 cosine similarity and 0.014 CKNNA scores, while REG’s class token reaches 0.458 and 0.514, respectively. This empirical evidence confirms REG’s superior capacity to preserve discriminative semantics compared to OLT’s limited representation capability in inference.

## B Analysis of training overhead in REG

We summarize the total training overhead in Tab. 7, reporting the costs required to reach the same performance upper bounds claimed in the original SiT and REPA papers. All experiments are conducted on 8 NVIDIA A40 GPUs. Our results show that REG requires only 110K training steps to reach the performance level of SiT trained for 7M steps, reducing GPU hours by 98.36%. Moreover, compared with REPA, REG achieves the performance of its 4M counterpart with only 170K iterations, reducing GPU hours by 95.72%. These results highlight the training efficiency of REG, demonstrating faster convergence and significantly lower training overhead compared to prior methods.

Table 7: Training overhead comparison. REG achieves comparable performance to other models while significantly reducing training time.

Model	FID↓	Training step↓	All GPU hours↓
SiT-XL/2	8.3	7M	2380
<b>+ REG (ours)</b>	<b>8.2</b>	<b>110K</b>	<b>39 (-98.36%)</b>
+ REPA	5.9	4M	1800
<b>+ REG (ours)</b>	<b>5.6</b>	<b>170K</b>	<b>77 (-95.72%)</b>

## C 256×256 ImageNet

Tab. 8 presents extended training results with CFG using the REPA’s same guidance interval [56], demonstrating REG’s excellent performance with a 1.40 FID at 480 epochs; it achieves better performance comparable to REPA [1] at 800 epochs while requiring fewer than 40% of the training iterations. In addition, Tab. 9 presents more specific performance details of SiT + REG, further highlighting its superior robustness and accelerated convergence. Tab. 10 presents quantitative performance metrics of SiT-XL + REG under varying classifier-free guidance scale  $w$ .

Table 8: Extended REG training on ImageNet 256×256 with CFG demonstrates progressive performance gains.

Method	Epochs	FID↓	sFID↓	IS↑	Pre.↑	Rec.↑
<i>Pixel diffusion</i>						
ADM-U [43]	400	3.94	6.14	186.7	0.82	0.52
VDM++ [44]	560	2.40	-	225.3	-	-
Simple diffusion [45]	800	2.77	-	211.8	-	-
CDM [46]	2160	4.88	-	158.7	-	-
<i>Latent diffusion, U-Net</i>						
LDM-4 [15]	200	3.60	-	247.7	0.87	0.48
<i>Latent diffusion, Transformer + U-Net hybrid</i>						
U-ViT-H/2 [47]	240	2.29	5.68	263.9	0.82	0.57
DiffiT* [48]	-	1.73	-	276.5	0.80	0.62
MDTV2-XL/2* [18]	1080	1.58	4.52	314.7	0.79	0.65
<i>Latent diffusion, Transformer</i>						
MaskDiT [49]	1600	2.28	5.67	276.6	0.80	0.61
SD-DiT [50]	480	3.23	-	-	-	-
DiT-XL/2 [16]	1400	2.27	4.60	278.2	0.83	0.57
SiT-XL/2 [2]	1400	2.06	4.50	270.3	0.82	0.59
+ REPA	800	1.42	4.70	305.7	0.80	0.65
+ REG (ours)	80	1.86	4.49	321.4	0.76	0.63
+ REG (ours)	160	1.59	4.36	304.6	0.77	0.65
+ REG (ours)	300	1.48	4.31	305.8	0.77	0.66
+ REG (ours)	480	1.40	4.24	296.9	0.77	0.66
+ REG (ours)	800	1.36	4.25	299.4	0.77	0.66

Table 9: More performance analysis of SiT + REG across model scales without CFG.

Model	#Params	Iter.	FID↓	sFID↓	IS↑	Prec.↑	Rec.↑
SiT-B/2 [2]	130M	400K	33.0	6.46	43.7	0.53	0.63
+ REPA	130M	400K	24.4	6.40	59.9	0.59	0.65
+ REG (ours)	132M	50K	64.7	9.47	23.2	0.40	0.51
+ REG (ours)	132M	100K	36.1	7.74	45.5	0.53	0.61
+ REG (ours)	132M	200K	22.1	7.19	72.2	0.60	0.63
+ REG (ours)	132M	400K	15.2	6.69	94.6	0.64	0.63
SiT-L/2 [2]	458M	400K	18.8	5.29	72.0	0.64	0.64
+ REPA	458M	400K	10.0	5.20	109.2	0.69	0.65
+ REG (ours)	460M	50K	30.1	8.92	52.6	0.58	0.57
+ REG (ours)	460M	100K	11.4	5.36	108.8	0.70	0.60
+ REG (ours)	460M	200K	6.6	5.16	145.4	0.73	0.63
+ REG (ours)	460M	400K	4.6	5.21	167.6	0.75	0.63
SiT-XL/2 [2]	675M	7M	8.3	6.32	131.7	0.68	0.67
+ REPA	675M	4M	5.9	5.73	157.8	0.70	0.69
+ REG (ours)	677M	50K	26.7	16.49	59.2	0.60	0.54
+ REG (ours)	677M	100K	8.9	5.50	125.3	0.72	0.59
+ REG (ours)	677M	200K	5.0	4.88	161.2	0.75	0.62
+ REG (ours)	677M	400K	3.4	4.87	184.1	0.76	0.64
+ REG (ours)	677M	1M	2.7	4.93	201.8	0.76	0.66
+ REG (ours)	677M	2.4M	2.2	4.79	219.1	0.76	0.66
+ REG (ours)	677M	4M	1.8	4.59	230.8	0.77	0.66

## D 512×512 ImageNet

To further validate REG’s effectiveness, we conduct experiments at 512×512 resolution following REPA’s protocol [1]. The RGB images are processed through the VAE [15] to yield 64×64×3 latents, with DINOv2 [23] providing both dense features and class token from 448×448 inputs. As demonstrated in Tab. 11, REG surpasses the performance of REPA trained for 200 epochs and SiT-XL/2 trained for 600 epochs in terms of FID at only 80 epochs, demonstrating its superior effectiveness.

Table 10: The results of SiT-XL + REG at 2.4M training iterations under varying classifier-free guidance scale  $w$ , employing the guidance interval method [56].

Model	#Params	Iter.	Interval	$w$	FID↓	sFID↓	IS↑	Prec.↑	Rec.↑
SiT-XL/2 [2]	675M	7M	[0, 1]	1.50	2.06	4.50	270.3	0.82	0.59
+ REG (ours)	675M	2.4M	[0, 0.8]	2.4	1.45	4.32	280.44	0.77	0.67
+ REG (ours)	675M	2.4M	[0, 0.85]	2.4	1.41	4.24	299.65	0.77	0.67
+ REG (ours)	675M	2.4M	[0, 0.9]	2.4	1.61	4.21	334.50	0.79	0.64
+ REG (ours)	675M	2.4M	[0, 0.85]	2.5	1.43	4.25	303.11	0.77	0.67
+ REG (ours)	675M	2.4M	[0, 0.85]	2.4	1.41	4.24	299.65	0.77	0.67
+ REG (ours)	675M	2.4M	[0, 0.85]	2.3	1.40	4.24	296.93	0.77	0.66
+ REG (ours)	675M	2.4M	[0, 0.85]	2.2	1.40	4.25	293.57	0.77	0.67

Table 11: Performance comparison on ImageNet 512×512 with CFG.

Model	Epochs	FID↓	sFID↓	IS↑	Pre.↑	Rec.↑
<i>Pixel diffusion</i>						
VDM++ [44]	-	2.65	-	278.1	-	-
ADM-G, ADM-U [43]	400	2.85	5.86	221.7	0.84	0.53
Simple diffusion (U-Net) [45]	800	4.28	-	171.0	-	-
Simple diffusion (U-ViT, L) [45]	800	4.53	-	205.3	-	-
<i>Latent diffusion, Transformer</i>						
MaskDiT [49]	800	2.50	5.10	256.3	0.83	0.56
DiT-XL/2 [16]	600	3.04	5.02	240.8	0.84	0.54
SiT-XL/2 [2]	600	2.62	4.18	252.2	0.84	0.57
+ REPA	80	2.44	4.21	247.3	0.84	0.56
+ REPA	100	2.32	4.16	255.7	0.84	0.56
+ REPA	200	2.08	4.19	274.6	0.83	0.58
+ REG (ours)	80	<b>1.68</b>	<b>3.87</b>	<b>306.9</b>	0.80	<b>0.63</b>

## E Experimental setup

**Hyperparameter setup.** Tab. 12 presents the hyperparameter configurations of SiT + REG across different model scales. Following REPA’s experimental protocol [1], we employ AdamW [57, 58] optimization with a batch size of  $1 \times 10^{-4}$  and adopt DINOv2-B [23] as the optimal alignment model, maintaining 250 denoising steps for all inference processes.

**CKNNA score.** We adopt REPA’s CKNNA computation [1], calculating scores exclusively between spatially averaged dense features from both the generative model and DINOv2-g representations [23]. To ensure fair comparison, the class token is explicitly excluded from all CKNNA calculations.

## F Limitations

Due to limitations in computing resources, we plan to conduct extended training of REG with additional iterations at higher resolutions and under varied experimental configurations.

## G Broader Impacts

REG provides a principled framework for rethinking discriminative-generative model integration, demonstrating how strategic utilization of pretrained vision representations can systematically enhance diffusion model performance while maintaining computational efficiency.

## H More visualization results

We present more visualization results of REG in Fig. 6 - 25 with CFG ( $w = 4.0$ ).

Table 12: Hyperparameter settings across different model scales.

Backbone	SiT-B	SiT-L	SiT-XL
<b>Architecture</b>			
#Params	132M	460M	677M
Input	$32 \times 32 \times 4$	$32 \times 32 \times 4$	$32 \times 32 \times 4$
Layers	12	24	28
Hidden dim.	768	1,024	1,152
Num. heads	12	16	16
<b>REG settings</b>			
$\beta$	0.03	0.03	0.03
$\lambda$	0.5	0.5	0.5
Alignment depth	4	8	8
$\text{sim}(\cdot, \cdot)$	cos. sim.	cos. sim.	cos. sim.
Encoder $\mathcal{E}_{VF}(I)$	DINOv2-B	DINOv2-B	DINOv2-B
<b>Optimization</b>			
Batch size	256	256	256
Optimizer	AdamW	AdamW	AdamW
lr	0.0001	0.0001	0.0001
$(\beta_1, \beta_2)$	(0.9, 0.999)	(0.9, 0.999)	(0.9, 0.999)
<b>Interpolants</b>			
$\alpha_t$	$1 - t$	$1 - t$	$1 - t$
$\sigma_t$	$t$	$t$	$t$
$w_t$	$\sigma_t$	$\sigma_t$	$\sigma_t$
Training objective	v-prediction	v-prediction	v-prediction
Sampler	Euler-Maruyama	Euler-Maruyama	Euler-Maruyama
Sampling steps	250	250	250

Figure 6: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Great white shark” (2).





Figure 7: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Bald eagle” (22).



Figure 8: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Great grey owl” (24).



Figure 9: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Macaw” (88).



Figure 10: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Sulphur-crested cockatoo” (89).



Figure 11: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Koala” (105).



Figure 12: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “American coot” (137).





Figure 13: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Lesser panda” (156).



Figure 14: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Border collie” (232).



Figure 15: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Timber wolf” (269).





Figure 16: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Polecat” (358).



Figure 17: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Lesser panda” (387).



Figure 18: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Giant panda” (388).



Figure 19: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Castle” (483).



Figure 20: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “China cabinet” (495).



Figure 21: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Convertible” (511).





Figure 22: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Bubble” (971).



Figure 23: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Geyser” (974).



Figure 24: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Lakeside” (975).





Figure 25: The visualization results of SiT-XL/2 + REG use CFG with  $w = 4.0$ , and the class label is “Volcano” (980).

## NeurIPS Paper Checklist

### 1. Claims

Question: Do the main claims made in the abstract and introduction accurately reflect the paper’s contributions and scope?

Answer: [\[Yes\]](#)

Justification: The abstract and introduction clearly outline the paper’s scope and contributions, with a strong focus on our key findings and experimental results.

Guidelines:

- The answer NA means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

### 2. Limitations

Question: Does the paper discuss the limitations of the work performed by the authors?

Answer: [\[Yes\]](#)

Justification: See the supplementary material.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.

- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

### 3. Theory assumptions and proofs

Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This paper does not include theoretical proofs.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

### 4. Experimental result reproducibility

Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

Answer: [Yes]

Justification: See the supplemental material and code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general, releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  - (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  - (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  - (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  - (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

## 5. Open access to data and code

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: See the supplemental material and code.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so “No” is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (<https://nips.cc/public/guides/CodeSubmissionPolicy>) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

## 6. Experimental setting/details

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

Answer: [Yes]

Justification: See the supplemental material and code.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.



- The full details can be provided either with the code, in appendix, or as supplemental material.

## 7. Experiment statistical significance

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [NA]

Justification: This line of experimentation represents a distinct research direction from ours.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

## 8. Experiments compute resources

Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

Answer: [Yes]

Justification: See the experiment part and supplemental material.

Guidelines:

- The answer NA means that the paper does not include experiments.
- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

## 9. Code of ethics

Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics <https://neurips.cc/public/EthicsGuidelines>?

Answer: [Yes]

Justification: Yes, our research methodology has been conducted in full accordance with the ethical guidelines established by NeurIPS.

Guidelines:

- The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.

- If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

#### 10. **Broader impacts**

Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

Answer: [\[Yes\]](#)

Justification: See the supplemental material.

Guidelines:

- The answer NA means that there is no societal impact of the work performed.
- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

#### 11. **Safeguards**

Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [\[NA\]](#)

Justification: Null.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

#### 12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: The construction of derivative datasets and subsequent research applications were predicated on obtaining express written consent from all originating data stewards. Each contributing entity formally validated their approval via email correspondence, unequivocally supporting our academic pursuits. These verified permission records, with all personally identifiable information redacted, will be disclosed alongside the published work.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, [paperswithcode.com/datasets](https://paperswithcode.com/datasets) has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

### 13. New assets

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: We have meticulously recorded the complete experimental methodology and manuscript development procedures.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

### 14. Crowdsourcing and research with human subjects

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: Null.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.



**15. Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: Null.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

**16. Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: Null.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (<https://neurips.cc/Conferences/2025/LLM>) for what should or should not be described.