

457 Appendix

458 A LII upper-bounds expected energy gap

459 We prove the following result to formalise the intuition that stable layers sit close to their energy
460 basins. **Lemma 1** states that for every layer ℓ

$$\mathbb{E}_x[\Delta\mathcal{E}^\ell(x)] \leq L_\ell \text{LII}^\ell + o(1), \quad (7)$$

461 where $\Delta\mathcal{E}^\ell(x)$ is the Hopfield-energy gap between the attention state reached on input x and the
462 global minimum, and L_ℓ is a finite layer-dependent constant. Thus, the median absolute deviation of
463 the operational mode—the *Layer Instability Index*—provides a linear upper bound on the expected
464 energy sub-optimality, with the residual term vanishing exponentially in sequence length. Layers with
465 a small LII^ℓ are therefore already near their optimal attractor and may be safely frozen or skipped
466 without degrading convergence or generalisation.

467 *Proof.* Fix a layer ℓ and write $K(x) = \bar{k}^\ell(x)$ for the operational mode of input x . Let

$$\tilde{k} = \text{median}_x K(x) \quad (8)$$

468 and recall that by Eq. (6) the expected per-head energy gap satisfies

$$f_\ell(k) := \mathbb{E}_{h,x} [\Delta\mathcal{E}_k^{\ell,h}(x)] \leq L_\ell e^{-\gamma k}, \quad (9)$$

469 and is L -Lipschitz with constant $L = L_\ell \gamma$.

470 **Step 1: Decompose around the median.** Using Lipschitz continuity, we have, for each input x ,

$$|f_\ell(K(x)) - f_\ell(\tilde{k})| \leq L |K(x) - \tilde{k}|. \quad (10)$$

471 Taking expectation over x and rearranging,

$$\mathbb{E}_x [f_\ell(K(x))] \leq f_\ell(\tilde{k}) + L \mathbb{E}_x |K(x) - \tilde{k}|. \quad (11)$$

472 **Step 2: Median absolute deviation.** By definition,

$$\text{LII}^\ell = \text{MAD}[K] = \text{median}_x |K(x) - \tilde{k}| \leq \mathbb{E}_x |K(x) - \tilde{k}|, \quad (12)$$

473 so the second term in (11) is bounded by $L \text{LII}^\ell$.

474 **Step 3: Bounding the residual term.** Applying Eq. (6) at $k = \tilde{k}$ gives

$$f_\ell(\tilde{k}) \leq L_\ell e^{-\gamma \tilde{k}}. \quad (13)$$

475 Because \tilde{k} is a median of token counts, it grows at least logarithmically with sequence length; hence
476 $L_\ell e^{-\gamma \tilde{k}} = o(1)$ and can be absorbed into the $o(1)$ term of the lemma.

477 **Step 4: Combine.** Substituting these bounds into Eq. (11) yields

$$\boxed{\mathbb{E}_x [\Delta\mathcal{E}^\ell(x)] \leq L \text{LII}^\ell + o(1).} \quad (14)$$

478 Thus, the desired inequality follows immediately. \square

479 B Information-beometric bound: LII^ℓ upper-bounds the fisher trace

480 We establish a theoretical bound showing that the Layer Instability Index (LII) upper-bounds the
481 trace of the Fisher Information Matrix (FIM) for transformer layers. The FIM trace characterises
482 the sensitivity of the loss to parameter updates, providing insights into learning dynamics. Through
483 careful analysis of gradients via softmax logits and exponential tail bounds of attention probabilities,
484 we derive a direct relationship between the Fisher trace and the dispersion of operational modes as
485 measured by LII.

486 **Preliminaries.** Let $p_i = \text{softmax}(\beta q^\top k_i)$ with inverse temperature β . Denote the per-example
 487 loss by $\mathcal{L} = \mathcal{L}(p)$ and define the layer-wise FIM

$$F^\ell = \mathbb{E}_x \left[\nabla_{\theta_\ell} \mathcal{L}(x) \nabla_{\theta_\ell} \mathcal{L}(x)^\top \right], \quad \theta_\ell \in \{W_Q^\ell, W_K^\ell, K^\ell\}. \quad (15)$$

488 Our goal is to bound $\text{tr } F^\ell$.

489 **Step 1: Gradient of the loss w.r.t. logits.** For the logits $z_i := \beta q^\top k_i$,

$$\frac{\partial \mathcal{L}}{\partial z_i} = \sum_j \frac{\partial \mathcal{L}}{\partial p_j} \frac{\partial p_j}{\partial z_i} = \beta(p_i - \hat{y}_i), \quad (16)$$

490 where \hat{y}_i is the “effective” target (one-hot for CE).

491 **Step 2: Fisher trace through the chain rule.** Let $J_z := \partial z / \partial \theta_\ell$. Then

$$\text{tr } F^\ell = \mathbb{E}_x \|J_z^\top \nabla_z \mathcal{L}(x)\|_2^2 \leq \frac{1}{\lambda_{\min}^2} \mathbb{E}_x \|\nabla_z \mathcal{L}(x)\|_2^2, \quad (17)$$

492 where λ_{\min} is the smallest singular value of J_z (assumed layer-dependent but strictly positive for
 493 typical initialisation).

494 Using (16),

$$\|\nabla_z \mathcal{L}\|_2^2 = \beta^2 \sum_i (p_i - \hat{y}_i)^2 \leq \beta^2 \sum_i p_i^2. \quad (18)$$

495 **Step 3: Relate $\sum_i p_i^2$ to residual mass r_k .** Let $k = \bar{k}^\ell(x)$ be the operational mode and $r_k :=$
 496 $1 - \sum_{i \leq k} p_i \leq 0.1$ (by definition of $\rho = 0.9$). Then

$$\sum_i p_i^2 = \sum_{i \leq k} p_i^2 + \sum_{i > k} p_i^2 \leq \sum_{i \leq k} p_i + \max_{i > k} p_i r_k \leq 0.9 + r_k^2. \quad (19)$$

497 Assuming an exponential tail $p_{i > k} \leq p_k e^{-\gamma(i-k)}$, $r_k \leq p_k / (e^\gamma - 1) \leq C e^{-\gamma k}$.

498 **Step 4: From $e^{-\gamma k}$ to LII^ℓ .** Taking expectation over inputs and using Jensen,

$$\mathbb{E}_x [e^{-\gamma k(x)}] \leq e^{-\gamma \text{median}(k)} (1 + \gamma \text{LII}^\ell), \quad (20)$$

499 hence

$$\sum_i p_i^2 \leq 0.9 + C' e^{-\gamma \text{median}(k)} (1 + \gamma \text{LII}^\ell). \quad (21)$$

500 **Step 5: Final bound.** Substituting into (17),

$$\text{tr } F^\ell \leq \frac{\beta^2(0.9 + C')}{\lambda_{\min}^2} (1 + \gamma \text{LII}^\ell) = C_\ell \text{LII}^\ell + C_{0,\ell}, \quad (22)$$

501 where C_ℓ and $C_{0,\ell}$ are layer-dependent constants. For practical purposes $C_{0,\ell}$ is negligible once LII
 502 exceeds 10^{-2} , yielding

$$\boxed{\text{tr } F^\ell \lesssim C_\ell \text{LII}^\ell} \quad (23)$$

503 as claimed.

504 **Connection to the 1-Wasserstein distance.** Sort the attention vector of layer ℓ and head h at step t ,
 505 $a_{1:N}^\downarrow(t)$, and define its empirical cumulative distribution function (CDF) $F_t(m) = \sum_{i=1}^m a_i^\downarrow(t)$.
 506 Because tokens are indexed by their rank, the earth-mover (1-Wasserstein) distance between two
 507 attention snapshots is simply

$$W_1(a(t), a(t')) = \sum_{m=1}^N |F_t(m) - F_{t'}(m)|. \quad (24)$$

Let $\rho = 0.9$ and let k_t be the minimal m such that $F_t(m) \geq \rho$ (operational mode). Then any deviation $|k_t - k_{t'}|$ shifts at least a residual mass $r = |F_t(k_{t'}) - \rho| \leq 1 - \rho = 0.1$ across $|k_t - k_{t'}|$ token positions, so

$$W_1(a(t), a(t')) \leq r |k_t - k_{t'}| \leq 0.1 |k_t - k_{t'}|. \quad (25)$$

Taking the median over t' in the sliding window and then the median over t gives

$$W_1^{\text{med}}(\ell) \leq 0.1 \text{LII}^\ell \quad (\text{A.1})$$

where $W_1^{\text{med}}(\ell)$ is the median Wasserstein distance between successive attention snapshots of layer ℓ . Equation (A.1) shows that ****LII** controls the earth-mover distance between attention distributions^{**}: A low LII implies the layer’s attention landscape hardly moves in Wasserstein space and is therefore safe to freeze. Via the Kantorovich–Rubinstein dual, the same bound controls the difference of *all* 1-Lipschitz observables of the attention measure, linking the energy and Fisher-flatness views in a common optimal-transport metric.

C Energy landscape-aware fine-tuning

Algorithm 1 details the complete training routine used in all experiments. After a short warm-up that estimates the Layer Instability Index (LII) for every block, layers whose instability falls below a user-defined threshold τ_{freeze} are frozen (`requires_grad=False`). Fine-tuning then proceeds on the remaining adaptive layers, incurring no further LII overhead.

Algorithm 1: Energy Landscape-Aware ViT Fine-Tuning

Input: pre-trained weights $\Theta^{(0)}$; dataset \mathcal{D} ; freeze threshold τ_{freeze} ; warm-up steps T ; LII window W

Warm-up phase: ; // estimate layer instability
for $t = 0$ **to** $T-1$ **do** // collect \bar{k} statistics
 sample mini-batch $(x, y) \sim \mathcal{D}$;
 forward and backward pass; update $\Theta^{(t+1)}$ with AdamW;
 update the circular buffer of size W and compute $\widehat{\text{LII}}^\ell$ for all layers ℓ ;

Freeze decision: ; // one-shot pruning of stable layers
foreach layer ℓ **do**
 if $\widehat{\text{LII}}^\ell < \tau_{\text{freeze}}$ **then**
 freeze(ℓ) // disable gradient updates

Consolidation phase: ; // train only adaptive layers
for $t = T$ **to** `max_steps` **do** // until convergence
 mini-batch $(x, y) \sim \mathcal{D}$; forward + backward pass on unfrozen layers only;

The algorithm runs in three stages: (i) *Warm-up* gathers a robust estimate of each layer’s variability via the median absolute deviation of its operational mode \bar{k} . (ii) *Freeze decision* is executed once, turning off gradient flow for layers whose LII indicates convergence to a low-energy basin. (iii) *Consolidation* fine-tunes the remaining layers, yielding substantial savings in memory and computation with no extra learnable parameters.

D Online update of the LII circular buffer

During warm-up we compute $\widehat{\text{LII}}_t^\ell$ for every layer on the fly. Algorithm 2 shows an eight-line Python reference implementation; it relies only on a layer-indexed deque of fixed capacity W (the sliding window size, default $W = 20$).

At each mini-batch, we compute the layer’s operational mode \bar{k}_t^ℓ (Sec. 3.2) and call `update_lii`. The deque acts as a circular buffer: the newest value is appended, the oldest is popped when the buffer overflows, and both operations are $\mathcal{O}(1)$. We then take the median of the window, followed by the median absolute deviation—exactly Eq. (2) but restricted to the latest W steps. The result is the online estimate $\widehat{\text{LII}}_t^\ell$ used in Alg. 1.

Algorithm 2: Update of layer instability index (LII) buffer

Input: layer ID layer_id ; new value \bar{k} ; buffer buf ; window size W **Output:** current LII_t^ℓ

```
buf[layer_id].append( $\bar{k}$ ) ; // 1. push newest value
if len(buf[layer_id]) > W then
    buf[layer_id].popleft() ; // 2. drop oldest (FIFO)
med = median(buf[layer_id]) ; // 3. running median
abs_dev = [abs(x - med) for x in buf[layer_id]] ;
return median(abs_dev) ; // 4. current  $\text{LII}_t^\ell$ 
```

Note: One dictionary, one deque per layer ($\text{maxlen} = W$).**Complexity:** $\mathcal{O}(1)$ per call, $\mathcal{O}(W)$ memory per layer.

Table 3: ImageNet-1k, ViT-B/16. \uparrow = higher is better, \downarrow = lower is better.

Method	Frozen layers	Trainable (%)	Top-1 (%) \uparrow	Latency (ms) \downarrow	Δ lat. (%)	Train time (h)
Full fine-tune	—	100.0	89.71	399.1	0.0	43.45
ELA-ViT-35%	3–6	67.2	89.17	330.8	−17.1	39.59 (−9.0)
ELA-ViT-50%	2–7	50.9	89.10	341.4	−16.7	38.81 (−10.7)
ELA-ViT-75%	0–7, 11	26.3	87.92	322.7	−12.3	35.09 (−19.2)

537 **Cold-start.** For $t < W$ the deque contains fewer than W elements; the function still returns a valid
538 LII based on the available prefix, ensuring that no additional initialisation logic is required.

539 **Efficiency.** The routine consumes negligible resources: $\mathcal{O}(LW)$ memory for L layers and $\mathcal{O}(1)$
540 extra time per iteration, contributing less than 1% overhead in all experiments (see App. B).

541 E ImageNet-1k: large-scale validation

542 We repeat the 15-epoch protocol on **ViT-B/16** using ImageNet-1k to test whether LII-guided freezing
543 holds at scale. Table 3 reports accuracy, latency, and wall-clock training time for three thresholds.
544 Latency is measured on 2,048 validation images with a single A100.

545 A **moderate freeze (35%)** already removes four layers, trims $\approx 9\%$ of training wall-time and 17%
546 of inference latency, while keeping accuracy within 0.54 pp of the full fine-tune. Freezing **half the**
547 **layers** halves gradient updates, maintains accuracy within 0.6 pp, and yields an 11% speed-up. Only
548 the **aggressive 75%** budget incurs a larger drop (1.8 pp) but delivers the greatest time saving (19%).
549 The results demonstrate that the Layer Instability Index (LII) effectively scales to ImageNet-1k,
550 highlighting an optimal operating region where *substantial speed improvements (exceeding 10%) are*
551 *achievable with minimal degradation in accuracy*. This validates the core hypotheses of our study
552 within a large-scale benchmark setting.

553 F Code Availability

554 The source code associated with this paper will be made publicly available upon acceptance for
555 publication.

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: Can be found in the section 1.

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: Can be found in the section 6.

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: [\[Yes\]](#)

Justification: Can be found in the section 3.

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: Can be found in the section 4.

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: Additional details and code repository URL are included in the supplemental material.

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: Can be found in the section 4.

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: Given that our experiments focus on training and fine-tuning Vision Transformer models to demonstrate improvements in accuracy and efficiency rather than testing hypotheses, statistical significance through error bars or formal significance testing is not directly applicable. Instead, we clearly report model performance metrics, including accuracy and computational improvements, consistently across multiple datasets and settings. These metrics reliably indicate performance differences without requiring traditional statistical significance testing.

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: Can be found in section 4

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: I confirm the research conform with the Code of Ethics.

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: [NA]

Justification: No potential societal impacts.

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: The research is not a high risk research.

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: All datasets and models used are explicitly mentioned and included in the reference.

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 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: [NA]

Justification: Not introduced new assets.

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: Not require human subjects.

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: Not require human subjects.

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: No LLM used for original, or non-standard component of the core methods.

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.