Predicting Influenza A Reassortment Potential Using Foundation Models and Genetic Algorithms for Pandemic Preparedness

Seeraja Konnur*
AI & Robotics Technology Park, I-Hub,
Indian Institute of Science,
Bengaluru, India
seeraja@artpark.in

Abstract

Influenza A virus (IAV) poses a persistent global threat due to its ability to evolve rapidly through reassortment. We present a computational framework that integrates DNABERT-2, a transformer-based foundation model for genomic sequences, with machine learning and genetic algorithms to predict reassortment events. Using H5N1 subtype sequences from the 2021–2022 U.S. outbreak, we generated segment-specific embeddings with DNABERT-2 to train a machine learning classifier capable of distinguishing reassortant from non-reassortant genotypes. Newly collected environmental samples are evaluated using this classifier to identify potential reassortants, while genetic algorithms generate novel segment combinations from these sequences to assess their likelihood of producing viable and potentially more virulent reassortants. This approach enables early detection of high-risk strains, offering a scalable tool for pandemic preparedness.

1 Introduction

Influenza A virus (IAV) represents one of the most significant ongoing threats to global public health, affecting both human and animal populations with devastating consequences. The virus continues to be a leading cause of respiratory infections worldwide, resulting in an estimated 36,000 deaths during typical endemic seasons in the United States alone (1). This substantial disease burden extends beyond human health, causing severe economic losses in agricultural sectors particularly the poultry and swine industries where outbreaks can lead to millions of dollars in losses through culling, trade restrictions, and reduced productivity. The persistent threat posed by Influenza A virus stems from its remarkable evolutionary adaptability, characterized by rapid mutation rates and genetic reassortment capabilities. These characteristics enable the virus to continuously evolve, evading host immune responses and existing therapeutic interventions. Two main phenomena drive this evolution: antigenic drift, involving the gradual accumulation of point mutations, and antigenic shift, which is caused by reassortment (2). In this study, we focus on the latter antigenic shift caused by reassortment. Influenza is a segmented RNA virus consisting of eight segments. Reassortment occurs when two or more different Influenza A viruses infect the same host cell and exchange viral RNA segments. This process can generate novel combinations of viral segments, potentially resulting in strains with enhanced transmissibility, virulence, or immune evasion capabilities. Historical pandemics—including the catastrophic 1918 H1N1 pandemic, the 1957 H2N2 Asian flu, and the 2009 H1N1 pandemic all

*LinkedIn: linkedin.com/in/seeraja-konnur-986794138/
GitHub: seerajakonnur.github.io/

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences.

emerged through reassortment events, underscoring the critical importance of understanding and predicting these evolutionary processes.

2 Challenge and innovation

Current influenza surveillance and vaccine development strategies face significant challenges due to the virus's unpredictable evolutionary trajectory. Annual vaccine formulations require predictions made months in advance, often resulting in mismatches between circulating strains and vaccine components. Furthermore, existing antiviral drugs show limited long-term effectiveness due to the rapid development of resistance mutations (3). The inability to accurately predict future viral variants hampers both vaccine design and antiviral drug development, highlighting the urgent need for innovative computational approaches to forecast viral evolution. Early detection and characterization of reassortment events are crucial for pandemic preparedness and response strategies. Environmental surveillance has emerged as a promising approach for monitoring viral circulation and diversity in various ecological niches. This study presents a novel computational framework that integrates the foundation model DNABERT-2 with genetic algorithms and environmental surveillance data to assess reassortment potential in circulating influenza viruses (4). By leveraging DNABERT-2, a foundation model developed for analyzing genomic sequence data, in combination with evolutionary algorithms, we aim to develop a predictive system capable of identifying reassortants in the environment and predicting the reassortment potential of circulating viruses. We aim to assess both existing and predicted reassortants for their potential risk. This approach represents a significant advancement in computational virology, offering the potential to transform routine environmental surveillance data into actionable intelligence for pandemic preparedness. By establishing quantitative risk scores for different reassortment scenarios, this system could provide early warning capabilities for public health authorities and inform targeted intervention strategies before pandemic emergence occurs. This paper presents work in progress from this ongoing research project, with the ultimate goal of developing a comprehensive surveillance and prediction system for influenza reassortment events. The prediction and assessment of influenza virus reassortment potential face several critical challenges. Traditional approaches to understanding viral evolution and reassortment rely predominantly on retrospective phylogenetic analyses, which while informative offer limited predictive power for future evolutionary events. Current computational methods in viral surveillance primarily use handcrafted features and use conventional machine learning techniques or statistical models, which often struggle to capture the complex, non-linear relationships present in viral genomic data (5). To address these limitations, we propose a novel approach that integrates foundation models with genetic algorithms to capture rich, non-linear representations particularly effective for modeling subtle reassortment patterns. At a later stage, we aim to incorporate the seasonality of the virus by including climatic variables such as rainfall, temperature, and humidity to better understand the influence of environmental factors on viral behavior (6).

3 Methods

This work-in-progress presents a four-stage computational pipeline designed to identify and assess influenza virus reassortment potential from environmental surveillance data. In order to generate a proof of concept, we started with the H5N1 subtype of Influenza A virus. H5N1 has emerged as one of the most devastating influenza subtypes with case fatality rates far exceeding those of seasonal influenza. Unlike many other subtypes, H5N1 exhibits a broad host range and a proven capacity to cross the species barrier, raising concerns of pandemic potential. Its evolutionary trajectory has been shaped by frequent reassortment events with co-circulating avian influenza viruses, leading to the emergence of novel genotypes with altered virulence, transmissibility, and immune escape properties. This makes H5N1 an ideal candidate to establish a proof-of-concept framework for reassortment potential prediction. Starting with H5N1 allows us to demonstrate the utility of predictive methods on a subtype where reassortment has repeatedly influenced viral evolution and where public health implications are immediate and significant. Our methodology begins by training a machine learning model on H5N1 reassorted and non-reassorted sequences, using DNABERT-2 embeddings for feature extraction to classify influenza A virus sequences as reassortants or non-reassortants. During routine environmental surveillance, newly sequenced influenza A viruses are processed through this trained model to identify potential reassortants. Critically, even sequences classified as non-reassortants retain the potential to undergo reassortment and generate more dangerous influenza variants. To proactively

assess this risk, we employ genetic algorithms to simulate reassortment using biological features that drive the process, as reassortment is not entirely random but governed by molecular compatibility constraints that serve as fitness functions to filter biologically viable reassortants from purely mathematical combinations. While genetic algorithms generate exclusively reassortant sequences, applying our binary classifier trained on natural reassortants versus non-reassortants is biologically justified through evolutionary pattern recognition. Natural reassortants represent "pre-validated" genomic combinations that survived evolutionary selection—successfully navigating host immune responses, demonstrating competitive fitness, and achieving population spread. These sequences encode the biological "rules" for successful reassortment, capturing molecular patterns including compatible segment combinations and functional genomic arrangements. Non-reassortant sequences provide the essential baseline, representing canonical genomic organization and helping the model identify what makes reassortants evolutionarily exceptional. When applied to GA-generated candidates, the model effectively asks: "Which of the possible reassortants exhibit the same molecular signatures of evolutionary success found in naturally viable reassortants?" The resulting probability scores create a biologically-grounded viability ranking system, where high-scoring candidates demonstrate strong similarity to evolutionary successful patterns, while low-scoring candidates lack these critical success signatures.

3.1 Feature extraction using foundation model

We used H5N1 clade 2.3.4.4b sequences from the United States during 2021–2022, a period marked by major reassortment events (7). This dataset captured both non-reassortant and reassortant genotypes circulating in the U.S. during this time. Specifically, it included non-reassortant genotypes such as A1, A2, and A3, as well as reassortant genotypes including B1.1, B1.2, B2, B3.1, B3.2, B4, B5, and additional minor reassortants that represent subsets of these major genotypes. For model development, we selected 120 non-reassortant sequences from genotype A1 as the negative training set and 119 reassortant sequences spanning genotypes B1.1, B1.2, B2, B3.1, B3.2, B4, and B5 as the positive training set, maintaining proportional representation across groups. Similarly, to evaluate model generalization on unseen data, non-reassortant genotypes A2 and A3 (25 sequences combined) were reserved exclusively for testing, while 30 minor reassortants were used as the reassortant test set. To capture segment-level patterns critical for reassortment detection, we employed a segment-specific feature extraction strategy using the DNABERT-2 foundation model. Instead of concatenating all eight influenza A genome segments into a single sequence, each segment (PB2, PB1, PA, HA, NP, NA, MP, and NS) was processed independently through DNABERT-2 to generate distinct embeddings. This design choice is particularly well-suited for reassortment analysis, as reassortment involves the exchange of individual genome segments between strains, making segment-level representations more biologically meaningful than whole-genome embeddings. For segments exceeding 400 base pairs, we implemented a chunking strategy with 100 bp overlaps to accommodate DNABERT-2's input length limitations, averaging embeddings across chunks to preserve segment integrity. The resulting segment-specific embeddings were then concatenated to construct a comprehensive genome-level representation that retains the distinct evolutionary history and functional characteristics of each segment. This methodology enables the model to capture both segment-specific signatures and inter-segment relationships essential for distinguishing reassortant from non-reassortant H5N1 viruses, leveraging DNABERT-2's transformer architecture to learn complex k-mer patterns and sequence motifs without the need for manual feature engineering.

3.2 Machine learning classifier

To classify reassortant versus non-reassortant H5N1 sequences, we used a Random Forest (RF) classifier trained on the DNABERT2-derived embeddings. RF was chosen for its robustness on small to medium datasets and its ability to handle non-linear relationships without extensive feature scaling. The training set embeddings were used to fit the RF model. Hyperparameters were optimized using a grid search with stratified 5-fold cross-validation to balance performance and prevent overfitting. The search space included the number of trees, maximum tree depth, minimum samples required for splitting, and minimum samples required at a leaf. The grid search was conducted with GridSearchCV from scikit-learn, using cross-validated accuracy as the primary scoring metric. The best-performing model from grid search was retrained on the full training set. The optimized Random Forest model was evaluated on an unseen test set comprising 25 non-reassortant sequences (genotypes A2 and A3) and 30 reassortant sequences (minor reassortants).

3.3 Genetic algorithm based candidate search

Influenza virus reassortment is not entirely random but is influenced by host species, viral subtypes, and segment combination (8). This part of the work is still in progress and requires further development before a full-scale genetic algorithm can be established. For the proof of concept, we initiated experiments using two non-reassortant parental genotypes reported in the same study: one example being the A1 genotype of H5N1, and the other the North American virus A/ruddy turnstone/Delaware Bay/210–212/2020 (H7N6). The fitness functions were constructed based on both the non-reassortants and the reassortants analyzed in this study. Since the reassortant segment combinations were already known, the fitness functions were designed to recover these combinations. Reassortants with an intact polymerase complex (PB2, PB1, PA derived from the same parent) were given higher fitness, consistent with the requirement for coordinated polymerase activity. Additional weight was given when the nucleoprotein (NP) originated from the same parent as the polymerase, reflecting their essential interaction in viral replication (9). Given that our analysis focuses on 2021-2022 H5N1 reassortants from the United States, where HA and NA segments consistently shared parental origin, co-inheritance of these surface glycoproteins was incorporated as an additional fitness parameter. Finally, complete parental genotypes received fitness penalties to prioritize reassortment. Using these criteria, we were able to generate reassortants that closely resembled the published B1 reassortant. Additional details and a pseudocode are provided in appendix A. In future work, additional biological fitness functions will be incorporated to improve the evaluation of reassortant viability, alongside expanding the number of parental sequences used in the algorithm.

4 Results

We applied t-distributed Stochastic Neighbor Embedding (t-SNE) to the trained data in order to project the high-dimensional embeddings into two dimensions. The visualization revealed distinct clustering between reassortant and non-reassortant sequences, highlighting DNABERT-2's ability to capture reassortment-related genomic features without task-specific fine-tuning. For classification, the optimal Random Forest configuration was `max_depth = 5, max_features = sqrt, min_samples_leaf = 5, min_samples_split = 5, n_estimators = 200`. On unseen data, the classifier achieved 100 % accuracy, with perfect precision, recall, and F1-scores for both classes. Statistical evidence supporting this performance, including p-value calculations, overfitting analysis, and biological justification, is presented in Appendix A. The prediction probabilities ranged from a minimum of 0.699 to a maximum of 1.000, with an average confidence of 0.960. Notably, all predictions exceeded a confidence threshold of 0.6, underscoring both the robustness and reliability of the model. The lowest prediction probability was for a reassortant which had different PB1 and PA segments as compared to the rest.

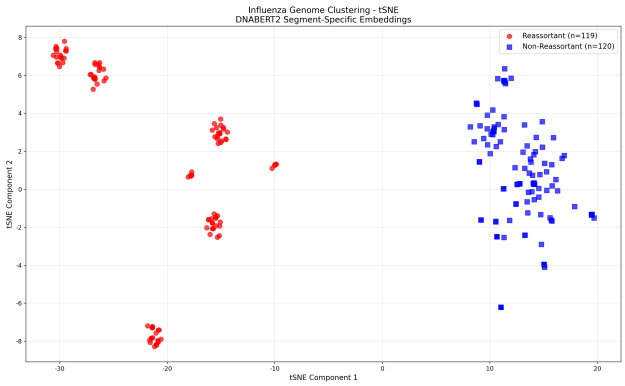


Figure 1: t-SNE visualization of DNABERT-2 embeddings showing distinct clustering of reassortant and non-reassortant influenza sequences

5 Future work

While this study focused on H5N1 clade 2.3.4.4b, the segment-specific DNABERT-2 framework can be extended to other influenza A subtypes. Additionally, we aim to enhance our genetic algorithm with enhanced biologically driven fitness functions incorporating nucleotide level features such as host adaptation signatures, segment compatibility complexes, and mutations associated with increased virulence to improve viable reassortant prediction from environmental samples.

References

[1] Taubenberger, J.K. & Kash, J.C. (2010). Influenza virus evolution, host adaptation, and pandemic formation. *Cell Host & Microbe*, **7**(6), 440–451. doi:10.1016/j.chom.2010.05.009. PMID: 20542248; PMCID: PMC2892379.

[2] Peiris, J.S., de Jong, M.D., & Guan, Y. (2007). Avian influenza virus (H5N1): a threat to human health. *Clinical Microbiology Reviews*, **20**(2), 243–267. doi:10.1128/CMR.00037-06. PMID: 17428885; PMCID: PMC1865597.

[3] Smyk, J.M., Szydłowska, N., Szulc, W., & Majewska, A. (2022). Evolution of Influenza Viruses—Drug Resistance, Treatment Options, and Prospects. *International Journal of Molecular Sciences*, **23**(20), 12244. doi:10.3390/ijms232012244. PMID: 36293099; PMCID: PMC9602850.

[4] Zhou, Z., Ji, Y., Li, W., Dutta, P., Davuluri, R.V., & Liu, H. (2023). DNABERT-2: Efficient Foundation Model and Benchmark For Multi-Species Genome. *arXiv preprint arXiv:2306.15006*.

[5] Yin, R., Zhou, X., Rashid, S., et al. (2020). HopPER: an adaptive model for probability estimation of influenza reassortment through host prediction. *BMC Medical Genomics*, **13**, 9. doi:10.1186/s12920-019-0656-7.

[6] He, Y., Liu, W.J., Jia, N., Richardson, S., & Huang, C. (2023). Viral respiratory infections in a rapidly changing climate: the need to prepare for the next pandemic. *EBioMedicine*, **93**, 104593. doi:10.1016/j.ebiom.2023.104593. PMID: 37169688; PMCID: PMC10363434.

[7] Youk, S., Torchetti, M.K., Lantz, K., Lenoch, J.B., Killian, M.L., Leyson, C., Bevins, S.N., Dilione, K., Ip, H.S., Stallknecht, D.E., Poulson, R.L., Suarez, D.L., Swayne, D.E., & Pantin-Jackwood, M.J. (2023). H5N1 highly pathogenic avian influenza clade 2.3.4.4b in wild and domestic birds: Introductions into the United States and reassortments, December 2021–April 2022. *Virology*, **587**, 109860. doi:10.1016/j.virol.2023.109860. PMID: 37572517.

[8] Ding, X., Ma, Y., Li, S., Liu, J., Qin, L., & Wu, A. (2025). Influenza virus reassortment patterns exhibit preference and continuity while uncovering cross-species transmission events. *Briefings in Bioinformatics*, **26**(3), bbaf233. doi:10.1093/bib/bbaf233.

[9] White, M.C., & Lowen, A.C. (2018). Implications of segment mismatch for influenza A virus evolution. *Journal of General Virology*, **99**(1), 3–16. doi:10.1099/jgv.0.000989. PMID: 29244017; PMCID: PMC5882089.

A Technical Appendices and Supplementary Material

Technical appendices with additional results, figures, graphs and proofs may be submitted with the paper submission before the full submission deadline (see above), or as a separate PDF in the ZIP file below before the supplementary material deadline. There is no page limit for the technical appendices.

A.1 Genetic Algorithms Fitness Functions

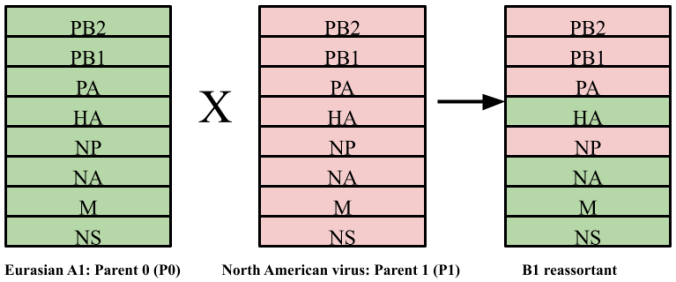
Below is the pseudocode for the fitness functions constructed in this study. These functions were designed specifically for H5N1 viruses of clade 2.3.4.4b. In future work, we aim to extend this framework to include additional clades and subtypes, as well as incorporate more refined fitness functions that capture detailed biological properties, structural constraints, and compatibility requirements of viable reassortants.

```
function fitness_function
1.  begin
2.  fitness ← 0
3.  genome ← I.segments // genome = [s0, s1, s2, s3, s4, s5, s6, s7]
4.  // order: [PB2, PB1, PA, HA, NP, NA, M, NS]
5.  // ---------- Polymerase complex integrity ----------
6.  polymerase ← [genome[0], genome[1], genome[2]]
7.  if all_same(polymerase) then
8.      fitness ← fitness + 100
9.      pol_parent ← polymerase[0]
10.
11.     // NP-Polymerase compatibility
12.     if genome[4] = pol_parent then
13.         fitness ← fitness + 50
14.     end if
15. else
16.     // Partial polymerase integrity (2 and 1 split)
17.     if count(polymerase, 0) = 2 or count(polymerase, 1) = 2 then
18.         fitness ← fitness + 35
19.     end if
20. end if
21. // ---------- HA NA functional pairing ----------
22. if genome[3] = genome[5] then
23.     fitness ← fitness + 40
24. end if
25. // ---------- Penalty for pure parental types ----------
26. if count(genome, 0) = 8 or count(genome, 1) = 8 then
27.     fitness ← fitness −70
28. end if
29. return fitness
30. end
```

Listing 1: Fitness Function for Influenza Reassortant Evaluation

Output:

```
[PB2:P1, PB1:P1, PA:P1, HA:P0, NP:P1, NA:P0, M:P0, NS:P0] Fitness: 190.0
Polymerase complex: Parent 1 (intact)
NP: Parent 1
HA-NA pair: Parent 0, Parent 0
Segments from Parent 0: 4, Parent 1: 4
```



B1 reassortant generated using genetic algorithm fitness functions

Figure 2:

A.2 Classification Accuracy

Multiple lines of evidence demonstrate that 100% accuracy reflects genuine biological signal rather than overfitting. The segment-specific approach is particularly well-suited for reassortment detection because influenza reassortment occurs through exchange of entire intact segments rather than point mutations, creating categorical genomic signatures. During model development, cross-validation analysis on training data showed zero overfitting gap (mean training score: 100.0%, mean validation score: 100.0%, gap: 0.0%), indicating excellent generalization capacity. This perfect generalization was confirmed on 55 completely unseen test samples from different genotypes, achieving 100.0% accuracy with high prediction confidence (mean = 0.96, all probabilities > 0.7). Statistical hypothesis testing indicates this performance is highly unlikely by chance ($p = 0.006$, binomial exact test). Since reassortment involves complete segment exchanges between North American and Eurasian genotypes, each segment retains its distinct evolutionary signature, enabling clear discrimination between reassorted and non-reassorted genomes. The combination of zero overfitting gap during training, uniformly high test prediction confidence, cross-genotype validation, and biological mechanism alignment demonstrates that segment-specific DNABERT-2 embeddings capture fundamental reassortment signatures.

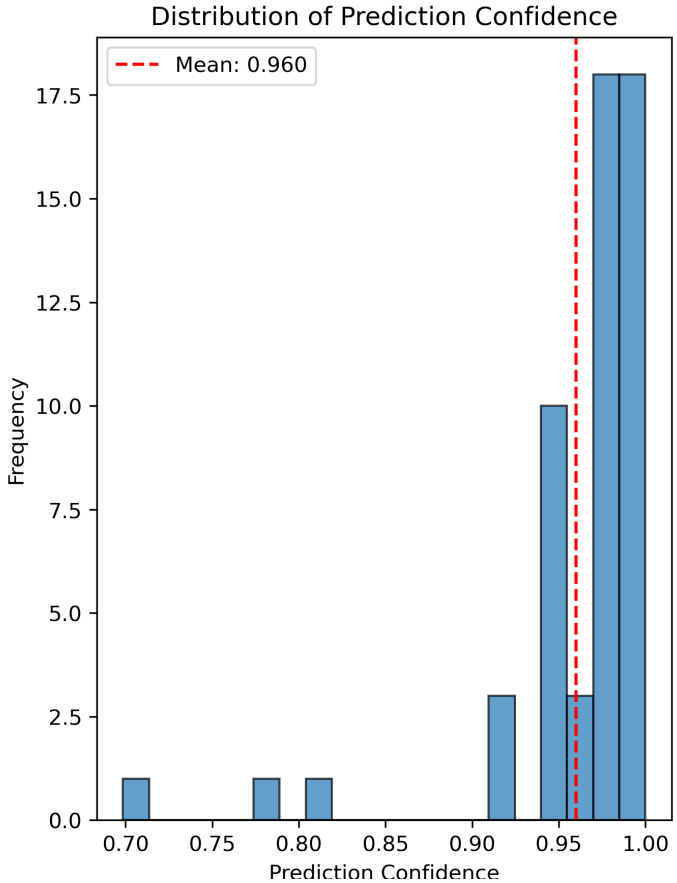


Figure 3: Distribution of prediction probabilities for reassortant and non-reassortant sequences on unseen data

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: **[TODO]**

   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: **[TODO]**

   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: **[TODO]**

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: **[TODO]**

   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: **[TODO]**

Justification: **[TODO]**

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: **[TODO]**

   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: [Yes]

   Justification: **[TODO]**

   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: **[TODO]**

   Justification: **[TODO]**

   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: **[TODO]**

   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: **[TODO]**

    Justification: **[TODO]**

    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: **[TODO]**

    Justification: **[TODO]**

    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: **[TODO]**

    Justification: **[TODO]**

    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: [TODO]

Justification: [TODO]

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: [TODO]

Justification: [TODO]

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: [TODO]

Justification: [TODO]

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: [TODO]

Justification: **[TODO]**

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.