Condition-Based Integration of In Vivo RNA Delivery Data Enables AI-Guided Ionizable Lipid Design

Black Goat
MOGAM Institute for Biomedical Research (MIBR)
Seoul 06730, Republic of Korea
kohyj0212@naver.com

Abstract

Lipid nanoparticles (LNPs) are widely used for RNA-based therapeutics, where ionizable lipids (ILs) play a critical role in enabling efficient cytoplasmic RNA delivery through pH-dependent mechanisms. However, exhaustive in vivo screening of ionizable lipid candidates is costly and time-consuming, motivating the development of AI-based models for predicting RNA transfection potency from lipid structure. A major challenge in this setting is the limited availability of in vivo data and the substantial heterogeneity across experimental conditions, which complicates direct aggregation of datasets in the absence of shared reference LNPs. In this work, we investigate condition-based integration of heterogeneous in vivo datasets and demonstrate that datasets sharing the same cargo type, delivered gene, target organ, and administration route can be meaningfully pooled despite differences in LNP composition. Leveraging this strategy, we train AI models that achieve strong within-condition predictive performance, enabling reliable ranking and prioritization of ionizable lipid candidates. Our results provide a practical framework for expanding training data for LNP design under data-limited in vivo settings and highlight both the potential and limitations of AI-driven prediction across heterogeneous experimental conditions.

1 Introduction

Lipid nanoparticles (LNPs) have become a central platform for RNA-based therapeutics and vaccines, enabling efficient delivery of messenger RNA (mRNA) and small interfering RNA (siRNA) to target tissues. Among the components of LNPs, ionizable lipids (ILs) play a particularly critical role due to their pH-dependent molecular properties, which facilitate endosomal escape and cytoplasmic release of RNA cargos. Consequently, the discovery of effective ionizable lipids and the rational design of LNP formulations are essential for achieving robust in vivo RNA delivery.

Despite their importance, systematic experimental evaluation of ionizable lipid candidates remains a major bottleneck. In vivo screening is inherently expensive and time-consuming, as each candidate lipid requires formulation optimization and animal testing. As a result, only a limited subset of the vast chemical space of ionizable lipids can be experimentally explored. Artificial intelligence (AI)-based predictive models have therefore attracted increasing interest as a means to prioritize promising lipid candidates and reduce experimental burden by predicting RNA transfection potency directly from molecular structure.

However, the development of accurate AI-based models is severely constrained by the limited availability of in vivo data. While aggregating datasets from multiple experimental studies could potentially increase training data size, such integration is complicated by substantial heterogeneity in experimental conditions. Delivery efficiency is strongly influenced by factors such as cargo type, delivered gene, target organ, and route of administration, leading to condition-dependent measurement scales and biological contexts. In the absence of shared reference ionizable lipids or reference LNP formulations, conventional normalization strategies are often insufficient, and indiscriminate pooling of heterogeneous datasets risks

Submitted to 1st 2026AI Co-Scientist Challenge Korea. Do not distribute.

introducing confounding effects.

In this work, we investigate the feasibility of condition-based integration of heterogeneous in vivo LNP delivery datasets. We hypothesize that although pooling data across fundamentally different experimental conditions is inappropriate, datasets sharing a minimal set of experimental parameters—cargo type, delivered gene, target organ, and administration route—can be meaningfully integrated even when LNP composition differs and no reference lipid is available. To address this question, we develop a condition-aware predictive modeling framework that leverages molecular representations of ionizable lipids and systematically evaluates performance under both random-split and condition-held-out settings. By doing so, we demonstrate that condition-based pooling enables strong within-condition prediction and candidate ranking, while also revealing clear limitations in cross-condition generalization. Together, these findings provide practical guidance for leveraging heterogeneous in vivo datasets in AI-driven LNP design under data-limited settings.

2 Results

2.1 Dataset characterization and experimental condition definition

We curated an in vivo LNP delivery dataset comprising 1,016 samples collected under diverse experimental settings [1]. Each sample was associated with a specific combination of cargo type (mRNA or siRNA), delivered gene, target organ, and route of administration, as well as corresponding ionizable lipid structures and LNP compositions. The delivery outcome was quantified using unnormalized delivery readouts, reflecting experimentally measured transfection potency without reference-based normalization.

To enable principled data integration, we defined an experimental condition as a unique combination of cargo type, delivered gene, target organ, and administration route. Under this definition, the dataset was partitioned into eight distinct experimental conditions, with sample sizes ranging from 22 to 260 (Table 1). Each condition exhibited markedly different target distributions in terms of mean, variance, and dynamic range. These differences underscore the challenge of directly comparing or normalizing delivery outcomes across conditions in the absence of reference LNPs.

Table 1: Summary of in vivo RNA delivery datasets grouped by experimental condition

Cargo type	Delivered gene	Target organ	Route	N
mRNA	hEPO	Muscle	IM	260
mRNA	FFL	Liver	IV	208
mRNA	FFL	Spleen	IV	176
mRNA	FFL	Muscle	IM	145
siRNA	FVII	Liver	IV	91
mRNA	Peptide barcode	Liver	IV	65
mRNA	FFL	Lung epithelium	IT	49
mRNA	FFL	Lung	IV	22
Total	-	-	-	1026

2.2 Baseline performance and impact of condition heterogeneity

As a baseline, we evaluated simple predictors that estimate delivery potency using global or condition-specific mean values. While condition-specific baselines reduced prediction error compared to a global mean, they failed to capture structure-dependent variations arising from differences in ionizable lipid chemistry and LNP composition. This observation highlights the need for models that can leverage molecular features while respecting condition-dependent measurement scales.

The pronounced heterogeneity in target distributions across conditions further emphasizes that naive aggregation or global normalization strategies are insufficient for integrating in vivo delivery data collected under diverse experimental settings.

2.3 Within-condition predictive performance

We first evaluated model performance using standard K-fold cross-validation, in which samples from the same experimental condition may appear in both training and test sets. Under this setting, the AI model achieved substantially improved performance relative to baseline predictors, with a mean root mean squared error (RMSE) of approximately 2.5 and a Spearman rank correlation of approximately. Importantly, strong rank correlation indicates that the model reliably captures relative differences in delivery potency among LNP formulations within the same experimental condition.

To further assess robustness to scale differences, we evaluated prediction accuracy using within-condition z-score normalization. The resulting zRMSE values remained close to unity, suggesting that the model effectively learns condition-specific structure–performance relationships rather than relying on absolute scale differences across conditions.

2.4 Cross-condition generalization analysis

To assess the model's ability to generalize across experimental conditions, we performed group-based cross-validation and leave-one-experiment-out (LOEO) evaluation, in which entire experimental conditions were held out during training. Under these more stringent settings, predictive performance degraded substantially, with rank correlations approaching zero or becoming negative in several conditions. This result indicates that the learned relationships between ionizable lipid structure and delivery potency do not readily transfer across conditions with different cargo types, target organs, or administration routes.

Importantly, this degradation was observed even when experimental-condition embeddings were explicitly handled to avoid information leakage, confirming that the limitation reflects genuine biological and experimental differences rather than modeling artifacts.

3 Discussion

Our results provide important insights into the integration and utilization of heterogeneous in vivo LNP delivery datasets for AI-based prediction. The strong within-condition performance demonstrates that, when a minimal set of experimental parameters is held constant, datasets differing only in LNP composition can be effectively pooled and leveraged to learn meaningful structure–performance relationships. Crucially, this is achieved without reliance on reference ionizable lipids or reference LNP formulations, which are often unavailable in practice.

At the same time, the pronounced degradation in cross-condition generalization highlights a fundamental limitation of current modeling approaches. Differences in cargo type, delivered gene, target organ, and administration route introduce substantial biological and technical variability that cannot be readily normalized away or inferred solely from lipid structure. These findings underscore the importance of explicitly accounting for experimental context when developing predictive models for in vivo RNA delivery.

From a practical perspective, our study suggests a realistic and scalable strategy for expanding training datasets under data-limited conditions. Rather than attempting to normalize and pool all available in vivo data, we advocate a condition-based integration approach in which datasets sharing key experimental parameters are aggregated, enabling robust within-condition ranking and prioritization of ionizable lipid candidates. Such models are well-suited for guiding experimental design by narrowing the search space of candidate lipids within a given delivery context.

Future work may extend this framework by incorporating explicit representations of experimental conditions into the model, enabling improved generalization across conditions. Additionally, integrating in vitro data or mechanistic descriptors may further enhance predictive power and interpretability. Overall, our findings establish principled boundaries for data integration in AI-driven LNP design and provide a foundation for more efficient exploration of ionizable lipid chemical space in RNA therapeutics.

References

[1] Witten, J., Raji, I., Manan, R.S. et al. (2025) Artificial intelligence-guided design of lipid nanoparticles for pulmonary gene therapy. Nat Biotechnol **43**, 1790–1799

AI Co-Scientist Challenge Korea Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [N/A] .
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[N/A] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "AI Co-Scientist Challenge Korea paper checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

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 accurately reflect the scope and contributions of the paper. The paper claims that heterogeneous in vivo RNA delivery datasets can be meaningfully integrated when core experimental factors (cargo type, delivered gene, target organ, and route of administration) are shared, even in the absence of reference

ionizable lipids or reference LNP formulations. These claims are directly supported by the experimental results, which demonstrate strong within-condition predictive and ranking performance, alongside explicitly documented limitations in cross-condition generalization. The paper clearly distinguishes between achieved results and aspirational goals, avoiding claims of universal generalization across experimental conditions.

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: The paper explicitly discusses several limitations of the proposed approach. In particular, it highlights the limited ability of the model to generalize across experimental conditions that differ in cargo type, target organ, or administration route, as demonstrated by group-based and leave-one-experiment-out evaluations. The study also acknowledges the data-limited nature of in vivo experiments, potential biases arising from heterogeneous experimental protocols, and the lack of reference LNPs for global normalization. These limitations are discussed in a dedicated section and reflected directly in the scope of the claims made.

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: The paper does not present formal theoretical results, theorems, or proofs. All contributions are empirical and methodological in nature, focusing on data integration strategies and predictive modeling evaluated through experiments.

   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: The paper provides sufficient detail to reproduce the main experimental results. The dataset construction, condition definitions, target

variables, model evaluation protocols (including K-fold, GroupKFold, and leave-one-experiment-out settings), and evaluation metrics are all clearly described. Even without direct release of code or raw data, the methodological description enables independent reproduction of the reported results using equivalent data and modeling frameworks.

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 AI Co-Scientist Challenge Korea 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: The paper does not provide open access to the full dataset or source code due to data ownership and confidentiality constraints. However, the paper includes detailed descriptions of data preprocessing, experimental conditions, model evaluation protocols, and performance metrics, providing a reasonable pathway for verification and partial reproduction of the results. The lack of open release is

not central to the contribution

of the work.

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, hyper-parameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: yes

   Justification: All essential experimental details required to understand and interpret the results are specified in the paper, including data splits, cross-validation strategies, evaluation metrics, and normalization procedures. Additional implementation details, such as model architecture and training protocols, are described at a level sufficient to contextualize the reported results, with further details available upon request or in supplemental material.

   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: The paper reports performance statistics aggregated across multiple folds and experimental conditions, including mean and standard deviation for RMSE, Spearman correlation, and zRMSE. Error bars reflect variability across data splits or held-out experimental conditions, and

the evaluation protocols used to compute these statistics are clearly described. These measures provide appropriate statistical context for the reported results.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: The paper specifies the computational resources used for training and evaluation, including the use of standard GPU-equipped workstations. Training time per experiment and overall computational requirements are modest and consistent with the scale of the dataset and model architecture, making reproduction feasible for most research environments.

   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://nips.cc/public/EthicsGuidelines?

   Answer: yes

   Justification: The research conducted in this paper conforms to the NeurIPS Code of Ethics. The study uses pre-existing experimental data, does not involve human subjects, and does not raise ethical concerns related to data collection, privacy, or misuse. Anonymity has been preserved where required.

   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: The paper discusses potential positive societal impacts, including accelerating the design of effective RNA delivery systems and reducing the cost and time associated with in vivo experimentation. Potential negative impacts, such as over-reliance on predictive models or misuse of predictions outside validated experimental conditions, are acknowledged. The paper emphasizes that the model is intended as a

decision-support tool rather
than a replacement for
experimental validation.

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 paper does
not release models or datasets
that pose a high risk of
misuse or dual-use concerns.
As such, no specific
safeguards are required.

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 existing assets used in the paper, including datasets and molecular representations derived from prior studies, are properly cited. The paper acknowledges original sources where applicable and respects the licensing and usage terms of the underlying data.

    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: The paper does not introduce new publicly released datasets, models, or code 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: The study does not involve crowdsourcing or research with 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: The paper does

    not involve human subjects or

    interventions requiring IRB

    approval.

    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.