
Delivery Is a Property of Sequence: An AI Co-Scientist Investigation into the Learnability Boundary and Non-Learnable Collapse of Biological Delivery

Anonymous Author

Abstract

Biological delivery is a central bottleneck in modern therapeutics, yet most existing studies frame it as a problem of performance prediction or optimization—asking which sequence or formulation delivers “better.” Such approaches have produced incremental gains, but they obscure a more fundamental question: *to what extent is biological delivery learnable from sequence and observable context, and where does learning structurally fail?*

In this study, we recast biological delivery as a **boundary-discovery problem** rather than a performance maximization task. We investigate delivery as a function of molecular sequence and context, focusing on identifying the **learnability boundary**—the regimes where predictive structure exists—and the conditions under which delivery exhibits **non-learnable collapse**, where prediction fails regardless of model complexity or data volume.

We adopt a dry-lab, design-driven methodology using a **Minimal Executable Study Design (MESD)** that deliberately exposes failure modes. Delivery is decomposed into stage-specific phenotypes (e.g., uptake versus endosomal escape), combined with controlled variation in observable context and intentionally unobserved hidden-state factors. Analysis is governed by pre-specified decision rules to distinguish learnable structure from measurement collapse, hidden-state effects, and non-linear regime transitions.

An **AI co-scientist** is integrated throughout the research process—not merely as an automation tool, but as a collaborative reasoning agent. The AI proposes research questions, alternative hypotheses, and experimental abstractions, while the human researcher critically evaluates, constrains, and ultimately determines the analytical framing and interpretation. This explicit separation of roles is documented as part of the research methodology.

Rather than reporting improved delivery performance, this work aims to produce a **map of where learning is possible and where it is not**. By making failure modes first-class scientific outcomes, we provide a principled framework for understanding the limits of sequence- and context-based learning in biological delivery, with implications for how future delivery research should be designed and evaluated.

1 Introduction

Biological delivery remains one of the most persistent bottlenecks in modern biotechnology. While advances in molecular design, formulation, and carrier systems have enabled incremental improvements, the majority of next-generation therapeutics fail not because their targets are unknown, but because the molecules cannot reliably reach the intended intracellular or intranuclear destinations. Despite decades of

work, delivery is still treated as an engineering problem whose success is measured primarily by performance metrics—higher uptake, stronger expression, or improved biodistribution.

Recent applications of machine learning and large language models (LLMs) to biological delivery have largely inherited this framing. Delivery is commonly modeled as a prediction or optimization task: given a molecular sequence or formulation, predict a scalar measure of delivery efficiency, or recommend modifications that improve performance. Although such approaches have produced useful heuristics, they implicitly assume that delivery is uniformly learnable from available data and that failure cases are merely artifacts of insufficient models or datasets. This assumption has rarely been examined directly.

However, empirical evidence from delivery research suggests a more complex reality. Delivery outcomes are often highly sensitive to experimental context, including cell type, passage number, serum conditions, receptor expression, and intracellular state. Small, poorly observed changes in these factors can lead to abrupt and irreproducible shifts in outcome. In practice, delivery failures are frequently treated as noise or experimental error, rather than as signals of structural limits to learning. As a result, existing datasets and models conflate three fundamentally different phenomena: genuinely learnable structure, measurement or standardization collapse, and biological regimes in which prediction may be intrinsically unreliable.

This work argues that the central question is therefore not *how to improve delivery performance*, but rather to *what extent biological delivery is learnable from sequence and observable context, and where learning structurally fails*. We propose that delivery should be studied as a **boundary-discovery problem**, in which the primary scientific objective is to map the boundary between learnable and non-learnable regimes, rather than to maximize a single performance metric.

A key obstacle to addressing this question is the way delivery is typically represented. Many studies reduce delivery to a single score, collapsing distinct biological processes—such as cellular uptake, endosomal escape, and subcellular targeting—into one aggregate outcome. This practice obscures where predictive structure exists and where it breaks down. Similarly, experimental context is often treated as nuisance variability rather than as a conditional input that may restore or destroy learnability. When context is incompletely observed, hidden-state variables can induce apparent randomness that no model can resolve, leading to what we refer to as **non-learnable collapse**.

To confront these issues, we adopt a design-driven, dry-lab approach that explicitly seeks failure. Instead of attempting to optimize delivery outcomes, we construct a **Minimal Executable Study Design (MESD)** that intentionally exposes regions where learning is expected to fail. Delivery is decomposed into stage-specific phenotypes (e.g., uptake versus endosomal escape), analyzed under controlled variation of observable context and deliberately unobserved hidden-state factors. Crucially, all interpretations are governed by pre-specified decision rules, preventing post-hoc rationalization of success or failure.

An important methodological contribution of this work is the explicit integration of an **AI co-scientist** throughout the research process. The AI is not used merely for automation or code generation, but as a collaborative reasoning agent that proposes research questions, alternative hypotheses, experimental abstractions, and analytical criteria. The human researcher retains full responsibility for evaluating these proposals, constraining their scope, rejecting inappropriate framings, and fixing the final interpretation. By documenting this interaction, we treat AI participation itself as part of the scientific method under investigation.

Rather than presenting improved delivery results, this study aims to produce a principled account of where learning is possible and where it is not. By elevating failure and collapse to first-class scientific outcomes, we provide a framework for understanding the limits of sequence- and context-based learning in biological delivery. This perspective has implications not only for delivery research, but for how AI-driven biological modeling should be designed, evaluated, and interpreted more broadly.

2 Related Work

Research on biological delivery spans multiple communities, including peptide- and protein-based delivery, nucleic acid delivery, nanoparticle systems, and, more recently, machine-learning–driven molecular design. While these efforts have produced valuable insights and incremental improvements, they largely share a common framing: delivery is treated as a problem of efficiency prediction, optimization, or mechanism identification. This section reviews representative strands of prior work and highlights their structural limitations from the perspective of learnability, reproducibility, and failure characterization.

CPP- and NLS-Based Delivery Studies

Cell-penetrating peptides (CPPs) and nuclear localization signals (NLSs) have long served as canonical motifs for studying intracellular and intranuclear delivery. CPP-focused studies typically examine how charge density, amphipathicity, or sequence patterns influence cellular uptake, often reporting scalar measures such as fluorescence intensity or transfection efficiency. Similarly, NLS studies investigate motif-driven nuclear transport, linking short sequence patterns to localization outcomes.

Although these studies have established that certain sequence features correlate with delivery-related phenotypes, they tend to emphasize positive cases - sequences that “work” - while treating failures as experimental noise. Moreover, delivery is frequently collapsed into a single outcome measure, obscuring the fact that uptake, endosomal escape, and subcellular targeting are distinct biological processes with different dependencies and failure modes.

Machine Learning for Delivery Prediction and Optimization

Recent years have seen increasing use of machine learning models to predict delivery outcomes or optimize molecular sequences and formulations. These approaches typically frame delivery as a supervised learning problem, mapping molecular descriptors or sequence embeddings to a delivery score. Variants of this approach include neural networks for uptake prediction, Bayesian optimization for formulation design, and generative models for proposing improved sequences.

While such models can capture correlations present in curated datasets, they implicitly assume that delivery is uniformly learnable given sufficient data and model capacity. Failure cases are often attributed to data sparsity or model inadequacy rather than examined as potential indicators of structural limits. As a result, these approaches provide limited insight into where learning is fundamentally constrained, particularly when delivery outcomes depend on poorly observed experimental context.

Context, Variability, and Reproducibility

Several studies have acknowledged that delivery outcomes are highly sensitive to experimental context, including cell type, culture conditions, serum composition, passage number, and intracellular state. However, context is frequently treated as a nuisance variable to be controlled or averaged out, rather than as a conditional input that may determine whether learning is possible at all. When contextual factors are incompletely recorded, hidden-state variables can induce apparent randomness, leading to irreproducible results across laboratories or experimental batches.

From a learnability perspective, this conflation of context and noise makes it difficult to distinguish between genuinely non-learnable regimes and failures arising from incomplete observation or standardization. Existing benchmarks rarely include the structured repetition, batch controls, or explicit hidden-state tracking needed to resolve this distinction.

Limitations of Performance-Centric Framing

Across these strands of work, a recurring limitation is the dominance of performance-centric framing. Delivery research is typically evaluated by how much a metric improves relative to a baseline, rather than by whether predictive structure exists across conditions. This emphasis encourages early termination once “good enough” performance is achieved and discourages systematic recording of failures. Consequently, collapse phenomena—whether due to measurement instability, hidden biological states, or non-linear regime shifts—are underreported and

underanalyzed.

Positioning of the Present Work

The present study departs from prior approaches by explicitly reframing biological delivery as a boundary-discovery problem. Rather than seeking improved delivery performance, we focus on identifying where learning is possible and where it fails, using pre-specified decision rules and a design that deliberately exposes collapse. In doing so, we treat failure as a first-class scientific outcome and aim to complement existing delivery research with a principled account of its learnability limits.

3 Research Questions and Hypotheses

The analysis of prior work highlights a recurring gap: while biological delivery is extensively studied, it is rarely examined through the lens of *learnability*. Existing approaches tend to assume that delivery outcomes are learnable given sufficient data and model capacity, and that failures primarily reflect experimental noise or incomplete optimization. This work instead treats learnability itself as the object of inquiry.

Accordingly, we formulate research questions that explicitly separate **learnable structure** from **non-learnable collapse**, and hypotheses that are designed to be *falsifiable by design* rather than supported by performance gains.

Research Questions

RQ1. To what extent is biological delivery learnable from sequence alone?

We ask whether molecular sequence contains predictive structure that generalizes across repetitions and batches, particularly for early-stage delivery phenotypes such as cellular uptake. This question targets the existence of intrinsic, sequence-encoded learnability independent of contextual information.

RQ2. How does observable context affect the learnability of delivery phenotypes?

We examine whether incorporating observable contextual variables (e.g., cell type, dose) expands the region of learnability, especially for downstream phenotypes such as endosomal escape. This question distinguishes between cases where context serves as a conditional input that restores learnability and those in which it does not.

RQ3. Under what conditions does delivery prediction undergo non-learnable collapse?

We investigate situations in which prediction fails despite controlled design and repeated measurement. These include failures arising from measurement instability, unobserved hidden-state variables, and non-linear or regime-dependent biological behavior. The goal is not to eliminate collapse, but to characterize and classify it.

RQ4. Can collapse be structured and mapped rather than treated as random failure?

We ask whether non-learnable collapse occurs in reproducible regimes—defined by combinations of sequence properties, context, and experimental conditions—or whether it appears irreducibly random. This question motivates the construction of a collapse map rather than a single performance score.

Hypotheses

The following hypotheses are formulated to correspond directly to the research questions above and are evaluated using pre-specified decision rules.

H1 (Sequence-only learnability).

Certain delivery phenotypes, particularly early-stage processes such as uptake, exhibit learnable structure that can be predicted from molecular sequence alone across repeated measurements.

H2 (Context-conditioned learnability).

For downstream delivery phenotypes, incorporating observable contextual variables expands the learnable regime relative to sequence-only models, indicating that learnability is conditional rather than intrinsic.

H3 (Design-driven boundary exposure).

A study design that explicitly varies sequence and context while enforcing repetition and batch structure exposes learnability boundaries more reliably than performance-driven optimization approaches.

H4 (Non-learnable collapse).

There exist regions of the delivery space in which prediction fails systematically, even under controlled design, due to measurement collapse, hidden-state effects, or non-linear regime transitions. These failures are structured and classifiable rather than random.

Together, these questions and hypotheses shift the focus of delivery research from *how well delivery can be optimized* to *where learning is possible and where it breaks down*. This framing enables failure to be treated as a meaningful scientific outcome and provides the basis for the study design and analysis framework described in the following sections.

4 Role of the AI Co-Scientist

In this study, artificial intelligence is not treated as a post hoc analysis tool or a writing assistant, but as an explicit **co-scientist** whose role is defined, constrained, and evaluated as part of the research methodology. The goal is not to demonstrate AI capability per se, but to examine how AI participation shapes scientific reasoning when its outputs are subjected to human judgment and correction.

AI as a Generative Research Partner

The AI co-scientist contributed primarily at the level of **hypothesis generation, problem reframing, and design abstraction**. Specifically, the AI proposed alternative framings of biological delivery, generated competing research questions, and articulated candidate hypotheses that spanned sequence-only, context-conditioned, and failure-driven perspectives. These proposals were often expansive and, in several cases, overly optimistic about the scope of learnability.

Rather than accepting these proposals directly, the human researcher evaluated them against the constraints of reproducibility, experimental realism, and Track 1's emphasis on boundary discovery. This evaluative step was essential in transforming broad AI-generated ideas into testable, constrained research questions.

AI-Proposed Hypotheses and Human Constraint

A recurring pattern throughout the research process was the AI's tendency to frame delivery as broadly learnable given sufficient data or model capacity. While useful as a starting point, this framing risked reducing the study to a performance-optimization paradigm. The human researcher explicitly rejected this assumption and reformulated the hypotheses to require falsifiability through failure.

For example, hypotheses suggesting uniform learnability were revised to distinguish between early-stage and downstream delivery phenotypes, and to require explicit decision rules for declaring non-learnable collapse. In this sense, AI-generated hypotheses functioned as **stress tests** for the researcher's framing, exposing implicit assumptions that required correction.

AI in Study Design and Failure Exposure

The AI co-scientist played a significant role in proposing abstractions for study design, including the decomposition of delivery into stage-specific phenotypes and the conceptual separation of observable context from hidden-state variables. However, the AI initially favored designs that maximized predictive coverage. These designs were deliberately constrained by the human researcher to produce a **Minimal Executable Study Design (MESD)** whose purpose was not optimization, but controlled exposure of failure modes.

This interaction illustrates a key principle of the present work: AI is effective at enumerating possibilities, but human judgment is required to impose epistemic discipline—deciding which possibilities should be excluded to make failure interpretable rather than accidental.

AI-Assisted Analysis and Pre-Specification

In the analysis phase, the AI proposed multiple evaluation metrics and interpretive lenses. Left unconstrained, these suggestions risked post hoc rationalization of outcomes. To prevent this, the human researcher fixed a set of pre-specified decision rules governing how learnability and collapse would be declared. AI outputs were then evaluated strictly within these constraints.

Importantly, instances where AI explanations conflicted with the pre-specified rules were treated as **negative evidence** rather than adjusted away. This ensured that AI participation did not introduce adaptive interpretation after observing outcomes.

Accountability and Role Separation

Throughout the study, the division of responsibility between AI and human researcher was explicitly maintained. The AI co-scientist generated hypotheses, abstractions, and analytical suggestions, but **all final decisions - including hypothesis acceptance, design constraints, interpretive rules, and conclusions - were made by the human researcher.**

This separation is central to the study's methodological contribution. By documenting where AI proposals were accepted, modified, or rejected, the research treats AI not as an oracle, but as a fallible reasoning agent whose influence must be audited. In doing so, the study aligns AI participation with the norms of scientific accountability rather than automation.

5 Role of the AI Co-Scientist

The study adopts a **Minimal Executable Study Design (MESD)** to investigate the learnability boundary of biological delivery. The MESD is intentionally not optimized for performance or scale. Instead, it is constructed to expose where learning is expected to succeed, where it degrades, and where it fails in a structured and interpretable manner. All design choices are fixed *a priori* and evaluated using pre-specified decision rules.

Design Principles

Three principles guide the MESD.

First, **failure is treated as an outcome**, not as noise to be eliminated. The design therefore includes conditions under which prediction is expected to break down. Second, **delivery is decomposed into stage-specific phenotypes**, rather than collapsed into a single score. Third, **context is partially observed by design**, enabling explicit distinction between observable conditioning effects and hidden-state-driven collapse.

Axes of Variation

The MESD is defined over three primary axes: sequence, context, and phenotype.

Sequence axis.

A finite library of molecular sequences is constructed to span regions where learnability is expected and regions where collapse is plausible. Sequences are grouped by qualitative properties (e.g., charge distribution, amphipathicity, hydrophobicity) and include both motif-based candidates and deliberately perturbed variants. A subset of sequences is designated as *bridge controls* and repeated across all batches to enable alignment and reproducibility assessment.

Context axis.

Context is divided into two classes:

- **Observable context**, such as cell type and dose, which is explicitly recorded and provided as conditional input in analysis.
- **Hidden-state context**, such as cellular passage or stress-related state, which is recorded as metadata but intentionally excluded from certain analytical models.

This separation allows the study to test whether adding observable context restores learnability and to identify regimes where hidden-state variables induce non-learnable collapse.

Phenotype axis.

Delivery is decomposed into stage-specific phenotypes rather than summarized by a single metric. At minimum, the design distinguishes between early-stage processes (e.g., cellular uptake) and downstream processes (e.g., endosomal escape). This decomposition is essential for identifying where predictive structure exists and where it breaks down across stages.

Repetition and Batch Structure

Reproducibility is enforced through explicit repetition across independent experimental batches. Measurements are repeated under nominally identical conditions, with bridge control sequences included in every batch. This structure supports separation of learnable variation from measurement instability and batch-induced artifacts.

Partial Factorial Execution

The MESD is not a full factorial design. Instead, a partial factorial execution is used to limit complexity while preserving interpretability. Coverage is concentrated near hypothesized boundaries, rather than evenly across the entire design space. This choice reflects the study's objective: mapping the boundary of learnability rather than estimating a global response surface.

Alignment with Hypotheses

The MESD is constructed to directly test the hypotheses defined in Section 3.

- H1 is evaluated through sequence-only analyses across repeated measurements.
- H2 is evaluated by introducing observable context as conditional input and examining whether learnability expands.
- H3 is supported by the design's ability to expose boundaries efficiently rather than by performance gains.
- H4 is evaluated by identifying regimes where prediction fails systematically despite controlled design and repetition.

Pre-Specification and Freezing of Design

All elements of the MESD - including axes of variation, repetition strategy, and interpretation criteria - are fixed prior to analysis. No adaptive modification of the design is permitted based on observed outcomes. This pre-specification is essential to prevent post hoc reinterpretation of failure as success and to ensure that collapse is meaningfully distinguished from noise.

6 Analysis Framework and Decision Rules

The analysis framework is designed to distinguish **learnable structure** from **non-learnable collapse** under controlled variation of sequence and context. Importantly, the framework does not optimize predictive accuracy. Instead, it defines *how* learnability and failure are declared, using criteria that are fixed prior to analysis and applied uniformly across conditions.

Analysis Objectives

The analysis pursues three objectives:

- (1) to determine whether predictive structure exists under specific informational constraints;
- (2) to identify conditions under which prediction fails despite controlled design and repetition; and
- (3) to classify failure modes in a way that is reproducible and interpretable.

Accordingly, all analyses are evaluated relative to **decision rules**, not relative to baseline performance.

Learnability Declaration

Learnability is declared only when predictive structure satisfies **both** of the following conditions:

- **Reproducibility across repetition.** Predictions must be consistent across independent batches and repetitions, as assessed using agreement and alignment measures anchored by bridge controls.
- **Stability under controlled variation.** Predictive relationships must persist when observable context is varied within the design envelope.

If either condition fails, the regime is not considered learnable, regardless of apparent performance on a single batch or subset.

Role of Context in Analysis

Analyses are conducted under progressively enriched information conditions:

- **Sequence-only analysis**, in which predictions are based solely on molecular sequence.
- **Sequence + observable context analysis**, in which recorded contextual variables (e.g., cell type, dose) are provided as conditional inputs.
- **Context-excluded analysis**, in which hidden-state variables are recorded but intentionally withheld from the model.

This progression enables explicit testing of whether observable context restores learnability and whether failures are attributable to unobserved state rather than model insufficiency.

Collapse Typology

Non-learnable collapse is not treated as a single outcome. Instead, failure modes are classified into the following categories, defined *a priori*:

- **C1: Measurement or standardization collapse.**

Prediction fails due to instability or inconsistency in measurement, as evidenced by poor agreement among bridge controls or batch misalignment.

- **C2: Hidden-state–driven collapse.**

Prediction fails when outcomes are dominated by unobserved biological state, such that adding

observable context does not restore reproducibility.

- **C3: Regime or non-linear collapse.**

Prediction fails due to abrupt transitions or non-linear behavior across sequence, context, or dose regimes, even when measurements are stable.

These categories are mutually exclusive at the level of primary attribution and are used to structure interpretation rather than to assign blame to models or data.

Decision Rules and Pre-Specification

All analytical decisions—including the declaration of learnability, the assignment of collapse categories, and the exclusion of interpretations—are governed by **pre-specified decision rules**. These rules are fixed prior to analysis and are not modified in response to observed outcomes.

When AI-generated explanations or hypotheses conflict with the decision rules, the rules take precedence. Such conflicts are treated as evidence of overgeneralization rather than as signals to relax interpretive criteria.

AI Participation in Analysis

The AI co-scientist contributes by proposing alternative analytical lenses, suggesting candidate metrics, and generating hypothetical interpretations. However, these suggestions are evaluated strictly within the confines of the pre-specified framework. The AI does not adapt the analysis criteria, redefine success, or reinterpret failure after outcomes are observed.

This constraint ensures that AI participation enhances exploratory reasoning without undermining epistemic discipline.

Interpretation Policy

Interpretation is limited to **mapping where learning holds and where it fails**. No claims are made regarding optimal delivery strategies, superior sequences, or improved performance. The analysis, therefore, yields a structured description of the learnability boundary and the distribution of collapse modes across the design space.

7 Failure Modes and Boundary Interpretation

The primary scientific contribution of this study lies not in improved delivery outcomes, but in the **interpretation of failure as structured information**. By design, the study exposes conditions under which learning succeeds and, more importantly, where it breaks down. This section synthesizes how the pre-defined failure modes (C1–C3) are interpreted as indicators of the learnability boundary rather than as experimental deficiencies.

Failure as a First-Class Outcome

In performance-centric delivery research, failure is typically treated as an artifact to be minimized or filtered out. In contrast, this work treats failure as a **first-class outcome**, on par with successful prediction. A failure is informative when it is reproducible, attributable, and interpretable under fixed analytical criteria. Under this perspective, the absence of learnable structure is itself a meaningful scientific result.

By enforcing repetition, batch structure, and pre-specified decision rules, the study distinguishes between random error and systematic collapse. This distinction enables failure to be analyzed rather than dismissed.

Interpreting Measurement and Standardization Collapse (C1)

Measurement or standardization collapse (C1) reflects conditions in which predictive failure arises

from instability in the measurement process itself. Such collapse is indicated by poor alignment across bridge controls or inconsistency across nominally identical repetitions. Importantly, C1 does not imply that delivery is intrinsically non-learnable; rather, it signals that the observational layer is insufficient to support learning.

From a boundary perspective, C1 delineates regions where improved standardization or instrumentation may be necessary before learnability can be meaningfully assessed. Treating these regions as “non-learnable” without qualification would conflate epistemic limits with measurement artifacts.

Interpreting Hidden-State–Driven Collapse (C2)

Hidden-state–driven collapse (C2) arises when unobserved biological variables dominate outcomes. In these regimes, prediction fails even when observable context is incorporated, indicating that the effective state space exceeds what is captured by the model inputs. Examples include shifts driven by cellular passage, stress responses, or intracellular conditions that are recorded but intentionally withheld from analysis.

C2 highlights a fundamental limit of sequence- and context-based learning under partial observability. From the perspective of boundary discovery, these regimes define areas where learnability may be restored only by expanding observation, not by increasing model complexity. Alternatively, they may represent domains where prediction is intrinsically constrained by biological variability.

Interpreting Regime and Non-Linear Collapse (C3)

Regime or non-linear collapse (C3) captures failures associated with abrupt transitions in behavior across sequence, context, or dose. In these cases, predictive relationships may hold locally but break down at regime boundaries, producing discontinuities that resist smooth modeling. Such collapse is not random; it reflects underlying non-linear biological dynamics.

C3 is particularly important for boundary interpretation, as it marks regions where learning is possible only within restricted domains. Attempting to extrapolate across these boundaries leads to systematic failure, regardless of data volume or model sophistication.

Mapping the Learnability Boundary

Taken together, C1–C3 define a **map of the learnability boundary** rather than a binary classification of success versus failure. The boundary is not a single line but a structured region shaped by measurement fidelity, observability of context, and biological regime behavior. This map provides a more nuanced understanding of delivery than aggregate performance metrics, revealing where learning is feasible, conditionally feasible, or fundamentally constrained.

By situating failure modes within this boundary framework, the study reframes delivery research as an exercise in epistemic mapping rather than optimization. The resulting interpretation shifts attention from “how to improve delivery” to “where improvement is theoretically possible and where it is not.”

Implications for AI-Driven Biological Research

The boundary-based interpretation has broader implications for AI-driven biology. It underscores the risk of conflating predictive accuracy with scientific understanding and highlights the importance of explicitly modeling the limits of learning. In this context, AI is most valuable not as an optimizer, but as a tool for **probing the structure of uncertainty and failure** under disciplined human control.

8 Limitations and Ethic

This study is intentionally constrained in scope. Its goal is not to demonstrate improved delivery performance or to propose deployable delivery strategies, but to interrogate the **limits of**

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

learnability in biological delivery under controlled abstraction. As such, several limitations are acknowledged, alongside ethical considerations specific to AI-assisted scientific research.

Methodological Limitations

First, the study adopts a **design-driven, dry-lab approach**. No wet-lab experiments, patient data, or clinical outcomes are involved. Consequently, the work does not validate delivery mechanisms *in vivo* or in clinical settings. The findings should be interpreted as epistemic insights into learnability boundaries rather than as evidence of practical efficacy.

Second, the **Minimal Executable Study Design (MESD)** prioritizes interpretability over coverage. By focusing on a limited set of sequences, contexts, and phenotypes, the design intentionally sacrifices breadth to expose failure modes. While this approach is suitable for boundary discovery, it does not aim to exhaustively characterize the delivery landscape.

Third, the decomposition of delivery into stage-specific phenotypes necessarily abstracts complex biological processes. Although this abstraction is essential for separating learnable from non-learnable regimes, it may omit interactions that become relevant at finer levels of biological detail.

Data and Reproducibility Constraints

All data used in this study are derived from **public sources or AI-generated synthetic abstractions**. No proprietary datasets, personal information, or protected biological data are included. While this choice ensures accessibility and ethical compliance, it also limits the study to phenomena that can be represented under these constraints.

Reproducibility is addressed through explicit repetition, batch structure, and bridge controls. However, reproducibility here is defined epistemically - consistency under controlled abstraction - rather than as laboratory-to-laboratory replication. The study does not claim that its abstractions fully capture the variability present in real-world experimental systems.

AI-Related Ethical Considerations

The use of an AI co-scientist introduces specific ethical considerations. The AI system participates in hypothesis generation, design abstraction, and analytical suggestion, but it does not conduct experiments, generate biological materials, or make autonomous scientific claims. All AI outputs are treated as provisional and are subject to human evaluation and constraint.

To prevent over-reliance on AI-generated reasoning, the study enforces **explicit role separation**: the AI proposes possibilities, while the human researcher determines validity, scope, and interpretation. Instances where AI-generated suggestions conflicted with pre-specified decision rules were documented and rejected rather than adapted.

The study also avoids generating or disseminating actionable biological protocols. By remaining at the level of abstract design and analytical framing, it mitigates risks associated with misuse or unintended application of biological knowledge.

Scope of Claims

This work does not claim to identify optimal delivery sequences, superior delivery methods, or actionable design rules for therapeutic development. It does not assert that non-learnable collapse is universal or immutable, nor does it preclude the possibility that expanded observation or alternative modeling paradigms could alter the boundaries of learnability.

All claims are limited to the interpretation of learnability under the defined abstractions and constraints.

9 Conclusion

This study reframes biological delivery from a performance-centric optimization problem into a **boundary-discovery problem** centered on learnability. Rather than asking how delivery can be

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

improved, we ask where learning from sequence and observable context is possible, where it conditionally holds, and where it structurally fails. By making these boundaries explicit, the work shifts attention from incremental gains to epistemic limits.

The proposed framework integrates a **Minimal Executable Study Design (MESD)**, stage-specific phenotypes, controlled context variation, and pre-specified decision rules to distinguish learnable structure from non-learnable collapse. Within this framework, failure is treated as a meaningful scientific outcome rather than as noise to be filtered or explained away. Measurement instability, hidden-state effects, and non-linear regime transitions are not conflated but instead classified as distinct collapse modes that define different regions of the learnability boundary.

A central methodological contribution of this work is the explicit incorporation of an **AI co-scientist** as part of the research process. By clearly separating AI-generated proposals from human judgment and constraint, the study demonstrates how AI can function as a collaborative reasoning agent without undermining scientific accountability. This interaction itself becomes an object of methodological scrutiny, highlighting both the strengths and limitations of AI-assisted scientific reasoning.

The contribution of this work is therefore not a new delivery mechanism or a superior predictive model, but a **principled way of asking the right questions** about biological delivery. By mapping where learning holds and where it does not, the framework provides a foundation for more disciplined future research—research that is explicit about its assumptions, transparent about its failures, and cautious in its claims.

More broadly, the boundary-discovery perspective articulated here may apply beyond biological delivery. In domains where AI models are increasingly used to predict complex biological phenomena, distinguishing between learnable structure and structural collapse is essential for preventing overconfidence and misinterpretation. Treating failure as knowledge is not a retreat from ambition, but a prerequisite for scientific progress.

References

- [1] Guidotti, G., Brambilla, L., & Rossi, D. (2017). Cell-penetrating peptides: From basic research to clinics. *Trends in Pharmacological Sciences*.
- [2] Stewart, M. P., Lorenz, A., Dahlman, J., & Sahay, G. (2016). Challenges in carrier-mediated intracellular delivery of proteins and nucleic acids. *Wiley Interdisciplinary Reviews: Nanomedicine and Nanobiotechnology*.
- [3] Torchilin, V. P. (2008). Tat peptide-mediated intracellular delivery of pharmaceutical nanocarriers. *Advanced Drug Delivery Reviews*.
- [4] Langer, R. (1998). Drug delivery and targeting. *Nature*.
- [5] Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*.

A Appendix / supplemental material

Optionally include supplemental material (complete proofs, additional experiments and plots) in appendix. All such materials **SHOULD** be included in the main submission

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: **[All claims in this paper are methodological and conceptual in nature. The paper does not claim performance improvements, empirical validation, or biological efficacy.]**

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 its limitations, including the use of a dry-lab, design-driven approach, the absence of wet-lab or clinical validation, and the intentional focus on learnability boundaries rather than performance outcomes. It also acknowledges abstraction limits arising from stage-wise phenotype decomposition and partial observability of biological context. All conclusions are scoped accordingly, and no claims beyond these limitations are 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: [N/A]

Justification: [The paper does not present formal theoretical results, mathematical theorems, or proofs. It focuses on conceptual framing, study design, and pre-specified decision rules for boundary discovery rather than analytical derivations. Accordingly, assumptions and proofs in the formal theoretical sense are not applicable.]

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: [N/A]

Justification: [The paper does not report experimental results or empirical performance outcomes. Its main claims and conclusions concern conceptual framing, study design, and pre-specified decision rules for identifying learnability boundaries and non-learnable collapse. Since no experimental results are presented, reproducibility of experimental outcomes is not applicable. Reproducibility is instead addressed at the level of design specification and analytical criteria.]

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 the supplemental material?

Answer: [N/A]

Justification: [The paper does not report experimental results or empirical evaluations that require executable code or datasets. Its contribution is conceptual and methodological, focusing on problem framing, study design, and pre-specified decision rules for boundary discovery rather than on producing experimental outcomes. Accordingly, there are no experimental results for which code or data reproduction is required. Public data sources and AI-generated synthetic abstractions are described for contextual grounding, but no executable pipelines are central to the paper's claims.]

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: [N/A]

Justification: [The paper does not include model training, empirical benchmarking, or experimental results that require specification of training or test settings. No data splits, hyperparameters, optimizers, or training procedures are involved in the paper's claims. Instead, the paper focuses on conceptual framing, study design (MESD), and pre-specified analytical decision rules for boundary discovery. Therefore, experimental settings and training details are not applicable.]

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: [N/A]

Justification: [The paper does not report experimental results, quantitative performance metrics, or empirical comparisons that would require statistical significance testing. No error bars, confidence intervals, or hypothesis tests are presented, as the paper's claims are conceptual and methodological in nature, focusing on boundary discovery, study design, and interpretation frameworks rather than on experimental outcomes. Therefore, statistical significance reporting is not applicable.]

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: [N/A]

Justification: [The paper does not include computational experiments, model training, or empirical evaluations that require specification of compute resources. No CPU/GPU configurations, memory requirements, or execution time are relevant to the paper's claims, which are conceptual and methodological in nature and focus on study design and analytical decision rules rather than experimental computation.]

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 authors have reviewed the NeurIPS Code of Ethics and confirm that the research fully conforms to it. The study is conducted as a dry-lab, design-driven investigation using public data and AI-generated synthetic abstractions only. No human subjects, patient data, or sensitive biological information are involved. The paper avoids unsafe or actionable biological protocols, preserves author anonymity, and clearly delineates human responsibility for all scientific judgments and conclusions.]

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 broader impacts at a conceptual level. Potential positive impacts include promoting more rigorous and responsible AI-driven scientific research by emphasizing learnability boundaries, reproducibility, and the treatment of failure as knowledge. Potential negative impacts are limited, as the work is foundational and does not propose deployable systems, biological protocols, or actionable methods. To mitigate misuse or misinterpretation, the paper explicitly avoids performance claims, experimental validation, and application-level guidance, and confines its contributions to abstract study design and analytical framing.]

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: [N/A]

Justification: [The paper does not release pretrained models, executable code, datasets scraped from the Internet, or any artifacts with a high risk of misuse or dual-use. Its contributions are conceptual and methodological, focusing on study design and analytical framing. No deployable systems, biological protocols, or reusable models are provided; therefore, safeguards for responsible release are not applicable.]

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 are properly credited and originate from publicly available sources. Public databases and literature resources (e.g., UniProtKB and PubMed) are cited as the sources of sequence data and literature metadata, and their terms of use permit academic research and analysis. No proprietary code, pretrained models, or restricted datasets are used or redistributed. AI-generated synthetic abstractions are derived from public information and are not released as standalone assets. Accordingly, all licenses and terms of use of existing assets are respected.]

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: [N/A]

Justification: [The paper does not release new datasets, code, models, or other reusable assets. While AI-generated synthetic abstractions are discussed conceptually as part of the study design, these are not released as standalone assets and are not intended for reuse. Accordingly, no new assets require documentation, licensing, consent procedures, or anonymized release.]

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: [N/A]

Justification: [The paper does not involve crowdsourcing, user studies, surveys, or research with human subjects. No participants are recruited, instructed, or compensated, and no human-generated data are collected. The study is conducted as a dry-lab, design-driven investigation using public data sources and conceptual abstractions only. Therefore, requirements related to human subjects or crowdsourcing do not apply.]

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: [N/A]

Justification: [The paper does not involve research with human subjects, crowdsourcing, surveys, or user studies. No participants are recruited, no human-generated data are collected, and no interventions or interactions with individuals occur. Accordingly, there are no participant risks to disclose and no requirement for IRB or equivalent ethical review.]

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.
-