Extrapolative Controlled Sequence Generation via Iterative Refinement

Vishakh Padmakumar, Richard Pang, He He, Ankur P. Parikh
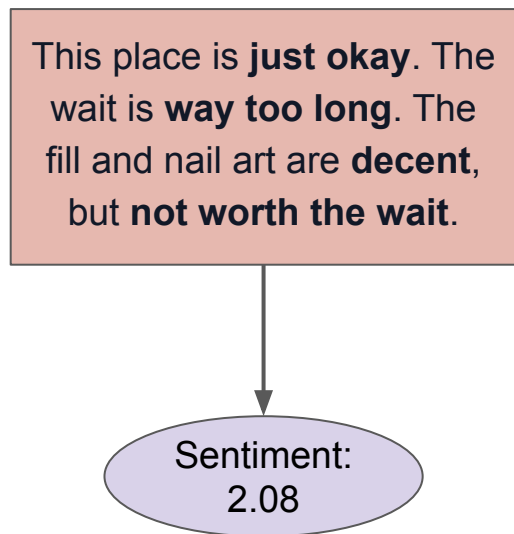
Traditional Controlled Sequence Generation

Traditional Controlled Sequence Generation
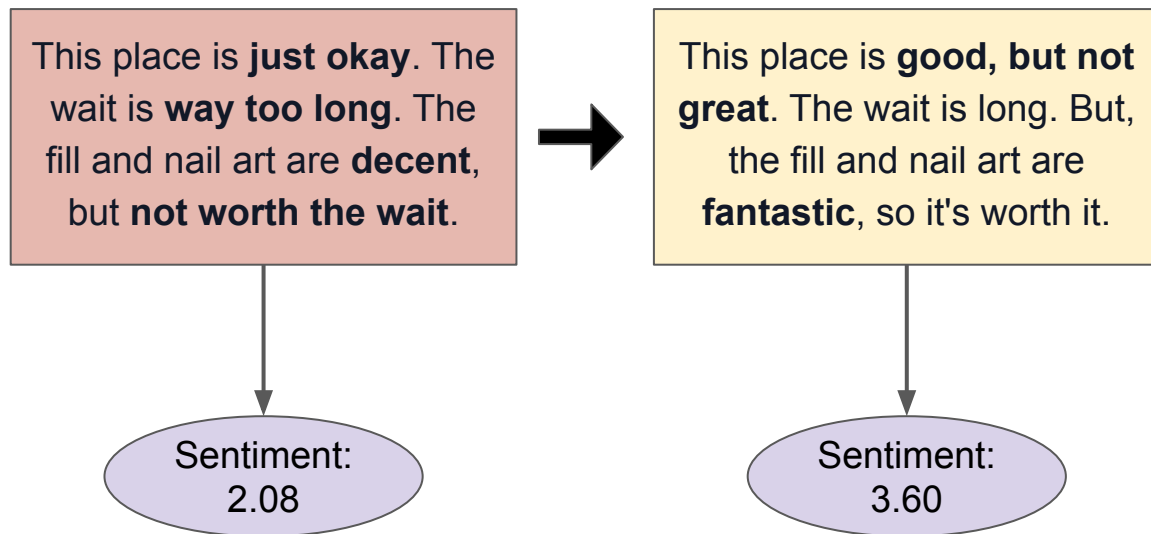
Example: Sentiment (Scale 1 to 5) associated with a review

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

Sentiment: 2.08

Traditional Controlled Sequence Generation

Problem: Modify the sentiment to be more positive i.e. 3.60

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

→

This place is **good, but not great**. The wait is long. But, the fill and nail art are **fantastic**, so it's worth it.

Sentiment: 2.08

Sentiment: 3.60

Extrapolative Controlled Sequence Generation

Problem: Modify the sentiment to be more positive

Training Data Range (2 to 4)

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

→

This place is **good, but not great**. The wait is long. But, the fill and nail art are **fantastic**, so it's worth it.

Sentiment: 2.08

Sentiment: 3.60

Extrapolative Controlled Sequence Generation

Problem: Modify the sentiment to be more positive *than seen during training*

Training Data Range (2 to 4)

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

This place is **good, but not great**. The wait is long. But, the fill and nail art are **fantastic**, so it's worth it.

Sentiment: 2.08

Sentiment: 3.60

Extrapolative Controlled Sequence Generation

Problem: Modify the sentiment to be more positive *than seen during training*

Training Data Range (2 to 4)

Test Time Extrapolation to Unseen Values

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

This place is **good, but not great**. The wait is long. But, the fill and nail art are **fantastic**, so it's worth it.

This place is **great**! The wait is **easy to overlook** when I leave with **the best** fill and nail art **ever**!!

Sentiment: 2.08

Sentiment: 3.60

Sentiment: 4.56

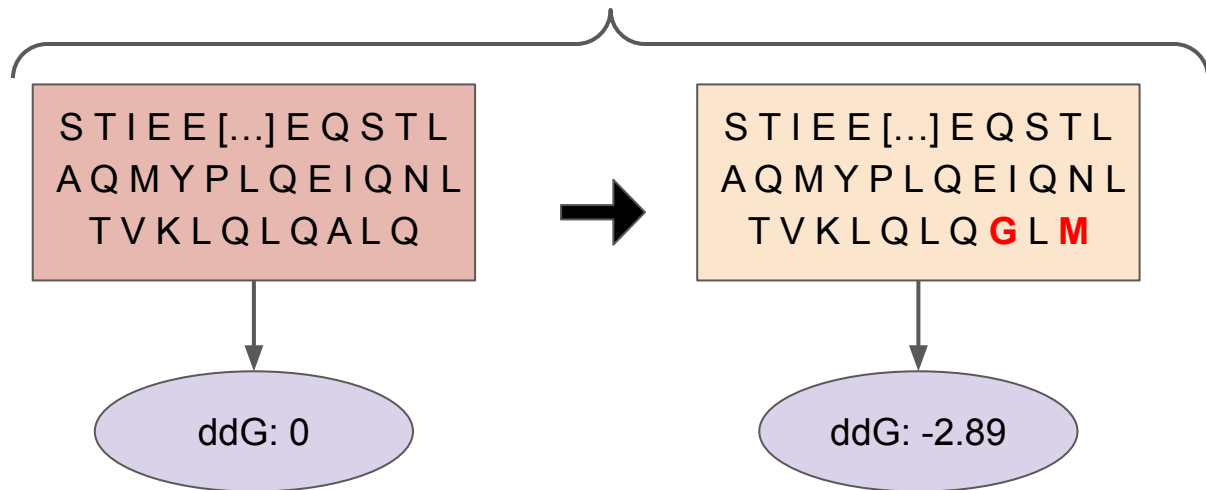# Motivation for Extrapolation Setting

In protein design (and other design problems), we might want to generate mutations more stable than those seen during training

Motivation for Extrapolation Setting

In protein design (and other design problems), we might want to generate mutations more stable than those seen during training

Motivation for Extrapolation Setting

In protein design (and other design problems), we might want to generate mutations more stable than those seen during training

Key Idea

Key Idea: Learn to make **local edits** to improve the attribute value at train time
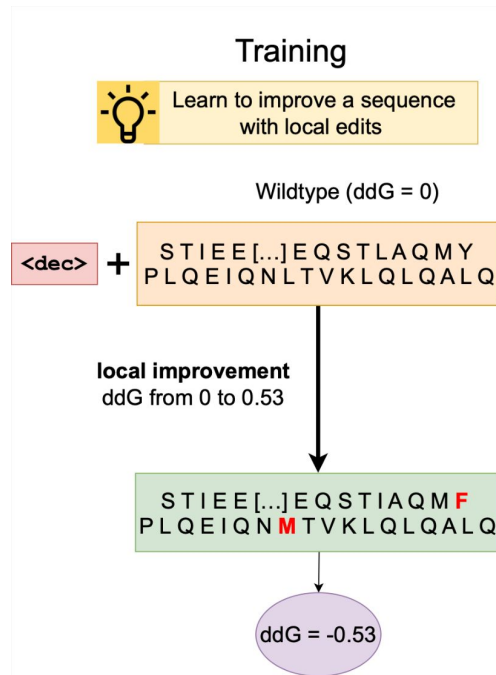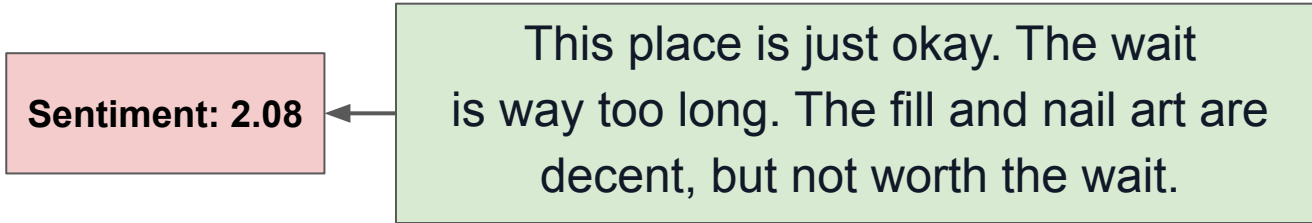
Key Idea: Learn to make **local edits** to improve the attribute value at train time, and apply **these in succession**
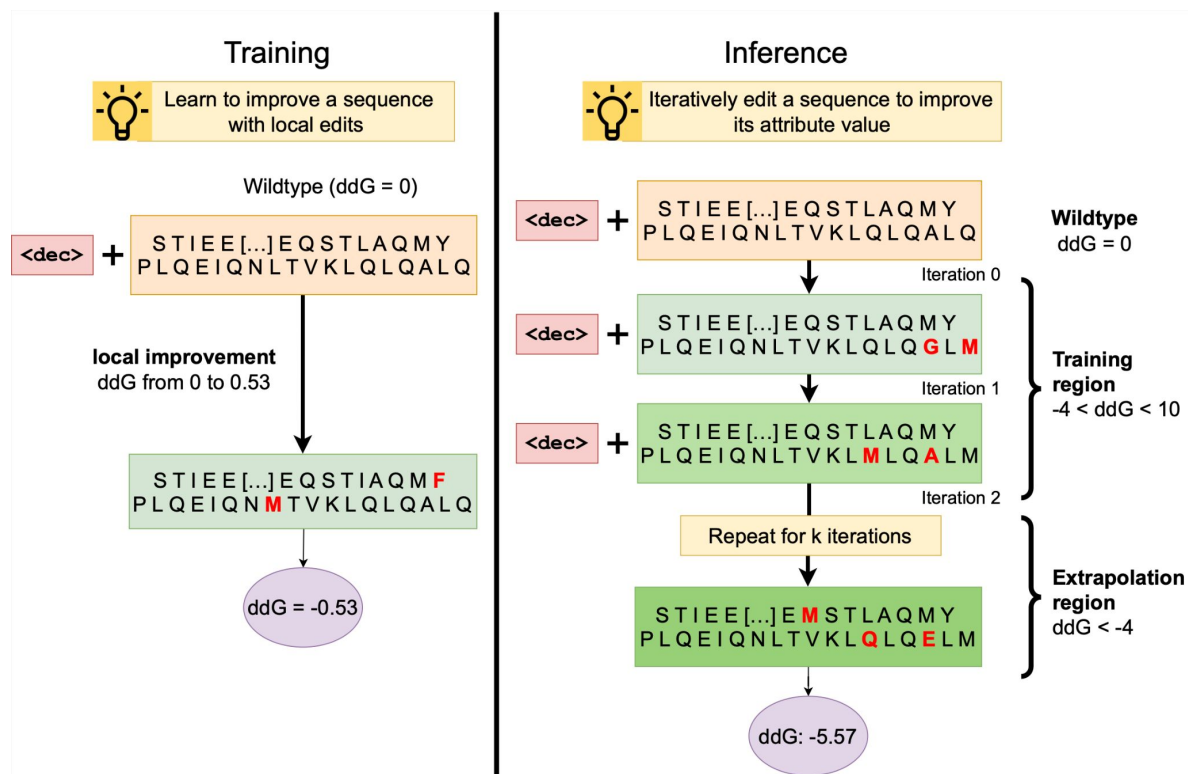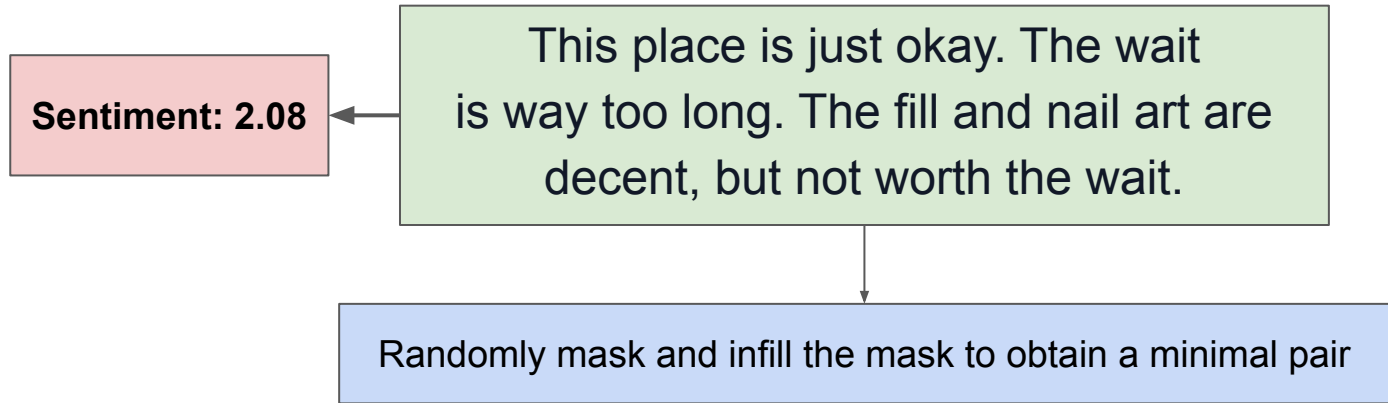
Training the Editor Model: Synthetic Data Creation

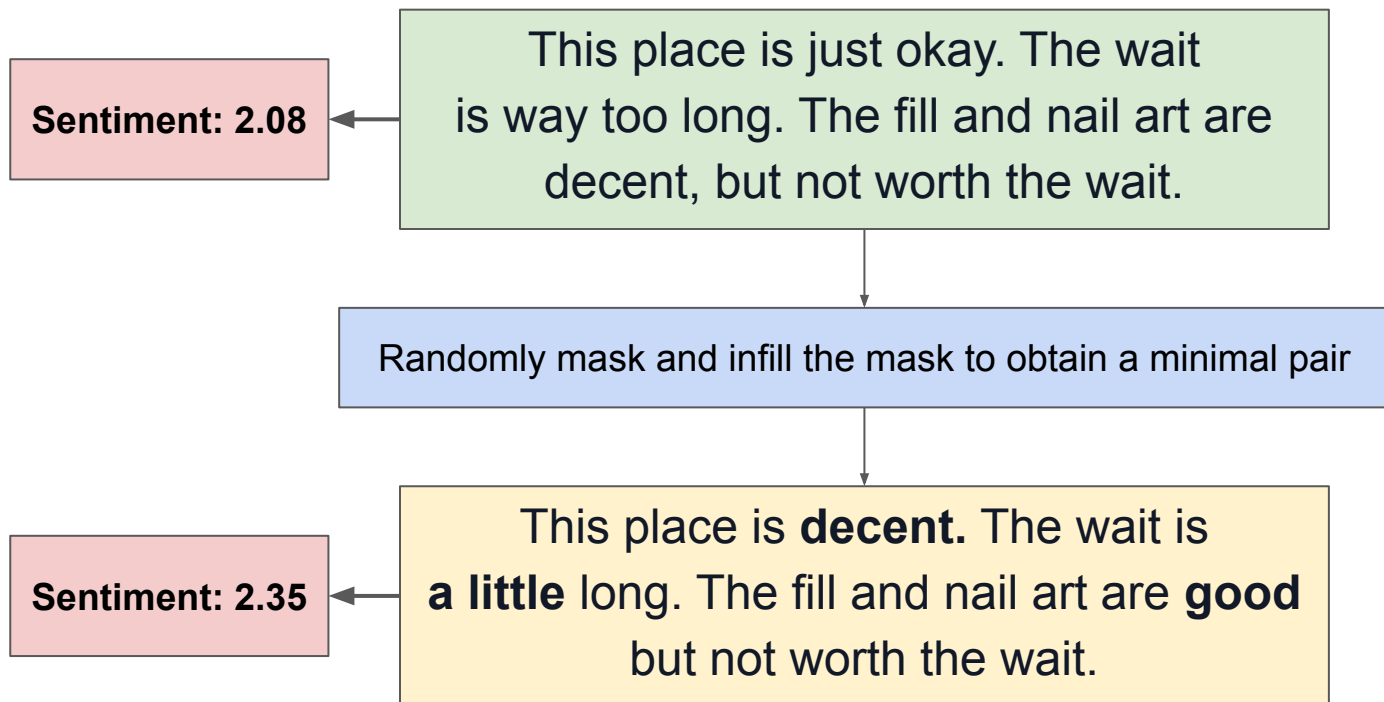# Training the Editor Model: Synthetic Data Creation

Sentiment: 2.08

This place is just okay. The wait is way too long. The fill and nail art are decent, but not worth the wait.

Training the Editor Model: Synthetic Data Creation

Sentiment: 2.08

This place is just okay. The wait is way too long. The fill and nail art are decent, but not worth the wait.

Randomly mask and infill the mask to obtain a minimal pair

Training the Editor Model: Synthetic Data Creation

Sentiment: 2.08

This place is just okay. The wait is way too long. The fill and nail art are decent, but not worth the wait.

Randomly mask and infill the mask to obtain a minimal pair

Sentiment: 2.35

This place is **decent.** The wait is **a little** long. The fill and nail art are **good** but not worth the wait.

Training the Editor Model: Fine-Tuning on Synthetic Pairs

Training the Editor Model: Fine-Tuning on Synthetic Pairs

Control Code   **Input**   **Output**

This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.

Training the Editor Model: Fine-Tuning on Synthetic Pairs

Control Code

Input

Output

<INC>

This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.

Training the Editor Model: Fine-Tuning on Synthetic Pairs

Control Code   **Input**   **Output**

<INC>

This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.

➡

This place is **decent.** The wait is **a little** long. The fill and nail art are **good** but not worth the wait.

<DEC>

This place is **decent.** The wait is **a little** long. The fill and nail art are **good** but not worth the wait.

➡

This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.

Iterative Inference: Example

This place is **just okay**. The wait is **way too long**. The fill and nail art are **decent**, but **not worth the wait**.

⬇

This place is **average at best**. The wait is **long**. The fill and nail art are decent, but nothing special.

⬇

This place is **good, but not great**. The wait is long. But, the fill and nail art are **fantastic**, so it's worth it.

⬇

This place is **pretty good**! The wait is **easy to ignore** when I leave with **awesome** fill and nail art!

⬇

This place is **great**! The wait is **easy to overlook** when I leave with **the best** fill and nail art **I've ever had**!

Experimental Details

- Results on 3 different tasks

  - **Natural Language:** Sentiment Control on the Yelp Dataset

  - **Protein Design:** ACE2 Stability and AAV fitness tasks

Experimental Details

- Results on 3 different tasks

  - **Natural Language:** Sentiment Control on the Yelp Dataset

  - **Protein Design:** ACE2 Stability and AAV fitness tasks

- **Evaluation Metric:** Success rate at editing examples to various target values

  in the training and extrapolation regions

Experimental Details

- Results on 3 different tasks

  - **Natural Language:** Sentiment Control on the Yelp Dataset

  - **Protein Design: ACE2 Stability** and AAV fitness tasks

- **Evaluation Metric:** Success rate at editing examples to various target values in the training and extrapolation regions

Task Details for ACE 2 Task

- Aim to generate mutations of the **ACE2 protein** with *higher* stability

Task Details for ACE 2 Task

- Aim to generate mutations of the **ACE2 protein** with *higher* stability

- Compare performance to two baselines: **Sampling** and **Genhance**

Task Details for ACE 2 Task

- Aim to generate mutations of the **ACE2 protein** with *higher* stability

- Compare performance to two baselines: **Sampling** and **Genhance**

  - Generate 10,000 mutants for the protein and compare the library of sequences

Task Details for ACE 2 Task

- Aim to generate mutations of the **ACE2 protein** with *higher* stability

- Compare performance to two baselines: **Sampling** and **Genhance**

  - Generate 10,000 mutants for the protein and compare the library of sequences

- Measure stability using the **FoldX** software

Experimental Results - ACE2 Stability Task
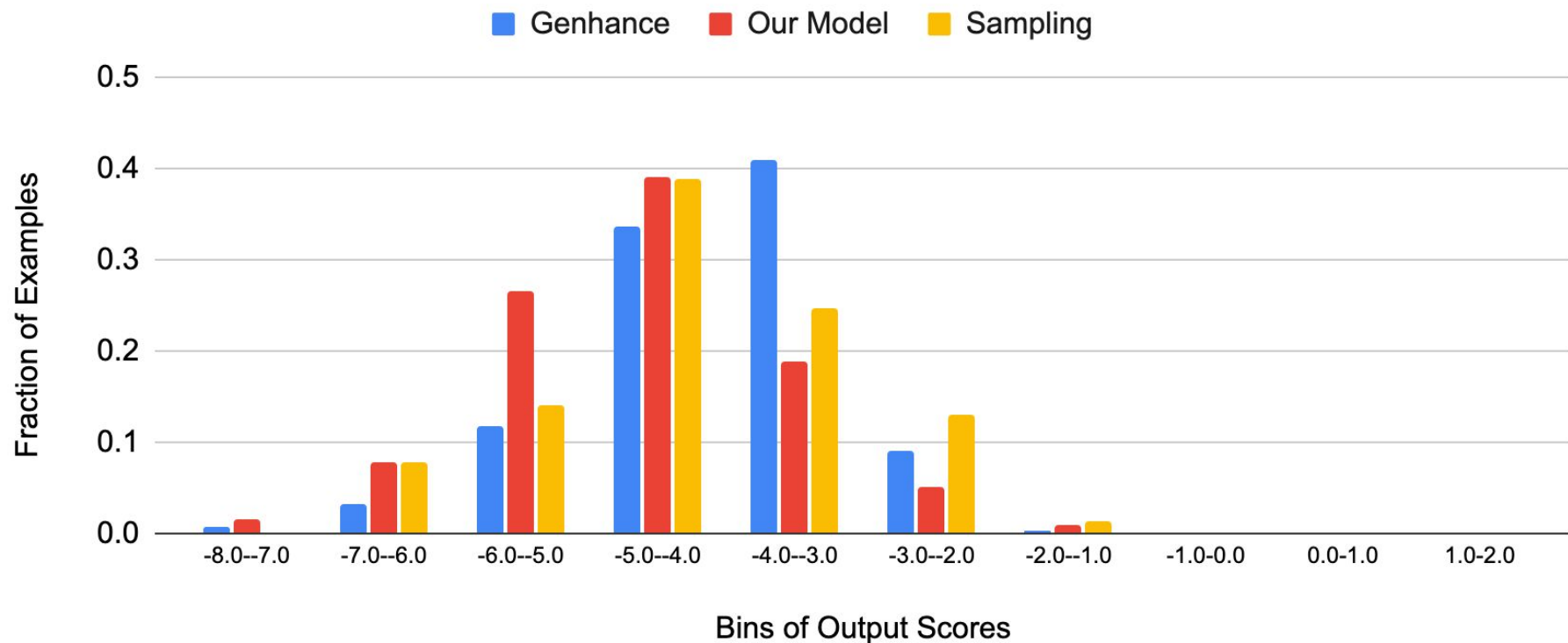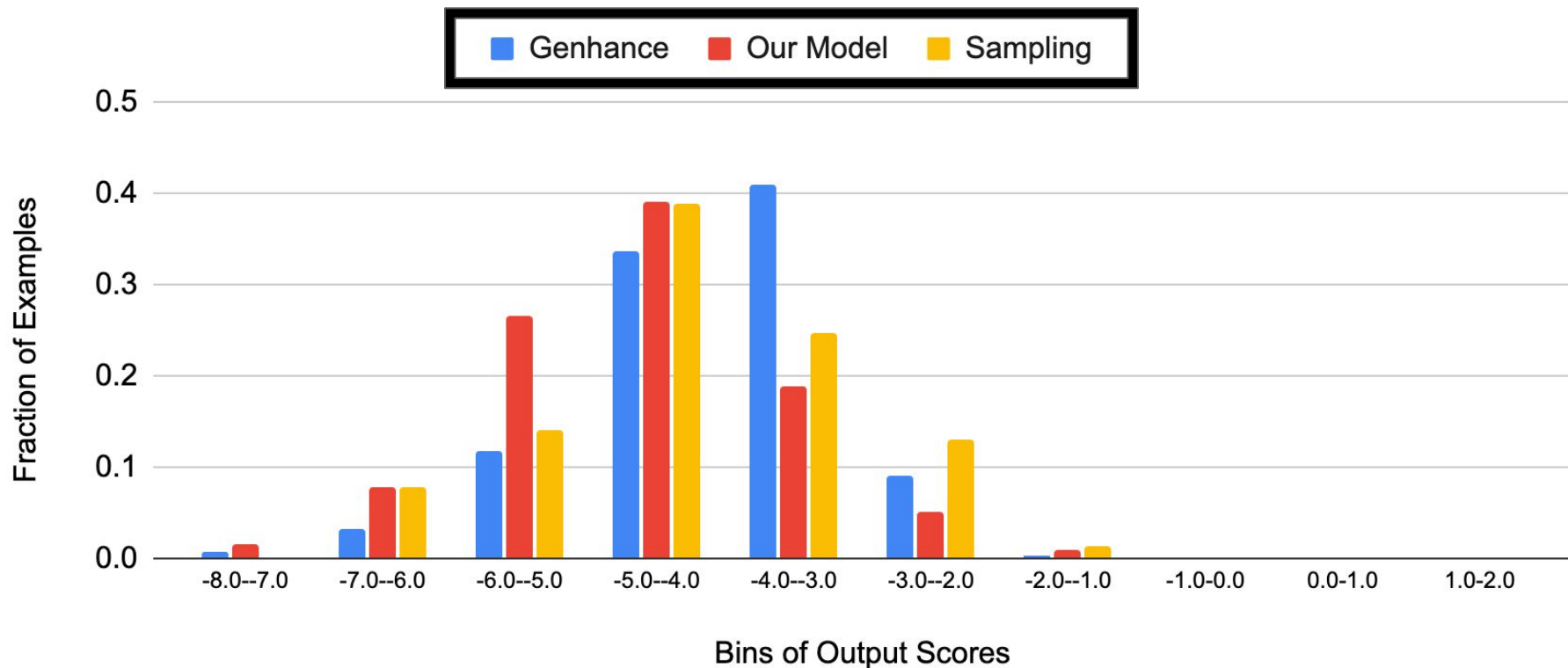
Experimental Results - ACE2 Stability Task
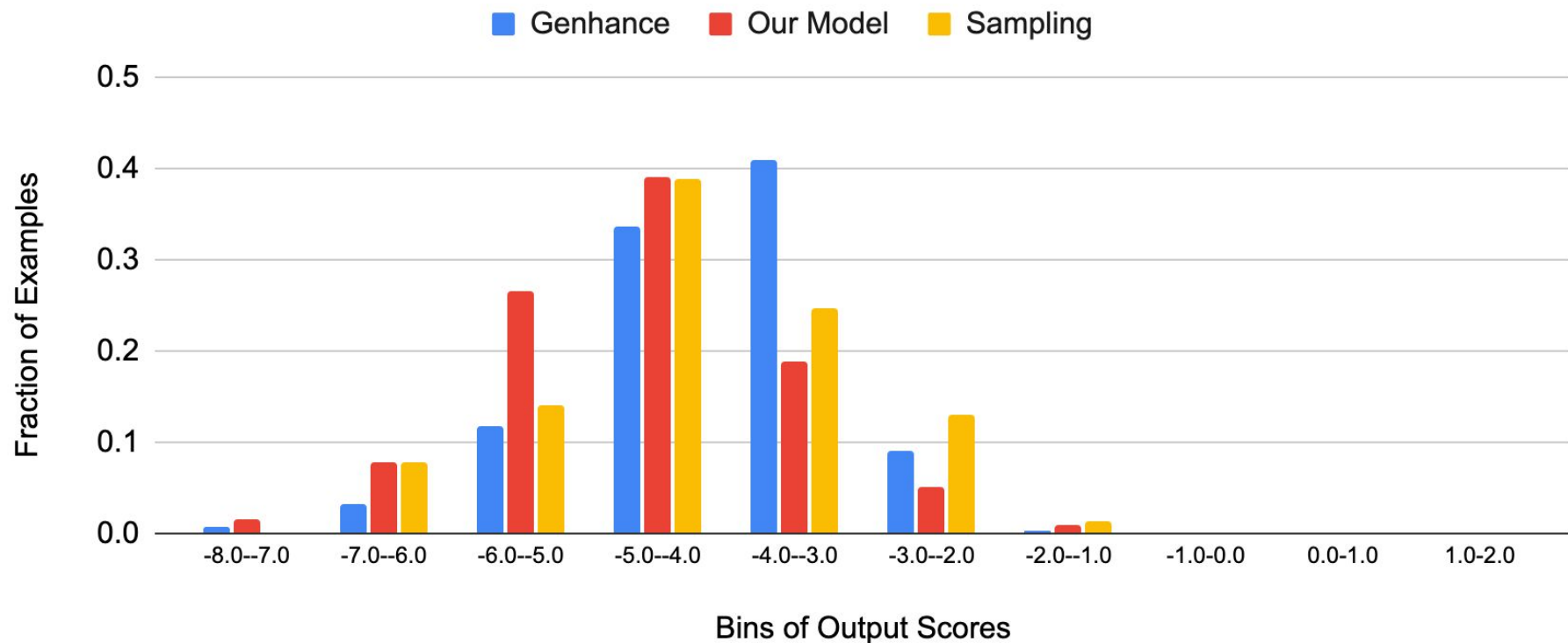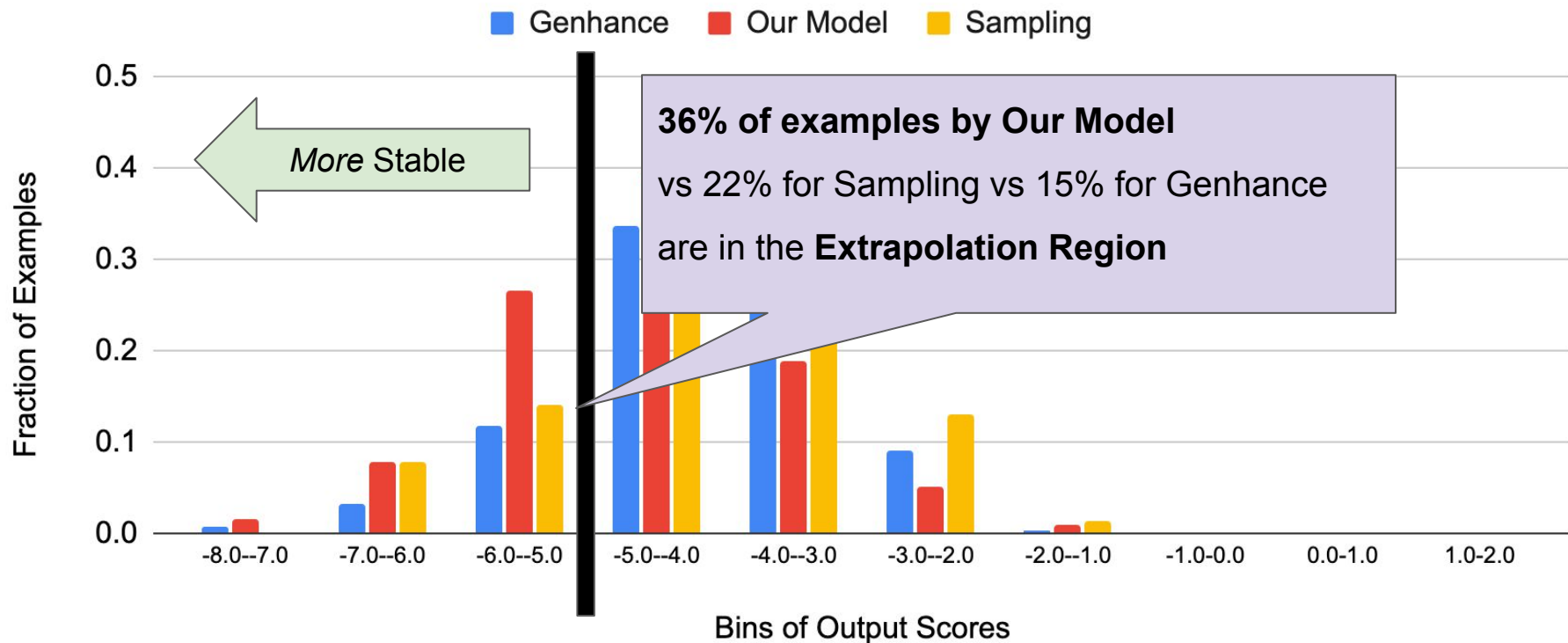
Experimental Results - ACE2 Stability Task
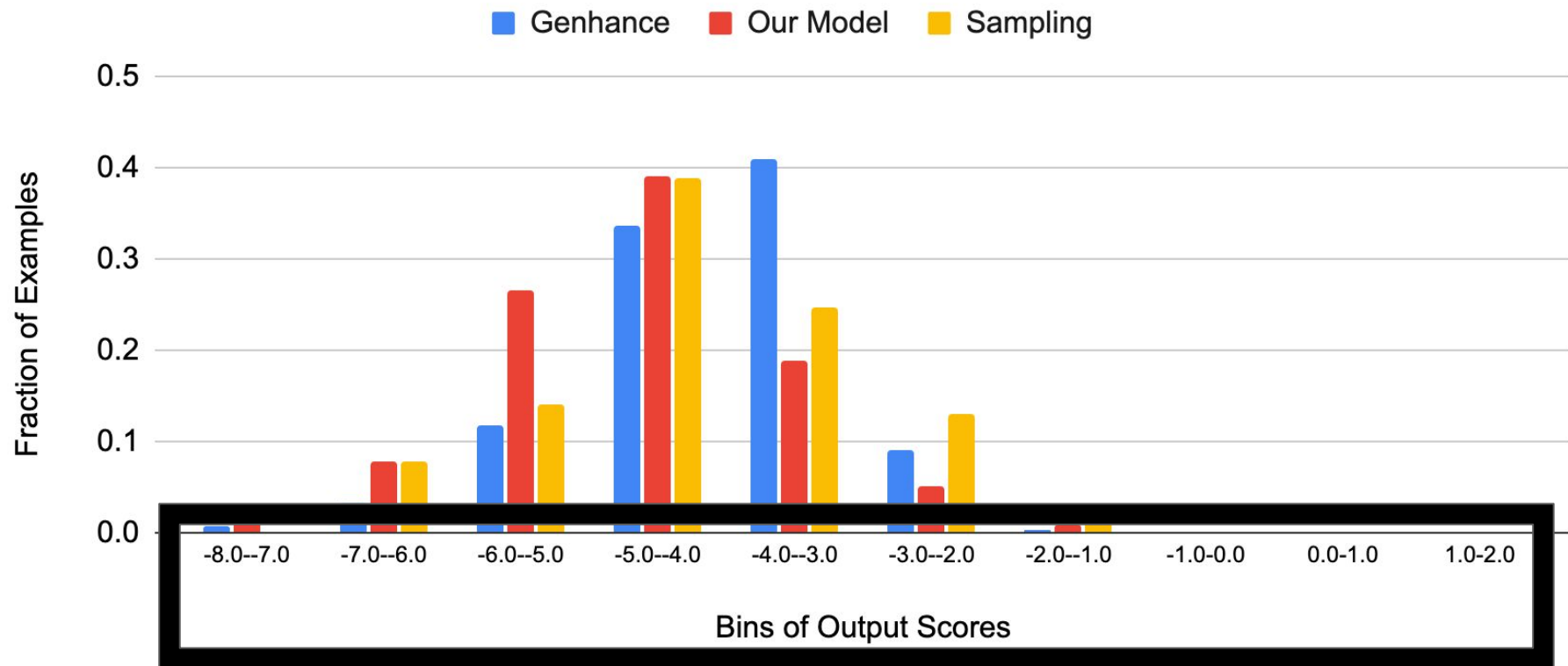
Experimental Results - ACE2 Stability Task
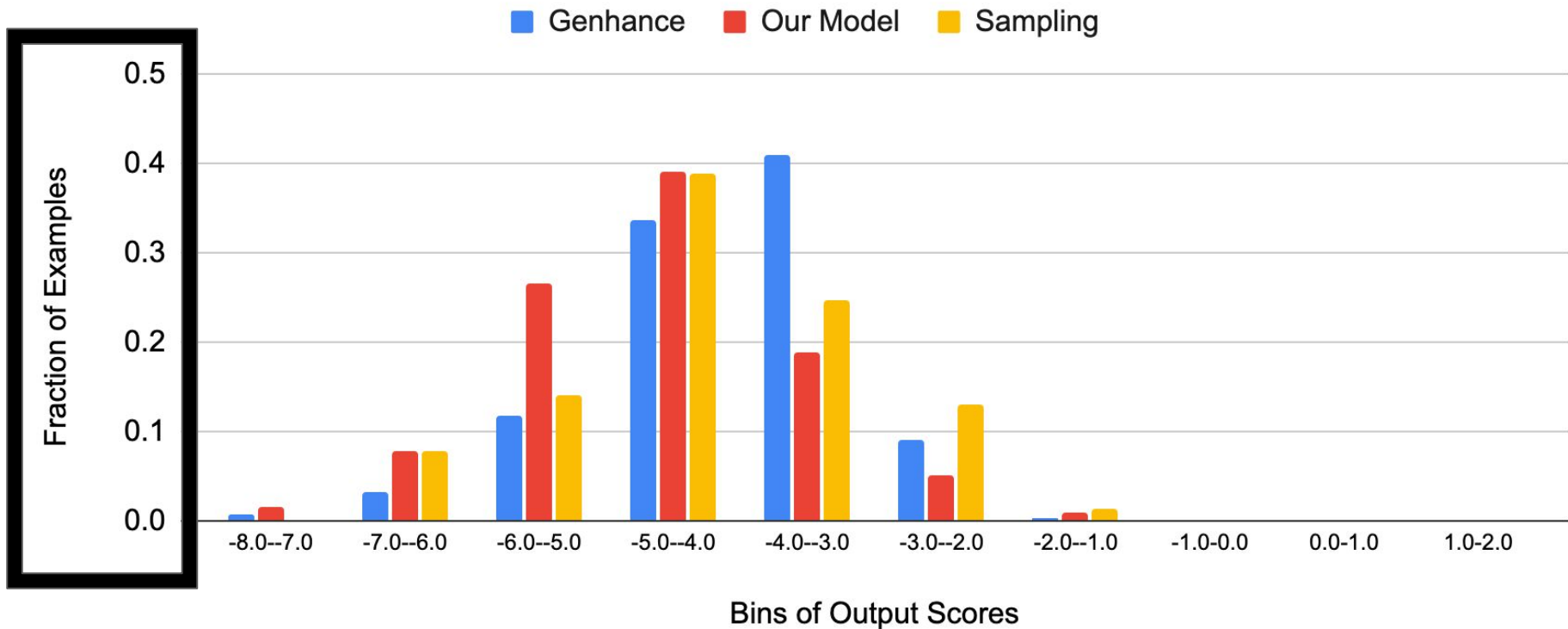
Experimental Results - ACE2 Stability Task

Iterative Editing Results in Better Extrapolation

Iterative Editing Results in Better Extrapolation

Iterative Editing Results in Higher Quality Candidates Even At the Head of the Distribution

	Iterative Sampling	Genhance	Our Model
All 10k Candidates	-4.326	-4.086	**-4.660**
Top 1k Candidates	-5.866	-6.030	**-6.575**
Top 100 Candidates	-6.413	-7.354	**-7.938**

Average ddG Score of Output Sequences

Takeaways

- Through iterative editing, we are able to generate *more* stable proteins and edit sequences to attribute values *beyond the training data range*
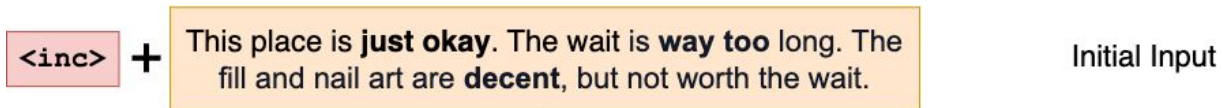
Takeaways

- For more details, including detailed results, inference methods, variation to

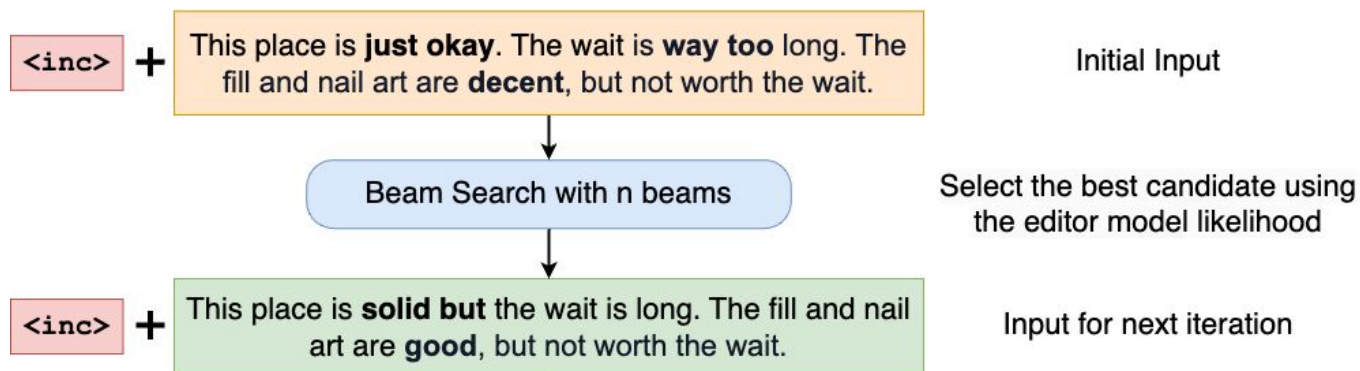 the generation hyperparameters, and more, check out our paper!
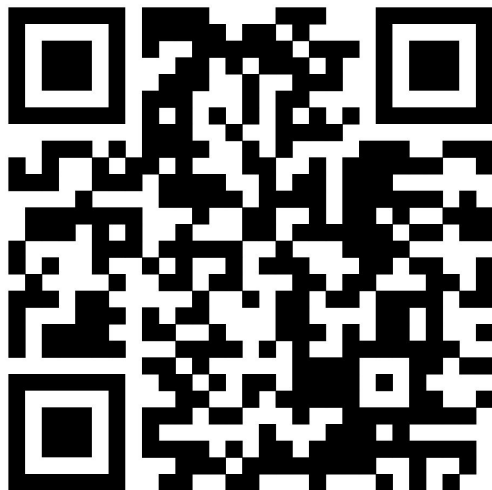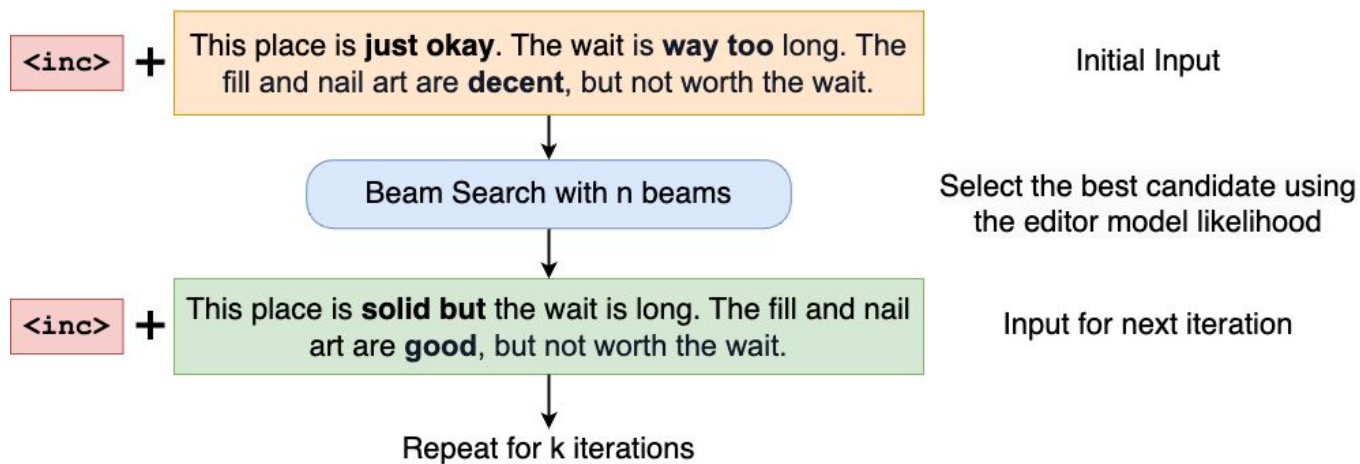
Iterative Inference

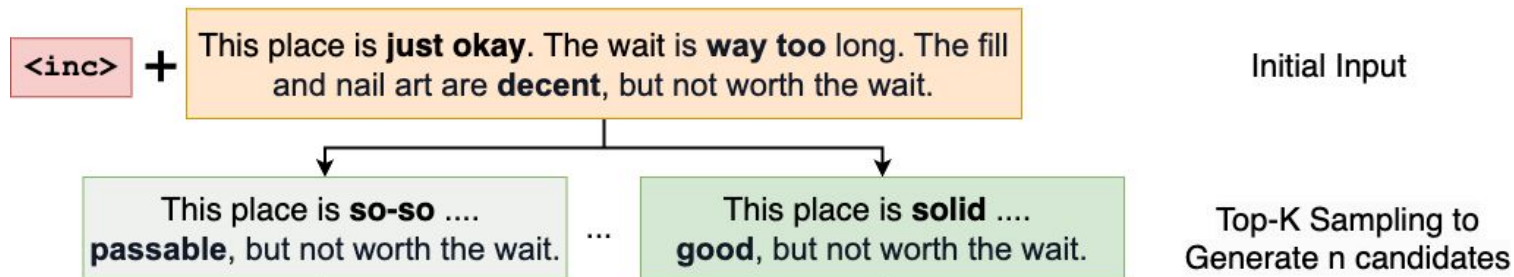# Iterative Inference: Guided by Editor Likelihood

<inc> + This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.   Initial Input

Iterative Inference: Guided by Editor Likelihood

Iterative Inference: Guided by Editor Likelihood

Iterative Inference: Guided by a Trained Scorer

- Based on the available training data, we can train a classifier f

Iterative Inference: Guided by a Trained Scorer
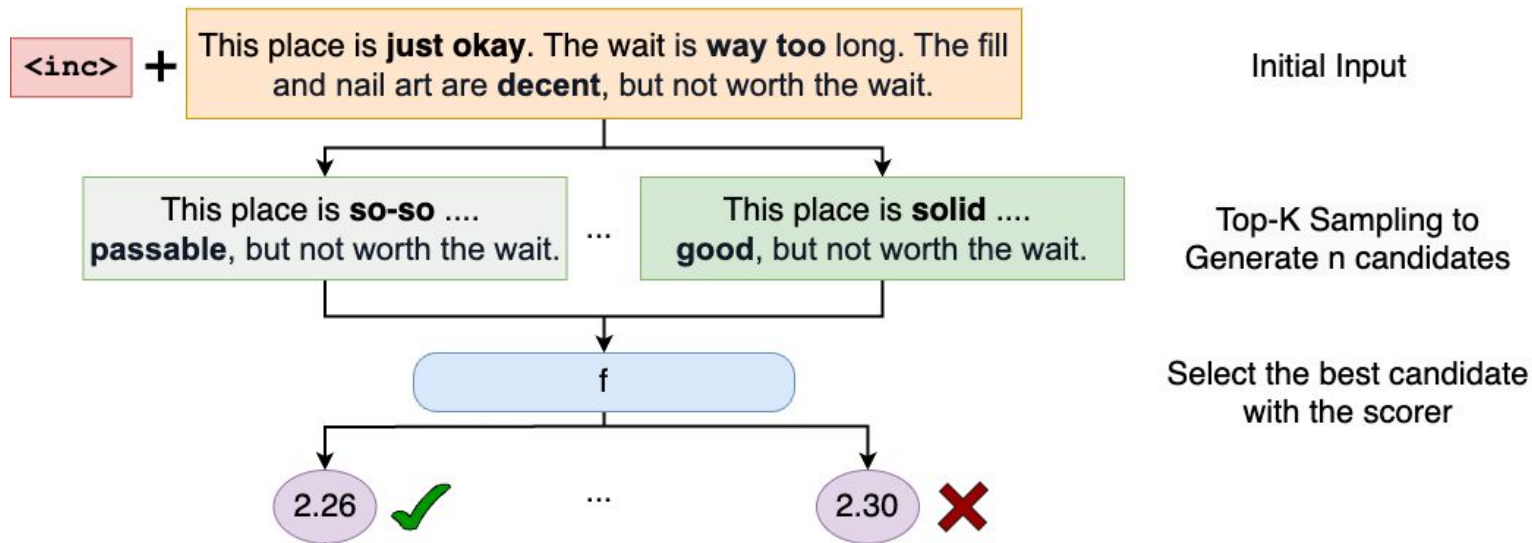
`<inc>` **+** This place is **just okay**. The wait is **way too** long. The fill and nail art are **decent**, but not worth the wait.
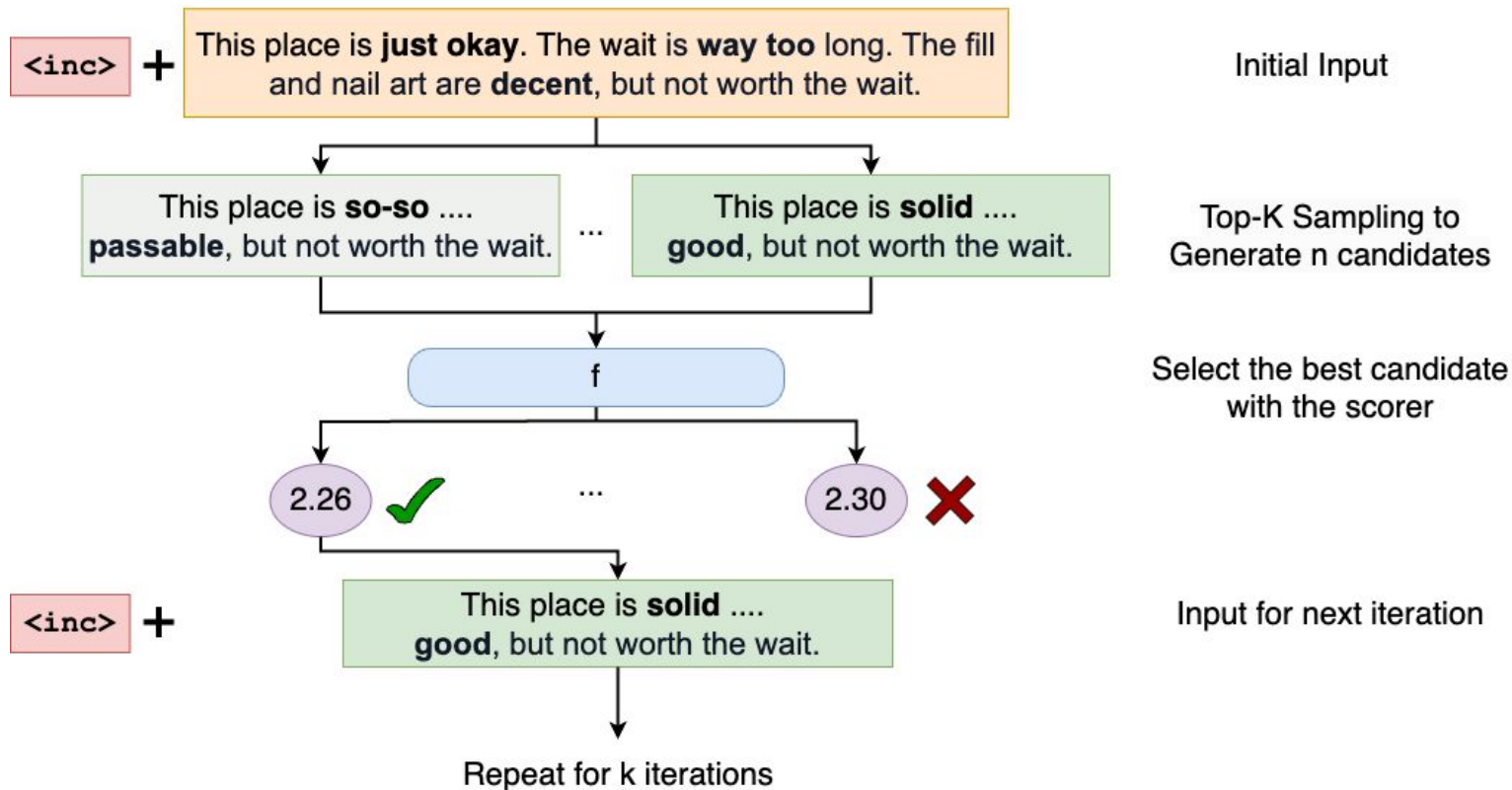
Initial Input

Iterative Inference: Guided by a Trained Scorer

Iterative Inference: Guided by a Trained Scorer

Iterative Inference: Guided by a Trained Scorer

Iterative Inference: Guided by a Trained Scorer