

Don't Settle Too Early: Self-Reflective Remasking for Diffusion Language Models

Zemin Huang, Yuhang Wang, Zhiyang Chen, Guo-Jun Qi

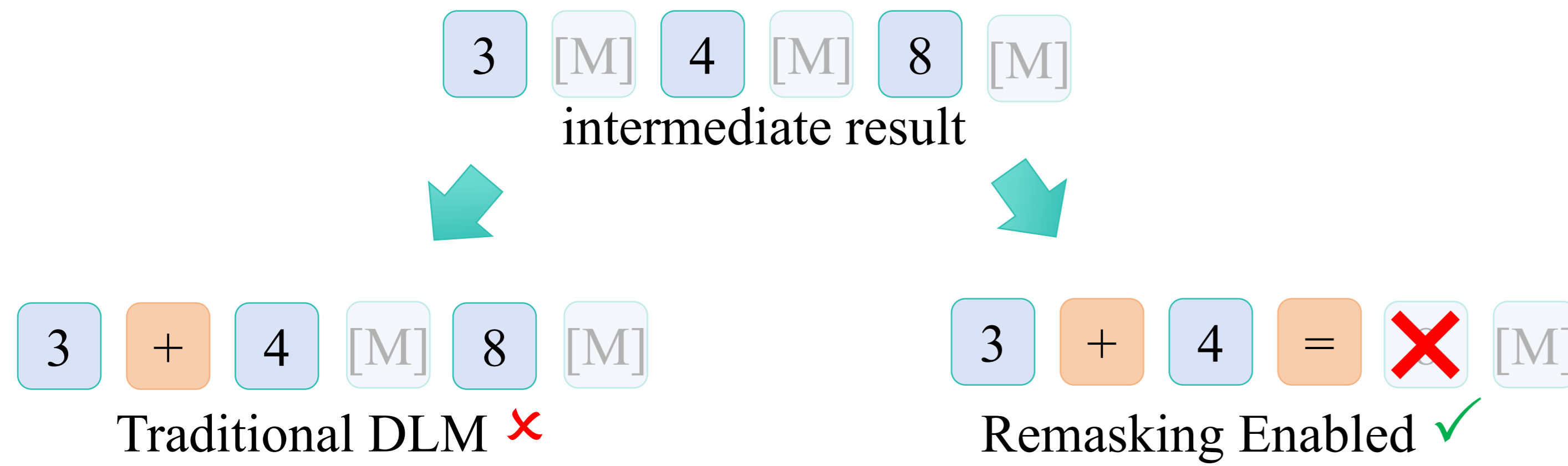


Model & Code Released!
maple-research-lab/RemeDi

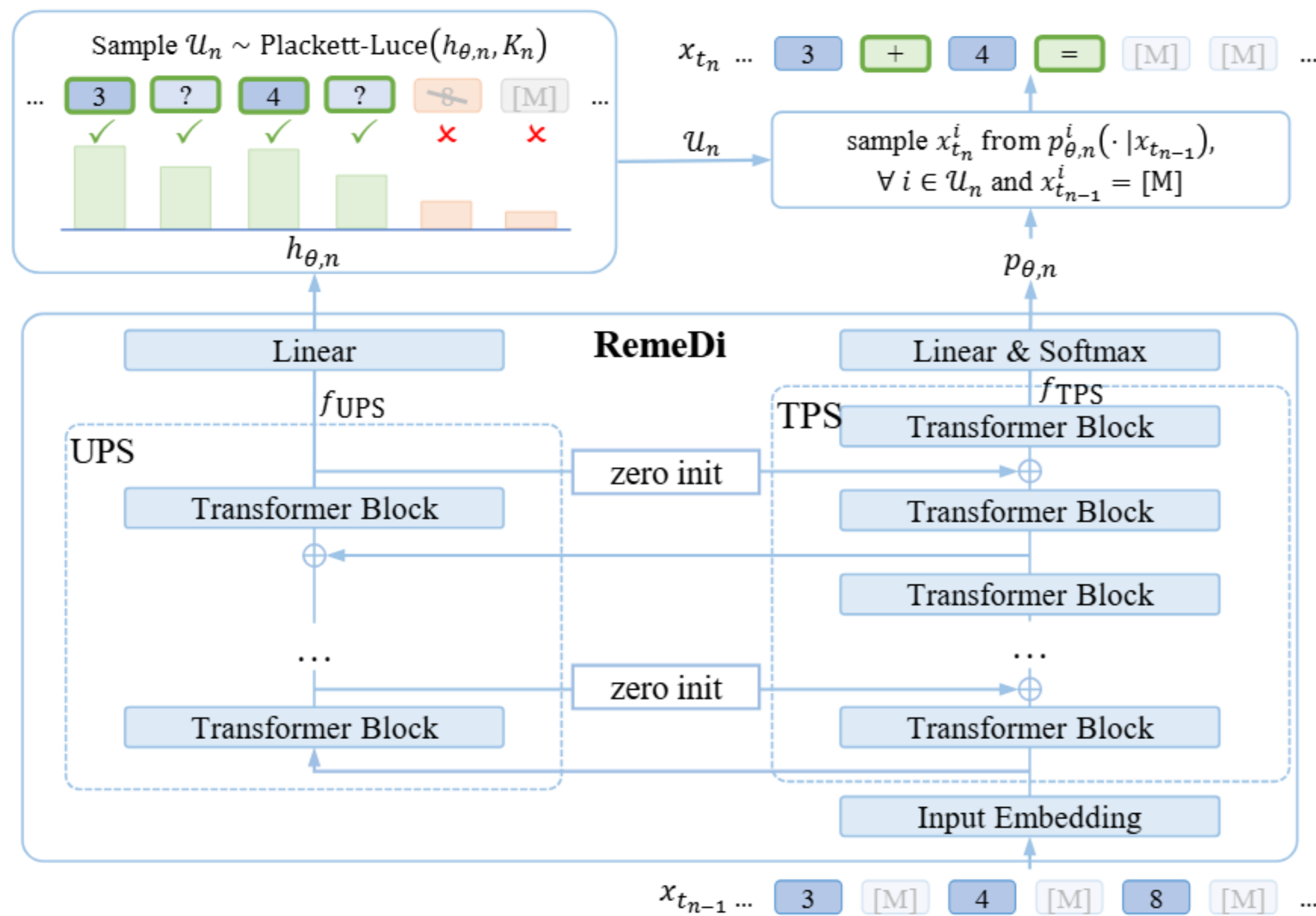


Motivation

Introduce *remasking* as a fundamental mechanism in mask-based DLMs to enable flexible text refinement during generation.



How to Achieve This?



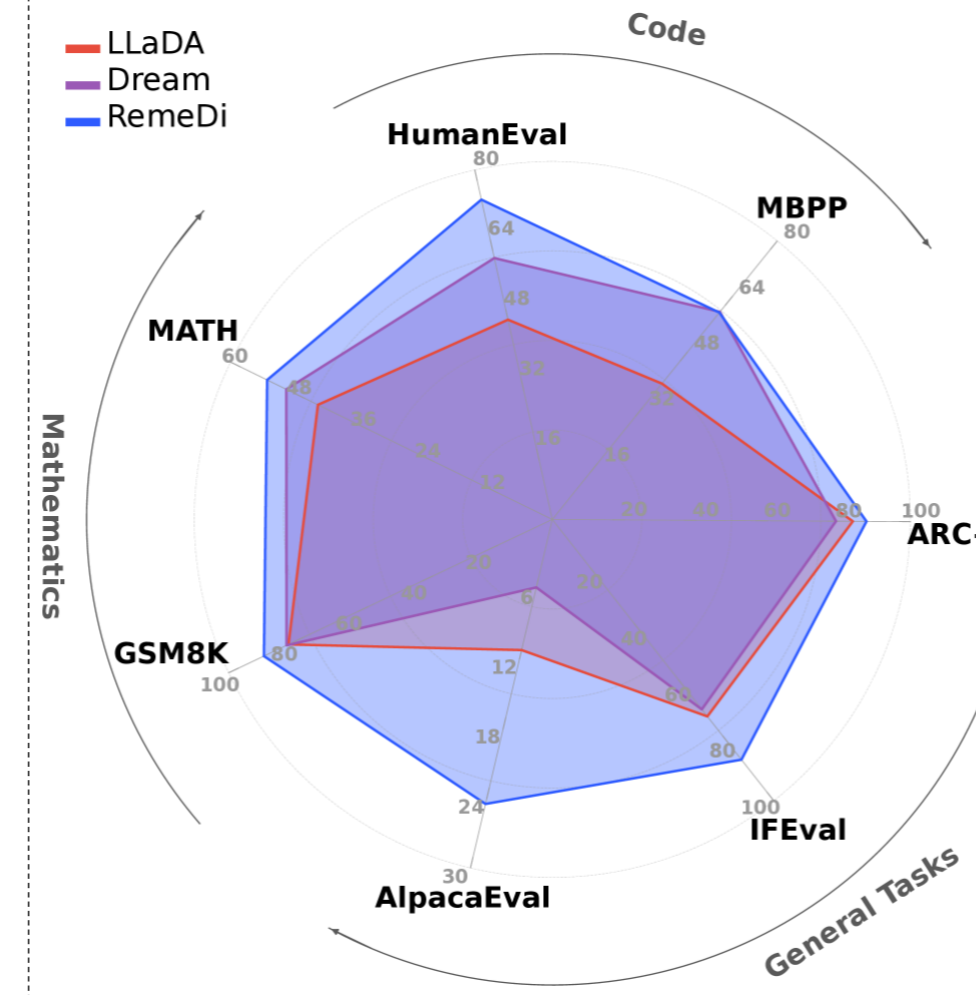
Dual-Stream Transformer

Unmasking Policy Stream +

predict confidence to rank tokens, identify low-quality tokens to be remasked

Token Prediction Stream

predict tokens for masked positions



Experiment Results

RemeDi achieves SOTA results!

Method	Hellaswag	ARC-C	IFEval	AlpacaEval
Diffusion Language Models				
Dream (Ye et al., 2025)	70.3	79.2	67.5	5.9
LLaDA (Nie et al., 2025)	69.7	83.9	70.0	11.2
LLaDA 1.5 (Zhu et al., 2025)	70.5	83.5	73.5	13.9
RemeDi (+ Remask SFT)	71.1	85.2	81.9	12.5
RemeDi (++) Remask RL)	72.2	87.7	85.4	24.8
Auto-regressive Models				
LLaMA2 7B (Touvron et al., 2023)	51.5	57.3	-	-
Deepseek 7B (Bi et al., 2024)	68.5	49.4	-	-
LLaMA3 8B (Dubey et al., 2024)	75.5	82.4	-	-

Method	GSM8K	Math MATH	GPQA	Code HumanEval	MBPP
Diffusion Language Models					
Dream (Ye et al., 2025)	82.1	49.6	30.6	59.8	59.6
LLaDA (Nie et al., 2025)	78.3	38.9	28.1	45.7	39.0
LLaDA + ReMDM (Wang et al., 2025a)	81.4	38.5	-	44.5	37.8
d1-LLaDA (Zhao et al., 2025)	82.1	-	-	37.8	44.7
wd1-LLaDA (Tang et al., 2025)	82.3	-	-	-	-
LLaDA 1.5 (Zhu et al., 2025)	83.3	42.6	36.9	52.4	42.8
LLaDOU (Huang et al., 2025)	88.1	44.6	-	59.1	51.6
RemeDi (+ Remask SFT)	86.3	51.4	32.6	71.3	57.8
RemeDi (++) Remask RL)	89.1	52.9	29.5	73.2	59.4
Auto-regressive Models					
LLaMA2 7B (Touvron et al., 2023)	14.6	2.5	28.4	12.8	20.8
MetaMath 7B (Yu et al., 2023)	66.5	19.8	-	-	-
CodeLLaMA 7B (Roziere et al., 2023)	-	-	-	34.8	44.4
Deepseek 7B (Bi et al., 2024)	63.0	15.8	-	48.2	35.2
DeepseekMath 7B (Shao et al., 2024)	88.2	51.7	-	-	-
DeepseekCoder 7B (Guo et al., 2024)	-	-	-	66.1	65.4
LLaMA3 8B (Dubey et al., 2024)	78.3	29.6	31.9	59.8	57.6
Gemma2 9B (Team, 2024)	76.7	44.3	32.8	68.9	74.9

Two-Stage Training Pipeline

Remask SFT: Construct noisy inputs, and teaches the model to detect "incorrect" tokens.

Remask RL: Uses outcome-based RL to optimize full generation trajectories toward higher rewards.

In-Depth Analysis

A Case of Revision during Generation

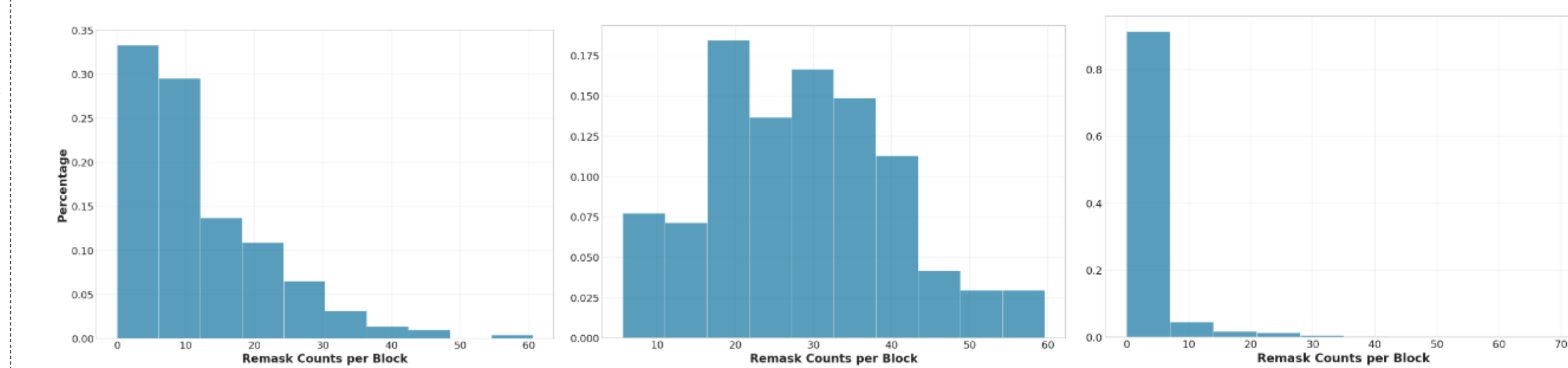


Figure 5: An example of refining text quality with remasking. Question: *What is the major outcome of the reaction between 4,4-dimethylcyclopent-1-enol and bromine?*

➤ *RemeDi performs diverse operations beyond correction, including inserting, deleting, merging, and splitting tokens.*

Remask Frequency across Different Datasets

Code > Math > General Text (per 32-token block)



(a) MATH-500 (11.81 ± 10.23) (b) HumanEval (28.52 ± 12.04) (c) Alpaca-Eval (2.78 ± 5.33)

Figure 4: Distribution of remasking frequencies per block across different evaluation datasets. The numbers in parentheses indicate the mean and standard deviation for each dataset.