UniForm: Segment-Aware GRPO for Joint Punctuation Restoration and Inverse Text Normalization

Junhao Gong[1] **Weiwei Li**[2*] **Rilin Chen**[2*]
[1]Peking University [2]Hunyuan Team, Tencent
gongjunhao.cs@gmail.com

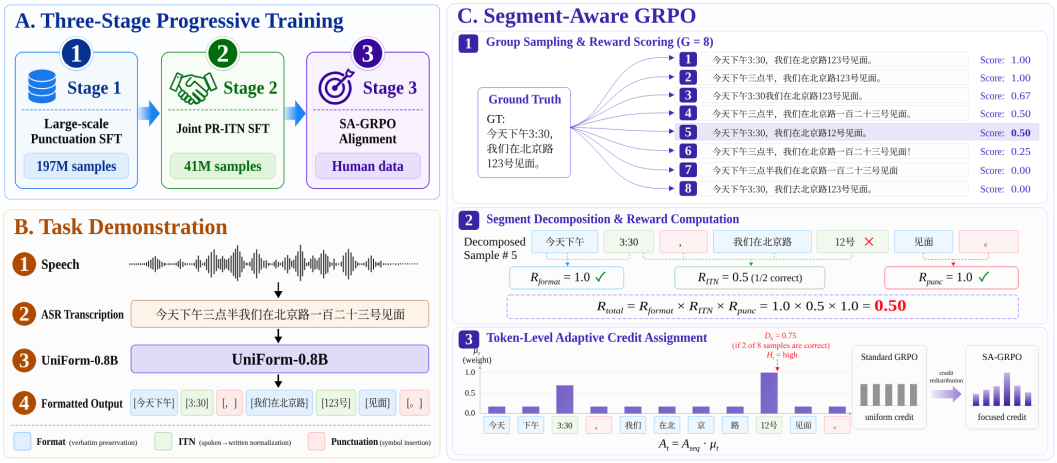

Figure 1: Overview of UniForm. **(A)** Three-stage progressive training. **(B)** Task demonstration. **(C)** Segment-Aware GRPO with token-level adaptive credit assignment.

Abstract

Punctuation Restoration (PR) and Inverse Text Normalization (ITN) are essential ASR post-processing tasks, conventionally tackled by separate models in a sequential pipeline with error propagation and suboptimal latency. UniForm, a unified framework, instead solves both jointly with one compact language model via three progressive stages: large-scale punctuation supervised fine-tuning (SFT), joint SFT of PR and ITN, and reinforcement learning (RL) alignment. For effective RL in this inherently low-entropy setting, Segment-Aware GRPO (SA-GRPO) couples two synergistic components: structured reward decomposition giving dedicated signals to distinct segment types, and token-level adaptive credit assignment redistributing gradients toward difficult tokens based on sampling consensus and model entropy. This eliminates manual sub-task weighting and enables fine-grained optimization for both tasks. Trained on over 238 million multilingual samples, UniForm-0.8B surpasses prior specialized models and larger general-purpose LMs on both tasks across diverse languages, domains, and code-switching scenarios while retaining real-time inference latency; analysis reveals a strong synergistic effect, with joint training mutually benefiting PR and ITN.

*Corresponding Authors.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

1 Introduction

Modern ASR systems have achieved remarkable accuracy in converting speech signals into text sequences [10, 28], yet their raw outputs remain far from readable. The transcriptions typically lack punctuation marks and retain spoken-form representations of numbers, dates, currencies, and other entities, such as rendering "two forty" rather than "2:40." Transforming these raw transcriptions into well-formatted written text is essential for downstream tasks including machine translation, information extraction, and summarization.

This transformation, commonly referred to as *text formatting* (TF), involves multiple sub-tasks. **Punctuation Restoration** (PR) inserts appropriate punctuation marks into unpunctuated text, enabling proper sentence segmentation and improving readability [6, 37]. **Inverse Text Normalization** (ITN), broadly construed, converts spoken-form expressions into their canonical written forms, including numerical expressions, dates, currencies, measurement units, abbreviations, and capitalization recovery [27, 40]. While full-fledged systems also address disfluency removal [16], PR and ITN are the most impactful for readability and are the focus of this work.

Conventionally, PR and ITN have been studied independently: PR as sequence labeling with pretrained encoders [3, 8, 20, 21], and ITN through WFSTs [32, 40] or neural methods [2, 23, 33]. In production systems, these are assembled into modular pipelines [39] with each component developed independently. However, this separation overlooks the strong interdependence between the two tasks: punctuation decisions near entities require compositional understanding that ITN provides, while ITN disambiguation often depends on sentence boundaries established by punctuation. Recent joint approaches [7, 16, 34] confirm the benefit of unified formulation, but still rely on classification-based architectures that cannot flexibly generate arbitrary written forms, or require modifications to the acoustic model itself.

Recent advances in compact language models [36] present an opportunity to approach TF differently: a generative model can produce punctuated and normalized text in one autoregressive pass. However, naively applying small LMs to this joint task faces non-trivial challenges. The low-entropy nature of punctuation decisions, where most inter-token positions require no punctuation and candidate outputs differ by only one or two symbols, makes reward signals sparse and poorly discriminated in standard reinforcement learning. Moreover, the scarcity of high-quality annotated data for certain languages such as Cantonese limits coverage without targeted data synthesis.

In this work, we propose **UniForm** (**Uni**fied **Form**atting), a framework that formulates joint PR and ITN as a single generative task solved by one compact language model with 0.8B parameters. UniForm adopts a three-stage progressive training paradigm: large-scale supervised fine-tuning on punctuation data, joint fine-tuning of PR and ITN, and reinforcement learning on expert-verified data to resolve ambiguities.

To enable effective RL in this low-entropy setting, we introduce Segment-Aware GRPO (SA-GRPO), an extension of Group Relative Policy Optimization [31] that decomposes each output into structurally distinct segment types with dedicated reward signals combined via multiplicative composition, and further introduces a token-level adaptive credit assignment mechanism that dynamically redistributes gradient signal toward difficult tokens.

Experiments on multilingual benchmarks demonstrate that UniForm-0.8B outperforms both task-specific baselines and larger general-purpose LMs while maintaining real-time inference latency. In summary, our contributions are outlined as follows:

- We develop a multi-granularity data construction pipeline producing over 238 million multilingual training samples through LLM-based synthesis and automated quality control.
- We propose a three-stage progressive training paradigm that enables a 0.8B LLM to achieve state-of-the-art performance on both tasks through large-scale SFT, joint PR-ITN SFT, and RL alignment.
- We introduce SA-GRPO, to our knowledge the first application of RL with structured reward decomposition to joint PR and ITN, combining multiplicative rewards with token-level adaptive credit assignment.
- We construct UniFormBench, a multilingual evaluation suite with Main, Dialect, and Foreign tiers.

2 Related Work

2.1 Text Formatting for ASR

Punctuation restoration has progressed from LSTM-based sequence labeling [37] to pre-trained Transformer encoders [1, 3, 6, 20] and multilingual models [9, 15, 42, 43], but these methods cannot transform tokens themselves. Joint truecasing and punctuation prediction has also been explored [26]. ITN has traditionally relied on WFST grammars [32, 40], with neural alternatives via sequence labeling [2, 27, 33], seq2seq [13, 23], streaming on-device approaches [11], LLM-based methods [4, 38], and rule-based libraries [22]. The mutual dependency between PR and ITN has motivated joint approaches through multi-head classification [34], hybrid token-classifier/seq2seq architectures [16], end-to-end speech-to-punctuated-text models [7, 24], spoken-to-written dual transcription [14], and fine-tuned LLMs [43], yet industrial pipelines [10, 39] still treat these as separate modules. Comprehensive readability improvement for ASR output has also been studied [19], and downstream tasks such as machine translation benefit significantly from intermediate punctuation restoration [5]. Our work differs by formulating text formatting as a single generative task solved by one compact LM with RL-based alignment.

2.2 Reinforcement Learning for Language Model Alignment

RLHF [25] has evolved from PPO [30] to DPO [29] and group-relative approaches such as GRPO [31], with DeepSeek-R1 [12] demonstrating RL-incentivized reasoning at scale. In ASR post-processing, LLM-based correction has been explored without explicit RL [35]. Our SA-GRPO adapts GRPO to text formatting outputs through segment-level reward decomposition, enabling fine-grained credit assignment where standard rewards lack discriminability.

3 Data Construction

We formulate text formatting as a sequence-to-sequence generation task. Given an unpunctuated, spoken-form input $X = (x_1, \ldots, x_n)$ produced by an ASR system, the model generates a formatted output $Y = (y_1, \ldots, y_m)$ containing appropriate punctuation marks and written-form entities. After all collection, filtering, and annotation stages, the final training corpus contains approximately 41.2M samples for joint supervised fine-tuning, in addition to 197M samples used for large-scale punctuation SFT in Stage 1. A complete, stage-by-stage account of the pipeline — including the synthesis modules, the filtering and correction rules, the ITN annotation schema, and the stage-wise use of each pool — is given in Appendix A.

3.1 Data Collection and Annotation Pipeline

LLM-based data synthesis. To supplement naturally occurring data, we design a structured synthesis pipeline that generates parallel spoken-form and written-form pairs. The pipeline is organized around seven linguistic modules covering punctuation rules, discourse markers, numerical and temporal expressions, currency and symbols, code-switching, named entities, and domain-specific terminology. Each module operates through two complementary routes. In Route A, a teacher LLM generates diverse scenario descriptions within a given domain, which then serve as prompts for parallel corpus generation. In Route B, a four-level hierarchical entity taxonomy guides the generation of entity-rich contexts. This dual-route architecture ensures that the synthetic data covers both diverse conversational scenarios and dense entity mentions.

Quality filtering and punctuation correction. Since our corpus is aggregated from heterogeneous sources, we apply a unified quality assurance pipeline. After exact-match deduplication (retaining $\sim$45.8% of raw samples), a teacher LLM evaluates each sample against comprehensive punctuation rules under the constraint that only punctuation symbols may be modified. Compliant samples ($\sim$87.5%) are retained directly; non-compliant samples receive corrected versions under the same constraints, recovering otherwise valuable training data. After this pipeline, the supervised fine-tuning corpus contains 41.2M samples.

ITN span annotation. Stage 2 requires ITN span annotations on the training corpus. We employ Qwen3.5-397B-A17B as a teacher model to identify ITN-eligible spans. Each annotated span

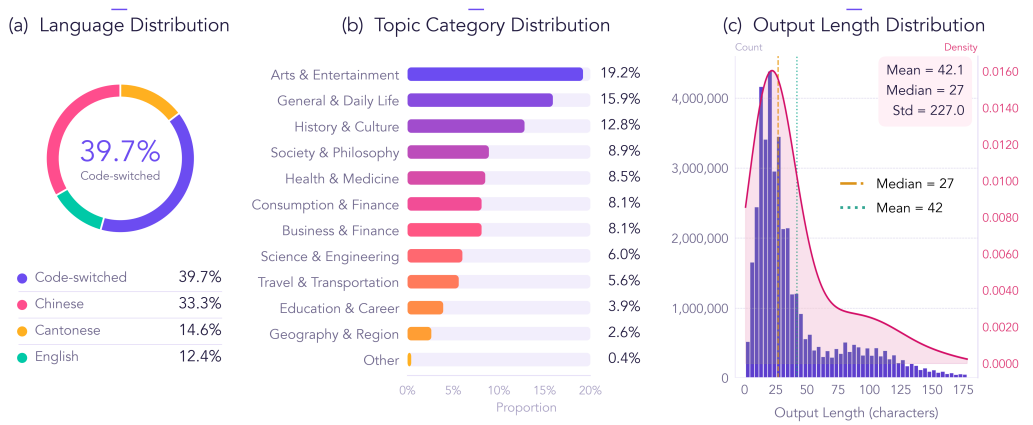

Figure 2: Distribution analysis of the final training corpus. **(a)** Language composition spanning four categories. **(b)** Topic category distribution across 13 knowledge domains. **(c)** Output length distribution (in characters).

records spoken and written forms, a candidate set for ambiguous cases, and an entity type label. After processing the full corpus, approximately 42.6% of samples contain non-trivial ITN content, providing the supervision signal for joint punctuation-ITN training.

Training–test deduplication. To prevent data leakage, we verify that UniFormBench does not overlap with the training corpus. The benchmark is built from disjoint source data, and residual overlap is identified by canonicalized exact matching, embedding-based retrieval with LLM-assisted verification, and ITN entity-instance retrieval. Because one benchmark instance typically matches multiple near-duplicate training samples, we remove the affected benchmark instances and leave the training corpus intact. Model details, thresholds, and decision criteria are given in Appendix D.

3.2 Final Data Composition

The resulting corpus comprises three components: (1) ∼7.2M real ASR transcriptions covering conversational, command, and dictation scenarios; (2) ∼28M samples combining entity-enriched text and quality-controlled corpora with verified punctuation; and (3) ∼6M synthetic Cantonese samples from our LLM synthesis pipeline. The distribution across languages, domains, and text lengths is shown in Figure 2.

4 Training Method

UniForm-0.8B is built upon Qwen3.5-0.8B [36]. The model takes an instruction prompt followed by the ASR text as input and generates the formatted output autoregressively. Training proceeds through three stages of increasing data quality and decreasing data volume, forming a natural curriculum from broad coverage to precise alignment, as illustrated in Figure 1(A).

4.1 Stage 1: Large-Scale Supervised Fine-Tuning on Punctuation

The first stage fine-tunes the pre-trained Qwen3.5-0.8B checkpoint on 197M samples focused exclusively on punctuation restoration. The sheer scale enables the model to internalize broad punctuation patterns across diverse domains and languages while preserving general language understanding.

4.2 Stage 2: Joint Punctuation-ITN Supervised Fine-Tuning

The second stage introduces ITN capability by fine-tuning the Stage 1 checkpoint on the 41.2M samples annotated with ITN span masks as described in Section 3. The model learns to simultaneously insert punctuation and transform spoken-form spans into their canonical written forms, while the

samples without ITN content provide continued punctuation supervision to prevent catastrophic forgetting. This joint formulation exploits the mutual dependency between the two tasks: ITN annotation implicitly encodes entity boundaries that benefit punctuation decisions, while sentence structure from punctuation helps resolve ambiguous ITN cases.

4.3 Stage 3: Segment-Aware GRPO

The third stage applies reinforcement learning to refine decisions on ambiguous cases. We select a subset of ITN-entity-dense samples that have been verified by domain experts, ensuring both high entity coverage and annotation quality. We introduce SA-GRPO, whose reward is tailored to text-formatting outputs.

GRPO backbone. GRPO [31] drops the critic: it samples G responses per prompt and normalizes their rewards into a group-relative advantage A_{seq} broadcast to every token. The policy is updated with the standard PPO-style clipped surrogate together with a KL penalty toward a frozen reference model. SA-GRPO keeps this objective and only replaces A_{seq} by the token-level A_t defined below. The full objective, notation, and zero-sum argument are given in Appendix B.

4.3.1 Segment Decomposition and Reward Design

A formatted output can be decomposed into three structurally distinct segment types, each requiring a different correctness criterion. Given a ground-truth output S and a model prediction S', we first strip all punctuation from both sequences to obtain pure-text representations, then identify three segment types: *invariant content* S_{format} consisting of characters that must remain unchanged between input and output, where any modification constitutes hallucination; *ITN spans* S_{ITN} consisting of subsequences that should be transformed according to normalization rules, where multiple valid written forms may exist; and *punctuation marks* S_{punc} consisting of symbols inserted by the model, which are evaluated separately through positional alignment.

We define three corresponding reward components:

$$R_{\text{format}} = \mathbb{1}[S'_{\text{format}} = S_{\text{format}}] \tag{1}$$

$$R_{\text{ITN}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}[S'_{\text{ITN},i} \in \mathcal{C}_i] \tag{2}$$

$$R_{\text{punc}} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{3}$$

where $R_{\text{format}} = 1$ iff all invariant content is preserved; R_{ITN} evaluates each of the n ITN spans against its set of valid candidates $\mathcal{C}_i$, averaging the per-span accuracy; and R_{punc} is the F1 score over positionally aligned punctuation symbols, with Precision $= |P_{\text{pred}} \cap P_{\text{gt}}|/|P_{\text{pred}}|$ and Recall $= |P_{\text{pred}} \cap P_{\text{gt}}|/|P_{\text{gt}}|$, where P_{gt} and P_{pred} denote the multisets of punctuation symbols at aligned positions. The positional mapping is established through sequence alignment on the pure-text representations.

4.3.2 Multiplicative Reward Composition

Given that all three components take values in $[0, 1]$, we compose them via multiplication:

$$R_{\text{total}} = R_{\text{format}} \cdot R_{\text{ITN}} \cdot R_{\text{punc}} \tag{4}$$

This design is motivated by two considerations. First, format preservation acts as a hard gate: since R_{format} is binary, any modification to invariant content immediately zeros out the total reward regardless of punctuation and ITN quality. Second, multiplication amplifies discriminability between near-identical candidates: since both R_{ITN} and R_{punc} concentrate near 1.0 in this low-entropy task, additive composition clusters total rewards tightly with weak gradient signals, whereas multiplicative composition creates proportionally larger gaps that enable more informative policy updates. We compare against additive and format-gated alternatives in Section 5.

4.3.3 Token-Level Adaptive Credit Assignment

While the multiplicative reward provides better sequence-level discrimination, the resulting advantage A_{seq} is still a scalar that gets uniformly broadcast to all response tokens. For example, a model

that misplaces only one comma receives the same gradient signal at all positions, including those it handles correctly.

To address this, we introduce a token-level advantage redistribution mechanism that dynamically concentrates gradient signal on tokens where the model struggles most, without requiring any manually specified sub-task weights. The mechanism operates on two complementary signals computed entirely from the mini-batch statistics.

Per-position span instantiation. For credit assignment, we instantiate the tripartite segments as an ordered sequence of spans $\{s_1, s_2, \ldots, s_K\}$ along the output. The only refinement is that punctuation is split into independent per-position spans rather than a single global span, since punctuation decisions at different positions are largely independent, allowing each location to receive its own difficulty estimate.

Empirical difficulty. For each span s_k, we compute its empirical difficulty D_k as the fraction of the G group samples that produce an incorrect output for that span:

$$D_k = 1 - \frac{N_{\text{correct}}(s_k)}{G} \tag{5}$$

where $N_{\text{correct}}(s_k)$ counts how many of the G sampled outputs satisfy the correctness criterion for span s_k, using the per-segment criteria above.

Intrinsic uncertainty. As a second signal source, we use the model's per-token predictive entropy:

$$H_t = H\big(p_\theta(\cdot \mid x_{<t})\big) \tag{6}$$

Entropy complements empirical difficulty: an empirically easy span may still contain high-entropy, fragile tokens.

Token-to-span alignment. Since the tokenizer's segmentation does not align with span boundaries, a single token may cover characters belonging to multiple spans. We establish a precise mapping from each token t to its covered spans via tokenizer offset mappings, which yield the character range $[c_{\text{start}}^{(t)}, c_{\text{end}}^{(t)})$ for each token. For the response text, each character position is annotated with its span identity through the sequence alignment computed during reward evaluation. The fractional attribution of token t to span k is then:

$$w_k^{(t)} = \frac{|\{c \in [c_{\text{start}}^{(t)}, c_{\text{end}}^{(t)}) : \text{span}(c) = k\}|}{c_{\text{end}}^{(t)} - c_{\text{start}}^{(t)}} \tag{7}$$

This character-level weighting naturally handles cross-span tokens without information loss.

Per-token importance and zero-sum redistribution. The two signals are fused into a per-token importance score:

$$\lambda_t = \alpha \cdot H_t + (1 - \alpha) \cdot \sum_k w_k^{(t)} \cdot D_k \tag{8}$$

where α controls the mixing proportion between entropy and empirical difficulty; H_t is the raw token entropy in nats and $D_k \in [0, 1]$. The token-level advantage is then obtained by modulating the sequence-level advantage with a zero-sum multiplier:

$$\mu_t = \begin{cases} 1 + \beta \cdot (\lambda_t - \bar{\lambda}), & \text{if } A_{\text{seq}} > 0 \\ 1 - \beta \cdot (\lambda_t - \bar{\lambda}), & \text{if } A_{\text{seq}} < 0 \\ 1, & \text{if } A_{\text{seq}} = 0 \end{cases} \tag{9}$$

where $\bar{\lambda} = \frac{1}{T}\sum_{t=1}^{T} \lambda_t$ is the mean importance over all T response tokens and β is a scaling factor. The final token-level advantage is $A_t = A_{\text{seq}} \cdot \mu_t$. Thus above-average-importance tokens are reinforced more when $A_{\text{seq}} > 0$ and penalized more when $A_{\text{seq}} < 0$. The zero-sum property $\sum_t (\mu_t - 1) = 0$ guarantees that the total gradient magnitude is preserved, maintaining training stability.

5 Experiments

5.1 Experimental Setup

5.1.1 Benchmarks

We construct **UniFormBench**, a multilingual evaluation suite organized into three tiers. **UniFormBench-Main** (62,559 samples) covers four language conditions (pure Chinese, pure English, Chinese-English code-switching, and Cantonese) drawn from real ASR transcriptions with LLM-assisted annotation and expert verification. **UniFormBench-Dialect** (32,762 samples) contains 31 distinct regional varieties, evaluating robustness to dialectal variation. **UniFormBench-Foreign** (31,741 samples) covers 8 world languages beyond Chinese and English, assessing cross-lingual generalization. Each tier provides punctuation-only, ITN-only, and joint evaluation splits. Appendix C lists the languages and dialects in each tier.

5.1.2 Baselines

We compare UniForm-0.8B against three categories of baselines. **(i) Punctuation-Only models:** CT-Transformer [3], the content-aware punctuation model from FunASR [10]; and RoBERTa-Punc, a RoBERTa-base model fine-tuned for punctuation restoration via token classification. **(ii) ITN-Only models:** NVIDIA NeMo ITN [40], a WFST-based inverse text normalization toolkit; FunASR ITN [10], a rule-based Chinese/English ITN system; Thutmose Tagger [2], a single-pass neural ITN tagger; and DuplexTagger + DuplexDecoder [18], a unified transformer-based duplex normalization framework. **(iii) Zero-Shot LLMs:** Qwen3.5-0.8B, Qwen3.5-35B-A3B, Qwen3.6-35B-A3B, and Qwen3-Omni-30B-A3B, evaluated with the same instruction prompt and output format constraints as UniForm.

Punctuation-only and ITN-only baselines are evaluated solely on their respective metrics, whereas UniForm-0.8B is evaluated on both tasks simultaneously, reflecting its unified design.

5.1.3 Metrics

We evaluate punctuation restoration using Precision (P), Recall (R), and F1 score at the punctuation-mark level. For ITN, we report Word Error Rate (WER) and Sentence Error Rate (SER) on ITN-specific spans, measuring character-level edit distance against reference normalizations. When evaluating with multiple valid ITN references (e.g., "27" vs. "Twenty-seven"), we adopt a best-match protocol that selects the reference minimizing edit distance.

5.1.4 Implementation Details

All training is conducted on NVIDIA H20 GPUs. Stage 1 fine-tunes Qwen3.5-0.8B on large-scale punctuation data with a learning rate of 1×10^{-5} using cosine decay and a batch size of 256 for a single epoch; this stage uses 32 GPUs and completes in approximately 38 hours. Stage 2 applies the same hyperparameters to the 41.2M ITN-annotated samples for joint training on 32 GPUs, requiring approximately 14 hours. For Stage 3, we sample $G = 8$ candidate outputs per input, compute group-relative advantages based on R_{total} with $\alpha = \beta = 0.5$, and train for 128 optimization steps on 32 GPUs, completing in approximately 8 hours.

5.2 Main Results and Ablation Studies

Table 1 presents the unified results. To assess stability, we repeat training with three random seeds and report mean $\pm$ 1-sigma sample standard deviations ($n = 3$) in Appendix F (Table 4); the main text reports seed 42.

5.2.1 Overall Performance

On the target domains of our training objective, namely Mandarin, English, Cantonese, and other Chinese dialects, UniForm-0.8B achieves the best overall punctuation F1 and competitive ITN performance across all three benchmarks. Despite having only 0.8B parameters, it surpasses dedicated classification-based punctuation models on these domains while simultaneously providing ITN capability that substantially outperforms both WFST-based approaches (NeMo ITN, FunASR ITN)

Table 1: Comprehensive results on UniFormBench. Each sub-benchmark reports five metrics: punctuation P↑/R↑/F1↑ and ITN WER↓/SER↓. "—" indicates the metric is not applicable. Per column, the best result is in **bold** and the second best is underlined.

Model / Configuration	UniFormBench-Main					UniFormBench-Dialect					UniFormBench-Foreign				
	P	R	F1	WER	SER	P	R	F1	WER	SER	P	R	F1	WER	SER
Punctuation-only models (evaluated on punc_only split)															
CT-Transformer	0.605	0.580	0.592	—	—	0.633	0.595	0.614	—	—	0.000	0.000	0.000	—	—
RoBERTa-Punc	0.658	0.610	0.633	—	—	0.674	0.639	0.656	—	—	0.000	0.000	0.000	—	—
ITN-only models (evaluated on itn_only split)															
NeMo ITN	—	—	—	15.18	28.51	—	—	—	12.74	43.76	—	—	—	3.92	47.68
FunASR ITN	—	—	—	0.84	7.75	—	—	—	3.19	35.60	—	—	—	**3.45**	**47.40**
Thutmose Tagger	—	—	—	1.24	8.71	—	—	—	3.35	35.76	—	—	—	6.73	53.27
DuplexTagger + DuplexDecoder	—	—	—	9.20	19.81	—	—	—	3.59	36.23	—	—	—	7.10	55.33
Zero-Shot LLMs (evaluated on combined split)															
Qwen3.5-0.8B	0.665	0.699	0.681	5.56	38.13	0.607	0.635	0.621	7.23	39.19	0.319	0.329	0.324	8.65	50.88
Qwen3.5-35B-A3B	0.722	0.742	0.732	6.00	37.39	0.651	0.671	0.661	6.40	39.96	0.342	0.345	0.343	11.55	48.09
Qwen3.6-35B-A3B	0.728	0.747	0.737	5.66	36.23	0.642	0.661	0.652	5.73	40.78	0.339	0.345	0.342	9.21	56.17
Qwen3-Omni-30B-A3B	0.699	0.725	0.712	5.93	35.34	0.631	0.658	0.644	7.59	37.12	0.414	0.422	0.418	9.61	53.48
Ablation A: Training Stage (evaluated on combined split)															
Stage 1 only (SFT-Punc)	0.813	0.669	0.734	1.57	9.55	0.749	0.659	0.702	3.28	36.20	0.421	0.567	0.483	8.32	67.80
Stage 1+2 (+ ITN SFT)	0.820	0.777	0.798	0.72	4.22	0.739	0.703	0.720	1.23	17.39	0.418	0.615	0.498	15.57	71.33
Ablation B: Reward Function Design															
+ DPO	0.023	0.607	0.045	31.24	33.38	0.011	0.495	0.021	46.39	63.83	0.001	0.036	0.002	38.30	96.12
+ SA-GRPO, Additive (w/ KL)	0.837	0.754	0.793	0.54	3.54	0.762	0.665	0.710	1.12	17.35	0.550	0.529	0.539	12.24	71.10
+ SA-GRPO, Additive (no KL)	0.837	0.755	0.794	0.39	**3.42**	0.768	0.669	0.715	1.08	**16.90**	**0.569**	0.562	0.566	7.47	63.02
+ SA-GRPO, Format-gated (w/ KL)	0.837	0.733	0.782	0.40	3.56	0.767	0.649	0.703	1.08	17.19	0.535	0.512	0.523	7.10	63.15
+ SA-GRPO, Format-gated (no KL)	0.832	0.756	0.792	0.64	3.68	0.755	0.656	0.702	1.27	17.22	0.509	0.480	0.494	7.59	62.57
+ SA-GRPO, Multiplicative (w/ KL)	0.812	**0.790**	**0.801**	**0.38**	3.54	0.756	0.689	0.721	**1.07**	17.13	0.541	**0.625**	**0.580**	7.02	61.52
Ablation C: Data Composition (Stage 2 setup)															
Chinese conv. + 1st proper noun	0.718	0.684	0.701	6.21	37.42	0.674	0.642	0.658	7.15	38.86	0.350	0.372	0.361	10.41	54.37
Chinese conv. + English	0.778	0.720	0.748	5.87	35.63	0.720	0.683	0.701	6.52	37.19	0.375	0.389	0.382	9.83	52.14
Chinese conv. + dialect	0.772	0.689	0.728	6.35	38.11	0.727	0.661	0.693	7.48	39.52	0.334	0.359	0.346	11.26	55.83
Chinese conv. + 2nd proper noun	0.799	0.754	0.776	5.72	34.89	0.750	0.714	0.732	6.38	36.47	0.396	0.410	0.403	9.15	50.26
All components (= Stage 2 PR-only)	0.797	0.766	0.781	5.54	34.52	0.751	**0.722**	**0.736**	6.17	35.81	0.405	0.417	0.411	8.94	49.63
Ablation D: Token-Level Credit Assignment															
Vanilla GRPO, scalar (w/ KL)	0.832	0.765	0.797	0.42	3.64	0.757	0.671	0.712	1.48	17.79	0.554	0.569	0.561	9.15	68.53
Vanilla GRPO, scalar (no KL)	**0.839**	0.750	0.792	0.39	3.49	**0.769**	0.658	0.709	1.08	17.28	0.548	0.520	0.533	8.10	62.36
Ablation E: Entropy Normalization															
SA-GRPO, normalized entropy	0.828	0.772	0.799	0.61	3.68	0.757	0.679	0.716	1.09	17.53	0.512	0.619	0.560	8.08	68.62
UniForm-0.8B — Full system (Stage 1 + Stage 2 + SA-GRPO)															
UniForm-0.8B (ours)	0.812	**0.790**	**0.801**	**0.38**	3.54	0.756	0.689	0.721	**1.07**	17.13	0.541	**0.625**	**0.580**	7.02	61.52

and specialized neural ITN systems (Thutmose Tagger, DuplexTagger + DuplexDecoder), serving as a drop-in replacement for the entire traditional multi-model pipeline. On the Zero-Shot Foreign subset, these ITN baselines attain slightly lower error rates than ours because they tend to abstain when the input exceeds a length limit or the language cannot be identified, whereas UniForm produces an output for every input; their lower rates thus reflect selective evaluation on a subset of inputs rather than stronger ITN capability, and UniForm remains substantially stronger on the target domains.

Within the training domains, UniForm-0.8B is robust to dialectal variation: on the dialect benchmark covering 31 distinct regional varieties, it maintains performance comparable to its main benchmark results, indicating resilience to the lexical and syntactic variation inherent in dialectal speech. Beyond the training domains, it generalizes to UniFormBench-Foreign, achieving meaningful F1 on all 8 foreign languages, whereas traditional punctuation models, designed exclusively for Chinese, cannot produce predictions on this subset. This suggests that the generative formulation enables better cross-lingual transfer. Among Zero-Shot LLMs, all tend to reformulate input rather than preserving it verbatim, resulting in substantially higher WER and SER that confirm task-specific training with content-preservation constraints is essential.

5.2.2 Ablation A: Training Stage Analysis

The progressive stages contribute cumulatively. Adding ITN supervision in Stage 2 substantially improves both punctuation F1 and ITN accuracy over Stage 1, with WER decreasing by more than half, confirming that joint training provides mutual benefit rather than capacity trade-off. Stage 3 with SA-GRPO further elevates F1 to the highest among all configurations with the best precision-recall balance. Notably, only a small number of optimization steps suffice to achieve consistent gains, underscoring the effectiveness of our reward decomposition in providing informative learning signals even in this low-entropy setting.

These gains are modest in punctuation F1 on Main because SFT on 238M samples already approaches the ceiling of this reference-based metric, and punctuation is inherently one-to-many: multiple

placements are linguistically valid while the benchmark provides a single reference. The RL data and the test set also come from different sources with different annotator groups, which limits in-distribution gains. The RL stage instead targets ITN ambiguity resolution and OOD robustness: on Main, ITN WER drops from 0.72% to 0.38% (47% relative) and SER from 4.22% to 3.54%, while on Foreign, F1 rises from 0.498 to 0.580 and WER falls from 15.57% to 7.02% (55% relative).

5.2.3 Ablation B: Reward Function Design

The multiplicative reward composition achieves the best overall F1. Additive and format-gated variants attain higher precision but lower recall, because additive rewards allow compensating for missing punctuation with high ITN accuracy, whereas multiplicative composition penalizes any deficiency proportionally, encouraging balanced performance across all segment types. The multiplicative formulation also achieves the best WER among all reward variants, confirming that its balanced optimization transfers to ITN quality as well.

Single-component rewards further confirm the necessity of the multiplicative composition (Appendix E, Table 3): training with R_{punc} alone raises Foreign F1 to 0.5924 but inflates WER to 15.60%, more than double the full model's 7.02%, while R_{ITN} alone marginally exceeds the full model's Main F1 (0.8012 vs. 0.8010) yet degrades WER from 0.38% to 0.52%. Such metric-specific volatility disappears under the combined reward, which attains the lowest WER on all three subsets.

DPO completely fails with F1 collapsing near zero, because chosen and rejected sequences differ by at most one or two punctuation marks, making their log-probabilities indistinguishable. SA-GRPO avoids this by computing explicit scalar rewards and normalizing advantages within each group. KL regularization provides a modest but consistent benefit by preventing over-optimization on the small Stage 3 dataset.

5.2.4 Ablation C: Data Composition

Data quality dominates over quantity in the Stage-2 PR-only ablation. The 1st proper noun batch yields the lowest global F1 with poor Cantonese performance, suggesting quality issues in this source. Dialect data severely degrades Cantonese and English performance, likely because expert inspection revealed that this batch lacks adequate punctuation annotations and contains fewer punctuation marks overall, biasing the model toward under-punctuation. English-only data also underperforms the 2nd proper noun batch despite similar volume. In contrast, the 2nd proper noun batch provides the strongest single-component gain, approaching full mixture performance and confirming that expert-checked entity-rich data outweighs larger volumes of lower-quality sources. Removing the 1st proper noun batch and training on the merged remaining data achieves the best Stage-2 punctuation performance.

5.2.5 Ablation D: Token-Level Credit Assignment

Table 1 isolates the contribution of token-level credit assignment: with the model, data, multiplicative reward, and KL regularization held fixed, replacing the uniform sequence-level advantage with our redistribution consistently improves F1 and lowers WER and SER on all three subsets. The improvements therefore do not arise solely from the multiplicative reward composition. They follow from the low-entropy setting that motivates SA-GRPO: when most token positions are trivially correct, a sequence-level scalar dilutes the learning signal over nearly deterministic tokens, whereas token-level redistribution concentrates it on the few uncertain positions that determine the output. The gains are largest on Dialect and the Zero-Shot Foreign OOD tier, indicating that fine-grained credit assignment becomes increasingly valuable under linguistic and distributional shift, consistent with the OOD robustness that the RL stage targets.

5.2.6 Ablation E: Entropy Normalization

We also evaluate the normalized-entropy variant $\widehat{H}_t = H_t / \log |V|$, which is intended to place entropy on the same $[0, 1]$ scale as empirical difficulty. As Table 1 shows, normalization yields no overall gain: F1 is nearly unchanged on Main and Dialect, while the Zero-Shot Foreign subset degrades in both F1 and WER. This is a scale effect rather than training instability. Since $\log |V| = 12.4215$ upper-bounds the entropy of a uniform vocabulary distribution while the SFT policy is near-deterministic here ($H_t \approx 0.06$ nats), $\widehat{H}_t$ collapses to ≈ 0.005 whereas $D_t \in [0, 1]$ is untouched; because μ_t depends

on mean-subtracted λ_t, rescaling only the entropy summand implicitly shifts the mixing weight α from 0.5 to ≈ 0.04. The entropy share of λ_t's variance accordingly drops from 0.3238 to 0.0405 and $\mathrm{corr}(H_t, \mu_t)$ from $|0.63|$ to $|0.07|$, degenerating SA-GRPO into difficulty-only reweighting. Appendix G gives the full diagnostics, training curves, and a closed-form prediction of this collapse. We therefore retain the unnormalized entropy used throughout the paper.

5.3 Inference Efficiency

Figure 3 visualizes the throughput–quality trade-off. All measurements use 8 NVIDIA H20 GPUs at batch concurrency 256: UniForm-0.8B and Qwen3.5-4B share the same vLLM serving configuration, CT-Transformer uses the environment recommended by FunASR, and the RoBERTa-Punc + WFST ITN pipeline is served with HuggingFace Transformers because this RoBERTa-Punc implementation cannot be served by our vLLM version due to architecture and framework constraints. The 2.6× gain is thus a system-level comparison against this complete pipeline rather than an isolated model-to-model benchmark. UniForm-0.8B outperforms RoBERTa-Punc in both speed and quality despite RoBERTa handling punctuation only—the advantage stems from performing both tasks in a single autoregressive pass rather than sequential pipeline execution. Compared to the zero-shot Qwen3.5-4B, UniForm-0.8B is an order of magnitude faster with substantially higher F1, confirming that task-specific compact models paired with efficient inference [17, 41] are far more practical for real-time deployment than general-purpose LLMs. With FP8 quantization, UniForm-0.8B can be deployed on a single consumer-grade GPU while sustaining high-concurrency serving, meeting the throughput demands of production ASR systems.

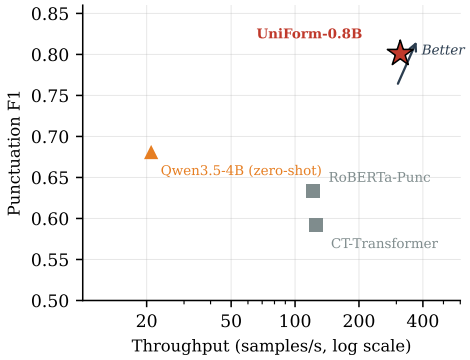

Figure 3: Speed–quality trade-off. In a system-level comparison, UniForm-0.8B achieves the highest F1 while being 2.6× faster than the full RoBERTa-Punc + WFST ITN pipeline and 14.9× faster than Qwen3.5-4B.

6 Limitations

First, despite cross-lingual generalization on UniFormBench-Foreign, the training data is predominantly Chinese and English; multilingual training with native-speaker annotations would likely improve performance on low-resource languages. Second, we cover only PR and ITN, whereas a complete ASR post-processing system should also address disfluency removal, error correction, and speaker attribution—integrating these into the unified framework remains future work. Third, the model is text-only; acoustic features via cross-modal attention could resolve ambiguous cases where prosodic cues are needed for punctuation placement or ITN disambiguation. In addition, two prior joint PR–ITN systems could not be evaluated under our protocol: Four-in-One [34] has no public weights or training code, and Universal-2-TF [16] is commercial-API-only.

7 Conclusion

We presented UniForm, a framework that formulates joint punctuation restoration and inverse text normalization as one generative task, instantiated with a 0.8B-parameter language model. It combines multi-granularity data construction, three-stage progressive training, and SA-GRPO with multiplicative rewards and token-level credit assignment, achieving state-of-the-art performance on multilingual benchmarks while maintaining real-time inference latency. Analyses further show that joint training benefits both tasks, multiplicative reward composition is essential in low-entropy RL, and data quality dominates over quantity. Future work will integrate acoustic features, extend the framework to broader post-processing tasks, and expand language coverage.

Acknowledgments and Disclosure of Funding

Funding in direct support of this work: computational resources and data support from Peking University and Tencent.

Additional revenues related to this work: travel support from the NeurIPS Foundation. The authors declare no other competing interests.

References

[1] Tanvirul Alam, Akib Khan, and Firoj Alam. Punctuation restoration using transformer models for high-and low-resource languages. In *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*, pages 132–142, 2020.

[2] Alexandra Antonova, Evelina Bakhturina, and Boris Ginsburg. Thutmose tagger: Single-pass neural model for inverse text normalization. *arXiv preprint arXiv:2208.00064*, 2022.

[3] Qian Chen, Mengzhe Chen, Bo Li, and Wen Wang. Controllable time-delay transformer for real-time punctuation prediction and disfluency detection. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8069–8073. IEEE, 2020.

[4] HyunJung Choi, Muyeol Choi, Yohan Lim, Minkyu Lee, Seon Hui Kim, Seung Yun, Donghyun Kim, and Sang-Hun Kim. Spoken-to-written text conversion with large language model. In *INTERSPEECH*, 2024.

[5] Tangsang Chongbang, Pranesh Pyara Shrestha, Amrit Sarki, and Anku Jaiswal. Mitigating structural noise in low-resource s2tt: An optimized cascaded nepali-english pipeline with punctuation restoration. *arXiv preprint arXiv:2602.21647*, 2026.

[6] Maury Courtland, Adam Faulkner, and Gayle McElvain. Efficient automatic punctuation restoration using bidirectional transformers with robust inference. In *Proceedings of the 17th International Conference on Spoken Language Translation*, pages 272–279, 2020.

[7] Can Cui, Imran Sheikh, Mostafa Sadeghi, and Emmanuel Vincent. End-to-end joint punctuated and normalized asr with a limited amount of punctuated training data. In *2025 33rd European Signal Processing Conference (EUSIPCO)*, pages 356–360. IEEE, 2025.

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)*, pages 4171–4186, 2019.

[9] Oliver Guhr1 Anne-Kathrin Schumann Frank and Bahrmann1 Hans-Joachim Böhme. Fullstop: Multilingual deep models for punctuation prediction. 2021.

[10] Zhifu Gao, Zerui Li, Jiaming Wang, Haoneng Luo, Xian Shi, Mengzhe Chen, Yabin Li, Lingyun Zuo, Zhihao Du, Zhangyu Xiao, et al. Funasr: A fundamental end-to-end speech recognition toolkit. *arXiv preprint arXiv:2305.11013*, 2023.

[11] Yashesh Gaur, Nick Kibre, Jian Xue, Kangyuan Shu, Yuhui Wang, Issac Alphanso, Jinyu Li, and Yifan Gong. Streaming, fast and accurate on-device inverse text normalization for automatic speech recognition. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 237–244. IEEE, 2023.

[12] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

[13] Luong Ho, Khanh Le, Vinh Pham, Bao Nguyen, Tan Tran, and Duc Chau. Dynamic context-aware streaming pretrained language model for inverse text normalization. *arXiv preprint arXiv:2505.24229*, 2025.

[14] Mana Ihori, Hiroshi Sato, Tomohiro Tanaka, Ryo Masumura, Saki Mizuno, and Nobukatsu Hojo. Transcribing speech as spoken and written dual text using an autoregressive model. In *Interspeech*, pages 461–465, 2023.

[15] Mohammad Javad Ranjbar Kalahroodi, Heshaam Faili, and Azadeh Shakery. Persianpunc: A large-scale dataset and bert-based approach for persian punctuation restoration. In *The Proceedings of the First Workshop on NLP and LLMs for the Iranian Language Family*, pages 105–113, 2026.

[16] Yash Khare, Taufiquzzaman Peyash, Andrea Vanzo, and Takuya Yoshioka. Universal-2-tf: Robust all-neural text formatting for asr. *arXiv preprint arXiv:2501.05948*, 2025.

[17] Woosuk Kwon, Zhuohan Li, Siyuan Zhuang, Ying Sheng, Lianmin Zheng, Cody Hao Yu, Joseph Gonzalez, Hao Zhang, and Ion Stoica. Efficient memory management for large language model serving with pagedattention. In *Proceedings of the 29th symposium on operating systems principles*, pages 611–626, 2023.

[18] Tuan Manh Lai, Yang Zhang, Evelina Bakhturina, Boris Ginsburg, and Heng Ji. A unified transformer-based framework for duplex text normalization. *arXiv preprint arXiv:2108.09889*, 2021.

[19] Junwei Liao, Sefik Eskimez, Liyang Lu, Yu Shi, Ming Gong, Linjun Shou, Hong Qu, and Michael Zeng. Improving readability for automatic speech recognition transcription. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 22(5):1–23, 2023.

[20] Xing Yi Liu and Homayoon Beigi. Efficient ensemble for multimodal punctuation restoration using time-delay neural network. In *2024 18th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, pages 1–6. IEEE, 2024.

[21] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*, 2019.

[22] Hung Vu Nguyen, Loan Do, Thanh Ngoc Nguyen, Ushik Shrestha Khwakhali, Thanh Pham, Vinh Do, Charlotte Nguyen, and Hien Nguyen. Vietnormalizer: An open-source, dependency-free python library for vietnamese text normalization in tts and nlp applications. *arXiv preprint arXiv:2603.04145*, 2026.

[23] Thai-Binh Nguyen, Quang Minh Nguyen, Quoc Truong Do, Chi Mai Luong, Alexander Waibel, et al. Adapitn: A fast, reliable, and dynamic adaptive inverse text normalization. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023.

[24] Jumon Nozaki, Tatsuya Kawahara, Kenkichi Ishizuka, and Taiichi Hashimoto. End-to-end speech-to-punctuated-text recognition. *arXiv preprint arXiv:2207.03169*, 2022.

[25] Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744, 2022.

[26] Raghavendra Pappagari, Piotr Żelasko, Agnieszka Mikołajczyk, Piotr Pęzik, and Najim Dehak. Joint prediction of truecasing and punctuation for conversational speech in low-resource scenarios. In *2021 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, pages 1185–1191. IEEE, 2021.

[27] Ernest Pusateri, Bharat Ram Ambati, Elizabeth Brooks, Ondrej Platek, Donald McAllaster, and Venki Nagesha. A mostly data-driven approach to inverse text normalization. In *INTERSPEECH*, pages 2784–2788. Stockholm, 2017.

[28] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pages 28492–28518. PMLR, 2023.

[29] Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

[30] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017.

[31] Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

[32] Richard Sproat, Alan W Black, Stanley Chen, Shankar Kumar, Mari Ostendorf, and Christina D Richards. Normalization of non-standard words. *Computer speech & language*, 15(3):287–333, 2001.

[33] Monica Sunkara, Chaitanya Shivade, Sravan Bodapati, and Katrin Kirchhoff. Neural inverse text normalization. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 7573–7577. IEEE, 2021.

[34] Sharman Tan, Piyush Behre, Nick Kibre, Issac Alphonso, and Shuangyu Chang. Four-in-one: a joint approach to inverse text normalization, punctuation, capitalization, and disfluency for automatic speech recognition. In *2022 IEEE Spoken Language Technology Workshop (SLT)*, pages 677–684. IEEE, 2023.

[35] Zhiyuan Tang, Dong Wang, Zhikai Zhou, Yong Liu, Shen Huang, and Shidong Shang. Chain of correction for full-text speech recognition with large language models. In *ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 19227–19231. IEEE, 2026.

[36] Qwen Team. Qwen3.5: Accelerating productivity with native multimodal agents, February 2026. URL `https://qwen.ai/blog?id=qwen3.5`.

[37] Ottokar Tilk and Tanel Alumäe. Bidirectional recurrent neural network with attention mechanism for punctuation restoration. In *Interspeech*, volume 3, page 9, 2016.

[38] Michel Wong, Ali Alshehri, Sophia Kao, and Haotian He. Polynorm: Few-shot llm-based text normalization for text-to-speech. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing: Industry Track*, pages 77–85, 2025.

[39] Kaituo Xu, Yan Jia, Kai Huang, Junjie Chen, Wenpeng Li, Kun Liu, Feng-Long Xie, Xu Tang, and Yao Hu. Fireredasr2s: A state-of-the-art industrial-grade all-in-one automatic speech recognition system. *arXiv preprint arXiv:2603.10420*, 2026.

[40] Yang Zhang, Evelina Bakhturina, and Boris Ginsburg. NeMo (Inverse) Text Normalization: From Development to Production. In *Proc. Interspeech 2021*, pages 4857–4859, 2021.

[41] Lianmin Zheng, Liangsheng Yin, Zhiqiang Xie, Chuyue Sun, Jeff Huang, Cody H Yu, Shiyi Cao, Christos Kozyrakis, Ion Stoica, Joseph E Gonzalez, et al. Sglang: Efficient execution of structured language model programs. *Advances in neural information processing systems*, 37:62557–62583, 2024.

[42] Xiliang Zhu, Chia-Tien Chang, Shayna Gardiner, David Rossouw, and Jonas Robertson. Resolving transcription ambiguity in spanish: A hybrid acoustic-lexical system for punctuation restoration. In *Proceedings of the Third Workshop on Understanding Implicit and Underspecified Language*, pages 33–41, 2024.

[43] Arturs Znotins, Normunds Gruzitis, and Roberts Darģis. From conversational speech to readable text: Post-processing noisy transcripts in a low-resource setting. In *Proceedings of the Tenth Workshop on Noisy and User-generated Text*, pages 143–148, 2025.

Appendix

A Details of the Data Construction Pipeline

This appendix expands Section 3 with a stage-by-stage account of how the training corpus is assembled. The pipeline turns heterogeneous raw text into the stage-specific training sets used by UniForm-0.8B. Table 2 summarizes the full chain; the subsections below give the details omitted from the main text for space.

Table 2: Overview of the data construction chain. Volumes are approximate.

Stage	Operation	Output / downstream use
(1) Source collection	Aggregate real ASR transcripts, entity-enriched text, curated corpora with verified punctuation, and synthetic Cantonese.	Three corpus types: real ASR transcriptions, entity-enriched and curated text with verified punctuation, and synthetic Cantonese.
(2) LLM synthesis	Seven linguistic modules $\times$ two routes (scene generation / entity integration) produce paired spoken-form and written-form samples.	$\sim$6M synthetic Cantonese samples; input is an unpunctuated spoken-form sentence, output its formatted written counterpart.
(3) Deduplication	Exact-match deduplication over the aggregated pool.	Retains $\sim$45.8% of raw samples; some sentences recur more than 500 times.
(4) Quality filtering & punctuation correction	A teacher LLM checks each sample against a comprehensive punctuation rule set under a content-preservation constraint; non-compliant samples are corrected rather than discarded.	41.2M SFT samples: $\sim$87.5% compliant retained directly, 12.5% corrected.
(5) ITN span annotation	Qwen3.5-397B-A17B identifies ITN-eligible spans by greedy left-to-right matching.	$\sim$42.6% of samples carry non-trivial ITN content.
(6) Stage-wise use	Route the pools to the three training stages.	Stage 1: 197M punctuation SFT; Stage 2: 41.2M joint punctuation–ITN SFT; Stage 3: expert-verified, ITN-entity-dense subset for SA-GRPO.
(7) Leakage control	Canonicalized exact matching, embedding retrieval with LLM verification, and ITN entity-instance retrieval against UniFormBench.	Removes affected benchmark instances; the training corpus is unchanged (Appendix D).

A.1 Source Data Collection

The corpus is built from four complementary source types, which after curation resolve into the three components reported in Section 3. First, real ASR transcriptions with human-corrected punctuation are drawn from production systems and cover conversational, command, and dictation scenarios. Second, a large pool of text specifically enriched with named entities — person names, organization names, product names, and technical terms — targets the well-known difficulty of placing punctuation at entity boundaries. Third, large-scale text corpora from news articles, books, and domain-specific publications with verified punctuation quality contribute broad domain coverage; together with the entity-enriched pool they form a single corpus component. Fourth, synthetic Cantonese samples are produced by the LLM synthesis pipeline described below, addressing the scarcity of properly punctuated Cantonese text.

A.2 LLM-Based Data Synthesis

The synthesis pipeline generates parallel *spoken-form* and *written-form* pairs: the input is an unpunctuated, spoken-form sentence and the output is its properly formatted written counterpart, with semantic content preserved. It is organized around seven linguistic modules, each targeting a distinct family of formatting decisions:

1. **Module 1 — Basic punctuation**: periods, commas, semicolons, colons, and quotation marks.
2. **Module 2 — Discourse markers and disfluencies**: interjections, filled pauses, and repetitions common in spontaneous speech.
3. **Module 3 — Numbers, dates, and time**: the ITN transformation of numerical and temporal expressions.
4. **Module 4 — Currency, mathematics, and symbols**: monetary amounts, mathematical expressions, emails, and URLs.
5. **Module 5 — Code-switching and units**: Chinese–English mixing patterns and measurement-unit normalization.
6. **Module 6 — Named entities**: proper nouns organized by a four-level hierarchical taxonomy.
7. **Module 7 — Domain-specific terminology**: specialized terms in technology, finance, medicine, and other verticals.

Each module operates through two complementary routes. In **Route A (scene generation)**, a teacher LLM produces diverse scenario seeds within a given domain; these seeds then serve as prompts for parallel corpus generation, yielding broad conversational coverage. In **Route B (entity integration)**, a four-level entity taxonomy (L1–L4 categories spanning domains such as technology, finance, entertainment, and daily life) guides the generation of entity-rich contexts. The dual-route design ensures that the synthetic data covers both diverse conversational scenarios and dense entity mentions.

In this work the pipeline is primarily applied to Cantonese, but it is language-agnostic: instantiating it for another target language requires only adapting the module prompts and the entity taxonomy.

A.3 Quality Filtering and Punctuation Correction

Because the corpus is aggregated from heterogeneous sources of varying reliability, we apply a unified quality-assurance pass before any downstream annotation. It both removes irredeemable samples and *corrects* recoverable punctuation errors rather than discarding them.

Exact-match deduplication. Exact-match deduplication retains approximately 45.8% of the raw samples. Certain subsets exhibit extreme duplication, with individual sentences recurring more than 500 times.

Rule-based assessment. A teacher LLM evaluates each surviving sample against a comprehensive punctuation rule set covering Chinese conventions, English conventions, Chinese–English mixed typesetting, and oral-form special cases. A hard constraint is that the teacher may modify *only* punctuation symbols: all text content — including dialectal expressions, ASR artifacts, and disfluencies — must be preserved verbatim. This guarantees that quality control never alters the semantic content that the model is asked to normalize.

Correction and quality scoring. Approximately 87.5% of samples are judged compliant and retained directly; each compliant sample additionally receives a data-quality score reflecting the structural complexity of its punctuation pattern, and samples falling below a threshold are filtered out as low-quality. The remaining 12.5% receive a corrected version generated under the same content-preservation constraint, recovering otherwise valuable training data. Analysis of the edits shows that the dominant correction types are comma insertions (28.8% of all edits), comma-to-period promotions at sentence boundaries (16.8%), and comma-to-question-mark corrections for interrogatives (10.0%), together with quotation-mark additions, colon replacements before reported speech, and enumeration-mark normalization. Diagnostic inspection of the flagged samples further surfaced recurring error categories such as incorrect book-title marks, inconsistent comma usage in parallel structures, and text modifications beyond the punctuation and ITN scope. After this stage the supervised fine-tuning corpus contains 41.2M samples.

A.4 ITN Span Annotation

Stage 2 requires explicit annotation of which text spans should undergo inverse text normalization. We employ Qwen3.5-397B-A17B as a teacher model to process the corpus and identify ITN-eligible spans by greedy left-to-right matching against its predictions. Each annotated span records four fields:

- the written-form substring in the target text;
- the corresponding spoken-form representation;
- a set of valid ITN candidates that captures legitimate ambiguity — e.g., the spoken form "one million" may be correctly written as "1,000,000", "1000000", or "1M";
- an entity-type label drawn from categories such as Number, Date, Currency, and Unit.

After annotation, approximately 42.6% of samples contain non-trivial ITN content; these supply the supervision signal for joint punctuation–ITN training, while the remaining samples continue to provide punctuation-only supervision. The candidate sets are also reused at evaluation time, where multiple valid ITN references are scored with a best-match protocol.

A.5 Stage-Wise Data Usage and Human-Curated RL Data

The assembled pools are routed to the three training stages in decreasing order of volume and increasing order of annotation quality.

Stage 1 (punctuation SFT). A separate large-scale corpus of 197M samples is used for punctuation-only supervised fine-tuning. Its breadth lets the model internalize punctuation patterns across Chinese, English, Cantonese, and code-switched text while preserving the backbone's general language understanding.

Stage 2 (joint punctuation–ITN SFT). The 41.2M-sample corpus carrying ITN span annotations is used for joint training. Samples with non-trivial ITN teach the two transformations simultaneously — punctuation insertion and spoken-to-written span conversion — while samples without ITN content continue to provide punctuation supervision and mitigate catastrophic forgetting.

Stage 3 (SA-GRPO). The reinforcement-learning stage uses a small, ITN-entity-dense subset verified by domain experts following detailed annotation guidelines. Compared with the machine-collected pools, this subset has longer average text length, carefully verified punctuation, and accurate ITN annotations; it is used to refine decision boundaries on ambiguous cases where the SFT stages may have absorbed conflicting patterns from heterogeneous data.

A.6 Corpus Statistics and Diagnostics

Figure 2 reports the distribution of the final corpus. The language axis spans four conditions — pure Chinese, pure English, Chinese–English code-switching, and Cantonese — and the topic axis covers 13 knowledge domains. The output-length distribution shows that machine-collected samples are typically short, whereas the human-annotated Stage 3 subset is substantially longer on average, reflecting its focus on ambiguous, entity-dense cases.

A diagnostic conducted during pipeline development further showed that the punctuation-error rate flagged by the teacher LLM is not uniform across domains: it is lowest for structured business and economic text and highest for spontaneous sports commentary.

B GRPO Objective and the SA-GRPO Formulation

This section gives the full training objective behind Stage 3 (Section 4.3); the main text keeps only a brief summary. For a prompt $q \sim P(Q)$, GRPO [31] samples a group of G responses $\{o_i\}_{i=1}^{G}$ from the behavior policy $\pi_{\theta_{\text{old}}}$ and scores each with the reward R_{total} defined in Section 4.3. Unlike PPO, it needs no learned critic: the group-relative advantage of response i, broadcast to all its tokens, is

$$A_{\text{seq}}^{(i)} = \frac{R_{\text{total}}^{(i)} - \text{mean}\big(\{R_{\text{total}}^{(j)}\}_{j=1}^{G}\big)}{\text{std}\big(\{R_{\text{total}}^{(j)}\}_{j=1}^{G}\big)}. \tag{10}$$

The policy is updated by maximizing the PPO-style clipped surrogate

$$\begin{aligned}\mathcal{J}_{\text{GRPO}}(\theta) = \mathbb{E}_{q,\,\{o_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}}\Bigg[\frac{1}{G}\sum_{i=1}^{G}\frac{1}{|o_i|}\sum_{t=1}^{|o_i|}\Big(\min\big[r_{i,t}A_{\text{seq}}^{(i)},\ \text{clip}\big(r_{i,t},1-\varepsilon,1+\varepsilon\big)A_{\text{seq}}^{(i)}\big] \\ - \gamma\, D_{\text{KL}}\big[\pi_\theta \,\|\, \pi_{\text{ref}}\big]\Big)\Bigg],\end{aligned} \tag{11}$$

where $o_{i,t}$ is the t-th token of o_i, $r_{i,t} = \pi_\theta(o_{i,t} \mid q, o_{i,<t}) / \pi_{\theta_{\text{old}}}(o_{i,t} \mid q, o_{i,<t})$ is the per-token importance ratio, ε is the clipping range, and γ is the KL coefficient toward a frozen reference policy π_{ref} ($\gamma = 0$ gives the no-KL variant in Section 5). Every token of o_i shares the same advantage $A_{\text{seq}}^{(i)}$. The main text suppresses the response index, writing A_{seq} and A_t for a single response of $T = |o_i|$ tokens.

SA-GRPO objective. SA-GRPO keeps Eq. (11) unchanged and replaces only the uniform advantage with the redistributed token advantage $A_{i,t} = A_{\text{seq}}^{(i)} \mu_{i,t}$, where $\mu_{i,t}$ is the zero-sum multiplier of Section 4.3 (written there as $A_t = A_{\text{seq}} \mu_t$ with the response index suppressed):

$$\mathcal{J}_{\text{SA-GRPO}}(\theta) = \mathbb{E}_{q,\, \{o_i\}_{i=1}^{G} \sim \pi_{\theta_{\text{old}}}} \Bigg[\frac{1}{G} \sum_{i=1}^{G} \frac{1}{|o_i|} \sum_{t=1}^{|o_i|} \Big(\min\big[r_{i,t} A_{i,t},\ \text{clip}\big(r_{i,t}, 1-\varepsilon, 1+\varepsilon\big) A_{i,t} \big] - \gamma\, D_{\text{KL}}\big[\pi_\theta \,\|\, \pi_{\text{ref}}\big] \Big) \Bigg]. \tag{12}$$

The clipped surrogate, importance ratio, and KL penalty are identical to GRPO; only the advantage is modified. Because $\sum_t (\mu_{i,t} - 1) = 0$, the substitution preserves each response's total advantage mass, leaving the overall update magnitude unchanged while concentrating the gradient on high-importance tokens.

C Language Composition of UniFormBench

We list the languages and dialects in each UniFormBench tier.

UniFormBench-Main. Four language conditions: pure Chinese, pure English, Chinese-English code-switching, and Cantonese.

UniFormBench-Foreign. Eight languages absent from training: French, German, Indonesian, Italian, Portuguese, Russian, Spanish, and Turkish.

UniFormBench-Dialect. Chinese regional varieties: Anhui dialects, Beijing Mandarin, Northeastern Mandarin, Southern Min (Hokkien), Guizhou dialect, Hangzhounese, Hebei dialect, Henanese, Hubei dialect, Xiang (Hunanese), Jiangsu dialects, Jinan dialect, Hakka, Lanzhou dialect, Liaoning dialect, Eastern Min, Nanchang dialect, Ningbonese, Qingdao dialect, Shaanxi dialect (Guanzhong), Shanghainese, Sichuanese, Suzhounese, Taiwanese Hokkien, Taiyuan dialect, Tianjin dialect, Wenzhounese, Southwestern Mandarin, Yinchuan dialect, Yunnan dialect, and Zhongyuan Mandarin.

D Training–Test Deduplication

To prevent data leakage, UniFormBench was constructed from initial data sources disjoint from those used for the training corpus. We then applied three complementary procedures that query the training corpus with each benchmark instance to identify residual overlap. Because one benchmark instance typically matches multiple near-duplicate training samples, we remove the affected benchmark instances rather than their more numerous training counterparts, which severs the same overlaps while preserving more training data; the training corpus itself is left unchanged.

Canonicalized exact matching. We removed punctuation and whitespace, lowercased the text, and performed exact matching on the resulting canonicalized strings.

Embedding-based retrieval and LLM-assisted verification. We encoded training and test samples with Qwen3-Embedding-0.6B and, for each test sample, retrieved training samples with cosine similarity above 0.95. Qwen3.5-397B-A17B then judged whether each retrieved pair was a duplicate differing only slightly in punctuation or ITN surface form, or instead showed substantial differences in punctuation annotation, ITN annotation, or contextual composition.

ITN entity-instance retrieval. For ITN instances involving product models, entity names, and similar entity-specific expressions, we searched the training set for samples containing the same

instance and inspected the retrieved candidates, removing the corresponding benchmark instance when a duplicate was confirmed.

Importantly, the mere occurrence of the same entity expression was not treated as evidence of duplication. In PR and ITN, different lexical arrangements and punctuation configurations may correspond to different ground-truth outputs and should therefore be treated as distinct task instances. We also retained semantically distinct uses of expressions that can function either as common words or as proper names, such as "titanic" as a common adjective versus the film title.

E Single-Component Reward Ablation

We train vanilla GRPO with each of the three reward components in isolation (with KL on), alongside the full SA-GRPO model and vanilla GRPO with all three components as a multiplicative scalar reward. Table 3 reports results on all three UniFormBench subsets. Single-component variants exhibit pronounced metric-specific volatility: a component that improves punctuation F1 on one subset can sharply worsen ITN WER, whereas the multiplicative combination stabilizes performance and attains the lowest WER on all three subsets.

Table 3: Single-component reward ablation on UniFormBench. Vanilla GRPO is trained with each reward component in isolation (KL on), alongside the full SA-GRPO model and vanilla GRPO with all three components as a multiplicative scalar reward.

Method	**UniFormBench-Main**					**UniFormBench-Dialect**					**UniFormBench-Foreign**				
	P	R	F1	WER	SER	P	R	F1	WER	SER	P	R	F1	WER	SER
SA-GRPO (full)	0.8120	0.7900	0.8010	0.38	3.54	0.7560	0.6890	0.7210	1.07	17.13	0.5410	0.6250	0.5800	7.02	61.52
Vanilla GRPO (all)	0.8319	0.7648	0.7969	0.42	3.64	0.7570	0.6713	0.7116	1.48	17.79	0.5543	0.5688	0.5615	9.15	68.53
R_{ITN} only	0.8131	0.7896	0.8012	0.52	4.08	0.7622	0.6698	0.7130	1.29	17.33	0.4235	0.6112	0.5003	12.87	70.63
R_{punc} only	0.8319	0.7620	0.7954	0.49	3.65	0.7651	0.6800	0.7200	1.10	17.03	0.6239	0.5640	0.5924	15.60	75.28
R_{format} only	0.8252	0.7670	0.7950	0.46	3.59	0.7559	0.6819	0.7170	1.44	19.59	0.3959	0.6179	0.4826	14.16	71.07

F Multi-Seed Stability

We repeat SA-GRPO training with three different random seeds (41, 42, and 43) and evaluate on all three UniFormBench subsets. Table 4 reports per-seed results together with the mean $\pm$ 1-sigma sample standard deviation ($n = 3$); the main text reports seed 42. The small deviations indicate stable performance across training seeds.

Table 4: Multi-seed stability of SA-GRPO on UniFormBench. Three independent training runs (seeds 41, 42, and 43), with mean $\pm$ 1-sigma sample standard deviation ($n = 3$). The main text reports seed 42.

Seed	**UniFormBench-Main**					**UniFormBench-Dialect**					**UniFormBench-Foreign**				
	P	R	F1	WER	SER	P	R	F1	WER	SER	P	R	F1	WER	SER
41	0.8104	0.7874	0.7988	0.37	3.52	0.7578	0.6830	0.7185	1.27	17.09	0.5474	0.6248	0.5835	7.18	61.49
42	0.8120	0.7900	0.8010	0.38	3.54	0.7560	0.6890	0.7210	1.07	17.13	0.5410	0.6250	0.5800	7.02	61.52
43	0.8318	0.7650	0.7970	0.38	3.54	0.7617	0.6782	0.7176	1.08	17.18	0.5436	0.6227	0.5805	6.93	61.56
Mean $\pm$ Std.	0.8181 $\pm$ 0.0119	0.7808 $\pm$ 0.0137	0.7989 $\pm$ 0.0020	0.3767 $\pm$ 0.0058	3.5333 $\pm$ 0.0115	0.7585 $\pm$ 0.0029	0.6834 $\pm$ 0.0054	0.7190 $\pm$ 0.0018	1.1400 $\pm$ 0.1127	17.1333 $\pm$ 0.0451	0.5440 $\pm$ 0.0032	0.6242 $\pm$ 0.0013	0.5813 $\pm$ 0.0019	7.0433 $\pm$ 0.1266	61.5233 $\pm$ 0.0351

G Mechanism Analysis of Entropy Normalization

G.1 Controlled Setup

To understand the failure of $\widehat{H}_t = H_t / \log |V|$, we run two otherwise identical one-epoch RL runs from the same SFT checkpoint (seed 42, $G = 8$, batch 1024, $\alpha = \beta = 0.5$, KL on) that differ only in the `normalize_entropy` switch; a full configuration diff confirms that all other hyperparameters, data, and the initial checkpoint coincide. Diagnostics of the `unified_dynamic` estimator are logged at every step of both arms.

G.2 Mechanism and Scale Mismatch

Why a global rescaling would be harmless, but this one is not. Because μ_t uses the mean-subtracted $\lambda_t - \bar{\lambda}$, multiplying the entire λ_t by a constant c yields $c(\lambda_t - \bar{\lambda})$ and would merely rescale

β; global changes of scale are therefore harmless. Normalization, however, divides only the entropy summand by $L = \log |V|$, leaving $D_t \in [0, 1]$ unchanged:

$$\lambda_t - \bar{\lambda} = \alpha\,(H_t - \bar{H}) + (1-\alpha)(D_t - \bar{D}) \;\longrightarrow\; \frac{\alpha}{L}(H_t - \bar{H}) + (1-\alpha)(D_t - \bar{D}).$$

The nominal $\alpha = 0.5$ thus becomes an effective ratio $\alpha_{\text{eff}} : (1-\alpha) = 1 : L \approx 1 : 12.4215$, i.e. $\alpha_{\text{eff}} \approx 0.04$ without re-tuning β.

Scale mismatch. $L = \log 248077 = 12.4215$ is the entropy of a *uniform* distribution over the vocabulary, an upper bound never approached in ASR text formatting. After SFT the policy is near-deterministic: actor token entropy averages 0.0636 / 0.0656 nats in the two arms and 0.0526 / 0.0514 over the last 20 steps, essentially identical, so normalization neither changes exploration nor causes entropy collapse. The normalized entropy actually fed into λ_t is only 0.0053, about 0.4% of the assumed $[0, 1]$ range and two orders of magnitude below D_t. Figure 4 shows the entropy trajectories and this gap, and Table 5 reports the diagnostics averaged over training.

Table 5: Unified-dynamic diagnostics averaged over 96 training steps for the two entropy-normalization arms. All values are computed from the logged per-step diagnostics.

Diagnostic	Unnormalized	Normalized
Normalizer L	1.0000	12.4215
Token entropy H_t (nats)	0.0636	0.0656
$H_t / \log \lvert V \rvert$	0.0051	0.0053
Entropy share of λ_t	0.3238	0.0405
$\mathrm{corr}(H_t, \mu_t)$, $A_{\text{seq}} > 0$	0.6343	0.0734
$\mathrm{corr}(H_t, \mu_t)$, $A_{\text{seq}} < 0$	-0.6264	-0.0787
$E\lvert\mu_t - 1\rvert$	0.0406	0.0279
$\mathrm{std}(A_t)/\mathrm{std}(A_{\text{seq}}) - 1$	0.0034	0.0017
$\mathrm{MAD}(\lambda_{\text{norm}})/\mathrm{MAD}(\lambda_{\text{raw}})$	0.6099	0.6146

G.3 Empirical Verification

Closed-form prediction and counterfactual check. Let $\rho = \alpha\sigma_H/(\alpha\sigma_H + (1-\alpha)\sigma_D)$ be the entropy share of λ_t's variance and $r = \rho/(1-\rho)$; normalization predicts $\rho' = (r/L)/(1 + r/L)$. From the unnormalized arm's measured $\rho = 0.3238$ ($r = 0.4788$) and $L = 12.4215$, this gives $\rho'_{\text{pred}} = 0.0371$, matching both the normalized arm's measured 0.0405 and the unnormalized arm's internal "if normalized" counterfactual 0.0375; conversely the normalized arm's "if unnormalized" counterfactual 0.3395 matches the unnormalized measurement 0.3238. Both directions close within 8%, and the counterfactual MAD ratio is nearly identical across arms (0.6099 vs. 0.6146), showing that the discrepancy is a deterministic scale effect present from the first step rather than a late-training divergence.

Three corroborating chains of evidence. (i) The entropy signal is erased: its share of λ_t's variance drops $0.3238 \to 0.0405$ (-87.5%). (ii) Entropy-aware alignment is switched off: $\mathrm{corr}(H_t, \mu_t)$ falls from 0.6343 to 0.0734 on positive-advantage samples and from -0.6264 to -0.0787 on negative-advantage samples, so μ_t is driven almost entirely by the coarse span-difficulty signal. (iii) Credit-assignment strength is halved: $E|\mu_t - 1|$ drops $0.0406 \to 0.0279$ and $\mathrm{std}(A_t)/\mathrm{std}(A_{\text{seq}}) - 1$ from 0.0034 to 0.0017. Together, these results show that normalization nearly switches off the entropy-aware component of SA-GRPO: with the entropy summand contributing almost nothing to λ_t, credit assignment degenerates into difficulty-only reweighting and the method falls back toward vanilla GRPO. Since training stays stable under this fallback (Table 6), the loss is confined to the entropy-driven gain, producing a mild degradation rather than a collapse. Figure 5 plots all four diagnostics over training.

Training dynamics data. Table 6 summarizes the training curves. Reward score, gradient norm, and KL stay close in the two arms ($\text{score} = 0.5122$ vs. 0.5118; $\text{grad_norm} = 3.235$ vs. 3.471; $\text{ppo_kl} = 0.00039$ vs. 0.00043 over the last 20 steps), confirming that normalization does not destabilize optimization. The mechanism is therefore a weakened credit-assignment signal, not a failure of training.

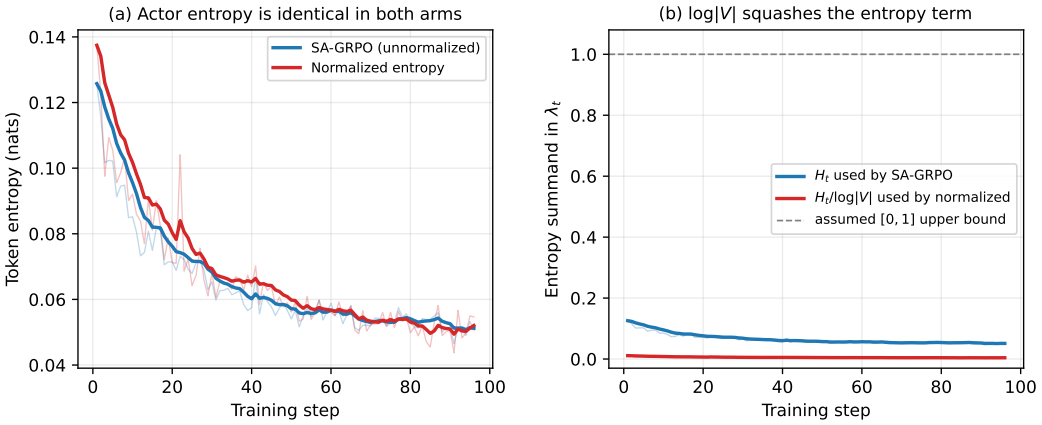

Figure 4: Training-time entropy in the two ablation arms. (a) Actor token entropy is nearly identical, ruling out differences in exploration or entropy collapse. (b) The entropy summand fed into λ_t stays near 0.0053 under normalization, far below the $[0, 1]$ upper bound assumed by the design (dashed line). Thin lines are per-step values; thick lines are EMA-smoothed.

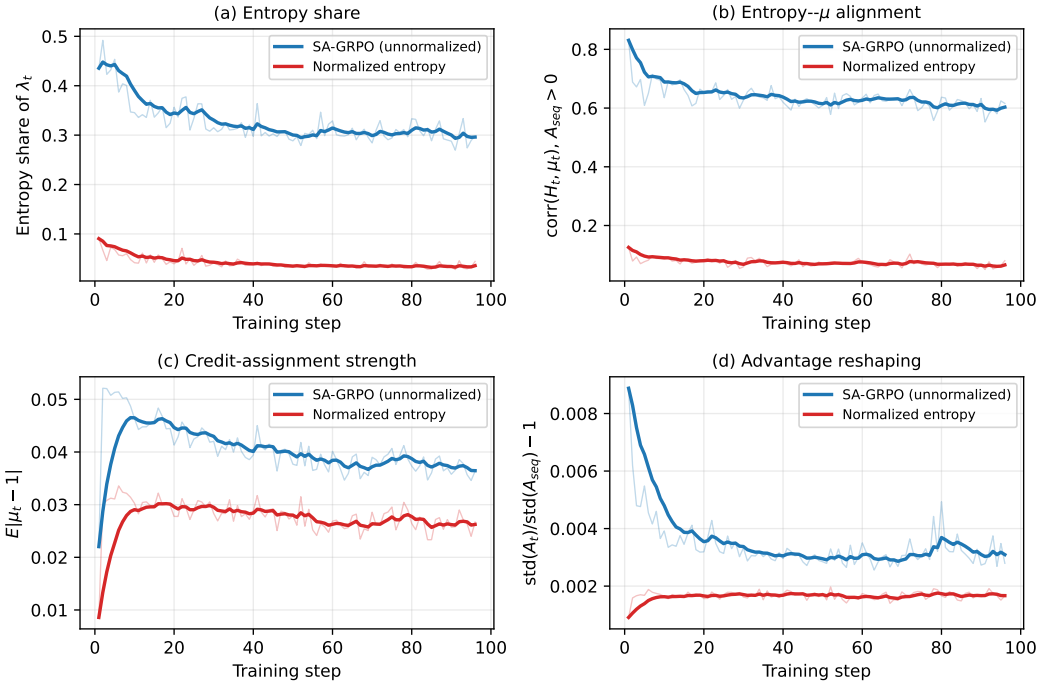

Figure 5: Unified-dynamic diagnostics over training: (a) entropy share of λ_t, (b) $\mathrm{corr}(H_t, \mu_t)$ for positive-advantage samples, (c) modulation strength $E|\mu_t - 1|$, (d) advantage reshaping $\mathrm{std}(A_t)/\mathrm{std}(A_{\mathrm{seq}}) - 1$. Thin lines are per-step values; thick lines are EMA-smoothed.

G.4 Limitations

These diagnostics come from a single seed, a single epoch, and one starting checkpoint. Training-time validation reward is noisy and nearly tied between arms (0.6544 vs. 0.6563 averaged over 20 evaluation points), so we do not use it as evidence of degradation, and we do not report benchmark scores from these auxiliary runs: the paper's results are those of the seed-42 runs in Table 1. Because normalization weakens the entropy-aware signal through a well-defined rescaling rather than by

Table 6: Training-curve summary for the two ablation arms (means over the indicated steps).

Quantity	Unnormalized	Normalized
Actor entropy, first 10 steps	0.0998	0.1052
Actor entropy, last 20 steps	0.0526	0.0514
Critic score, last 20 steps	0.5122	0.5118
Gradient norm, last 20 steps	3.235	3.471
Policy loss, last 20 steps	0.0094	0.0117
PPO KL, last 20 steps	0.00039	0.00043

destabilizing optimization, the effect should be read as a mild degradation rather than a training failure.

H Qualitative Examples

This section presents a selection of examples illustrating the outputs of UniForm-0.8B on the three tiers of UniFormBench. Each example shows the raw, unpunctuated ASR transcript followed by the jointly punctuated and inverse-text-normalized output produced in a single autoregressive pass. The examples cover diverse punctuation patterns—enumerations, book-title marks, quotations, colons, semicolons, and question or exclamation marks—as well as ITN conversions such as spoken numerals to digits, case restoration, and acronym normalization. The three boxes below correspond to UniFormBench-Main (four language groups), UniFormBench-Dialect (31 regional varieties), and the zero-shot UniFormBench-Foreign tier (eight languages).

UniFormBench-Main — 10 examples per language group

Mandarin Chinese

In: 按照加涅对学习结果的分类不属于智能学习的是一语言信息二具体概念三定义概念四高级规则哪一个不属于

Out: 按照加涅对学习结果的分类，不属于智能学习的是：一、语言信息；二、具体概念；三、定义概念；四、高级规则。哪一个不属于？

In: 能以命题形式进行思维表明儿童的认知发展已进入一感知运动阶段二前运算阶段三具体运算阶段四形式运算阶段

Out: 能以命题形式进行思维，表明儿童的认知发展已进入：一、感知运动阶段；二、前运算阶段；三、具体运算阶段；四、形式运算阶段。

In: 我要看的是巴啦啦小魔仙那个嗯有珠子的 1000 个你能帮我打出来吗那个珠珠要放在他包里面

Out: 我要看的是《巴啦啦小魔仙》那个，嗯，有珠子的1000个，你能帮我打出来吗？那个珠珠要放在他包里面。

In: 海通证券嗯早晨 9:15～9:25 竞价排行榜排序榜在哪里可以查找

Out: 海通证券，嗯，早晨 9:15～9:25 竞价排行榜、排序榜在哪里可以查找？

In: 趁我女朋友去超市购物我就吃了些棒棒糖眼巴巴地说我也想吃你们叹了口气还是给我买了

Out: 趁我女朋友去超市购物，我就吃了些棒棒糖，眼巴巴地说：“我也想吃。”你们叹了口气，还是给我买了。

In: 我给女朋友分享她的照片女朋友说你怎么能随便发领导的照片怎么回复啊

Out: 我给女朋友分享她的照片，女朋友说：“你怎么能随便发领导的照片？”怎么回复啊？

In: 双世宠妃 2 你俩是属狗的吗天天给我撒狗粮

Out: 《双世宠妃2》，你俩是属狗的吗？天天给我撒狗粮。

In: 根据告别童年这一课补充一下教学目标教学重难点教学资源和教法需要

Out: 根据《告别童年》这一课，补充一下教学目标、教学重难点、教学资源和教法需要。

In: 安徒生丹麦童话作家代表作有卖火柴的小女孩丑小鸭夜莺海的女儿等这句话对吗

Out: 安徒生，丹麦童话作家，代表作有《卖火柴的小女孩》《丑小鸭》《夜莺》《海的女儿》等，这句话对吗？

In: 植物大战僵尸花园战争 2 钢甲大嘴花

Out: 《植物大战僵尸：花园战争2》，钢甲大嘴花。

Cantonese

In: 请问 senior 年龄学历工作经验如何只想作个参考 thanks
Out: 请问 senior 年龄、学历、工作经验如何？只想作个参考，thanks。

In: 寻日兴起去超市买零食仲有百力滋童星面 m and m's
Out: 寻日兴起去超市买零食，仲有百力滋、童星面、M&M's。

In: 4 日 3 晚永州行~行程想去零陵古城桐子坳阳明山九嶷山路线行程如何安排
Out: 4日3晚永州行~行程想去零陵古城、桐子坳、阳明山、九嶷山，路线行程如何安排？

In: 李宁一切皆可皆有可能商标图案
Out: 李宁"一切皆可，皆有可能"商标图案。

In: 叶问 3 有吴千语呢个演员
Out: 《叶问3》有吴千语呢个演员？

In: 伦敦陷落和冰海陷落里面男主角系咪同一个人系边个国家
Out: 《伦敦陷落》和《冰海陷落》里面男主角系咪同一个人？系边个国家？

In: 用英文表达我将自己变靓方法例如有做面部护理染发剪发同埋节食然后我仲会 train 我肌肉线条添
Out: 用英文表达我将自己变靓方法，例如有做面部护理、染发、剪发同埋节食，然后我仲会 train 我肌肉线条添。

In: 我俾自己贴上呢标签电影自由职业爱上网乐观缺爱 ing
Out: 我俾自己贴上呢标签：电影、自由职业、爱上网、乐观、缺爱 ing。

In: 钓鱼钓虾捉蟹都是享受过程西贡有鱼排可以钓鱼同 BBQ 应该几好玩
Out: 钓鱼、钓虾、捉蟹都是享受过程，西贡有鱼排可以钓鱼同 BBQ，应该几好玩。

In: 好多位可以试如 Chief Architect Assistant Clerical Officer SSA
Out: 好多位可以试，如 Chief Architect、Assistant Clerical Officer、SSA。

English

In: hello goodbye Hello my name is Maggie What's your name
Out: Hello, goodbye. Hello, my name is Maggie. What's your name?

In: did you leave a tip No we don't have to We've already paid the service charge
Out: Did you leave a tip? No, we don't have to. We've already paid the service charge.

In: it's it's going well What fresh things did you meet today
Out: It's, it's going well. What fresh things did you meet today?

In: how are you today I'm fine thank you
Out: How are you today? I'm fine, thank you.

In: i mean in which city are you Because you said you have twenty two degrees sunny and warm
Out: I mean, in which city are you? Because you said you have 22°, sunny and warm.

In: at two thirty class starts at three o'clock
Out: At 2:30, class starts at 3:00.

In: why do you wanna see his chat window It's because you don't trust me That's why
Out: Why do you wanna see his chat window? It's because you don't trust me. That's why.

In: all right all right Hmm if I feel the life is so boring how can I do
Out: All right, all right. Hmm, if I feel the life is so boring, how can I do?

In: in Shenzhen right Right around the time when you guys launched uh clash royale
Out: In Shenzhen, right? Right around the time when you guys launched, uh, Clash Royale.

In: OK we can we can debate about Should college education be free
Out: OK, we can, we can debate about: Should college education be free?

Chinese–English Mix

In: 星际争霸二虚空之遗大结局当最后的 CG 出现雷诺坐在酒吧凯瑞甘在身后出现的时候这个画面配个什么音乐最好
Out: 《星际争霸2：虚空之遗》大结局，当最后的 CG 出现，雷诺坐在酒吧，凯瑞甘在身后出现的时候，这个画面配个什么音乐最好？

In: 以下哪个定位的主体拥有重大价值 i p r 选项一核心公司选项二业务中心选项三海外子公司
Out: 以下哪个定位的主体拥有重大价值？IPR：选项 1，核心公司；选项 2，业务中心；选项 3，海外子公司。

In: 从认知发展的阶段来看二至七岁正处于 a 感知运动阶段 B 具体运算阶段 c 前运算阶段 d 形式运算阶段
Out: 从认知发展的阶段来看，2 至 7 岁正处于：A. 感知运动阶段；B. 具体运算阶段；C. 前运算阶段；D. 形式运算阶段。

In: 三氧化二铝和氢氧化钠在水中反应生成四氢氧化氯化钠的本质以及 n a a l o h 四这个化合物怎么读怎么形成呢
Out: Al2O3和NaOH在水中反应生成四氢氧化氯化钠的本质，以及 NaAl(OH)4 这个化合物怎么读、怎么形成呢？

In: 我现在定制子实体纯提物的 pH 是 4 五怎么能调成 5 用柠檬酸柠檬酸钠调
Out: 我现在定制子实体纯提物的 pH 是 4.5，怎么能调成 5？用柠檬酸、柠檬酸钠调。

In: 在面试过程当中面试官问到 h r b p 你具备什么样的硬性优势和软性优势该怎么回答
Out: 在面试过程当中，面试官问到 HRBP：“你具备什么样的硬性优势和软性优势？”该怎么回答？

In: steam 版怪物猎人荒野点击开始游戏之后没有任何反应不能正常启动是什么原因
Out: Steam 版《怪物猎人：荒野》点击开始游戏之后没有任何反应，不能正常启动是什么原因？

In: 骁龙 Gen1 gen 二 gen 三哪一个芯片的口碑比较好例如它发热低功耗低手机不发烫等等
Out: 骁龙 Gen1、Gen2、Gen3，哪一个芯片的口碑比较好？例如它发热低、功耗低、手机不发烫等等。

In: 下列关于地球的表述不正确的是 A 地壳平均厚度约 17 千米 B 地幔厚度约 3900 千米 C 初始地球平均温度超过 1000 摄氏度 D 地核厚度约 3400 千米
Out: 下列关于地球的表述不正确的是：A.地壳平均厚度约 17 千米；B.地幔厚度约 3900 千米；C.初始地球平均温度超过 1000 摄氏度；D.地核厚度约 3400 千米。

In: 通过买入 p b 零点四以下股息率百分之四以上的港股 30 只分散买长期可以获得多少的年化收益率
Out: 通过买入 PB 0.4 以下、股息率 4% 以上的港股，30 只分散买，长期可以获得多少的年化收益率？

UniFormBench-Dialect — 2 examples per regional variety

Anhui dialects

In: 哎哟喂哎哟咋搞诶哎哟妈嘞这如今一百万在我们手头上我们咋讲得清呢

Out: 哎哟喂，哎哟，咋搞诶？哎哟妈嘞！这如今100万在我们手头上，我们咋讲得清呢？

In: 子富啊你们嘞又年轻又没小伢拖累出去拼几年照哦不要像我一样的四十岁喽下岗了再出去拼那就不容易咯

Out: 子富啊，你们嘞又年轻又没小伢拖累，出去拼几年照哦，不要像我一样的40岁喽，下岗了再出去拼，那就不容易咯。

Beijing Mandarin

In: 昨儿个我溜达 s k p 瞅见一包儿好家伙标价五万八你说这玩意儿值当么皮子倒是挺软和可这价儿够买一平米房了我寻思半天也没敢下手那售货员还一个劲儿跟我唠限量款听得我直嘬牙花子咱普通老百姓整这玩意儿干啥背出去都怕人惦记不如多置办两件实在衣裳穿

Out: 昨儿个我溜达 SKP 瞅见一包儿，好家伙，标价5万8！你说这玩意儿值当么？皮子倒是挺软和，可这价儿够买1平米房了。我寻思半天也没敢下手，那售货员还一个劲儿跟我唠“限量款”，听得我直嘬牙花子。咱普通老百姓整这玩意儿干啥？背出去都怕人惦记，不如多置办两件实在衣裳穿。

In: 得嘞坏咧那麻利儿扔了呗它那上头保质期还写着十二个月呢我看这刚几天呀刚刚一俩月就就那嘛了

Out: 得嘞，坏咧，那麻利儿扔了呗，它那上头保质期还写着12个月呢，我看这刚几天呀，刚刚一俩月就就那嘛了。

Northeastern Mandarin

In: 紧接脚儿哪吒跟我吧说那你演的那出哭戏咋那老真呢我说我滴溜的眼药水儿呗

Out: 紧接脚儿，哪吒跟我吧说：“那你演的那出哭戏咋那老真呢？”我说：“我滴溜的眼药水儿呗。”

In: 俺在岫岩卖衣裳都干了三十年了老主顾十里八村哪旮旯都有你咋没上俺店里来呢快请进

Out: 俺在岫岩卖衣裳都干了30年了，老主顾十里八村哪旮旯都有。你咋没上俺店里来呢？快请进！

Southern Min (Hokkien)

In: 各位乡亲恁遮咧收看咧是我爱闽南语欢迎逐个登录咱咧官方微博QZTV 我爱闽南语阮提者意见建议

Out: 各位乡亲，恁遮咧收看咧是《我爱闽南语》。欢迎逐个登录咱咧官方微博——QZTV 我爱闽南语，阮提者意见建议。

In: 抑搁一诶说法因迄芥菜拢是食霜冻菜啦较讲愈寒时阵是愈好食啊六月迄去割菜哦根本是假有心六月根本毋是咧割菜诶时阵啦

Out: 抑搁一诶说法：因迄芥菜拢是食霜冻菜啦。较讲，愈寒时阵是愈好食。啊，6月迄去割菜哦，根本是假有心；6月根本毋是咧割菜诶时阵啦。

Guizhou dialect

In: 要是四岁勒时候四岁勒时候他他把他屋头勒东西全部卖完咯卖完咯然后他四岁他又搞不成哪样咯只有卖东西维持生活咯嘛他把他屋头东西全部卖完咯以后

Out: 要是4岁勒时候，4岁勒时候他，他把他屋头勒东西全部卖完咯；卖完咯，然后他4岁，他又搞不成哪样咯，只有卖东西维持生活咯嘛——他把他屋头东西全部卖完咯以后。

In: 嗯要是嗯我们勒边有货咯要是说你买我们屋头嘞热水器的话我们可以免费送你五双过膝袜

Out: 嗯，要是嗯我们勒边有货咯，要是说你买我们屋头嘞热水器的话，我们可以免费送你5双过膝袜。

Hangzhounese

In: 侬格点水果都烂特嘞还摆勒最上头卖辩不是明摆着坑人伐

Out: 侬格点水果都烂特嘞，还摆勒最上头卖，辩不是明摆着坑人伐？

In: 老师侬对阿拉小伢儿介关心阿拉真当过意弗去要牢好谢谢侬哉

Out: 老师侬对阿拉小伢儿介关心，阿拉真当过意弗去，要牢好谢谢侬哉！

Hebei dialect

In: 具体情况下还得看咋开你是玩越野咧还是打沙滩上还是草原上综合油耗都不一样大哥们呐还有这 3.5 是最最省油的呃11 个油到 13 个油3.5 全时四驱

Out: 具体情况下还得看咋开：你是玩越野咧？还是打沙滩上还是草原上？综合油耗都不一样，大哥们呐！还有，这3.5是最最省油的——呃，11个油到13个油——3.5全时四驱。

In: 这车搁这也算台二手车虽说性价比俺觉着还中啊俺姐俩刚咋儿算了算四十九咧你如今三十九都勿要

Out: 这车搁这也算台二手车，虽说性价比俺觉着还中啊，俺姐俩刚咋儿算了算49咧，你如今39都勿要。

Henanese

In: 咦这原神里头还有啥子多人游戏嘞咋我之前都注意到过咧

Out: 咦，这《原神》里头还有啥子多人游戏嘞？咋我之前都注意到过咧？

In: 咋会中嘞我感觉给六星好评七星好评也中啊九星好评

Out: 咋会中嘞？我感觉给6星好评、7星好评也中啊，9星好评！

Hubei dialect

In: 阿凡达哦阿凡达得第一部好看咧

Out: 《阿凡达》哦，《阿凡达》得第一部好看咧。

In: 几多本科生送外卖的占百分之三十是么事概念这些书读到最后就是压根就是最后得工作的你吓我蛮吓人咧

Out: 几多本科生送外卖的占30%，是么事概念？这些书读到最后就是压根就是最后得工作的。你吓我，蛮吓人咧！

Xiang (Hunanese)

In: 就证明咧它虽然落哒一哈子雪但是嘞它的气温还是得好低晓啵咯它要零度以下才结冰我记得我们细伢子那阵子啊

Out: 就证明咧，它虽然落哒一哈子雪，但是嘞，它的气温还是得好低。晓啵咯？它要0度以下才结冰。我记得我们细伢子那阵子啊。

In: 所以讲我现在咯反正我基本上就是几个水果咧苹果梨子咧西红柿咧石榴咧我别个东西我不蛮吃得啵我嘞我呃心脏也不蛮好

Out: 所以讲我现在咯，反正我基本上就是几个水果咧：苹果、梨子咧，西红柿咧，石榴咧。我别个东西我不蛮吃得啵。我嘞，我呃心脏也不蛮好。

Jiangsu dialects

In: 分派任务也是个技术活侬弗能硬邦邦地甩过去讲舀个做埃个伊做我一般是先摸清爽各人擅长哪块哪个做事体细发哪个脑子活络然后凑到跟前去用商量个口气讲哎舀块侬熟拨给侬我放心侬看大概要几天噶么一来俚拉感觉是自家接个活弗是被派个做起来也有劲多哉有时候俚拉自家还会主动加码讲舀块我再顺带弄一记

Out: 分派任务也是个技术活，侬弗能硬邦邦地甩过去讲：“舀个做，埃个伊做。”我一般是先摸清爽各人擅长哪块，哪个做事体细发，哪个脑子活络。然后凑到跟前去，用商量个口气讲：“哎，舀块侬熟，拨给侬我放心，侬看大概要几天？”噶么一来，俚拉感觉是自家接个活，弗是被派个，做起来也有劲多哉。有时候俚拉自家还会主动加码，讲：“舀块我再顺带弄一记。”

In: 等消息葛几天我时不时就摸出手机来望望生怕漏脱银行葛电话我屋里人看我心神不定葛就说我侬这个人就是沉不住气人家讲勒等几天就等几天呗侬急有啥用啊我想想也是急也急弗来就强迫自家勿去想舀桩事体

Out: 等消息葛几天，我时不时就摸出手机来望望，生怕漏脱银行葛电话。我屋里人看我心神不定葛，就说我：“侬这个人就是沉不住气，人家讲勒等几天就等几天呗，侬急有啥用啊？”我想想也是，急也急弗来，就强迫自家勿去想舀桩事体。

Jinan dialect

In: 哎对咧嗯咋说你看过路人女主呗

Out: 哎，对咧，嗯，咋说，你看过《路人女主》呗？

In: 知否里头演了个如兰俺知否俺稀罕看呗因着它就是那种宅斗剧

Out: 《知否》里头演了个如兰，俺《知否》俺稀罕看呗，因着它就是那种宅斗剧。

Hakka

In: 净系同阿姐在做副校长旁边打王者我两个差唔多就搞掂了

Out: 净系同阿姐在做，副校长旁边打王者，我2个差唔多就搞掂了。

In: 涯女朋侪爱过生日咧能推荐一兜适合十九岁妹个生日礼物涯今年二十六岁

Out: 涯女朋侪爱过生日咧，能推荐一兜适合19岁妹个生日礼物？涯今年26岁。

Lanzhou dialect

In: 他们说撒啊呀晚上十一点咋着尼你要是啊特别在外头显摆呗哎戴个啥子大金链子小手表那可能就让人抢掉呢

Out: 他们说撒，啊呀，晚上11点咋着尼？你要是啊特别在外头显摆呗，哎，戴个啥子大金链子、小手表，那可能就让人抢掉呢。

In: 就像那个山上做下的那种玉米啊土豆啊这些烤下的也有煮下的你爱咋选咋选呗

Out: 就像那个山上做下的那种玉米啊、土豆啊这些，烤下的也有煮下的，你爱咋选咋选呗。

Liaoning dialect

In: 俺寻思这攒钱的事儿得好好唠唠你瞅现在这钱毛得跟啥似滴不攒点儿真不行我每个月工资一到手就先抠出来两千块存死期剩下滴才敢可劲儿造咱这旮沓老话儿说吃不穷穿不穷算计不到才受穷这话在理儿你平常咋整滴也整点儿存钱招儿不

Out: 俺寻思这攒钱的事儿得好好唠唠，你瞅现在这钱毛得跟啥似滴，不攒点儿真不行。我每个月工资一到手就先抠出来2000块存死期，剩下滴才敢可劲儿造。咱这旮沓老话儿说：“吃不穷，穿不穷，算计不到才受穷”，这话在理儿。你平常咋整滴？也整点儿存钱招儿不？

In: 哎呀妈呀今儿个这天儿溜达可忒得劲儿了咱俩搁中央大街那溜达溜达呗道两边儿那些个老建筑瞅着就稀罕人石板路踩上去咯噔咯噔的旁边卖糖葫芦的老太太吆喝得可带劲儿了你说咱从防洪纪念塔那头往松花江边儿溜达顺道还能整根马迭尔冰棍儿那玩意儿啃着贼拉爽

Out: 哎呀妈呀，今儿个这天儿溜达可忒得劲儿了！咱俩搁中央大街那溜达溜达呗，道两边儿那些个老建筑瞅着就稀罕人。石板路踩上去咯噔咯噔的，旁边卖糖葫芦的老太太吆喝得可带劲儿了。你说咱从防洪纪念塔那头往松花江边儿溜达，顺道还能整根马迭尔冰棍儿，那玩意儿啃着贼拉爽！

Eastern Min

In: 故呢做乜唔去关注一下大叔杂货铺咧

Out: 故呢做乜唔去关注一下《大叔杂货铺》咧?

In: 许场大雨连续落咧四个钟头一下仔让老巴嘎侬都野难适应

Out: 许场大雨连续落咧4个钟头，一下仔让老巴嘎侬都野难适应。

Nanchang dialect

In: 就两三个呃有两个咧暗恋个人咯都追到咯你又唔是不知道

Out: 就两三个呃，有2个咧，暗恋个人咯，都追到咯，你又唔是不知道。

In: 掰些 app 啊什哩小红书抖音高头看下人家推荐个有什哩好恰个我觉得就差不多咧

Out: 掰些 APP 啊，什哩《小红书》、《抖音》高头看下人家推荐个有什哩好恰个，我觉得就差不多咧。

Ningbonese

In: 老师侬真当是活菩萨阿拉小囡现在读书劲头交关足全靠侬哉

Out: 老师侬真当是活菩萨，阿拉小囡现在读书劲头交关足，全靠侬哉！

In: 望见天边掰团乌云哉伐搞勿好要出龙卷风嘞

Out: 望见天边掰团乌云哉伐？搞勿好要出龙卷风嘞！

Qingdao dialect

In: 你且看看这音乐可不只俺大人听就是三岁两岁的小小嘎也愿意听咧

Out: 你且看看，这音乐可不只俺大人听，就是3岁2岁的小小嘎也愿意听咧。

In: 两三千就杠好咧反正俺欢喜把这个华为的 vivo 俺就是稀罕选这俩款他们说小米的也杠好

Out: 两三千就杠好咧，反正俺欢喜把这个华为的、vivo，俺就是稀罕选这俩款，他们说小米的也杠好。

Shaanxi dialect (Guanzhong)

In: 中医一个疗程要调一开始七次咦呀不对上次你不是给我说你去中医时候调理的时候大夫给你咋说的一个疗程喝喝不是一天喝几包呃几包中药

Out: 中医1个疗程要调，一开始7次。咦呀，不对！上次你不是给我说，你去中医时候调理的时候，大夫给你咋说的？1个疗程喝……喝不是，一天喝几包？呃，几包中药。

In: 她写一哈她奏说是手疼手疼不想写胳膊疼胳膊疼不想写其实奏是手腕疼她不会用劲她奏会把乃个攥滴很实后来在这搭才写咧两天然后老师奏给我说回去写字滴时候你瞅一哈你要时刻提醒她给她说

Out: 她写一哈，她奏说是手疼手疼不想写，胳膊疼胳膊疼不想写，其实奏是手腕疼。她不会用劲，她奏会把乃个攥滴很实。后来在这搭才写咧2天，然后老师奏给我说："回去写字滴时候，你瞅一哈，你要时刻提醒她、给她说。"

Shanghainese

In: 分蛋糕个辰光又乱了一歇小囡要掰朵花要伊个字大人呢就讲随便随便一小块就好了我阿爷拿第一块蛋糕伊勿大舍得吃就用调羹舀了一点点放勒嘴巴里伊讲味道蛮好就是忒浪费钞票我爷叔就讲阿爸一生一世就一趟八十岁肉麻钞票讲得大家侪笑了蛋糕蛮好吃勿甜勿腻我阿爷也吃了小半块

Out: 分蛋糕个辰光又乱了一歇。小囡要掰朵花，要伊个字；大人呢，就讲："随便随便，一小块就好了。"我阿爷拿第一块蛋糕，伊勿大舍得吃，就用调羹舀了一点点，放勒嘴巴里。伊讲："味道蛮好，就是忒浪费钞票。"我爷叔就讲："阿爸，一生一世就一趟80岁，肉麻钞票。"讲得大家侪笑了。蛋糕蛮好吃，勿甜勿腻，我阿爷也吃了小半块。

In: 我就立勒旁边看看心里向勒算掰只包要几钿啦后来听到售货员讲价钿我吓煞特了要五万块以上我朋友倒好像呒没啥感觉还在问有啥颜色好换我偷偷碰碰伊讲侬真个要买啊掰能贵个物事背出去勿怕弄脏脱啊伊朝我眨眨眼轻手轻脚讲弗关侬事我自家个铜钿我只好瘪瘪嘴继续立勒旁边陪伊

Out: 我就立勒旁边看看，心里向勒算：掰只包要几钿啦？后来听到售货员讲价钿，我吓煞特了，要5万块以上！我朋友倒好像呒没啥感觉，还在问有啥颜色好换。我偷偷碰碰伊，讲："侬真个要买啊？掰能贵个物事，背出去勿怕弄脏脱啊？"伊朝我眨眨眼，轻手轻脚讲："弗关侬事，我自家个铜钿。"我只好瘪瘪嘴，继续立勒旁边陪伊。

Sichuanese

In: 而且这二十袋都是混装嘞有榨菜心儿有萝卜丁丁儿还有榨菜丝儿吃起都老下饭咯

Out: 而且这20袋都是混装嘞，有榨菜心儿、有萝卜丁丁儿、还有榨菜丝儿，吃起都老下饭咯。

In: 硬是不像在真个生气我赶忙问道爷爷这把剑是咋个回事哦

Out: 硬是不像在真个生气，我赶忙问道："爷爷，这把剑是咋个回事哦？"

Suzhounese

In: 讲得对极哉我俚现在每日下班转来就专门用苏州闲话搭侬子讲白相教伊认物事比方讲台子弗说桌子杌子弗说凳子困觉弗说睡觉最有趣是教伊讲晏歇会伊总归讲成晏歇见还要加个拜拜真是让人啼笑皆非阿有啥好法子教小囡学方言

Out: 讲得对极哉！我俚现在每日下班转来就专门用苏州闲话搭侬子讲白相，教伊认物事。比方讲"台子"弗说桌子，"杌子"弗说凳子，"困觉"弗说睡觉。最有趣是教伊讲"晏歇会"，伊总归讲成"晏歇见"，还要加个拜拜，真是让人啼笑皆非。阿有啥好法子教小囡学方言？

In: 是咯是咯我俚做导游个最清爽弗过游客来仔总问为啥今年个竹子看起来蔫头耷脑我只能实话实说天时勿对哉呀不过讲起来西山头北面片混交林倒还争气榉树麻栎长得还算登样特别是几棵百年银杏叶子还是碧绿生青看来老树到底根基深

Out: 是咯是咯，我俚做导游个最清爽弗过。游客来仔总问："为啥今年个竹子看起来蔫头耷脑？"我只能实话实说："天时勿对哉呀。"不过讲起来，西山头北面片混交林倒还争气，榉树、麻栎长得还算登样。特别是几棵百年银杏，叶子还是碧绿生青，看来老树到底根基深。

Taiwanese Hokkien

In: 啊毋过你按呢想抑是三四冬嘛可能有一寡仔伤长因为结婚著是要冲动三四冬你会恁著会想讲哦阮已经足稳定啊敢有需要结婚

Out: 啊毋过你按呢想，抑是三四冬嘛可能有一寡仔伤长，因为结婚著是要冲动。三四冬你会恁著会想讲："哦，阮已经足稳定啊，敢有需要结婚？"

In: 台东花莲是伤远啦欲按怎去下港爱坐车到遐爱四五点钟足诶

Out: 台东花莲是伤远啦，欲按怎去下港爱坐车到遐爱4、5点钟，足诶。

Taiyuan dialect

In: 你你想俺们现在的人绝对不可能哎新华字典跟有声书你觉见是买平板呀还是买新华字典呀

Out: 你，你想俺们现在的人绝对不可能。哎，《新华字典》跟有声书，你觉见是买平板呀，还是买《新华字典》呀？

In: 因它人工智能哇它不是靠那些大数据来收揽信息嘞这忽儿哇收揽信息哇嗯可难咧因这忽儿信息太多咧嗯他要收揽那七信息哇得从那么多信息里头筛拣些有用的信息

Out: 因它人工智能哇，它不是靠那些大数据来收揽信息嘞；这忽儿哇，收揽信息哇，嗯，可难咧，因这忽儿信息太多咧；嗯，他要收揽那7信息哇，得从那么多信息里头筛拣些有用的信息。

Tianjin dialect

In: 介是气候上的毛病你瞅你那一点雪都没下我们这嘎达雪都快埋过膝盖了你说咋整

Out: 介是气候上的毛病。你瞅，你那一点雪都没下；我们这嘎达，雪都快埋过膝盖了。你说咋整？

In: 然后一点儿都不带动弹的我就跟我当家的说我说咱就甭提别的咧打你二百斤的分量你背一百斤的物件儿

Out: 然后一点儿都不带动弹的，我就跟我当家的说，我说：“咱就甭提别的咧，打你200斤的分量，你背100斤的物件儿。”

Wenzhounese

In: 阿芬姐今日哪能又告假哉伊舜个月侪第三回哉

Out: 阿芬姐今日哪能又告假哉？伊舜个月侪第三回哉！

In: 你囡儿想学唱歌啊让渠去学好嘞侬温州人最会培养艺术人才个

Out: 你囡儿想学唱歌啊？让渠去学好嘞，侬温州人最会培养艺术人才个！

Southwestern Mandarin

In: 有句话咋个说嘞都有点儿想不醒豁咯富在深山有远亲穷在闹市无人问哦对喽就是这句

Out: 有句话咋个说嘞？都有点儿想不醒豁咯：“富在深山有远亲，穷在闹市无人问。”哦，对喽，就是这句。

In: 现在我又遭起新问题了就是那个两千块钱的转账记录哈我弄落教了应该是被我删咯噻我现在只有那个五百块钱的转账记录不我不晓得那些删了的账单能不能重新找出来哦

Out: 现在我又遭起新问题了，就是那个2000块钱的转账记录哈，我弄落教了，应该是被我删咯噻。我现在只有那个500块钱的转账记录。不，我不晓得那些删了的账单能不能重新找出来哦？

Yinchuan dialect

In: 对啵现在他就预制咧一个实际上就系那个四季鲜嘛还系那个哪搭我去过一哈哦

Out: 对啵？现在他就预制咧一个，实际上就系那个《四季鲜》嘛，还系那个哪搭？我去过一哈哦。

In: 表演将二十分钟个节目完咧人家接着泼水泼水都随住阿达音乐音乐也响起来咧

Out: 表演将20分钟个节目完咧，人家接着泼水，泼水都随住阿达音乐，音乐也响起来咧。

Yunnan dialect

In: 他说认不得咋个整嘛他根本就没得麻将瘾才不克打现在倒是牌也不打麻将也不克瞧咯

Out: 他说认不得咋个整嘛，他根本就没得麻将瘾，才不克打，现在倒是牌也不打、麻将也不克瞧咯。

In: 好生学个一年我瞧瞧给有哪样成果如果没得哪样成果我就不干这行了要是他们都不教的话那些老师还说了不要频繁的跳槽

Out: 好生学个1年，我瞧瞧给有哪样成果；如果没得哪样成果，我就不干这行了——要是他们都不教的话。那些老师还说了：不要频繁的跳槽。

Zhongyuan Mandarin

In: 今儿个瞅见有说弄啥嘞嗯可别喷那么多咧俺也没恁多事儿咧说过咧都算球咧恁也都该歇歇咧

Out: 今儿个瞅见有说弄啥嘞，嗯，可别喷那么多咧，俺也没恁多事儿咧，说过咧都算球咧，恁也都该歇歇咧。

In: 工钱都拖了一年多咧四个月嘞俺咋弄类

Out: 工钱都拖了一年多咧，4个月嘞，俺咋弄类？

UniFormBench-Foreign (zero-shot) — 2 examples per language, 8 languages

German (de)

In: die questreihe beginnt damit dass du mit dem npc sprichst der am eingang der höhle steht und er gibt dir eine karte mit den koordinaten x gleich fünf und y gleich minus drei woraufhin du dich auf den weg machen musst um das verlorene schwert zu finden das angeblich unendliche kraft besitzt

Out: Die Questreihe beginnt damit, dass du mit dem NPC sprichst, der am Eingang der Höhle steht, und er gibt dir eine Karte mit den Koordinaten x=5 und y=-3, woraufhin du dich auf den Weg machen musst, um das verlorene Schwert zu finden, das angeblich unendliche Kraft besitzt.

In: beim skispringen ist die windmessung extrem wichtig denn wenn der wind plötzlich dreht kann das den sprung weit verkürzen und den springer in große gefahr bringen weshalb der wettkampf oft pausiert wird

Out: Beim Skispringen ist die Windmessung extrem wichtig, denn wenn der Wind plötzlich dreht, kann das den Sprung weit verkürzen und den Springer in große Gefahr bringen, weshalb der Wettkampf oft pausiert wird.

Spanish (es)

In: el análisis del electrocardiograma muestra una elevación significativa del segmento ST en las derivaciones V uno a V cuatro lo cual sugiere fuertemente un infarto agudo de miocardio en la pared anterior por eso es vital activar el protocolo de cateterismo cardíaco de inmediato para restaurar el flujo sanguíneo coronario

Out: El análisis del electrocardiograma muestra una elevación significativa del segmento ST en las derivaciones V1 a V4, lo cual sugiere fuertemente un infarto agudo de miocardio en la pared anterior; por eso, es vital activar el protocolo de cateterismo cardíaco de inmediato para restaurar el flujo sanguíneo coronario.

In: si quieres encontrar unos jeans Levi's 501 originales a buen precio te recomiendo que vayas a las tiendas outlet que hay a las afueras de la ciudad porque en el centro comercial todo está a precio completo y apenas hay descuentos reales salvo en las rebajas de enero y julio que es cuando todo el mundo se vuelve loco

Out: Si quieres encontrar unos jeans Levi's 501 originales a buen precio, te recomiendo que vayas a las tiendas outlet que hay a las afueras de la ciudad, porque en el centro comercial todo está a precio completo y apenas hay descuentos reales, salvo en las rebajas de enero y julio, que es cuando todo el mundo se vuelve loco.

French (fr)

In: le volleyeur a réalisé un contre bloquant efficace au filet empêchant l attaque puissante de l équipe adverse de toucher le sol et permettant à son propre équipe de lancer une contre attaque rapide qui s est conclue par un smash irrésistible dans le coin opposé du terrain pour gagner le set

Out: Le volleyeur a réalisé un contre bloquant efficace au filet, empêchant l'attaque puissante de l'équipe adverse de toucher le sol, et permettant à son propre équipe de lancer une contre-attaque rapide, qui s'est conclue par un smash irrésistible dans le coin opposé du terrain pour gagner le set.

In: le framework Django gère très bien la sécurité par défaut mais tu as désactivé la protection CSRF dans tes vues API ce qui expose ton application à des attaques par injection de script inter site et c'est une faille critique que les auditeurs de sécurité ne manqueront pas de souligner lors du prochain audit

Out: Le framework Django gère très bien la sécurité par défaut, mais tu as désactivé la protection CSRF dans tes vues API, ce qui expose ton application à des attaques par injection de script inter-site, et c'est une faille critique que les auditeurs de sécurité ne manqueront pas de souligner lors du prochain audit.

Indonesian (id)

In: dalam buku motivasinya andi f noya menulis bahwa mimpi besar tidak akan pernah terwujud jika kita hanya tidur siang saja karena impian membutuhkan aksi nyata dan kerja keras yang konsisten setiap hari tanpa mengenal lelah hingga tujuan tersebut benar benar tercapai di depan mata kita semua

Out: Dalam buku motivasinya, Andi F. Noya menulis bahwa mimpi besar tidak akan pernah terwujud jika kita hanya tidur siang saja, karena impian membutuhkan aksi nyata dan kerja keras yang konsisten setiap hari, tanpa mengenal lelah, hingga tujuan tersebut benar-benar tercapai di depan mata kita semua.

In: wasit sudah meniup peluit untuk menghentikan permainan karena pemain nomor sepuluh itu dianggap melakukan pelanggaran berat di area kotak penalti dan akibatnya tim lawan mendapatkan hadiah tendangan penalti yang bisa mengubah skor sementara menjadi dua satu padahal tadi mereka sempat tertinggal nol satu sebelum babak kedua dimulai

Out: Wasit sudah meniup peluit untuk menghentikan permainan, karena pemain nomor 10 itu dianggap melakukan pelanggaran berat di area kotak penalti, dan akibatnya tim lawan mendapatkan hadiah tendangan penalti yang bisa mengubah skor sementara menjadi 2-1, padahal tadi mereka sempat tertinggal 0-1 sebelum babak kedua dimulai.

Italian (it)

In: ho visto un tutorial su youtube che spiegava come fare editing video professionale usando solo il telefono e devo dire che il risultato finale era così impressive che ho deciso di scaricare quell'app subito

Out: Ho visto un tutorial su YouTube che spiegava come fare editing video professionale usando solo il telefono, e devo dire che il risultato finale era così impressive che ho deciso di scaricare quell'app subito.

In: l'uso di antenne MIMO multiple permette di sfruttare la diversità spaziale per aumentare la capacità del canale senza richiedere più spettro di frequenza ma richiede algoritmi di elaborazione del segnale complessi per separare i flussi di dati sovrapposti e se la correlazione tra le antenne è troppo alta a causa di una spaziatura insufficiente il vantaggio del MIMO si riduce drasticamente

Out: L'uso di antenne MIMO multiple permette di sfruttare la diversità spaziale per aumentare la capacità del canale senza richiedere più spettro di frequenza, ma richiede algoritmi di elaborazione del segnale complessi per separare i flussi di dati sovrapposti; e se la correlazione tra le antenne è troppo alta a causa di una spaziatura insufficiente, il vantaggio del MIMO si riduce drasticamente.

Portuguese (pt)

In: a louis vuitton acabou de abrir uma loja pop up no centro da cidade e a fila para entrar já está enorme desde cedo pois todos querem ver de perto as novas malas de viagem desenhadas em colaboração com aquele artista japonês famoso

Out: A Louis Vuitton acabou de abrir uma loja pop-up no centro da cidade, e a fila para entrar já está enorme desde cedo, pois todos querem ver de perto as novas malas de viagem desenhadas em colaboração com aquele artista japonês famoso.

In: precisamos fazer um debug no código embedded que controla o display lcd porque os pixels estão piscando aleatoriamente e suspeitamos que o buffer de memória esteja sendo sobrescrito por uma interrupção mal tratada no timer do microcontrolador arm cortex m quatro o que gera instabilidade visual constante

Out: Precisamos fazer um debug no código embedded que controla o display LCD, porque os pixels estão piscando aleatoriamente, e suspeitamos que o buffer de memória esteja sendo sobrescrito por uma interrupção mal tratada no timer do microcontrolador ARM Cortex-M4, o que gera instabilidade visual constante.

Russian (ru)

In: я поставил телефон на silent mode но вдруг придёт важное уведомление от server о том что база данных перегружена запросами свыше тысячи в секунду и администратор должен немедленно войти в систему иначе произойдёт crash и мы потеряем все логи за последний месяц что станет настоящей катастрофой для нашего отдела аналитики

Out: Я поставил телефон на silent mode, но вдруг придёт важное уведомление от server о том, что база данных перегружена запросами свыше 1000 в секунду, и администратор должен немедленно войти в систему, иначе произойдёт crash, и мы потеряем все логи за последний месяц, что станет настоящей катастрофой для нашего отдела аналитики.

In: система контроля версий показала что последний коммит сломал сборку проекта потому что кто то забыл закоммить файл конфигурации окружения и теперь локально у всех разработчиков приложение работает а на сервере падает с ошибкой отсутствия переменных среды

Out: Система контроля версий показала, что последний коммит сломал сборку проекта, потому что кто-то забыл закоммить файл конфигурации окружения, и теперь локально у всех разработчиков приложение работает, а на сервере падает с ошибкой отсутствия переменных среды.

Turkish (tr) — near-match

The two cases below are the closest Turkish predictions: the first differs from the reference by one extra comma, the second by one missing apostrophe and one extra comma; all other punctuation and ITN are correct.

In: pizzanızın hamuru normal mi yoksa tam buğday unundan mı yapılıyor ayrıca üzerine serpiştirdiğiniz baharat karışımının içinde hayvansal kökenli herhangi bir katkı maddesi bulunuyor mu bu detaylar benim için kritik

Out: Pizzanızın hamuru normal mi, yoksa tam buğday unundan mı yapılıyor? Ayrıca, üzerine serpiştirdiğiniz baharat karışımının içinde hayvansal kökenli herhangi bir katkı maddesi bulunuyor mu? Bu detaylar benim için kritik.

In: distributed systemde clock skew nedeniyle event sıralaması karışıyor logical clock veya vector clock mekanizmalarını entegre etmezsek veri tutarlılığını sağlayamayız

Out: Distributed system'de clock skew nedeniyle event sıralaması karışıyor, logical clock veya vector clock mekanizmalarını entegre etmezsek veri tutarlılığını sağlayamayız.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction accurately reflect the paper's contributions and scope, including the three-stage training paradigm, SA-GRPO, and empirical results.

   Guidelines:

   - The answer [N/A] 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 [N/A] 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 Discussion section addresses limitations including the lack of acoustic information and identifies cases where text-only input is insufficient for correct formatting.

   Guidelines:

   - The answer [N/A] 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 include theoretical results; it is an empirical systems paper.

Guidelines:

- The answer [N/A] 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: We provide all necessary information for reproducibility including model architecture, training hyperparameters, data construction pipeline details, and evaluation protocols in Sections 3–5.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: Subject to internal corporate review, the code and data are not publicly available at present. Upon approval, we intend to release the evaluation data and evaluation code after removing all project-sensitive information.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.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://neurips.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) necessary to understand the results?

Answer: [Yes]

Justification: Section 5.1 provides full implementation details including learning rates, batch sizes, GPU configurations, and training durations for all three stages.

Guidelines:

- The answer [N/A] 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: Evaluation is deterministic given fixed weights and a fixed evaluation environment (greedy decoding), whereas training is stochastic; we report mean $\pm$ 1-sigma sample standard deviation over three training seeds in Appendix F. These bars describe seed variability only: compute limits precluded enough runs for formal significance testing, which we do not claim, and we make no claim of strict reproducibility across different hardware or software environments.

Guidelines:

- The answer [N/A] 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: Section 5.1 specifies GPU type (NVIDIA H20), number of GPUs (32), and training time for each stage.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics in all respects.

   Guidelines:

   - The answer [N/A] 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: [N/A]

Justification: The work focuses on ASR post-processing (punctuation and text normalization), which is a standard NLP utility with no foreseeable negative societal impact.

Guidelines:

- The answer [N/A] means that there is no societal impact of the work performed.
- If the authors answer [N/A] 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., pre-trained language models, image generators, or scraped datasets)?

Answer: [N/A]

Justification: The model and data are not publicly released at this time, and the text formatting task poses no high risk for misuse.

Guidelines:

- The answer [N/A] 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 (pre-trained models, baseline systems, evaluation toolkits) are properly cited with references to their original papers.

Guidelines:

- The answer [N/A] 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: [No]

    Justification: The new benchmark (UniFormBench) and model are not yet accompanied by standalone documentation; we plan to provide this upon public release.

    Guidelines:

    - The answer [N/A] 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 or research with human subjects.

    Guidelines:

    - The answer [N/A] 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 requiring IRB approval.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: The model we propose and train is an LLM, which is an important, original component of our core method. LLMs are also used as teacher models in our data construction pipeline.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.